

Dominique FOATA ET Aimé FUCHS

**CALCUL DES
PROBABILITÉS**

Cours, exercices et problèmes corrigés

SECONDE ÉDITION : 1998

ISBN 2 10 007547 0

Dunod, Paris

TABLE DES MATIÈRES

Préface	ix
Préface de la première édition	x
Liste des symboles utilisés	xiii
CHAPITRE PREMIER. Le langage des probabilités	1
Un exemple. Le triplet fondamental. Suites infinies d'évènements. Compléments et exercices.	
CHAPITRE 2. Les évènements	7
Les algèbres. Les tribus. Les systèmes de Dynkin. Les classes monotones. Compléments et exercices.	
CHAPITRE 3. Espaces probabilisés	15
Probabilités. Propriétés. Formule de Poincaré et inégalité de Boole. Autres propriétés. Identités binomiales. Compléments et exercices.	
CHAPITRE 4. Probabilités discrètes. Dénombrements	25
Probabilités discrètes. Équirépartition sur les espaces finis. Ensem- bles finis. Formules classiques de dénombrement. Le principe de réflexion. Compléments et exercices (problème des rencontres, le chevalier de Méré, boules et urnes).	
CHAPITRE 5. Variables aléatoires	43
Application réciproque. Fonctions mesurables. Variables aléa- toires. Loi de probabilité d'une variable aléatoire. Fonction de répartition d'une variable aléatoire réelle. La fonction de masse et les discontinuités de la fonction de répartition. Tribu engendrée par une variable aléatoire. Compléments et exercices.	
CHAPITRE 6. Probabilités conditionnelles. Indépendance	53
Probabilité conditionnelle. Systèmes complets d'évènements. Pro- babilités définies par des probabilités conditionnelles. Évènements indépendants. Indépendance de classes d'évènements. Variables aléatoires indépendantes. Compléments et exercices (tirages avec et sans remise).	
CHAPITRE 7. Variables aléatoires discrètes. Lois usuelles	67
Variables aléatoires discrètes. La loi binomiale. La loi hyper- géométrique. La loi géométrique. La loi de Poisson. Complé- ments et exercices (problème des boîtes d'allumettes de Banach, poissonisation, le paradoxe de l'inspection).	

CHAPITRE 8. Espérance mathématique. Valeurs typiques 79

Transformation de variables aléatoires. Indépendance. Convolution des lois de probabilité discrètes. Espérance mathématique. Moments. Covariance. Le coefficient de corrélation linéaire. L'inégalité de Tchebychev. Les inégalités relatives aux moments dans le cas fini. Médiane, écart moyen minimum. Compléments et exercices.

CHAPITRE 9. Fonctions génératrices 99

Définitions. Propriétés. Sommes de variables aléatoires. Le théorème de continuité. Compléments et exercices.

CHAPITRE 10. Mesures de Stieltjes-Lebesgue. Intégrale des variables aléatoires réelles 113

Mesures. Mesures de Stieltjes-Lebesgue sur la droite. Mesure de probabilité induite par une fonction de répartition. Mesures de Stieltjes-Lebesgue sur \mathbb{R}^n . Variables aléatoires réelles. Intégrale d'une variable aléatoire réelle par rapport à une mesure. Exemples. Propriétés de l'intégrale. Théorèmes de convergence. Compléments et exercices (comment probabiliser l'ensemble des suites infinies du jeu de « pile » ou « face »).

CHAPITRE 11. Espérance mathématique. Lois absolument continues 129

Espérance mathématique d'une variable aléatoire. Mesures de probabilité produit et théorème de Fubini. Intégrale de Lebesgue. Lois de probabilité absolument continues. Les trois types de fonctions de répartition. Convolution. Compléments et exercices.

CHAPITRE 12. Variables aléatoires à deux dimensions ; espérance conditionnelle. Lois normales 141

Définitions et premières propriétés. Loi de probabilité absolument continue, densité de probabilité. Loi de probabilité conditionnelle, espérance mathématique conditionnelle, régression. Règles de calcul concernant les espérances conditionnelles. La loi normale à deux dimensions. Compléments et exercices.

CHAPITRE 13. Fonction génératrice des moments ; fonction caractéristique 159

Introduction. Propriétés élémentaires. Moments. Fonction caractéristique. Seconde fonction caractéristique. Fonction génératrice d'un vecteur aléatoire. Propriété fondamentale. Compléments et exercices.

CHAPITRE 14. Les principales lois de probabilité (absolument continues)	177
La loi uniforme sur $[0, 1]$. La loi uniforme sur $[a, b]$. La loi normale ou de Laplace-Gauss. La loi Log-normale. La loi exponentielle. La première loi de Laplace. La loi de Cauchy. La loi gamma. La loi bêta. Les lois arcsinus. Compléments et exercices.	
CHAPITRE 15. Lois de probabilité de fonctions de variables aléatoires	195
Cas à une dimension. Cas à deux dimensions. Loi de probabilité d'une fonction de deux variables aléatoires. Compléments et exercices.	
CHAPITRE 16. Convergences stochastiques	205
Convergence en loi ou convergence étroite. Convergence en probabilité. Convergence en moyenne d'ordre $r > 0$. Convergence presque sûre. Comparaison des divers types de convergence. Convergence en loi de variables aléatoires à valeurs entières et absolument continues. Convergence étroite et convergence presque sûre. La convergence en loi d'un point de vue fonctionnel. Le théorème de Paul Lévy. Compléments et exercices.	
CHAPITRE 17. Loi des grands nombres	225
La loi faible des grands nombres. La loi forte des grands nombres. Les lemmes de Borel-Cantelli. Compléments et exercices.	
CHAPITRE 18. Le rôle central de la loi normale ; le théorème « central limit »	237
Aperçu historique. Le théorème «central limit». Le théorème «central limit» et la formule de Stirling. Le théorème de Lindeberg. Le théorème de Liapounov. Compléments et exercices.	
CHAPITRE 19. La loi du logarithme itéré	251
Notations et lemmes préliminaires. Loi forte des grands nombres et théorème de Hardy-Littlewood. La loi du logarithme itéré.	
CHAPITRE 20. Applications des probabilités : problèmes résolus	259
Le problème des rencontres revisité. Un problème de temps d'attente. Acheminement du courrier par voie hiérarchique. Fractions continues. Une application de la formule de Bernstein. Le modèle de la diffusion d'Ehrenfest. Vecteurs uniformément répartis sur la sphère-unité de \mathbb{R}^n . Un problème de probabilité géométrique.	
Solutions des exercices	281
Index	325

PRÉFACE

Cette nouvelle édition du livre conserve le même esprit que l'édition précédente : les éléments du calcul des probabilités sont exposés dans le corps des chapitres et les exercices proposés reçoivent des solutions souvent détaillées. L'ouvrage comporte vingt chapitres (le précédent en comportait dix-neuf). Les dix-neuf premiers chapitres n'ont pas été fondamentalement modifiés. Seules des améliorations locales ont été apportées : par exemple, des démonstrations plus élégantes de résultats ont été incluses, les coquilles notées par les auteurs ou relevées par les lecteurs dans la première édition ont été corrigées. Plus important, plusieurs nouveaux exercices ont été ajoutés. Enfin, certains exercices de la première édition qui n'avaient pas de solutions détaillées en ont maintenant une.

Un chapitre 20 intitulé *Applications des probabilités : problèmes résolus* a été ajouté. La solution de ces problèmes fait appel aux différentes techniques et méthodes présentées dans le livre, souvent de façon simultanée. Ce ne sont plus des *exercices*, mais des *problèmes inhabituels*, totalement résolus, qui fournissent également une *ouverture* vers d'autres branches des mathématiques.

Nous avons également ajouté la référence à l'ouvrage récent de Daniel Revuz sur la théorie de l'intégration et apporté dans le corps du texte plusieurs appels à référence à ce livre.

Pour la rédaction de cette nouvelle édition, nous avons bénéficié des remarques et suggestions de la part de plusieurs collègues. Nous remercions tout particulièrement Jean-Pierre Dion, qui a fait une lecture approfondie du précédent livre et a ainsi relevé plusieurs imperfections qui ont été corrigées dans cette nouvelle édition. Anatole Joffe a continué de nous faire bénéficier de toute son érudition. Nos deux collègues Wilbur Jonsson et Volker Strehl, qui ont bien voulu se charger, respectivement, des traductions anglaise et allemande du présent ouvrage, nous ont fait part de remarques judicieuses, quant au contenu mathématique lui-même. Nous remercions enfin d'autres lecteurs attentifs et notamment Edith Kosmanek, Michel Valadier.

Strasbourg, le 20 mars 1998

Dominique FOATA
Aimé FUCHS

PRÉFACE DE LA PREMIÈRE ÉDITION

Ce cours de probabilités s'adresse aux étudiants de licence de mathématique (bac+3) des Universités. Ils y trouveront aussi des éléments du calcul des probabilités qui ne sont développés qu'en maîtrise (bac+4). On suppose de leur part une maîtrise des techniques de l'analyse mathématique telle qu'elle est enseignée dans les deux premières années des Universités ou dans les classes préparatoires aux Grandes Écoles et tout particulièrement des techniques sur les séries numériques et les séries de puissances.

Ce bagage mathématique étant supposé acquis, nous avons délibérément pris le parti de développer, au début de ce livre, en fait durant les neuf premiers chapitres, une théorie des probabilités *discrètes*, reposant sur la seule technologie des séries, tout en distillant quelques notions plus avancées. Il est plus aisé de la sorte de s'initier aux idées probabilistes et de les reprendre ensuite dans le contexte d'une théorie de la mesure.

Comme le dit Pierre Cartier, la théorie de la mesure constitue pour les probabilités une hygiène indispensable. Autrefois, c'est-à-dire dans les années soixante, on pouvait disposer du livre de poche de Bauer [1], qui, en quelques pages, donnait ces règles d'hygiène. L'ouvrage n'est plus en librairie, il est remplacé par le premier tome du traité [2] du même auteur, dans lequel on retrouve un texte moins concis que celui du livre de poche. On peut encore trouver un excellent exposé des bases mathématiques du calcul des probabilités dans le manuel de Neveu [9].

De façon générale, les étudiants de licence suivent, parallèlement à un cours de probabilités, un cours de théorie de l'intégration. L'expérience montre qu'il faut un certain temps pour que les notions développées dans un tel cours soient bien utilisées dans d'autres matières. Il nous a donc paru indispensable de donner (dans les chapitres 10 et 11) des éléments de la théorie de la mesure et de l'intégration, pour les appliquer ensuite au traitement de l'espérance mathématique et des autres notions probabilistes dans le cas général.

Les chapitres ultérieurs traitent des variables aléatoires à plusieurs dimensions ; on y présente une théorie de l'espérance conditionnelle pour les variables aléatoires absolument continues, ainsi qu'un exposé sur les lois normales à plusieurs dimensions. On y trouve aussi un traitement de la fonction génératrice des moments, une étude approfondie des principales lois de probabilités avec une indication des domaines d'activité dans lesquels on les rencontre, enfin un exposé sur les convergences stochastiques, la loi des grands nombres, le théorème « central limit » et la loi du logarithme itéré.

Cet ouvrage comporte de nombreux exercices, traditionnels comme le fameux problème des boîtes d'allumettes de Banach ou plus originaux, comme la « poissonisation » (voir chap. 7). La plupart reçoivent une solution détaillée.

Les théorèmes, propositions, lemmes ont une numérotation dépendant du paragraphe dans lequel ils apparaissent. Dans le chapitre 6, § 1, par exemple, on trouvera successivement le Théorème 1.1, puis la Proposition 1.2. Les remarques écrites à la suite l'une de l'autre sont numérotées 1, 2, ... sans référence au paragraphe qui les contient. Les définitions ne sont pas numérotées. Les formules centrées ont une numérotation entre parenthèses ne dépendant que du paragraphe dans lequel elles se trouvent. Les renvois aux théorèmes, propositions, lemmes, sont faits sous la forme : *cf.* Théorème 1.1 du chap. 6, par exemple.

Il y a aujourd'hui beaucoup de traités de probabilités, en beaucoup de langues. Il ne nous est pas possible de les citer tous. Nous ne pouvons pas cependant ne pas rendre hommage au premier d'entre eux, à celui qui a tant fait pour populariser le sujet et qui est toujours le premier succès mondial de librairie pour les ouvrages de mathématique, le fameux livre de Feller [3]. On y traite seulement de probabilités discrètes, mais avec un talent qui est resté inégalé.

En langue française, citons les ouvrages de Métivier [7] et Rényi [10], qui présupposent un bagage mathématique semblable à celui qui est demandé ici. Citons également la version anglaise du second ouvrage, entièrement repensée, contenant une étude tout à fait passionnante sur les fondements des probabilités [11]. Signalons aussi l'ouvrage récent de Grimmett et Stirzaker [4], aussi du niveau de la licence de mathématique; enfin, en langue italienne, l'élégant manuel de notre collègue Letta [6].

Il nous paraît utile enfin de citer en référence quelques traités de théorie de la mesure et de l'intégration, comme, par exemple, le livre de Munroe [8] ou celui de Jean [5].

Plusieurs collègues bienveillants nous ont apporté leurs conseils et nous ont fait part de remarques et corrections dans la lecture des premières versions du présent texte. Qu'il nous soit permis de remercier tout particulièrement Philippe Artzner, Milos Dostal, Xavier Fernique, Bernard Heinkel, Elisabeth Khalili, Giorgio Letta, Gérard Rauch, Gian-Carlo Rota, Raymond Sérout.

L'aide la plus efficace, la plus chaleureuse, la plus instructive aussi, nous a été apportée par Anatole Joffe, qui a bien voulu tester une première version de ce livre dans son enseignement des probabilités, lors de son séjour à Strasbourg au printemps 1995. C'est grâce à lui et à son talent que nous avons pu enfin boucler l'ouvrage qui était resté fragmentaire durant plusieurs années.

Mme Martine Lemonnier, directrice d'édition scientifique chez Masson a fait une relecture très précieuse du manuscrit, tout comme M. Sinnou David, le directeur scientifique de la Collection, pour la partie mathématique et nous ont suggéré plusieurs améliorations. Enfin, Mme Geneviève Bignet a réalisé une lecture du document final et nous a signalé plusieurs passages qui méritaient un meilleur traitement typographique. Nous les remercions tous trois très chaleureusement.

RÉFÉRENCES

- [1] Bauer (Heinz). — *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie*, Band I. — Berlin, Walter De Gruyter & Co., Sammlung Götschen Band 1216/1216a, 1964. Traduction anglaise : *Probability Theory and Elements of Measure Theory*. New York, Academic Press, 1971.
- [2] Bauer (Heinz). — *Maß und Integrationstheorie*, 2. Auflage. — Berlin, Walter De Gruyter & Co., 1990.
- [3] Feller (William). — *An Introduction to Probability and its Applications*, vol. 1, 3rd Edition. — New York, John Wiley & Sons, 1968.
- [4] Grimmett (G.R.) and Stirzaker (D.R.). — *Probability and Random Processes*, 2 vol., (with problems and solutions). — Oxford, Clarendon Press, 1992.
- [5] Jean (R.). — *Mesure et Intégration*. — Montréal, Presses de l'Université du Québec, 1975.
- [6] Letta (Giorgio). — *Probabilità elementare*. — Bologna, Zanichelli, 1993.
- [7] Métivier (Michel). — *Notions fondamentales de la théorie des probabilités*. — Paris, Dunod, 1972.
- [8] Munroe (M.E.). — *Introduction to Measure and Integration*. — Reading, Mass., Addison-Wesley, seconde édition, 1970.
- [9] Neveu (Jacques). — *Bases mathématiques du calcul des probabilités*. — Paris, Masson, 1964; réédition : 1970.
- [10] Rényi (Alfred). — *Calcul des probabilités avec un appendice sur la théorie de l'information*. — Paris, Dunod, 1966.
- [11] Rényi (Alfred). — *Foundations of Probability*. — San Francisco, Holden-Day, Inc., 1970.
- [12] Revuz (Daniel). — *Mesure et intégration*. — Paris, Hermann (Collection Méthodes), 1994.

Strasbourg, le 5 janvier 1996

Dominique FOATA

Aimé FUCHS

Département de mathématique

Université Louis Pasteur

7, rue René-Descartes,

F-67084 Strasbourg

courr. élect. : foata@math.u-strasbg.fr

LISTE DES SYMBOLES UTILISÉS

- $(\Omega, \mathfrak{A}, P)$: triplet fondamental, chap. 1 § 2
 $\mathfrak{P}(\Omega)$: l'ensemble de tous les sous-ensembles de Ω , chap. 1 § 2
 \emptyset : évènement impossible, chap. 1 § 2
 Ω : évènement certain, chap. 1 § 2
 $A \subset B$: A implique B , chap. 1 § 2
 $A \cap B$ ou AB : conjonction de A et B , chap. 1 § 2
 $A \cup B$: union de A et B , chap. 1 § 2
 $A^c = \Omega \setminus A$: évènement contraire à A , chap. 1 § 2
 $A + B = A \cup B$ (si $A \cap B = \emptyset$) : chap. 1 § 2
 $A \setminus B$: différence de A et B , chap. 1 § 2
 $\bigcup_n A_n$: "au moins l'un des A_n se réalise", chap. 1 § 2
 $\bigcap_n A_n$: "tous les A_n se réalisent", chap. 1 § 2
 $\liminf_n A_n$: "tous les A_n se réalisent après un certain indice", chap. 1 § 3
 $\limsup_n A_n$: "une infinité d'évènements A_n se réalisent", chap. 1 § 3
 I_A : indicateur de A , chap. 1 § 3
 \mathcal{P} : l'ensemble de tous les intervalles semi-ouverts de la droite, chap. 2 § 1
 $\mathbb{R} =] - \infty, +\infty[$: la droite réelle, chap. 2 § 1
 $\mathfrak{T}(\mathcal{C})$: tribu engendrée par \mathcal{C} , chap. 2, § 2
 \mathcal{B} ou \mathcal{B}^1 : tribu des boréliens de la droite, chap. 2, § 2
 \mathcal{B}^n : tribu des boréliens de \mathbb{R}^n , chap. 2, § 2
 (Ω, \mathfrak{A}) : espace probabilisable, chap. 2, § 2
 $\mathfrak{D}(\mathcal{C})$: système de Dynkin engendré par \mathcal{C} , chap. 2, § 3
 $\mathfrak{M}(\mathcal{C})$: classe monotone engendrée par \mathcal{C} , chap. 2, § 4
 $(\Omega, \mathfrak{A}, P)$: espace probabilisé, chap. 3, § 1
 $(a)_n$: factorielle montante, chap. 3, § 5
 ${}_pF_q \left(\begin{matrix} a_1, \dots, a_p \\ b_1, \dots, b_q \end{matrix} ; x \right)$: la fonction hypergéométrique, chap. 3, § 5
 ε_{ω_0} : mesure de probabilité singulière en ω_0 , chap. 4, § 1
 $\text{card } A$ ou $|A|$: cardinal ou cardinalité de A , chap. 4, § 3
 $\mathbb{N} = \{0, 1, \dots\}$: ensemble des entiers naturels, chap. 4, § 3
 $\mathbb{N}^* = \{1, 2, \dots\}$, chap. 4, § 3
 $[n]$: l'intervalle $\{1, 2, \dots, n\}$, chap. 4, § 3
 $\binom{p}{n}$: coefficient binomial, chap. 4, § 4
 $\binom{p}{n_1, n_2, \dots, n_k}$: coefficient multinomial, chap. 4, § 4
 X^{-1} : application réciproque, chap. 5, § 1
 $\{X \in B\}$: l'évènement $\{\omega \in \Omega : X(\omega) \in B\}$, chap. 5, § 3
 $P(A_1, A_2)$: probabilité que A_1 et A_2 se produisent tous deux, chap. 5, § 3

- P_X ou $\mathcal{L}(X)$: loi de probabilité de X , chap. 5, § 4
 $F(x) = P\{X \leq x\}$: fonction de répartition de X , chap. 5, § 5
 $\pi(x)$: fonction de masse de X , chap. 5, § 6
 S_X : support de la variable aléatoire X , chap. 5, § 6
 $\mathfrak{T}(X)$: tribu engendrée par la variable aléatoire X , chap. 5, § 6
 $P\{\cdot | A\}$: mesure de probabilité conditionnelle A étant donné, chap. 6, § 1
 $B(n, p)$: loi binomiale, chap. 7, § 2
 $H(n, N, M)$: loi hypergéométrique, chap. 7, § 3
 $G(p)$: loi géométrique, chap. 7, § 4
 $\overline{\mathbb{R}} = [-\infty, +\infty]$: droite achevée, chap. 7, § 4
 π_λ ou $\mathcal{P}(\lambda)$: loi de Poisson, chap. 7, § 5
 $P * Q$: produit de convolution de P par Q , chap. 8, § 3
 $\mathbb{E}[X]$: espérance mathématique de la variable discrète X , chap. 8, § 4
 ${}_a m_r$: moment d'ordre r centré en a chap. 8, § 5
 $\text{Var } X$: variance de X , chap. 8, § 5
 $\sigma(X)$: écart-type de X , chap. 8, § 5
 e_r : écart absolu d'ordre r , chap. 8, § 5
 $\text{Cov}(X, Y)$: covariance de X et Y , chap. 8, § 6
 $r(X, Y)$: coefficient de corrélation linéaire de (X, Y) , chap. 8, § 7
 \mathcal{M} : ensemble de mesures de probabilité de support \mathbb{N} , chap. 9, § 1
 \mathcal{M}' : classe des variables aléatoires dont la loi appartient à \mathcal{M} , chap. 9, § 1
 $\mathcal{M}(s)$: ensemble des séries de puissance en bijection avec \mathcal{M} , chap. 9, § 1
 $G_P(s)$: fonction génératrice de la mesure de probabilité P , chap. 9, § 1
 $G_X(s)$: fonction génératrice de la variable aléatoire X , chap. 9, § 1
 $(\Omega, \mathfrak{A}, \mu)$: espace mesuré, chap. 10, § 1
 $F\{\cdot\}$: mesure de Lebesgue-Stieltjes induite par F , chap. 10, § 2
 λ^1 ou λ : mesure de Lebesgue sur la droite réelle, chap. 10, § 2
 λ^n : mesure de Lebesgue sur \mathbb{R}^n , chap. 10, § 4
 $\overline{\mathcal{B}}$: tribu borélienne achevée, chap. 10, § 5
 $\int X d\mu$: intégrale de X par rapport à la mesure μ , chap. 10, § 6
 $\mathbb{E}[X]$: espérance mathématique de la variable aléatoire X , chap. 11, § 1
 $\int X d\lambda$ ou $\int X(x) dx$: intégrale de Lebesgue de X , chap. 11, § 3
 $\int X dF$: intégrale de Stieltjes-Lebesgue de X , chap. 11, § 4
 $f_{X,Y}(x, y)$: densité conjointe de (X, Y) , chap. 12, § 2
 $f_{Y|X}(\cdot | x)$: densité de Y conditionnelle à $\{X = x\}$, chap. 12, § 3
 $\mathbb{E}[Y | X = x]$: espérance de Y conditionnelle à $\{X = x\}$, chap. 12, § 3
 $\mathbb{E}[Y | X]$: espérance mathématique de Y en X , chap. 12, § 3
 ${}^t A$: transposée de la matrice A , chap. 12, § 5
 $\mathcal{N}_2(0, \Gamma)$: loi normale centrée à deux dimensions, chap. 12, § 5
 $\mathcal{N}_2(\mu, \Gamma)$: loi normale à deux dimensions, chap. 12, § 5
 $g_X(u)$: fonction génératrice des moments de X , chap. 13, § 1
 $\varphi_X(t)$: fonction caractéristique de X , chap. 13, § 4
 $\psi_X(t)$: seconde fonction caractéristique de X , chap. 13, § 5
 κ_n : le $n^{\text{ième}}$ cumulants de X , chap. 13, § 5

$g(u, v)$: fonction génératrice des moments de (X, Y) , chap. 13, § 6

$\mathcal{N}(0, 1)$: loi normale centrée, réduite, chap. 14, § 3

$\mathcal{N}(\mu, \sigma)$: loi normale, chap. 14, § 3

$\mathcal{E}(\lambda)$: loi exponentielle, chap. 14, § 5

$C(0, 1)$: loi de Cauchy, chap. 14, § 7

$\Gamma(p, \lambda)$: loi gamma, chap. 14, § 8

$B(r, s)$: loi bêta, chap. 14, § 9

$X_n \xrightarrow{\mathcal{L}} X$: X_n converge en loi vers X , chap. 16, § 1

$X_n \xrightarrow{p} X$: X_n converge en probabilité vers X , chap. 16, § 2

$X_n \xrightarrow{p.s.} X$: X_n converge presque sûrement vers X , chap. 16, § 4

CHAPITRE PREMIER

LE LANGAGE DES PROBABILITÉS

L'objet des trois premiers chapitres est de donner la description complète et les propriétés du triplet fondamental $(\Omega, \mathfrak{A}, P)$, maintenant adopté par tous les probabilistes pour une première adéquation mathématique de la notion de hasard. D'abord, il faut décrire les évènements¹ attachés au phénomène considéré où intervient le hasard. C'est l'objet de ce présent chapitre. Ces évènements apparaîtront comme des sous-ensembles du premier élément Ω du triplet. Pour permettre une manipulation aisée des évènements, nous introduirons, au chapitre suivant, essentiellement une famille remarquable \mathfrak{A} d'évènements, qui obéit à des règles algébriques précises. Dans le troisième chapitre enfin, nous ferons l'étude du troisième élément P (la *mesure de probabilité*) et pourrons alors donner une pondération aux évènements appartenant à la classe \mathfrak{A} .

1. Un exemple. — Imaginons une compagnie pétrolière, disposant d'un certain nombre de navires et désirant faire une étude rigoureuse de l'arrivée de ses pétroliers à leur port d'attache. Les aléas de la mer font que l'arrivée de ces pétroliers ne peut être prévue à l'heure près. Pour permettre l'analyse de ces arrivées, la compagnie doit considérer *a priori* une classe d'évènements liés à ces arrivées, comme «pas d'arrivée le vendredi entre 8 heures et 10 heures», ou bien «les bateaux 1 et 2 arrivent le 8 janvier», ou bien «le pétrolier n° 2 tombe en panne en mer». Elle considère également une famille d'évènements plus ponctuels ou élémentaires — on les appelle plus volontiers *épreuves* — comme «le pétrolier n° 2 arrive le vendredi 8 janvier à 9 heures». Appelons ω cette épreuve. Il est clair que l'évènement A «pas d'arrivée le

¹ Voilà un mot qu'on rencontrera souvent dans cet ouvrage. On notera attentivement l'accentuation du mot. Comme nous l'a signalé le professeur Charles Muller de la Faculté des Lettres de Strasbourg, «la graphie avec accent aigu sur l'«e» de la deuxième syllabe n'a de justification ni phonétique, ni étymologique. Elle vient du fait qu'au dix-septième siècle, on n'utilisait l'accent grave que sur «a» (à) et sur «u» (comme dans «où», qui est, par ailleurs, le seul mot de la langue française comportant cette lettre accentuée!).» M. Muller ajoute : «Quand pour la troisième édition (1740) on décida de mettre un accent grave sur des mots comme «père», «funèbre», «allègre», etc., l'imprimeur de l'Académie, Coignard, n'avait pas fondu assez d'«e» avec accent grave. Dans de nombreuses pages, il ne mit que des «e» avec accent aigu. Lors de la quatrième édition (1762), quelques mots furent oubliés, dont *évènement*, où subsista l'accent aigu, lequel devint au dix-neuvième siècle un objet de vénération puriste (également *allègrement*, *allègement*, etc.). Désormais la forme *évènement* est la norme, l'autre restant correcte jusqu'à disparition de ses derniers adorateurs.»

vendredi entre 8 heures et 10 heures » ne se réalise pas dans l'épreuve ω , mais que l'évènement B « au moins l'un des bateaux 1 et 2 arrive le 8 janvier » se réalise en ω .

Pour étudier ce phénomène des arrivées, on est donc amené à introduire un *ensemble fondamental* Ω et à l'identifier à l'ensemble des couples (n, t) , où n désigne le numéro d'un pétrolier de la compagnie et t l'instant d'arrivée au port du pétrolier. L'épreuve ω considérée précédemment est alors l'élément $(2, (8, 9))$, où $(8, 9)$ désigne le 8 janvier, 9 heures. L'évènement B est identifié à l'ensemble de tous les éléments (n, t) , tels que $n = 1$ ou 2 et $(8, 0) \leq t \leq (8, 24)$. Dans cet exemple, on voit donc que l'appartenance « $\omega \in B$ » équivaut au fait que « B s'est réalisé dans l'épreuve ω ». Le langage d'*appartenance* provient de la théorie des ensembles, celui de la *réalisation d'évènements* du langage des probabilités. L'équivalence qui vient d'être décrite permet de passer constamment d'un langage à l'autre, comme il est précisé dans les paragraphes suivants.

2. Le triplet fondamental. — Ci-dessus, on a pu expliciter complètement l'ensemble fondamental Ω . Dans la plupart des cas cependant, on se borne à supposer l'existence d'un tel ensemble. Celui-ci fait partie d'un triplet $(\Omega, \mathfrak{A}, P)$ ayant les propriétés suivantes :

- a) Ω est un ensemble non vide appelé *ensemble fondamental*; ses éléments ω sont appelés *épreuves*. L'ensemble Ω contient l'ensemble de tous les résultats possibles de l'expérience aléatoire.
- b) \mathfrak{A} est une classe de sous-ensembles de Ω appelés *évènements*; un évènement est un fait attaché à l'expérience aléatoire et susceptible ou non de se produire. Si ω est une épreuve et A un évènement, on dit que *l'évènement A se réalise dans l'épreuve ω si et seulement si ω appartient à A* .

Dans les cas élémentaires, la classe \mathfrak{A} des évènements est la plupart du temps égale à $\mathfrak{P}(\Omega)$, l'*ensemble de toutes les parties de Ω* (une notation qui est constamment utilisée par la suite), de sorte que tout sous-ensemble de Ω peut être considéré comme évènement. En particulier, tout singleton, c'est-à-dire tout sous-ensemble de la forme $\{\omega\}$, où ω est un élément de Ω , est un évènement appelé *évènement élémentaire*. A un stade plus élaboré de la théorie, il convient de ne pas supposer que la classe \mathfrak{A} soit confondue avec $\mathfrak{P}(\Omega)$, de sorte qu'un singleton peut ne pas être un évènement; la notion d'évènement élémentaire est plutôt à proscrire. Toutefois, on cherche à faire en sorte que la classe \mathfrak{A} obéisse à des règles algébriques commodes, comme celles d'une *tribu*.

- c) P est une *pondération* sur la classe \mathfrak{A} des évènements. Dans le modèle que nous adoptons, P est une *mesure de probabilité* sur \mathfrak{A} . Le nombre $P(A)$ est appelé la *probabilité de A* .

Le modèle que nous venons de proposer permet de traduire dans le langage du calcul des probabilités la plupart des notions de la théorie des ensembles. En particulier, les opérations *logiques* sur les *évènements* vont maintenant être

des opérations *ensemblistes* sur les *parties* d'un ensemble. On utilise parfois simultanément le langage de la théorie des ensembles et celui des probabilités. Les notions suivantes sont d'un usage courant :

1) La partie vide \emptyset est un évènement appelé *évènement impossible*; il ne se réalise dans aucune épreuve;

2) La partie pleine Ω est un évènement appelé *évènement certain*; il se réalise dans toute épreuve;

3) Soient A et B deux évènements. On dit que l'évènement A *entraîne* l'évènement B , si B se réalise (exactement) dans toute épreuve où A se réalise; autrement dit, si l'on a $A \subset B$.

4) Deux évènements A et B sont *équivalents*, si A entraîne B et B entraîne A ; autrement dit, si l'on a $A = B$.

5) La *conjonction* de deux évènements A et B est l'évènement qui se réalise exactement dans toute épreuve où A et B se réalisent tous deux. C'est donc leur *intersection* $A \cap B$.

6) La *réunion* de deux évènements A et B est l'évènement qui se réalise exactement dans toute épreuve où l'un au moins des évènements A , B se réalise. C'est donc $A \cup B$.

7) L'évènement *contraire* d'un évènement A est l'évènement qui se réalise exactement dans toute épreuve où A ne se réalise pas. C'est donc le *complémentaire* de A , que l'on note A^c ($= \Omega \setminus A$).

8) Deux évènements A et B sont *incompatibles* (ou *disjoints*) si leur conjonction est l'évènement impossible, autrement dit, si $A \cap B = \emptyset$.

Notations. — Lorsque les évènements A et B sont incompatibles, on notera $A + B$, plutôt que $A \cup B$, leur *réunion*. De même, on écrit $\sum_i A_i$ la réunion d'évènements A_i , deux à deux incompatibles. On utilise aussi la notation AB pour $A \cap B$.

La *différence* entre deux évènements A et B , dans cet ordre, notée $A \setminus B$, est l'évènement $A \cap B^c$. Lorsque $A \supset B$, on parle plus volontiers de la *différence propre* $A \setminus B$. Une suite d'évènements A_n est notée (A_n) ($n \geq 1$) ou simplement (A_n) . S'il s'agit d'une suite *finie*, on écrit, par exemple, (A_i) ($i = 1, 2, \dots, n$).

3. Suites infinies d'évènements

3.1. *Limites de suites d'évènements.* — Soit (A_n) une suite d'évènements. Dans la théorie des ensembles, la réunion $\bigcup_n A_n$ (on écrit encore $\bigcup_{n=1}^{\infty} A_n$) de la suite (A_n) est l'ensemble des éléments $\omega \in \Omega$ ayant la propriété « il existe un entier n tel que $\omega \in A_n$ ». Dans le langage des probabilités, compte tenu de la transcription ci-dessus, l'évènement $\bigcup_n A_n$ est l'évènement « l'un au moins des A_n se réalise ». De même, $\bigcap_n A_n$ est l'évènement « tous les A_n se réalisent ».

Si la suite (A_n) est *croissante*, c'est-à-dire si l'on a $A_n \subset A_{n+1}$ pour tout n , la réunion $\bigcup_n A_n$ de la suite (A_n) est aussi appelée la *limite* de la suite. On écrit alors $\bigcup_n A_n = \lim_n A_n$, ou encore $A_n \uparrow \lim_n A_n$.

De même, si la suite (A_n) est *décroissante*, à savoir si $A_n \supset A_{n+1}$ pour tout n , l'intersection $\bigcap_n A_n$ de la suite est encore appelée la *limite* de la suite. On écrit $\bigcap_n A_n = \lim_n A_n$ et $A_n \downarrow \lim_n A_n$.

Lorsqu'une suite (A_n) est ou bien croissante, ou bien décroissante, on dit qu'elle est *monotone*.

3.2. *Limites inférieure et supérieure.* — Une suite (A_n) d'évènements étant donnée, on peut toujours définir la limite inférieure et la limite supérieure de cette suite. La *limite inférieure* de la suite (A_n) est définie comme l'ensemble, noté $\liminf_n A_n$, de tous les éléments ω de Ω qui *appartiennent à tous les A_n sauf à un nombre fini d'entre eux*.

De même, la *limite supérieure* de la suite (A_n) est l'ensemble, noté $\limsup_n A_n$, de tous les éléments ω de Ω qui *appartiennent à A_n pour une infinité d'indices n* .

Exprimons les limites inférieure et supérieure en fonction des symboles union et intersection.

PROPOSITION 3.2.1. — *Soit (A_n) une suite de parties d'un ensemble Ω . Alors*

$$\liminf_n A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m; \quad \limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m;$$

$$\liminf_n A_n \subset \limsup_n A_n.$$

Démonstration. — Dire que ω appartient à tous les A_n sauf à un nombre fini, c'est dire qu'à partir d'un rang n , il appartient à l'intersection $B_n = \bigcap_{m \geq n} A_m$. Par conséquent, il existe un entier n tel que $\omega \in B_n$, ce qui prouve la première identité.

Le fait pour l'élément ω d'appartenir à une infinité de parties A_n veut dire qu'aussi loin qu'on peut aller dans la suite des entiers, disons à l'indice n , cet élément appartient toujours à la réunion $C_n = \bigcup_{m \geq n} A_m$. Par conséquent, ω appartient à l'intersection de la suite (C_n) , ce qui établit la seconde identité.

L'inclusion est banale, car si ω appartient à $\liminf_n A_n$, il appartient à tous les A_n à partir d'un certain rang, donc à une infinité d'évènements A_n . \square

Posons $A_* = \liminf_n A_n$ et $A^* = \limsup_n A_n$. On vérifie immédiatement les relations :

$$(3.2.1) \quad (A_*)^c = \limsup_n A_n^c \quad \text{et} \quad (A^*)^c = \liminf_n A_n^c.$$

Définition. — Si $\liminf_n A_n = \limsup_n A_n$, on dit que la suite (A_n) a une *limite* et on écrit :

$$(3.2.2) \quad \lim_n A_n = \liminf_n A_n = \limsup_n A_n.$$

On dit encore que la suite (A_n) *tend vers* $A = \lim_n A_n$ ou *converge vers* A .

PROPOSITION 3.2.2. — Lorsque la suite (A_n) est monotone, les relations (3.2.2) sont vérifiées.

Démonstration. — Si (A_n) est croissante, alors $\bigcap_{m=n}^{\infty} A_m = A_n$, d'où $\liminf_n A_n = \bigcup_{n=1}^{\infty} A_n = \lim_n A_n$. D'autre part, pour tout $n \geq 1$, on a $\bigcup_{m=n}^{\infty} A_m = \bigcup_{m=1}^{\infty} A_m$. D'où $\limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \bigcup_{m=1}^{\infty} A_m = \lim_n A_n$.

Comme la proposition est vraie pour les suites croissantes, elle est aussi vraie pour les suites décroissantes, par application des relations (3.2.1). \square

Dans le langage des probabilités, si l'on considère les termes de la suite (A_n) comme des évènements, l'ensemble $\liminf_n A_n$ est alors l'évènement «tous les A_n se réalisent après un certain rang». De même, $\limsup_n A_n$ est l'évènement «il se produit une infinité d'évènements A_n ». Ce dernier évènement est souvent considéré en calcul des probabilités, en particulier dans la théorie des évènements récurrents. Les Anglo-saxons l'écrivent $\{A_n, \text{i.o.}\}$ («i.o.» signifie «infinitely often», une infinité de fois).

3.3. *Indicatrice d'un évènement.* — Une fonction très utilisée est l'*indicatrice* I_A d'un évènement A . C'est une fonction sur Ω dans le doubleton $\{0, 1\}$ définie par :

$$I_A(\omega) = \begin{cases} 1, & \text{si } \omega \in A; \\ 0, & \text{si } \omega \notin A. \end{cases}$$

COMPLÉMENTS ET EXERCICES

1. — Soient A, B, C trois évènements. Exprimer en fonction de A, B, C et des opérations ensemblistes les évènements ci-après :

- a) A seul se produit ;
- b) A et C se produisent, mais non B ;
- c) les trois évènements se produisent ;
- d) l'un au moins des évènements se produit ;
- e) deux évènements au moins se produisent ;
- f) un évènement au plus se produit ;
- g) aucun des trois évènements ne se produit ;
- h) deux évènements exactement se produisent ;
- i) pas plus de deux évènements ne se produisent.

2. — Soit Ω l'ensemble des couples mariés d'une ville donnée. On considère les évènements :

- A «l'homme a plus de quarante ans» ;
- B «la femme est plus jeune que l'homme» ;
- C «la femme a plus de quarante ans».

a) Interpréter en fonction de A, B, C l'évènement «le mari a plus de quarante ans, mais non sa femme».

b) Décrire en langage ordinaire les évènements $A \cap B \cap C^c, A \setminus (A \cap B), A \cap B^c \cap C, A \cup B$.

c) Vérifier que $A \cap C^c \subset B$.

3. — Étant donné une suite d'évènements (A_i) ($i = 1, 2, \dots$), montrer que l'on a

$$\bigcup_{i=1}^{\infty} A_i = \sum_{n=1}^{\infty} \left(A_n \setminus \bigcup_{i=0}^{n-1} A_i \right),$$

où l'on pose $A_0 = \emptyset$. Ceci signifie que toute réunion peut s'exprimer sous forme de réunion *disjointe*.

4. — Soient A et B deux évènements et (A_n) une suite d'évènements définis par $A_n = A$ ou B suivant que n est pair ou impair. Montrer que

$$\liminf_n A_n = A \cap B \quad \text{et} \quad \limsup_n A_n = A \cup B.$$

5. — Soient A, B avec ou sans indices des évènements. Vérifier les formules concernant les indicatrices :

a) $I_{\Omega} \equiv 1; I_{\emptyset} \equiv 0$;

b) $I_A(\omega) \leq I_B(\omega)$ pour tout ω dans Ω si et seulement si $A \subset B$;

c) $I_{A \cap B} = I_A I_B; I_{A \cup B} = I_A + I_B - I_{A \cap B}$;

d) $I_{A^c} = 1 - I_A; I_{A \setminus B} = I_A(1 - I_B)$.

e) Soient $A_* = \liminf_n A_n$ et $A^* = \limsup_n A_n$. Alors

$$I_{A_*} = \liminf_n I_{A_n} \quad \text{et} \quad I_{A^*} = \limsup_n I_{A_n}.$$

6. — Soient trois évènements E, F, G auxquels on associe les deux évènements

$$A = (E \cup F) \cap G, \quad B = E \cup (F \cap G).$$

a) Parmi les deux évènements A, B , l'un entraîne l'autre; lequel?

b) Trouver une condition nécessaire et suffisante, portant sur E et G , pour que l'on ait $A = B$.

7. — Étant donné deux évènements A, B , on désigne par $A \triangle B$ l'évènement consistant en le fait que l'un exactement des évènements A, B se réalise; cet évènement est appelé la *différence symétrique* des évènements A, B . Donnons-nous trois évènements A, B, C de Ω .

a) Montrer que : $(A \triangle B) \cup (A \triangle B^c) = \Omega$.

b) Trouver une condition nécessaire et suffisante pour que l'on ait :

$$(A \triangle B) \cap (A \triangle C) = A \triangle (B \cup C).$$

CHAPITRE 2
LES ÉVÈNEMENTS

Dans la construction du modèle probabiliste, on se donne un ensemble non vide Ω et on cherche à remplir conjointement les deux conditions suivantes :

- 1) attribuer une probabilité à tout sous-ensemble de Ω ;
- 2) respecter quelques règles de calcul simple, en premier lieu la *règle d'additivité*.

Il se trouve que pour des raisons mathématiques (que l'on peut résumer en disant qu'il existe des sous-ensembles de Ω de nature extrêmement compliquée) on ne peut satisfaire à ces deux exigences à la fois, du moins lorsque Ω a la puissance du continu. L'idée est alors de ne pas attribuer une probabilité à toute partie $A \in \mathfrak{P}(\Omega)$, mais seulement aux parties appartenant à une certaine classe \mathfrak{A} , en général strictement contenue dans $\mathfrak{P}(\Omega)$. Si l'on munit cette classe de quelques propriétés algébriques commodes, on peut alors remplir la seconde condition. Ce sont les propriétés d'*algèbre* et de *tribu* qui se sont avérées les plus efficaces. Leurs axiomes et leurs propriétés élémentaires sont donnés dans le présent chapitre.

Les *systèmes de Dynkin* et les *classes monotones*, dont l'étude est aussi abordée ici, sont surtout des outils techniques.

1. Les algèbres

Définition. — Soient Ω un ensemble fondamental et \mathfrak{A} un sous-ensemble de $\mathfrak{P}(\Omega)$. On dit que \mathfrak{A} est une *algèbre (de Boole)*, si \mathfrak{A} satisfait aux axiomes suivants :

- (A1) $\Omega \in \mathfrak{A}$;
- (A2) $A \in \mathfrak{A}, B \in \mathfrak{A} \Rightarrow A \cup B \in \mathfrak{A}$;
- (A3) $A \in \mathfrak{A} \Rightarrow A^c \in \mathfrak{A}$.

Conséquences

- (A4) $\emptyset \in \mathfrak{A}$;
- (A5) $A \in \mathfrak{A}, B \in \mathfrak{A} \Rightarrow A \cap B \in \mathfrak{A}$;
- (A6) $A_1, A_2, \dots, A_n \in \mathfrak{A} \Rightarrow \bigcup_{i=1}^n A_i \in \mathfrak{A}$ et $\bigcap_{i=1}^n A_i \in \mathfrak{A}$.

Ces trois propriétés sont des conséquences immédiates des trois axiomes. La propriété (A4) découle de (A1) et (A3), la propriété (A5) de l'identité $A \cap B = (A^c \cup B^c)^c$, de (A2) et de (A3). Enfin, (A6) se déduit de (A2) par récurrence sur n .

De façon équivalente, il est commode de dire qu'une classe \mathfrak{A} de sous-ensembles de Ω est une algèbre, si elle contient l'élément Ω et si elle est *stable par réunion finie* et par *passage au complémentaire*.

Un exemple d'algèbre est donné par la classe des réunions finies d'intervalles semi-ouverts de la droite, comme il est décrit de façon précise dans l'exemple suivant.

Exemple. — Soit \mathcal{P} l'ensemble de tous les intervalles semi-ouverts de la droite réelle qui sont de la forme :

$$\begin{aligned} &] - \infty, a' [; \quad [a, b [; \quad [a'', +\infty [; \\ -\infty < a \leq b < +\infty; \quad -\infty < a' \leq +\infty; \quad -\infty < a'' < +\infty. \end{aligned}$$

PROPOSITION 1.1. — *La famille \mathfrak{A} de tous les sous-ensembles de \mathbb{R} qui sont des réunions finies d'intervalles appartenant à \mathcal{P} est une algèbre.*

Pour démontrer cette proposition, on vérifie tout d'abord (et de façon immédiate) les points suivants :

- 1) le complémentaire d'un intervalle de \mathcal{P} appartient à \mathfrak{A} ;
- 2) $\Omega = \mathbb{R} =] - \infty, +\infty [$ et $\emptyset = [a, a [$ appartiennent à \mathfrak{A} ;
- 3) la réunion de deux éléments de \mathfrak{A} appartient à \mathfrak{A} ;
- 4) l'intersection de deux intervalles de \mathcal{P} appartient à \mathfrak{A} .

Il en résulte que si $A = \bigcup_i I_i$ et $B = \bigcup_j J_j$ sont deux éléments de \mathfrak{A} , leur intersection $A \cap B = \bigcup_{i,j} I_i \cap J_j$ est encore un élément de \mathfrak{A} . Par conséquent, le complémentaire $A^c = \bigcap_i I_i^c$ d'un élément $A = \bigcup_i I_i$ de \mathfrak{A} appartient encore à \mathfrak{A} . \square

Remarque. — On note que tout élément A de l'algèbre précédente peut toujours s'exprimer comme réunion finie d'intervalles de \mathcal{P} , *disjoints deux à deux*.

2. Les tribus. — Dans la définition des tribus donnée ci-après, on reprend les trois axiomes des algèbres, mais on modifie seulement le deuxième. On ne traite plus des seules réunions finies, mais des réunions *dénombrables*.

Définition. — Soient Ω un ensemble fondamental et \mathfrak{A} un sous-ensemble de $\mathfrak{P}(\Omega)$. On dit que \mathfrak{A} est une *tribu* (on dit encore *σ -algèbre*, *σ -corps*, *corps de Borel*), si \mathfrak{A} satisfait aux axiomes suivants :

- (T1) $\Omega \in \mathfrak{A}$;
- (T2) si (A_n) ($n = 1, 2, \dots$) est une suite d'éléments appartenant à \mathfrak{A} , alors la réunion $\bigcup_{n=1}^{\infty} A_n$ appartient aussi à \mathfrak{A} ;
- (T3) $A \in \mathfrak{A} \Rightarrow A^c \in \mathfrak{A}$.

On peut encore dire qu'une tribu est une classe de parties de Ω , qui contient Ω et qui est stable par *réunion dénombrable* et par *passage au complémentaire*.

Les deux propriétés suivantes sont des conséquences immédiates de ces trois axiomes (*cf.* les vérifications précédentes pour les propriétés concernant l'algèbre) :

(T4) $\emptyset \in \mathfrak{A}$;

(T5) si (A_n) ($n = 1, 2, \dots$) est une suite d'éléments appartenant à \mathfrak{A} , alors l'intersection $\bigcap_{n=1}^{\infty} A_n$ est aussi dans \mathfrak{A} .

Remarque. — Une tribu \mathfrak{A} est aussi une algèbre. Il suffit, en effet, partant de deux éléments A et B de \mathfrak{A} , de former la suite $A_1 = A$, $A_2 = B$ et $A_n = \emptyset$ pour $n \geq 3$. L'axiome (T2) entraîne que la réunion $A \cup B = \bigcup_{n=1}^{\infty} A_n$ est encore dans \mathfrak{A} . L'axiome (A2) est donc ainsi vérifié.

Exemples. — Pour tout Ω non vide, le doubleton $\{\Omega, \emptyset\}$ et l'ensemble $\mathfrak{P}(\Omega)$ de toutes les parties de Ω sont des tribus. On convient d'associer toujours cette dernière tribu à Ω , lorsque ce dernier ensemble est fini ou dénombrable.

Contrairement au cas des algèbres (*cf.*, par exemple, la Proposition 1.1), il est plus difficile de donner une description explicite de tous les éléments appartenant à une tribu non triviale.

PROPOSITION 2.1 (tribu engendrée par une classe de parties). — *Soit \mathcal{C} un ensemble de parties de Ω . Alors il existe une et une seule tribu $\mathfrak{T}(\mathcal{C})$ ayant les propriétés suivantes :*

(i) $\mathfrak{T}(\mathcal{C}) \supset \mathcal{C}$;

(ii) *si \mathfrak{T}' est une tribu contenant \mathcal{C} , alors elle contient aussi $\mathfrak{T}(\mathcal{C})$.*

La tribu $\mathfrak{T}(\mathcal{C})$ est appelée tribu engendrée par \mathcal{C} .

Démonstration. — Vérifions tout d'abord que toute intersection d'une famille *non vide* de tribus est encore une tribu. En effet, si (\mathfrak{T}_i) est une famille non vide de tribus de Ω , l'ensemble Ω étant dans chacune des tribus est encore dans leur intersection $\bigcap_i \mathfrak{T}_i$. De même, on vérifie que les axiomes (T2) et (T3) étant vrais pour chacune des tribus \mathfrak{T}_i , sont encore satisfaits pour l'intersection.

On remarque ensuite que la famille des tribus qui contiennent \mathcal{C} n'est pas vide, puisque $\mathfrak{P}(\Omega)$ en fait partie. On considère alors l'intersection de toutes les tribus qui contiennent \mathcal{C} . C'est une tribu d'après la première partie de la démonstration. Elle satisfait ensuite aux deux propriétés (i) et (ii) et par construction elle est évidemment unique. \square

Exemples

1) Si la classe \mathcal{C} est elle-même une tribu, alors la tribu engendrée par \mathcal{C} est identique à \mathcal{C} .

2) Soit A un sous-ensemble de Ω , alors la tribu engendrée par la classe $\{A\}$, réduite au seul élément A , est $\{\emptyset, A, A^c, \Omega\}$.

3) Soient A et B deux sous-ensembles disjoints de Ω . La tribu engendrée par le doubleton $\{A, B\}$ se compose des huit sous-ensembles (non nécessairement distincts) : $\emptyset, A, B, A + B, A^c, B^c, A^c \cap B^c, \Omega$.

Définition. — On appelle *tribu borélienne* de la droite réelle \mathbb{R} , la tribu engendrée par la classe des intervalles fermés bornés $\{[a, b] : a \leq b\}$. Cette tribu est notée \mathcal{B}^1 . Ses éléments sont appelés les *boréliens* de la droite.

Comme on peut le vérifier (*cf.* exercice 6), la tribu borélienne peut être engendrée par beaucoup d'autres classes de sous-ensembles de \mathbb{R} .

Définition. — On appelle *tribu borélienne de \mathbb{R}^n* la tribu, notée \mathcal{B}^n , engendrée par les pavés fermés

$$\{(x_1, x_2, \dots, x_n) : a_i \leq x_i \leq b_i, i = 1, 2, \dots, n\}.$$

Définition. — On appelle *espace probabilisable* (ou *espace mesurable*) tout couple (Ω, \mathfrak{A}) formé d'un ensemble non vide Ω et d'une tribu \mathfrak{A} de parties de Ω . Les éléments de \mathfrak{A} sont appelés *événements*.

Exemples

1) Le couple $(\Omega, \mathfrak{P}(\Omega))$ est un espace probabilisable. C'est l'espace probabilisable qu'on associe toujours à Ω , lorsque Ω est au plus dénombrable.

2) Le couple $(\mathbb{R}^n, \mathcal{B}^n)$ est un espace probabilisable.

3. Les systèmes de Dynkin. — Avec les systèmes de Dynkin, on dispose d'un nouvel outil pour démontrer qu'une classe de parties est une tribu. Comme indiqué dans la Proposition 3.1, il suffit de partir d'un système de Dynkin et de s'assurer qu'il est stable par intersection finie. Les systèmes de Dynkin seront essentiellement utilisés dans l'étude de l'indépendance des classes d'événements. Pour une première lecture, on peut se contenter de lire la définition et les énoncés des deux propositions suivantes.

Définition. — Soient Ω un ensemble fondamental et \mathfrak{D} une classe de parties de Ω . On dit que \mathfrak{D} est un *système de Dynkin*, s'il satisfait aux axiomes suivants :

(D1) $\Omega \in \mathfrak{D}$;

(D2) $A \in \mathfrak{D}, B \in \mathfrak{D}, A \supset B \Rightarrow A \setminus B \in \mathfrak{D}$;

(D3) si (A_n) ($n = 1, 2, \dots$) est une suite d'éléments de \mathfrak{D} , *disjoints deux à deux*, alors la réunion (disjointe) $\sum_{n=1}^{\infty} A_n$ est aussi dans \mathfrak{D} .

On peut encore dire qu'un système de Dynkin de Ω est une classe de parties, qui contient l'élément Ω et qui est stable par *différence propre* et par *réunion dénombrable disjointe*.

PROPOSITION 3.1. — *Une tribu est toujours un système de Dynkin. Pour qu'un système de Dynkin \mathfrak{D} soit une tribu, il faut et il suffit qu'il soit aussi stable par intersection finie, c'est-à-dire qu'il satisfasse encore à l'axiome :*

(If) $A \in \mathfrak{D}, B \in \mathfrak{D} \Rightarrow A \cap B \in \mathfrak{D}$.

Démonstration. — La première partie de la proposition est évidente. Il suffit donc de démontrer qu'un système de Dynkin qui est stable par intersection finie est aussi une tribu. Partons d'un tel système \mathfrak{D} . D'abord, les axiomes (T1) et (T3) sont vérifiés, puisqu'en particulier $A^c = \Omega \setminus A$. D'autre part, \mathfrak{D} est stable par réunion finie, puisque si A et B sont dans \mathfrak{D} , l'intersection $A \cap B$ et la différence propre $A \setminus (A \cap B)$ le sont aussi, donc encore la réunion disjointe :

$$A \cup B = (A \setminus (A \cap B)) + B.$$

Si maintenant (A_n) est une suite d'éléments de \mathfrak{D} , toutes les réunions finies $B_n = A_1 \cup \dots \cup A_n$ sont aussi dans \mathfrak{D} . On peut alors écrire, en posant $B_0 = \emptyset$,

$$\bigcup_{n=1}^{\infty} A_n = \sum_{n=1}^{\infty} (A_n \setminus B_{n-1}),$$

ce qui montre bien que la réunion (dénombrable) est encore dans \mathfrak{D} . \square

Comme pour les tribus, on se convainc qu'étant donné une classe de parties \mathcal{C} , il existe un système de Dynkin unique, contenant \mathcal{C} et contenu dans tout système de Dynkin contenant \mathcal{C} . On l'appelle le *système de Dynkin engendré par \mathcal{C}* et on le note $\mathfrak{D}(\mathcal{C})$.

PROPOSITION 3.2. — *Soit \mathcal{C} un ensemble de parties de Ω stable par intersection finie. Alors*

$$\mathfrak{D}(\mathcal{C}) = \mathfrak{T}(\mathcal{C}).$$

Démonstration. — Comme toute tribu est un système de Dynkin, on a déjà l'inclusion $\mathfrak{D}(\mathcal{C}) \subset \mathfrak{T}(\mathcal{C})$. Pour prouver l'inclusion inverse, il suffit de montrer que $\mathfrak{D}(\mathcal{C})$ est aussi une tribu. D'après la proposition précédente, il suffit même de montrer que $\mathfrak{D}(\mathcal{C})$ est lui-même stable par intersection finie.

Prenons un élément A de $\mathfrak{D}(\mathcal{C})$ et formons l'ensemble, noté $\mathfrak{I}(A)$, de toutes les parties B de Ω telles que $B \cap A \in \mathfrak{D}(\mathcal{C})$. La classe $\mathfrak{I}(A)$ est un système de Dynkin, car elle contient Ω et elle est stable par différence propre et par réunion dénombrable disjointe. Or si E est dans \mathcal{C} , on a $F \cap E \in \mathcal{C}$ pour tout $F \in \mathcal{C}$; d'où $\mathcal{C} \subset \mathfrak{I}(E)$ et $\mathfrak{D}(\mathcal{C}) \subset \mathfrak{I}(E)$ pour tout $E \in \mathcal{C}$. La dernière inclusion se traduit encore par : pour tout $A \in \mathfrak{D}(\mathcal{C})$ et tout $E \in \mathcal{C}$, on a $A \cap E \in \mathfrak{D}(\mathcal{C})$. Il en résulte l'inclusion $\mathcal{C} \subset \mathfrak{I}(A)$ et $\mathfrak{D}(\mathcal{C}) \subset \mathfrak{I}(A)$ pour tout $A \in \mathfrak{D}(\mathcal{C})$. Ceci entraîne bien que $\mathfrak{D}(\mathcal{C})$ est stable par intersection finie. \square

4. Les classes monotones. — Ce sont des outils techniques comme les systèmes de Dynkin. Pour la compréhension de cet ouvrage, on peut se contenter de la lecture de la définition et des deux propositions suivantes.

Définition. — Une classe non vide \mathfrak{M} de sous-ensembles de Ω est dite *monotone* si, pour toute suite monotone (A_n) d'éléments de \mathfrak{M} (c'est-à-dire pour toute suite croissante ou décroissante d'éléments de \mathfrak{M}), on a $\lim_n A_n \in \mathfrak{M}$ (on dit encore que \mathfrak{M} est stable par passage à la limite monotone.)

Comme pour les tribus et les systèmes de Dynkin, on vérifie que toute intersection de classes monotones est encore une classe monotone et qu'étant donnée une classe \mathcal{C} de parties de Ω , il existe une et une seule classe monotone contenant \mathcal{C} et contenue dans toute classe monotone contenant \mathcal{C} . On l'appelle *classe monotone engendrée par \mathcal{C}* et on la note $\mathfrak{M}(\mathcal{C})$.

PROPOSITION 4.1. — *Toute tribu est une classe monotone. Toute algèbre monotone est une tribu.*

Démonstration. — La première partie de la proposition est évidente. Pour la seconde partie, considérons une algèbre monotone \mathfrak{A} et une suite (A_n) ($n = 1, 2, \dots$) d'éléments de \mathfrak{A} . Alors toute réunion finie $B_n = \bigcup_{i=1}^n A_i$ est dans \mathfrak{A} , par l'axiome (A2). Comme $B_n \uparrow \bigcup_{i=1}^{\infty} A_i$, on a aussi $\bigcup_{i=1}^{\infty} A_i \in \mathfrak{A}$, puisque \mathfrak{A} est aussi une classe monotone. \square

La proposition suivante joue pour les classes monotones le rôle joué par la Proposition 3.2 pour les systèmes de Dynkin.

PROPOSITION 4.2. — *Si \mathfrak{A} est une algèbre, on a :*

$$\mathfrak{T}(\mathfrak{A}) = \mathfrak{M}(\mathfrak{A}).$$

De là, si une classe monotone contient une algèbre \mathfrak{A} , elle contient aussi la tribu $\mathfrak{T}(\mathfrak{A})$ engendrée par \mathfrak{A} .

Démonstration. — D'après la proposition précédente, il suffit de démontrer que $\mathfrak{M} = \mathfrak{M}(\mathfrak{A})$ est une tribu. Formons pour tout $A \in \mathfrak{M}$ tel que $A^c \in \mathfrak{M}$ la classe $K(A)$ de tous les sous-ensembles B tels que $B^c \in \mathfrak{M}$ et $A \cup B \in \mathfrak{M}$. Une telle classe est non vide, puisque A en fait partie. D'autre part, si B et B^c sont dans \mathfrak{M} , alors les relations $B \in K(A)$ et $A \in K(B)$ sont équivalentes.

Montrons que $K(A)$ est une classe monotone. Pour cela, prenons une suite monotone (B_n) d'éléments de $K(A)$. Alors $(\lim_n B_n)^c = \lim_n B_n^c \in \mathfrak{M}$ et $A \cup \lim_n B_n = \lim_n A \cup B_n \in \mathfrak{M}$.

D'autre part, comme on a l'inclusion $\mathfrak{A} \subset \mathfrak{M}$ et que \mathfrak{A} est une algèbre, le complémentaire A^c est dans \mathfrak{M} , dès que A est dans \mathfrak{A} . Il s'ensuit que la classe $K(\mathfrak{A})$ contient \mathfrak{A} , puisque B^c et $A \cup B$ sont dans \mathfrak{A} , donc dans \mathfrak{M} . Comme \mathfrak{M} est la classe engendrée par \mathfrak{A} , on a donc l'inclusion $\mathfrak{M} \subset K(\mathfrak{A})$.

Pour tout $B \in \mathfrak{M}$, on a $B^c \in \mathfrak{M}$. On peut donc former la classe monotone $K(B)$. Pour tout $A \in \mathfrak{A}$, il vient $B \in K(A)$ et par suite aussi $A \in K(B)$. D'où $\mathfrak{A} \subset K(B)$ et $\mathfrak{M} \subset K(B)$. Cette inclusion, étant vraie pour tout $B \in \mathfrak{M}$, implique que \mathfrak{M} satisfait les axiomes (A2) et (A3). C'est donc une algèbre monotone, donc une tribu. \square

COMPLÉMENTS ET EXERCICES

1. — Comment peut-on introduire la notion d'algèbre engendrée par une classe de parties?

2. — On considère l'ensemble à trois éléments $\Omega = \{a, b, c\}$. Déterminer la tribu engendrée par la partie $\{a, b\}$.

3. — Soit Ω un ensemble de cinq éléments notés a, b, c, d, e . On considère les deux familles $\mathfrak{F}_1, \mathfrak{F}_2$ de parties de Ω : $\mathfrak{F}_1 = \{\emptyset, \{a\}, \{b, c, d, e\}, \Omega\}$ et $\mathfrak{F}_2 = \{\emptyset, \{a\}, \{b\}, \{a, b\}, \{c, d, e\}, \{a, c, d, e\}, \{b, c, d, e\}, \Omega\}$.

- a) Vérifier que \mathfrak{F}_1 et \mathfrak{F}_2 sont des tribus.
- b) Construire l'algèbre (de Boole) \mathfrak{F}_3 engendrée par la classe formée par les deux sous-ensembles $\{a\}$ et $\{c, d\}$.
- c) Vérifier que \mathfrak{F}_3 est une tribu.
- d) Construire la tribu \mathfrak{F}_4 engendrée par $\mathfrak{F}_2 \cup \mathfrak{F}_3$.

4. — Soient Ω un ensemble et $\Pi = (A_n)_{n \geq 1}$ une partition (dénombrable) de Ω , c'est-à-dire une famille de parties de Ω vérifiant :

$$A_n \neq \emptyset \text{ pour tout } n, \quad \bigcup_{n \geq 1} A_n = \Omega, \quad A_i \cap A_j = \emptyset \text{ pour tout } i \neq j.$$

Une tribu \mathfrak{A} sur Ω est dite engendrée par la partition Π si tous les éléments de Π sont des éléments de \mathfrak{A} et si tout élément de \mathfrak{A} est la réunion (finie ou dénombrable) d'éléments de Π , c'est-à-dire si l'on a :

$$\mathfrak{A} = \left\{ \bigcup_{n \in T} A_n : T \in \mathfrak{P}(\mathbb{N}^*) \right\}.$$

- a) Montrer que toute tribu \mathfrak{A} sur un ensemble *dénombrable* Ω est engendrée par une partition du type ci-dessus.
- b) Existe-t-il une tribu qui ait une infinité dénombrable d'éléments ?

5. — Soit A_1, \dots, A_n un système de n ($n \geq 1$) parties d'un ensemble non vide Ω . Donner une description de l'algèbre \mathfrak{A} engendrée par la classe $\{A_1, \dots, A_n\}$. Donner une majoration pour le cardinal de \mathfrak{A} .

6. — Montrer que la tribu borélienne \mathcal{B}^1 de \mathbb{R} est engendrée par l'une des classes suivantes, où a et b sont des réels vérifiant $-\infty < a \leq b < \infty$:

- a) $C_1 = \{]a, b[\}$; b) $C_2 = \{[a, b[\}$; c) $C_3 = \{]a, b \}$;
- d) $C_4 = \{[a, +\infty[\}$; e) $C_5 = \{]-\infty, a[\}$; f) $C_6 = \{]a, +\infty[\}$;
- g) $C_7 = \{]-\infty, a \}$; h) $C_8 = \{ \text{réunions finies d'intervalles semi-ouverts à droite, c'est-à-dire appartenant à } \mathcal{P} \}$ (cf. Proposition 1.1);
- i) $C_9 = \{ \text{ouverts de la droite} \}$; j) $C_{10} = \{ \text{fermés de la droite} \}$; cet inventaire n'est pas exhaustif!

7. — Soit \mathcal{C} la classe des parties d'un ensemble non vide Ω réduites à un seul élément. Montrer que la tribu engendrée par \mathcal{C} est l'ensemble $\mathfrak{P}(\Omega)$ des parties de Ω , si et seulement si Ω est au plus dénombrable.

8. — On a vu que l'intersection de deux tribus est encore une tribu; en revanche, la réunion de deux tribus n'est pas nécessairement une tribu. Fournir un exemple de deux tribus dont la réunion n'est pas une tribu.

ESPACES PROBABILISÉS

On part maintenant d'un espace probabilisable (Ω, \mathfrak{A}) et on cherche à munir les événements, c'est-à-dire les éléments de la tribu \mathfrak{A} , d'une pondération utilisant pleinement les propriétés des tribus. On obtient ainsi un triplet $(\Omega, \mathfrak{A}, P)$, appelé *espace probabilisé*. L'idée de modéliser une expérience aléatoire au moyen d'un tel triplet (non nécessairement unique!) a marqué un tournant décisif dans le développement du calcul des probabilités. Elle est due essentiellement à Kolmogorov.¹

1. Mesures de probabilité

Définition. — Soit (Ω, \mathfrak{A}) un espace probabilisable. On appelle *mesure de probabilité* P sur \mathfrak{A} toute application qui associe à tout événement A un nombre $P(A)$, appelé *probabilité de A* , satisfaisant aux axiomes :

(P1) $0 \leq P(A) \leq 1$ pour tout $A \in \mathfrak{A}$;

(P2) $P(\Omega) = 1$;

(P3) si (A_n) est une suite d'événements de \mathfrak{A} , deux à deux incompatibles (*i.e.*, $A_i \cap A_j = \emptyset$ si $i \neq j$), alors

$$P\left(\sum_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

(P3) s'appelle l'axiome de σ -additivité de P .

Définition. — On appelle *espace probabilisé* tout triplet $(\Omega, \mathfrak{A}, P)$, où (Ω, \mathfrak{A}) est un espace probabilisable et où P est une mesure de probabilité sur \mathfrak{A} (ou, de façon plus précise, sur (Ω, \mathfrak{A})).

Remarque. — Du point de vue de l'analyse, une mesure de probabilité n'est autre qu'une mesure positive bornée, dont la valeur en Ω vaut 1 (voir chap. 10).

Remarque. — Le nombre $P(A)$, la probabilité de l'événement A , intervient dans des relations du type $P(A) = p$. On lit souvent cette expression comme : *la probabilité que A se produise est égale à p* .

¹ Kolmogorov (A.N.). — *Grundbegriffe der Wahrscheinlichkeitsrechnung*. — Berlin, Springer, 1933.

2. Propriétés. — Donnons une liste de propriétés élémentaires sur les mesures de probabilité, simples à établir, mais qui sont d'un usage constant dans l'évaluation effective de probabilités d'évènements. On suppose donné un espace probabilisé $(\Omega, \mathfrak{A}, P)$; les lettres A, B , avec ou sans indices sont des évènements appartenant à \mathfrak{A} .

PROPOSITION 2.1

- 1) $P(\emptyset) = 0$.
- 2) Soient $n \geq 2$ et (A_i) ($i = 1, 2, \dots, n$) une suite de n évènements deux à deux incompatibles. Alors

$$P\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i).$$

- 3) Si A et B sont incompatibles, alors $P(A + B) = P(A) + P(B)$.
- 4) On a : $P(A) + P(A^c) = 1$.
- 5) On a $P(A_1) + \dots + P(A_n) = 1$ pour toute partition (A_1, \dots, A_n) de Ω .
- 6) Si A et B sont deux évènements quelconques, alors : $P(A \cup B) = P(A) + P(B) - P(AB)$ (utilisant la notation $AB = A \cap B$).
- 7) Si A entraîne B , c'est-à-dire, si $A \subset B$, alors

$$P(A) \leq P(B) \quad \text{et} \quad P(B \setminus A) = P(B) - P(A).$$

Démonstration. — Pour 1) prendre la suite (A_n) avec $A_1 = \Omega$ et $A_n = \emptyset$ pour $n \geq 2$ et utiliser la σ -additivité de P . Pour 2) prendre $A_i = \emptyset$ pour $i \geq n + 1$ et appliquer la σ -additivité et la propriété 1) juste démontrée. La propriété 3) est un cas particulier de 2). Pour 4), on note que $A + A^c = \Omega$ et on applique 3) et l'axiome (P2). Pour 5) il suffit d'appliquer 3) et l'axiome (P2).

Dans la propriété 6) on utilise les décompositions : $A \cup B = A + (B \setminus A)$ et $B = AB + (B \setminus A)$. On en tire : $P(A \cup B) = P(A) + P(B \setminus A)$ et $P(B) = P(AB) + P(B \setminus A)$ et l'identité cherchée. Enfin, pour 7) on note que $A \subset B$ entraîne $B = A + (B \setminus A)$, d'où $P(B) = P(A) + P(B \setminus A)$. Par suite, $P(A) \leq P(B)$ et $P(B \setminus A) = P(B) - P(A)$. \square

3. Formule de Poincaré et inégalité de Boole. — La formule de Poincaré s'applique à une suite (A_i) ($i = 1, 2, \dots, n$) d'évènements pour lesquels on connaît *a priori* les probabilités des conjonctions d'évènements $P(A_{i_1} \cdots A_{i_k})$. On peut ainsi évaluer la probabilité de l'évènement $\bigcup_{i=1}^n A_i$ « au moins l'un des A_i se réalise » ou de l'évènement $\bigcap_{i=1}^n A_i^c$ « aucun des A_i ne se réalise ».

PROPOSITION 3.1 (Formule de Poincaré). — Soient $n \geq 2$ et (A_i) ($i = 1, 2, \dots, n$) une suite d'évènements. Alors

$$P(A_1 \cup \dots \cup A_n) = \sum_i P(A_i) - \sum_{i < j} P(A_i A_j) + \dots + (-1)^{n-1} P(A_1 \cdots A_n),$$

ou bien
$$P(A_1 \cup \dots \cup A_n) = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P(A_{i_1} \cdots A_{i_k}),$$

la dernière sommation étant faite sur toutes les suites strictement croissantes (i_1, \dots, i_k) de k entiers extraits de l'intervalle $[1, n]$.

Démonstration. — La formule est déjà vraie pour $n = 2$: c'est la propriété 6) de la Proposition 2.1. Par récurrence sur n on a

$$\begin{aligned} P(A_1 \cup \dots \cup A_n) &= P((A_1 \cup \dots \cup A_{n-1}) \cup A_n) \\ &= P(A_1 \cup \dots \cup A_{n-1}) + P(A_n) \\ &\quad - P((A_1 \cap A_n) \cup \dots \cup (A_{n-1} \cap A_n)) \\ &= \sum_{k=1}^{n-1} (-1)^{k-1} \sum' P(A_{i_1} \dots A_{i_k}) \\ &\quad - \sum_{k=1}^{n-1} (-1)^{k-1} \sum'' P(A_{i_1} \dots A_{i_k} A_n) + P(A_n), \end{aligned}$$

où les sommations \sum' et \sum'' sont faites sur toutes les suites strictement croissantes (i_1, \dots, i_k) de k entiers extraits de l'intervalle $[1, n-1]$. On obtient ainsi une sommation sur toutes les suites de k entiers de $[1, n]$, puisque la première (resp. la seconde) sommation se compose de toutes les suites ne contenant pas n (resp. contenant n à l'exclusion de la suite (n) de longueur 1) et que $P(A_n)$ vient compléter le terme qui manquait. \square

Remarque. — On utilise fréquemment la formule de Poincaré sous la forme

$$P(A_1^c \cap \dots \cap A_n^c) = \sum_{k=0}^n (-1)^k \sum P(A_{i_1} \dots A_{i_k}),$$

avec toujours $1 \leq i_1 < \dots < i_k \leq n$ et $0 \leq k \leq n$ en convenant que le terme de la sommation correspondant à $k = 0$ est égal à 1.

PROPOSITION 3.2 (Inégalité de Boole). — *Pour toute suite d'évènements (A_n) ($n = 1, 2, \dots$), on a :*

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n).$$

Démonstration. — On utilise la décomposition

$$\bigcup_{n=1}^{\infty} A_n = \sum_{n=1}^{\infty} \left(A_n \setminus \bigcup_{i=0}^{n-1} A_i \right) \quad (A_0 = \emptyset).$$

Or la réunion du second membre est disjointe et $A_n \setminus \bigcup_{i=0}^{n-1} A_i \subset A_n$, d'où

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P\left(A_n \setminus \bigcup_{i=0}^{n-1} A_i\right) \leq \sum_{n=1}^{\infty} P(A_n). \quad \square$$

Il faut noter que l'inégalité ne donne quelquefois aucune information, en particulier lorsque la série du second membre est de somme supérieure à 1 ou diverge!

4. Autres propriétés. — La propriété de σ -additivité est en fait une propriété de continuité. Celle-ci doit être convenablement interprétée en termes de suites d'évènements, d'une part, de suites de nombres réels, d'autre part.

PROPOSITION 4.1. — *Soit (A_n) une suite monotone d'évènements. Alors*

$$P(\lim_n A_n) = \lim_n P(A_n).$$

Démonstration. — Il faut noter que les deux opérateurs «lim» s'appliquent dans deux sens différents. Supposons, d'abord, la suite (A_n) croissante. Avec $A_0 = \emptyset$ on a :

$$\begin{aligned} P(\lim_n A_n) &= P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\sum_{n=1}^{\infty} (A_n \setminus A_{n-1})\right) \\ &= \sum_{n=1}^{\infty} P(A_n \setminus A_{n-1}) = \lim_n \sum_{k=1}^n P(A_k \setminus A_{k-1}). \end{aligned}$$

Or

$$\sum_{k=1}^n P(A_k \setminus A_{k-1}) = \sum_{k=1}^n (P(A_k) - P(A_{k-1})) = P(A_n).$$

D'où

$$P(\lim_n A_n) = \lim_n P(A_n).$$

Si la suite (A_n) est décroissante, la suite $(A_1 \setminus A_n)$ ($n = 2, 3, \dots$) est croissante. D'où

$$\begin{aligned} P(A_1) - P(\lim_n A_n) &= P(A_1 \setminus \lim_n A_n) \\ &= P(\lim_n (A_1 \setminus A_n)) = \lim_n P(A_1 \setminus A_n) \\ &= \lim_n (P(A_1) - P(A_n)) = P(A_1) - \lim_n P(A_n). \end{aligned}$$

D'où

$$P(\lim_n A_n) = \lim_n P(A_n). \quad \square$$

Cette proposition admet une réciproque que voici.

PROPOSITION 4.2. — *Soit P une fonction d'ensembles, définie sur la tribu \mathfrak{A} , à valeurs dans l'intervalle fermé $[0, 1]$. On suppose, de plus, qu'elle satisfait aux propriétés 1), 2) et 3) ou 1), 2) et 3') suivantes :*

- 1) $P(\Omega) = 1$;
- 2) *l'additivité finie* : $P(A + B) = P(A) + P(B)$, pour tout couple A, B disjoints, appartenant à \mathfrak{A} ;
- 3) *pour toute suite décroissante (A_n) d'éléments de \mathfrak{A} tendant vers \emptyset , on a* : $\lim_n P(A_n) = 0$;
- 3') *pour tout $A \in \mathfrak{A}$ et toute suite croissante d'éléments tendant vers A , on a* : $\lim_n P(A_n) = P(A)$.

Alors P est une mesure de probabilité sur \mathfrak{A} .

Démonstration. — Soit (A_n) une suite d'éléments de \mathfrak{A} disjoints deux à deux. Il suffit de prouver la σ -additivité. Posons $A = \sum_{n=1}^{\infty} A_n$, $B_n = \sum_{k=1}^n A_k$ et $C_n = A \setminus B_n$. La suite (C_n) est décroissante et tend vers \emptyset . De là, si P satisfait la condition 3), on a $\lim_n P(C_n) = 0$ et donc

$$P(A) = P(B_n) + P(C_n) = \sum_{k=1}^n P(A_k) + P(C_n), \quad \text{pour tout } n.$$

D'où
$$P(A) = \sum_{n=1}^{\infty} P(A_n).$$

De même, la suite (B_n) est croissante et tend vers A . Si donc P satisfait la condition 3'), on a :

$$P(A) = P\left(\lim_n \sum_{k=1}^n A_k\right) = \lim_n \sum_{k=1}^n P(A_k) = \sum_{k=1}^{\infty} P(A_k). \quad \square$$

5. Identités binomiales. — Dans l'évaluation des probabilités de certains évènements, on fait usage de nombreuses identités contenant des coefficients binomiaux. Ces identités se présentent sous des aspects variés, mais bien souvent elles ne sont que de simples conséquences de deux formules classiques bien connues, d'abord l'*identité binomiale*, ensuite l'*identité de Chu-Vandermonde*. Pour établir ces identités, il est commode de se placer dans le contexte des fonctions hypergéométriques. Nous nous proposons de donner ici quelques éléments sur ces fonctions et de décrire quelques techniques de calcul les concernant.

5.1. *Les factorielles montantes.* — Soient a un nombre réel ou complexe et n un entier positif. La *factorielle montante* $(a)_n$ est définie par :

$$(a)_n = \begin{cases} 1, & \text{si } n = 0; \\ a(a+1) \cdots (a+n-1), & \text{si } n \geq 1. \end{cases}$$

Pour prolonger la définition de $(a)_n$ à tout entier $n \in \mathbb{Z}$, on utilise la fonction gamma dont nous rappelons la définition. Pour a complexe telle que $\Re a > 0$ on définit $\Gamma(a) = \int_0^{+\infty} e^{-t} t^{a-1} dt$; pour a complexe différent de tout entier nul ou négatif, on définit $\Gamma(a)$ par

$$\Gamma(a) = \frac{\Gamma(a+n)}{(a)_n},$$

en prenant n entier tel que $n + \Re a > 0$. On peut ainsi prolonger la définition de $(a)_n$ pour tout entier $n \in \mathbb{Z}$ en posant :

$$(a)_n = \frac{\Gamma(a+n)}{\Gamma(a)}.$$

Notons les propriétés suivantes, banales, mais fort utiles dans les calculs :

$$(a)_{i+j} = (a)_i (a+i)_j; \quad (a)_n = (-1)^n (1-n-a)_n.$$

D'autre part, $(-m)_n = 0$ si m, n sont des entiers positifs tels que $n > m$.

Le *coefficient binomial* est défini pour tout nombre complexe a et tout entier $n \geq 0$ par

$$\binom{a}{n} = \frac{a(a-1)\cdots(a-n+1)}{n!},$$

ou, en utilisant la notation des factorielles montantes,

$$\binom{a}{n} = (-1)^n \frac{(-a)_n}{n!}.$$

Notons que la définition du coefficient binomial comme le rapport des produits de factorielles $a!/(n!(a-n)!)$ n'a de sens que si a est lui-même un entier positif.

5.2. *Les fonctions hypergéométriques.* — Soient p, q des entiers positifs et (a_1, \dots, a_p) et (b_1, \dots, b_q) deux suites de nombres réels. Si aucun des b_i n'est un entier négatif ou nul, on définit la *fonction hypergéométrique* avec p et q paramètres, comme étant la série entière en x complexe :

$${}_pF_q \left(\begin{matrix} a_1, \dots, a_p \\ b_1, \dots, b_q \end{matrix}; x \right) = \sum_{n \geq 0} \frac{(a_1)_n \cdots (a_p)_n}{(b_1)_n \cdots (b_q)_n} \frac{x^n}{n!}.$$

Cette série converge pour tout x complexe, lorsque $p \leq q$ et pour $|x| < 1$, lorsque $p = q + 1$. Lorsque $p = 0$ (resp. $q = 0$), on indique l'absence de paramètres par un trait horizontal dans l'expression de F .

On note que si l'un des paramètres a_i du numérateur est un entier négatif ou nul $-m$, la série hypergéométrique est un *polynôme* en x de degré au plus égal à m , tous les termes $(-m)_n$ étant nuls pour $n \geq m + 1$.

La plupart des fonctions élémentaires (cf. exercice 13) s'expriment à l'aide des fonctions hypergéométriques. On ne trouve, en effet, dans leurs développements en série, que des coefficients rationnels. Il est ainsi facile d'obtenir les paramètres des fonctions hypergéométriques correspondantes. On a, par exemple, la série exponentielle

$$\exp x = {}_0F_0 \left(\begin{matrix} - \\ - \end{matrix}; x \right) = \sum_{n \geq 0} \frac{x^n}{n!};$$

et l'identité binomiale

$$(5.2.1) \quad (1-x)^{-a} = {}_1F_0 \left(\begin{matrix} a \\ - \end{matrix}; x \right) = \sum_{n \geq 0} (a)_n \frac{x^n}{n!} \quad (|x| < 1).$$

Pour établir cette dernière, on considère le développement en série $f_a(x) = \sum_{n \geq 0} (a)_n (x^n/n!)$ et l'on cherche des relations, d'une part entre $f_a(x)$ et sa dérivée $f'_a(x)$, d'autre part entre $f_a(x)$ et $f_{a+1}(x)$. On en déduit une équation différentielle simple et, par intégration, on obtient $f_a(x) = (1-x)^{-a}$ (cf. exercice 12).

5.3. *L'identité de Chu-Vandermonde.* — L'identité binomiale donne la sommation de la série ${}_1F_0$. L'identité de Chu-Vandermonde permet de

sommer la série ${}_2F_1$ en $x = 1$, lorsque l'un des paramètres du numérateur est un entier négatif et que la série est en fait un polynôme :

$$(5.2.2) \quad {}_2F_1\left(\begin{matrix} -n, a \\ c \end{matrix}; 1\right) = \frac{(c-a)_n}{(c)_n}, \quad c \notin -\mathbb{N}.$$

Il est curieux d'observer le nombre impressionnant d'identités binomiales qui se ramènent à (5.2.2).

Pour établir (5.2.2), on part de $(1-x)^{-(a+b)} = (1-x)^{-a}(1-x)^{-b}$ et on développe en utilisant l'identité binomiale. En prenant le coefficient de x^n dans chacun des deux membres, on en déduit

$$\frac{(a+b)_n}{n!} = \sum_{0 \leq k \leq n} \frac{(a)_k (b)_{n-k}}{k! (n-k)!} \quad (n \geq 0),$$

d'où l'on tire

$$\begin{aligned} \frac{(a+b)_n}{(b)_n} &= \sum_{0 \leq k \leq n} \frac{(a)_k (n-k+1)_k}{(b+n-k)_k k!} = \sum_{0 \leq k \leq n} \frac{(a)_k (-n)_k}{(1-b-n)_k k!} \\ &= {}_2F_1\left(\begin{matrix} a, -n \\ 1-b-n \end{matrix}; 1\right), \end{aligned}$$

ou encore

$$\frac{(c-a)_n}{(c)_n} = {}_2F_1\left(\begin{matrix} -n, a \\ c \end{matrix}; 1\right), \quad c \notin -\mathbb{N}.$$

5.4. *Une manipulation de l'identité de Chu-Vandermonde.* — Pour établir l'une des identités qui généralisent la formule de Poincaré, par exemple, celle qui permet de calculer la probabilité pour qu'au moins r événements parmi n événements donnés se réalisent, on est amené à prouver l'identité :

$$\sum_{k=0}^{l-r} (-1)^k \binom{r+k-1}{r-1} \binom{l}{r+k} = 1 \quad (r < l).$$

On exprime d'abord les coefficients binomiaux en fonction des factorielles montantes, de sorte que le premier membre devient :

$$\begin{aligned} \sum_{k=0}^{l-r} (-1)^k \frac{(r)_k}{k!} (-1)^{r+k} \frac{(-l)_{r+k}}{(r+k)!} &= (-1)^r \frac{(-l)_r}{r!} \sum_{k=0}^{l-r} \frac{(r)_k}{k!} \frac{(-l+r)_k}{(1+r)_k} \\ &= (-1)^r \frac{(-l)_r}{r!} {}_2F_1\left(\begin{matrix} -(l-r), r \\ r+1 \end{matrix}; 1\right) \\ &= \binom{l}{r} \frac{(1)_{l-r}}{(r+1)_{l-r}} = \frac{l!}{r! (l-r)! (r+1) \cdots l} = 1. \quad \square \end{aligned}$$

COMPLÉMENTS ET EXERCICES

1. — Procédons à deux jets consécutifs d'une pièce de monnaie *parfaite*; le triplet $(\Omega, \mathfrak{A}, P)$ suivant est généralement utilisé pour décrire cette expérience. On prend pour Ω l'ensemble dont les éléments sont les quatre issues possibles de l'expérience : (P, P) , (P, F) , (F, P) , (F, F) , où (P, F) , par exemple, désigne l'issue consistant à amener « pile » au premier jet et « face » au second. Pour \mathfrak{A} on prend $\mathfrak{P}(\Omega)$ et pour P la mesure de probabilité uniforme sur $(\Omega, \mathfrak{P}(\Omega))$, à savoir celle définie par $P(\{\omega\}) = 1/4$ pour tout $\omega \in \Omega$. Soient alors les deux évènements : A « amener “pile” au premier jet »; B « amener “face” au second jet ». Décrire A , B comme éléments de $\mathfrak{P}(\Omega)$; calculer $P(A \cup B)$.

2. Différence entre « lancer une pièce n fois de suite » et « lancer simultanément n pièces ». — Le présent exercice consiste à construire des triplets $(\Omega, \mathfrak{A}, P)$ associés aux deux expériences suivantes :

a) On procède à n jets consécutifs d'une pièce de monnaie parfaite : $\Omega = \{P, F\} \times \dots \times \{P, F\} = \{P, F\}^n$, où $\{P, F\}$ désigne l'ensemble contenant les deux éléments P (« pile ») et F (« face »), $\mathfrak{A} = \mathfrak{P}(\Omega)$, P est l'équirépartition sur Ω . On a $\text{card } \Omega = 2^n$, $\text{card } \mathfrak{A} = 2^{2^n}$, $P(\{\omega\}) = 1/2^n$ pour tout $\omega \in \Omega$.

b) On procède à un jet simultané de n pièces de monnaie *parfaites* et *indiscernables*. On prend $\Omega = \{\omega_0, \omega_1, \dots, \omega_n\}$, où ω_k ($k = 0, 1, \dots, n$) désigne l'épreuve consistant à amener k « pile » sur les n pièces de monnaie lancées; on prend, de plus, $\mathfrak{A} = \mathfrak{P}(\Omega)$. On a $\text{card } \Omega = n + 1$ et $\text{card } \mathfrak{A} = 2^{n+1}$. Comme mesure de probabilité, il est contre-indiqué de prendre l'équirépartition. Un peu de réflexion montre qu'il convient de prendre P telle que, pour $k = 0, 1, \dots, n$, on ait $P(\{\omega_k\}) = \binom{n}{k}/2^n$. L'expérience b) est bien plus pauvre que a). Citer des évènements associés à l'expérience a) qui n'ont plus aucun sens dans b).

3. — On procède à deux jets consécutifs d'un dé *parfait*. Construire un triplet $(\Omega, \mathfrak{A}, P)$ qui décrit cette expérience. On considère ensuite les deux évènements : A « la somme des points amenés par les deux jets est impaire »; B « le point 1 est amené au moins une fois ».

a) Interpréter les évènements : $A \cap B$, $A \cup B$, $A \cap B^c$.

b) Calculer leurs probabilités.

4. — Montrer que la formule de Poincaré (*cf.* Proposition 3.1) est encore vraie en échangeant les signes « \cup » et « \cap »; en d'autres termes, montrer que l'on a :

$$P(A_1 \cap \dots \cap A_n) = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P(A_{i_1} \cup \dots \cup A_{i_k}).$$

5. — Soit $(\Omega, \mathfrak{A}, P)$ un espace probabilisé. La *différence symétrique* entre deux éléments $A, B \in \mathfrak{A}$ est définie par : $A \Delta B = (A \cap B^c) \cup (A^c \cap B)$. Montrer que :

- a) la fonction $d(A, B) = P(A \Delta B)$ est une distance sur \mathfrak{A} ;
- b) $|P(A) - P(B)| \leq P(A \Delta B)$.

6. — Soit (A_n) ($n \geq 1$) une suite d'évènements associés à un espace probabilisé $(\Omega, \mathfrak{A}, P)$. Alors on a les « inégalités de Fatou » :

- (1) $P(\liminf_n A_n) \leq \liminf_n P(A_n)$;
- (2) $\limsup_n P(A_n) \leq P(\limsup_n A_n)$;

Posons $A_* = \liminf_n A_n$, $A^* = \limsup_n A_n$. Si $A_* = A^*$, on dit que la suite (A_n) ($n \geq 1$) est *convergente* et sa limite, notée $\lim_n A_n$, est la valeur commune de A_* et de A^* . Dans ce cas, les deux inégalités (1) et (2) montrent que la suite numérique $(P(A_n))$ ($n \geq 1$) admet une limite (au sens usuel) et que l'on a : $P(\lim_n A_n) = \lim_n P(A_n)$. Construire un exemple où chacune des inégalités de Fatou (1), (2) est une inégalité *stricte*.

7. — Soit Ω un ensemble infini (par exemple \mathbb{N}) et soit \mathfrak{A} la classe des parties de Ω qui sont finies ou cofinies (c'est-à-dire ayant des complémentaires finis).

- a) Montrer que \mathfrak{A} est une *algèbre*, mais non une σ -algèbre.
- b) Montrer que \mathfrak{A} est l'algèbre engendrée par la classe \mathcal{C} des parties de Ω réduites à un seul élément.
- c) On désigne par P l'application de \mathfrak{A} dans \mathbb{R}^+ définie par :

$$P(A) = \begin{cases} 0, & \text{si } A \text{ est fini;} \\ 1, & \text{si } A \text{ est cofini.} \end{cases}$$

Montrer que P est *simplement* additive sur \mathfrak{A} , mais n'est pas σ -additive sur \mathfrak{A} (c'est-à-dire si, pour toute suite (A_n) d'éléments deux à deux disjoints de \mathfrak{A} , on a $\sum_{n \geq 1} A_n \in \mathfrak{A}$, mais non nécessairement $P(\sum_{n \geq 1} A_n) = \sum_{n \geq 1} P(A_n)$.)

8. — Soit Ω un ensemble infini non dénombrable (par exemple \mathbb{R}) et soit \mathfrak{A} la classe des parties de Ω qui sont dénombrables ou codénombrables (c'est-à-dire complémentaires de sous-ensembles dénombrables). Ici « dénombrable » signifie « fini ou infini dénombrable ».

- a) Montrer que \mathfrak{A} est une tribu.
- b) Montrer que \mathfrak{A} est la tribu engendrée par la classe \mathcal{C} des parties de Ω réduites à un seul élément.
- c) On désigne par P l'application de \mathfrak{A} dans \mathbb{R}^+ définie par :

$$P(A) = \begin{cases} 0, & \text{si } A \text{ est dénombrable;} \\ 1, & \text{si } A \text{ est codénombrable.} \end{cases}$$

Montrer que P est une mesure de probabilité sur \mathfrak{A} . [Si « dénombrable » signifie « infini et dénombrable », alors \mathfrak{A} n'est même pas une algèbre !]

9. — Soient (A_1, \dots, A_n) ($n \geq 0$) une suite de parties de Ω et \mathfrak{A} l'algèbre engendrée par la classe $\{A_1, \dots, A_n\}$. Soient, en outre, (c_1, \dots, c_m) une suite de réels et (B_1, \dots, B_m) une suite d'éléments de \mathfrak{A} ($m \geq 0$). On considère l'inégalité

$$(1) \quad \sum_{k=1}^m c_k P(B_k) \geq 0,$$

où P est une mesure de probabilité sur \mathfrak{A} . Alors les deux propriétés suivantes sont équivalentes :

- a) L'inégalité (1) est vraie pour toute mesure de probabilité P sur \mathfrak{A} .
- b) L'inégalité (1) est vraie pour toute P vérifiant $P(A_i) = 0$ ou 1 , pour tout $i = 1, 2, \dots, n$.

10. — On pose $S_0^n = 1$, $S_k^n = \sum_{1 \leq i_1 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k})$ ($1 \leq k \leq n$)

et on désigne par V_n^r (resp. W_n^r , resp. $\text{Imp}(A_i)$) la probabilité pour que r exactement (resp. pour qu'au moins r , resp. pour qu'un nombre impair) des événements A_1, \dots, A_n se réalisent ($0 \leq r \leq n$). A l'aide de l'exercice 9 rétablir la formule de Poincaré, puis démontrer les formules

$$V_n^r = \sum_{k=0}^{n-r} (-1)^k \binom{r+k}{k} S_{r+k}^n; \quad W_n^r = \sum_{k=0}^{n-r} (-1)^k \binom{r+k-1}{k} S_{r+k}^n;$$

$$\text{Imp}(A_i) = \sum_{j=1}^n (-1)^{j-1} 2^{j-1} S_j^n.$$

11. — Soit (Ω, \mathfrak{A}) un espace probabilisable. Considérons une application P de \mathfrak{A} dans \mathbb{R}^+ vérifiant :

- a) $0 \leq P(A) \leq 1$ pour tout $A \in \mathfrak{A}$;
- b) $P(\Omega) = 1$;
- c) pour tout entier $n \geq 1$ et tout n -uplet (A_1, \dots, A_n) d'éléments de \mathfrak{A} deux à deux disjoints, l'identité

$$P(A_1 \cup \dots \cup A_n) = P(A_1) + \dots + P(A_n).$$

(Propriété d'additivité simple.)

Montrer que si (A_n) ($n \geq 1$) est une suite d'éléments deux à deux disjoints, et si $A = \bigcup_{n \geq 1} A_n$, alors $P(A) \geq \sum_{n \geq 1} P(A_n)$. (L'axiome P3 de σ -additivité remplace, dans cette dernière formule, l'inégalité par l'égalité.)

12. — Établir l'identité binomiale (5.2.1) en utilisant les indications données dans le texte.

13. — Après avoir obtenu le développement en série de Taylor, au voisinage de 0, des fonctions $\sin x$, $\cos x$, $\ln(1+x)$, $\text{arctg } x$, $\text{arcsin } x$, exprimer ces développements en série à l'aide des fonctions hypergéométriques.

CHAPITRE 4

PROBABILITÉS DISCRÈTES. DÉNOMBREMENTS

Dans ce chapitre on étudie les mesures de probabilité portées par un ensemble fini ou dénombrable. Lorsque l'ensemble a des propriétés géométriques ou algébriques supplémentaires, on exploite celles-ci pour en déduire des évaluations précises de probabilités d'évènements. Les techniques d'évaluation relèvent souvent de l'analyse combinatoire. Il nous a donc paru utile de donner un exposé détaillé des dénombrements classiques.

1. Mesures de probabilité discrètes. — Soient (Ω, \mathfrak{A}) un espace probabilisable et ω_0 un élément de Ω ; on appelle *mesure de probabilité singulière* en ω_0 la mesure de probabilité, notée ε_{ω_0} , qui, à tout évènement A , fait correspondre la valeur :

$$\varepsilon_{\omega_0}(A) = \begin{cases} 1, & \text{si } \omega_0 \in A; \\ 0, & \text{si } \omega_0 \notin A. \end{cases}$$

On dit encore que la *masse unité est concentrée en ω_0* . La mesure ε_{ω_0} est aussi appelée *mesure de Dirac* en ω_0 .

Définition. — Soit $((\alpha_n, \omega_n))$ ($n = 1, 2, \dots$) une suite infinie d'éléments de $\mathbb{R} \times \Omega$ telle que

(i) $\alpha_n \geq 0$ pour tout $n = 1, 2, \dots$;

(ii) $\sum_{n=1}^{\infty} \alpha_n = 1$.

L'application qui à tout évènement A fait correspondre la valeur

$$P(A) = \sum_{n: \omega_n \in A} \alpha_n,$$

est appelée *mesure de probabilité discrète* (portée par les éléments ω_n pondérés par les poids α_n).

La sommation précédente est faite sur un ensemble au plus dénombrable et les nombres α_n sont positifs. Il n'y a donc aucune ambiguïté dans cette définition. Il est commode de noter cette mesure discrète : $P = \sum_n \alpha_n \varepsilon_{\omega_n}$.

Remarque. — Toute mesure de probabilité sur un espace Ω fini ou dénombrable est discrète. Cependant on peut toujours munir un espace (Ω, \mathfrak{A}) , où Ω a la puissance du continu, d'une mesure de probabilité discrète.

Par exemple, si λ est un réel strictement positif, on peut munir $(\Omega, \mathfrak{A}) = (\mathbb{R}, \mathcal{B}^1)$ de la mesure de probabilité discrète :

$$\pi_\lambda = \sum_{n=0}^{\infty} \frac{e^{-\lambda} \lambda^n}{n!} \varepsilon_n,$$

puisque $e^{-\lambda} \lambda^n / n! \geq 0$ et $\sum_{n=0}^{\infty} e^{-\lambda} \lambda^n / n! = 1$. Cette mesure de probabilité est appelée *loi de Poisson de paramètre λ* . Ce sont les entiers positifs qui portent toute la masse. On dit que \mathbb{N} est le *support* de la mesure.

2. Équirépartition sur les espaces finis. — Soient N un entier strictement positif et $\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}$ un ensemble fini. L'*équirépartition sur Ω* est définie comme la mesure de probabilité :

$$P = \sum_{n=1}^N \frac{1}{N} \varepsilon_{\omega_n} = \frac{1}{N} \sum_{n=1}^N \varepsilon_{\omega_n}.$$

En particulier, $P(\{\omega_n\}) = 1/N$ pour tout $n = 1, 2, \dots, N$. En désignant par $\text{card } A$ le cardinal de l'ensemble A , on a, de plus, la formule :

$$P(A) = \frac{1}{N} \sum_{n=1}^N \varepsilon_{\omega_n}(A) = \frac{\sum_{\omega_n \in A} 1}{\sum_{\omega_n \in \Omega} 1},$$

d'où

$$P(A) = \frac{\text{card } A}{\text{card } \Omega}.$$

On retrouve la définition de Laplace de la probabilité, à savoir que la probabilité de A est égale au *nombre de cas « favorables »* (à A) *divisé par le nombre de cas possibles*. Il faut bien noter que cette définition ne s'applique qu'aux espaces *finis*, sur lesquels on a imposé l'*équirépartition*.

Si on se trouve dans ce cas, l'évaluation de la probabilité d'un évènement se ramène à des évaluations de cardinaux d'ensembles finis. Il importe donc de passer en revue l'algèbre combinatoire de ces ensembles.

3. Ensembles finis. — Pour définir la notion d'ensemble fini, on se réfère à l'ensemble $\mathbb{N}^* = \mathbb{N} \setminus \{0\} = \{1, 2, \dots\}$ des entiers naturels. Les sous-ensembles de \mathbb{N}^* qui interviennent constamment par la suite sont les intervalles $\{1, 2, \dots, n\}$ et $\{m+1, m+2, \dots, n\}$, que l'on note respectivement $[n]$ et $[m+1, n]$. Par convention, $[n] = \emptyset$ si $n = 0$. Rappelons les propriétés suivantes de l'ensemble des entiers naturels.

PROPRIÉTÉS 3.1. — Soient n et p deux entiers strictement positifs.

(i) Il existe une bijection de l'intervalle $[n]$ sur l'intervalle $[p]$, si et seulement si $n = p$.

(ii) Pour tout sous-ensemble non vide A de $[n]$, il existe une bijection de A sur un intervalle $[p]$ tel que $p \leq n$.

(iii) Il existe une bijection de $[p]$ sur l'intervalle $[n+1, n+p]$.

On dit qu'un ensemble non vide A est *fini*, s'il existe une bijection d'un intervalle $[n]$ de \mathbb{N}^* sur A . Une telle bijection est appelée *numérotation* de A .

Si A est un ensemble fini, il ne peut être mis en bijection, d'après la propriété (i), qu'avec un seul intervalle de la forme $[n]$. Cet entier est donc déterminé de façon unique. On l'appelle le *cardinal* de A (ou la *cardinalité* de A). On le note $\text{card } A$ ou $|A|$ (si aucune confusion n'est à craindre). On dit encore que A *contient* n *éléments*, ou que *le nombre des éléments de A est n* . On convient que l'ensemble vide est fini et l'on pose $|\emptyset| = 0$. La proposition suivante est une conséquence immédiate de la définition du cardinal et des Propriétés 3.1.

PROPOSITION 3.2

- (i) *Toute partie d'un ensemble fini est finie.*
- (ii) *Deux ensembles finis A et B ont même cardinal, si et seulement s'il existe une bijection de A sur B .*

Si l'on sait qu'un ensemble A est fini, on ne sait pas pour autant le mettre explicitement en correspondance biunivoque avec un ensemble $[n]$. La construction de ces correspondances repose sur les propriétés algébriques ou géométriques que peuvent avoir ces ensembles. Les formules de la somme et du produit, données ci-dessous, sont fondamentales. Ce sont, en fait, les seules véritables formules de dénombrement.

PROPOSITION 3.3 (formule de la somme). — *Si A et B sont deux ensembles finis, disjoints, leur réunion $A + B$ est finie et l'on a :*

$$(3.1) \quad |A + B| = |A| + |B|.$$

Démonstration. — Par hypothèse, on a deux bijections $\varphi : [n] \rightarrow A$ et $\psi : [p] \rightarrow B$. On définit une bijection $\theta : [n+p] \rightarrow A+B$ de la façon suivante. La restriction de θ à $[n]$ est φ . Ensuite, la restriction de θ à $[n+1, n+p]$ est le produit de composition d'une bijection de $[n+1, n+p]$ sur $[p]$ (cf. Propriété 3.1 (iii)) avec ψ . \square

Par récurrence sur k , si A_1, A_2, \dots, A_k sont des ensembles finis et disjoints deux à deux, alors

$$(3.2) \quad |A_1 + \dots + A_k| = |A_1| + \dots + |A_k| \quad (k \geq 1).$$

Lorsque A et B sont finis, mais non nécessairement disjoints, on a la formule dite des « quatre cardinaux » :

$$(3.3) \quad |A \cup B| + |A \cap B| = |A| + |B|.$$

On retrouve une formule déjà obtenue pour les probabilités et qui se démontre de la même façon.

La formule pour les cardinaux qui correspond à la formule de Poincaré pour les probabilités porte le nom de formule du *principe d'inclusion-exclusion*. Elle a la même allure et se démontre, évidemment, de la même façon. En fait, si P désigne l'équipartition sur un ensemble fini Ω et si les ensembles considérés sont des sous-ensembles de Ω , on a $|A| = P(A)|\Omega|$. Il n'y a donc pas lieu de redémontrer cette formule. Citons-la pour référence.

Soient $n \geq 2$ et A_1, A_2, \dots, A_n des ensembles quelconques, éventuellement vides, mais finis. Alors

$$|A_1 \cup \dots \cup A_n| = \sum_i |A_i| - \sum_{i < j} |A_i \cap A_j| + \dots + (-1)^{n-1} |A_1 \cap \dots \cap A_n|,$$

ou bien

$$(3.4) \quad |A_1 \cup \dots \cup A_n| = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} |A_{i_1} \cap \dots \cap A_{i_k}|.$$

Remarque. — Soient A et B deux ensembles finis de cardinal n et p , respectivement. La formule de la somme peut être reformulée de façon intuitive en une *règle de la somme* : «si un objet a peut être choisi de n façons et un objet b de p autres façons, il y a $(n+p)$ façons de choisir soit a , soit b ».

Cet énoncé conserve une légère ambiguïté et c'est justement le langage de la théorie des ensembles qui permet de lever celle-ci. Dans la formule suivante, on voit apparaître le produit cartésien $A \times B$ constitué par l'ensemble de tous les couples (a, b) où a (resp. b) appartient à A (resp. à B).

PROPOSITION 3.4 (formule du produit). — *Si A et B sont deux ensembles finis (non nécessairement disjoints), alors le produit cartésien $A \times B$ est fini et l'on a :*

$$(3.5) \quad |A \times B| = |A| \cdot |B|.$$

Démonstration. — En conservant les mêmes notations que dans la Proposition 3.3, on construit une bijection θ de $[np] = \{1, 2, \dots, np\}$ sur $A \times B$ de la façon suivante. D'abord $[np] = \bigcup_{1 \leq k \leq p} [n(k-1) + 1, nk]$. Il existe ensuite une bijection de $[n(k-1) + 1, nk]$ sur $[n]$ et une bijection de $[n]$ sur le sous-ensemble $A_k = \{(\varphi(1), \psi(k)), (\varphi(2), \psi(k)), \dots, (\varphi(n), \psi(k))\}$ de $A \times B$. Soit θ_k le produit de composition de ces deux bijections ($1 \leq k \leq p$). Comme les ensembles A_k sont disjoints deux à deux et de réunion $A \times B$, on définit θ comme étant l'application dont la restriction à $[n(k-1) + 1, nk]$ est θ_k . La formule (3.2) donne alors :

$$|A \times B| = |A_1| + \dots + |A_p| = np = |A| \cdot |B|. \quad \square$$

Remarque. — De même, la *règle du produit* s'énonce : «si un objet a peut être choisi de n façons et qu'ensuite l'objet b peut être choisi de p façons, la paire (a, b) , prise dans cet ordre, peut être choisie de np façons».

La formule du produit se prolonge au cas de n ($n \geq 2$) ensembles finis. On obtient pour toute suite B_1, B_2, \dots, B_n d'ensembles finis la formule :

$$(3.7) \quad |B_1 \times B_2 \times \dots \times B_n| = |B_1| \times |B_2| \times \dots \times |B_n|.$$

Autrement dit, si B_i a pour cardinal p_i pour $i = 1, 2, \dots, n$, le nombre de suites ordonnées (b_1, b_2, \dots, b_n) , où chaque b_i appartient à B_i ($i = 1, 2, \dots, n$) est égal à $p_1 p_2 \dots p_n$.

Exemple. — Si on lance une pièce de monnaie et un dé à six faces, l'ensemble Ω que l'on doit associer à cette expérience suivant les principes du chapitre 1 est le produit cartésien $A \times B$, où $A = \{\text{pile, face}\}$ et $B = \{1, 2, 3, 4, 5, 6\}$. Il est de cardinal $2 \times 6 = 12$.

4. Formules classiques de dénombrement. — Dans ce paragraphe figure une liste d'ensembles finis pour lesquels on connaît des cardinaux explicites.

4.1. *Suites quelconques de longueur n .* — Dans la formule (3.7), prenons tous les ensembles B_i ($i = 1, 2, \dots, n$) égaux au même ensemble B . On obtient :

$$(4.1.1) \quad |B^n| = |B|^n.$$

Ainsi l'ensemble de toutes les suites (b_1, b_2, \dots, b_n) de longueur n , où chaque b_i est pris dans B a pour cardinal $|B|^n$. Si $|B| = p$, de telles suites étaient appelées dans le langage traditionnel «arrangements avec répétitions de p objets pris n à n », deux des éléments b_i et b_j pouvant être égaux pour $i \neq j$. Dans le langage fonctionnel, on peut encore dire que l'ensemble B^A de toutes les applications d'un ensemble A , de cardinal n , dans un ensemble B , de cardinal p , a pour cardinal p^n .

Exemple. — Le nombre de mots possibles (non nécessairement prononçables) de cinq lettres est égal à $26^5 = 11.881.376$, et non pas 5^{26} .

Exemple. — Supposons que l'on ait r boules numérotées $1, 2, \dots, r$ que l'on répartit «au hasard» dans n urnes numérotées $1, 2, \dots, n$ et que l'on s'intéresse aux différentes répartition de ces r boules. Les urnes et boules étant différenciées (ou discernables), l'ensemble des répartitions peut être assimilé à l'ensemble de toutes les suites (x_1, x_2, \dots, x_r) , où x_i est le numéro de l'urne qu'occupe la boule i ($1 \leq i \leq r$). Il y a donc n^r telles répartitions.

4.2. *Ensembles des parties d'un ensemble.* — Supposons que les p éléments d'un ensemble B de cardinal p soient numérotés b_1, b_2, \dots, b_p . Toute partie A de B est complètement caractérisée par la donnée d'une suite (x_1, x_2, \dots, x_p) où pour $1 \leq i \leq p$ chaque x_i est égal à 1 ou à 0, suivant que b_i appartient ou non à A .

L'application qui envoie toute partie A de B sur la suite associée (x_1, x_2, \dots, x_p) est donc bijective. Par conséquent l'ensemble $\mathfrak{P}(B)$ de toutes les parties de B a même cardinal que l'ensemble $\{0, 1\}^p$ de toutes les suites (x_1, x_2, \dots, x_p) , de longueur p , où chaque x_i est égal à 0 ou à 1. De là :

$$(4.2.1) \quad |\mathfrak{P}(B)| = |\{0, 1\}^p| = 2^p = 2^{|B|}.$$

Exemple. — Soit B un groupe de sept individus. Le nombre de comités que l'on peut former à partir de ces sept personnes, y compris le comité vide et les comités réduits à une seule personne, est égal à $2^7 = 128$.

4.3. *Suites de termes distincts.* — Supposons toujours B de cardinal $p \geq 1$. Une suite (c_1, c_2, \dots, c_n) est dite (n, p) -*injective*, si elle est de longueur n , si tous ses termes c_i sont pris dans B (de cardinal p) et si tous les c_i sont *distincts*. (Traditionnellement, une telle suite est appelée « arrangement sans répétition de p éléments pris n à n ».) Soit $\mathfrak{I}(n, p)$ l'ensemble de toutes les suites (n, p) -injectives et soit $I(n, p)$ son cardinal. Naturellement, $\mathfrak{I}(n, p)$ est vide si $p < n$. Si $n = p$, les suites (p, p) -injectives sont les *numérotations* de l'ensemble B . On dit encore les *permutations* de B .

PROPOSITION 4.3.1. — Si $0 \leq n \leq p$, le nombre $I(n, p)$ de suites (n, p) -injectives est donné par :

$$(4.3.1) \quad I(n, p) = \frac{p!}{(p-n)!} = p(p-1) \cdots (p-n+1).$$

En particulier, le nombre de permutations d'un ensemble de cardinal p est égal à :

$$(4.3.2) \quad I(p, p) = p!$$

Démonstration. — Soit $c = (c_1, c_2, \dots, c_p)$ une numérotation de B . La suite initiale $c' = (c_1, c_2, \dots, c_n)$ de cette suite c est une suite (n, p) -injective. Notons $\mathfrak{I}(c')$ l'ensemble des numérotations (d_1, d_2, \dots, d_p) de B telles que $(d_1, d_2, \dots, d_n) = (c_1, c_2, \dots, c_n) = c'$. L'ensemble $\mathfrak{I}(p, p)$ est la réunion de tous les $\mathfrak{I}(c')$, où c' varie dans $\mathfrak{I}(n, p)$.

Il est clair que les ensembles $\mathfrak{I}(c')$ sont deux à deux disjoints, d'où

$$(4.3.3) \quad \mathfrak{I}(p, p) = \sum_{c'} \mathfrak{I}(c') \quad (c' \in \mathfrak{I}(n, p)).$$

D'autre part, pour construire tous les éléments de $\mathfrak{I}(c')$, il suffit de se donner toutes les suites $c'' = (d_{n+1}, d_{n+2}, \dots, d_p)$ de longueur $(p-n)$, d'éléments distincts appartenant à $B' = B \setminus \{c_1, c_2, \dots, c_n\}$ et les juxtaposer à (c_1, c_2, \dots, c_n) . Or B' a pour cardinal $(p-n)$. Par conséquent, $|\mathfrak{I}(c')| = I(p-n, p-n)$. Les ensembles $\mathfrak{I}(c')$ ont donc même cardinal. La formule (4.3.3) donne alors :

$$I(p, p) = I(n, p)I(p-n, p-n).$$

Comme $I(1, p) = p$ et $I(0, p) = 1$, il vient $I(p, p) = pI(p-1, p-1)$. D'où $I(p, p) = p!$ et $I(n, p) = p!/(p-n)!$ \square

Remarque. — Dans le langage fonctionnel, $p(p-1) \cdots (p-n+1)$ est le cardinal de l'ensemble des *injections* d'un ensemble de cardinal n dans un ensemble de cardinal p .

Exemple. — Vingt-cinq chevaux participent à la course du tiercé. Un tiercé est une suite $(n=3, p=25)$ -injective. Il y a donc $25 \times 24 \times 23 = 13.800$ tiercés possibles.

Exemple. — Les anagrammes du mot TRaine, dont les six lettres sont distinctes, sont les permutations d'un ensemble de six éléments. Le nombre d'anagrammes est donc égal à : $6! = 120$.

Les anniversaires. — Parmi n personnes en présence, quelle est la probabilité pour qu'au moins deux d'entre elles aient leur anniversaire le même jour ? Pour répondre à ce problème, on convient de ne pas prendre en charge les personnes nées un 29 février. On convient ensuite que quel que soit le jour imposé, à l'exclusion du 29 février, la probabilité pour qu'une personne prise au hasard ait son anniversaire ce jour-là est égale à $1/365$. Pour formaliser le problème, on prend pour Ω l'ensemble des suites (x_1, \dots, x_n) où chaque x_i varie dans l'ensemble des 365 jours de l'année considérés et l'équirépartition sur Ω .

La probabilité cherchée est égale à :

$$P_n = 1 - \frac{I(n, 365)}{365^n} = 1 - \frac{365 \times 364 \times \dots \times (365 - n + 1)}{365^n}$$

Quelques valeurs de P_n sont consignées dans le tableau suivant :

n	2	10	15	22	23	32	35	41	55
P_n	0,003	0,12	0,25	0,48	0,51	0,75	0,81	0,90	0,99

Par exemple, dans une classe de vingt-trois élèves, un professeur peut parier, avec plus de cinquante chances de succès sur cent, qu'au moins deux élèves ont leur anniversaire le même jour (*cf.* Fig. 1).

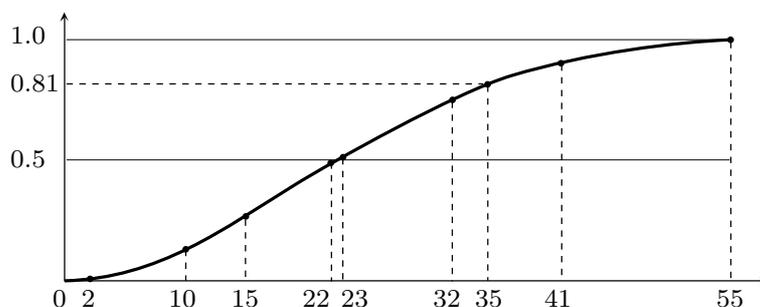


Fig. 1

4.4. *Parties d'un ensemble.* — Les coefficients binomiaux $\binom{p}{n}$ ($n \geq 0$, $p \geq 0$) sont définis par la relation de récurrence :

$$(4.4.1) \quad \binom{p}{n} = \binom{p-1}{n-1} + \binom{p-1}{n} \quad (n, p \geq 1),$$

avec les conditions initiales : $\binom{p}{0} = 1$ pour tout $p \geq 0$ et $\binom{0}{n} = 0$ pour tout $n \geq 1$. Leur représentation dans un tableau avec p comme indice de ligne et n comme indice de colonne n'est autre que le *triangle de Pascal* :

n	0	1	2	3	4	5	6
p							
0	1						
1	1 1						
2	1 2 1						
3	1 3 3 1						
4	1 4 6 4 1						
5	1 5 10 10 5 1						
6	1 6 15 20 15 6 1						

A partir de la récurrence précédente, on obtient la valeur de $\binom{p}{n}$ sous la forme bien connue :

$$(4.4.2) \quad \binom{p}{n} = \frac{p!}{n!(p-n)!} \quad (0 \leq n \leq p)$$

et $\binom{p}{n} = 0$ si n n'appartient pas à l'intervalle $[0, p]$.

PROPOSITION 4.4.1. — Soit $0 \leq n \leq p$; le nombre de parties de cardinal n d'un ensemble de cardinal p est égal à $\binom{p}{n}$.

Démonstration. — Soit B de cardinal p . Pour obtenir une suite (n, p) -injective (c_1, c_2, \dots, c_n) , il suffit de se donner la partie $\{c_1, c_2, \dots, c_n\}$ de B , puis une permutation de ces n éléments. Par conséquent, avec $b_{n,p}$ désignant le nombre de parties de B de cardinal n , on obtient $I(n, p) = b_{n,p}I(n, n)$, soit $b_{n,p} = (p!/(p-n)!)/n! = \binom{p}{n}$. \square

Exemple. — Combien y a-t-il de tirages de cinq cartes d'un jeu de cinquante-deux? Si B est l'ensemble de toutes les cartes, un tirage de cinq cartes n'est autre qu'une partie de B , de cardinal 5. Par conséquent, le nombre cherché est égal à $\binom{52}{5} = \frac{52 \times 51 \times 50 \times 49 \times 48}{5 \times 4 \times 3 \times 2 \times 1} = 2.395.120$.

Exemple. — Combien y a-t-il de mains de cinq cartes d'un jeu de bridge contenant exactement deux as, deux rois et une dame? Soient A (resp. B) l'ensemble de toutes les paires (non ordonnées) de deux as (resp. deux rois) et C l'ensemble des dames. Tout tirage de cinq cartes avec la distribution ci-dessus est une suite (a, b, c) , où $a \in A$, $b \in B$ et $c \in C$. D'après la formule du produit, le nombre de tirages demandé est égal à $t = |A| \times |B| \times |C|$. Comme $|A| = |B| = \binom{4}{2} = 6$ et $|C| = 4$, on trouve $t = 6 \times 6 \times 4 = 144$.

Supposons donné un ordre total $b_1 < b_2 < \dots < b_p$ sur les p éléments de l'ensemble fini B . Lorsqu'on prend une partie A de B , de cardinal n ($n \leq p$), on peut lui faire correspondre la suite croissante $(c_1 < c_2 < \dots < c_n)$ de ses n éléments, et ceci de façon bijective. On peut donc énoncer le corollaire suivant.

COROLLAIRE. — Le nombre de suites strictement croissantes $(c_1 < c_2 < \dots < c_n)$, de longueur n , où les termes sont extraits d'un ensemble de cardinal p , est égal à $\binom{p}{n}$.

4.5. *Suites croissantes.* — Dans la proposition suivante, on construit explicitement une bijection entre un ensemble de suites croissantes (au sens large) et un ensemble de suites strictement croissantes. On applique enfin le corollaire précédent.

PROPOSITION 4.5.1. — *Soient n et p deux entiers positifs et B un ensemble totalement ordonné de cardinal p . Alors le nombre de suites croissantes $(c_1 \leq c_2 \leq \dots \leq c_n)$ (au sens large), de longueur n , dont les c_i sont extraits de B est égal à $\binom{p+n-1}{n}$.*

Démonstration. — Sans restreindre la généralité, on peut prendre B comme étant l'intervalle $[p]$. Soit $(c_1 \leq c_2 \leq \dots \leq c_n)$ une telle suite croissante. On lui fait correspondre la suite $d = (d_1 < d_2 < \dots < d_n)$, définie par :

$$d_1 = c_1; \quad d_2 = c_2 + 1; \quad d_3 = c_3 + 2; \quad \dots; \quad d_n = c_n + n - 1.$$

La suite d ainsi formée est naturellement strictement croissante. De plus,

$$1 \leq d_1 < d_2 < \dots < d_n \leq p + n - 1.$$

L'application $c \mapsto d$ envoie bijectivement l'ensemble des suites croissantes décrites dans l'énoncé de la proposition sur l'ensemble des suites strictement croissantes, de longueur n , dont les termes sont pris dans l'intervalle $[p + n - 1]$. Comme le cardinal de l'ensemble de ces dernières suites est égal à $\binom{p+n-1}{n}$ d'après le précédent corollaire, ce nombre est aussi le nombre des suites croissantes, de longueur n , dont les termes sont pris entre 1 et p . \square

On trouve encore un coefficient binomial dans le dénombrement suivant, mais les paramètres sont différents.

PROPOSITION 4.5.2. — *Le nombre de suites (x_1, x_2, \dots, x_n) qui sont solutions en nombres entiers positifs de l'équation*

$$(4.5.1) \quad x_1 + x_2 + \dots + x_n = p \quad (n \text{ et } p \text{ fixés})$$

est égal au coefficient binomial $\binom{p+n-1}{p}$.

Démonstration. — Soit $x = (x_1, x_2, \dots, x_n)$ une telle solution. Associons-lui la suite croissante $y = (y_1, y_2, \dots, y_n)$ définie par

$$y_1 = 1 + x_1; \quad y_2 = 1 + x_1 + x_2; \quad \dots; \quad y_{n-1} = 1 + x_1 + \dots + x_{n-1}; \\ y_n = 1 + x_1 + \dots + x_{n-1} + x_n = 1 + p.$$

On a : $1 \leq y_1 \leq y_2 \leq \dots \leq y_{n-1} \leq 1 + p$. Réciproquement, si y est une telle suite, on définit $x = (x_1, x_2, \dots, x_n)$ par :

$$x_n = 1 + p - y_{n-1}; \quad x_{n-1} = y_{n-1} - y_{n-2}; \quad \dots; \quad x_2 = y_2 - y_1; \quad x_1 = y_1 - 1.$$

Il y a donc bijection entre les solutions x de l'équation (4.5.1) et les suites croissantes, au sens large, de longueur $(n - 1)$, dont les termes sont pris dans $[1 + p]$. Le cardinal de l'ensemble des solutions x est donc égal à

$$\binom{(1+p) + (n-1) - 1}{n-1} = \binom{p+n-1}{n-1} = \binom{p+n-1}{p}. \quad \square$$

4.6. *Coefficients multinomiaux.* — Supposons donnés deux entiers p et k tels que $1 \leq k \leq p$, ainsi qu'une suite d'entiers (n_1, n_2, \dots, n_k) satisfaisant à :

$$(4.6.1) \quad n_1 \geq 0, n_2 \geq 0, \dots, n_k \geq 0 \quad \text{et} \quad n_1 + n_2 + \dots + n_k = p.$$

Un *coefficient multinomial* est un nombre de la forme : $\frac{p!}{n_1! n_2! \dots n_k!}$. On le note : $\binom{p}{n_1, n_2, \dots, n_k}$. Lorsque $k = 2$, on a $n_1 + n_2 = p$ et on retrouve naturellement le coefficient binomial $\binom{p}{n_1} = \frac{p!}{n_1! (p - n_1)!}$.

PROPOSITION 4.6.1. — *Le nombre de suites de longueur p , contenant n_1 fois 1, n_2 fois 2, \dots , n_k fois k , les n_i satisfaisant les relations (4.6.1), est égal au coefficient multinomial $\binom{p}{n_1, n_2, \dots, n_k}$.*

Démonstration. — Notons $C(n_1, n_2, \dots, n_k)$ l'ensemble des suites contenant exactement n_1 fois 1, \dots , n_k fois k , puis considérons la suite

$$a = (1_1, 1_2, \dots, 1_{n_1}, 2_1, 2_2, \dots, 2_{n_2}, \dots, k_1, k_2, \dots, k_{n_k}),$$

de longueur $n_1 + n_2 + \dots + n_k = p$ et désignons par A l'ensemble des $p!$ réarrangements (permutations) de a .

Prenons un réarrangement b de la suite a et lisons les termes de ce réarrangement b de la gauche vers la droite en écrivant d'abord les indices des lettres 1. On obtient une permutation $\sigma_1 = (i_1, i_2, \dots, i_{n_1})$, de longueur n_1 . De même, la lecture des indices des lettres 2, de la gauche vers la droite, fournit une permutation $\sigma_2 = (j_1, j_2, \dots, j_{n_2})$, de longueur n_2 , et ainsi de suite. ... Prenons note de ces k permutations $\sigma_1, \sigma_2, \dots, \sigma_k$ et effaçons tous les indices dans la suite b . On obtient une suite c de l'ensemble $C(n_1, \dots, n_k)$. Il est clair que l'application qui envoie b sur $(c; \sigma_1, \sigma_2, \dots, \sigma_k)$ est bijective. En fait, $(c; \sigma_1, \sigma_2, \dots, \sigma_k)$ est un simple codage de la suite b . Or le nombre des suites $(c; \sigma_1, \sigma_2, \dots, \sigma_k)$ est égal à $|C(n_1, \dots, n_k)| n_1! n_2! \dots n_k!$. Comme $|A| = p!$, on obtient bien la formule annoncée. \square

Exemple. — Le nombre d'anagrammes du mot VASSAL, mot qui contient deux lettres A, deux lettres S, une lettre V et une lettre L, est égal à $\binom{6}{2, 2, 1, 1} = 6! / (2! 2! 1! 1!) = 180$. Le nombre d'anagrammes du mot BERLIET est égal à $7! / 2! = 2.520$. Parmi eux, se trouve le mot LIBERTÉ.

La formule binomiale admet une extension multinomiale exprimée dans la proposition suivante.

PROPOSITION 4.6.2. — *Soient z_1, z_2, \dots, z_k des nombres complexes (ou des éléments pris dans un anneau commutatif). On a l'identité multinomiale :*

$$(4.6.2) \quad (z_1 + z_2 + \dots + z_k)^p = \sum \binom{p}{n_1, n_2, \dots, n_k} z_1^{n_1} z_2^{n_2} \dots z_k^{n_k},$$

où la sommation est étendue à l'ensemble des suites (n_1, n_2, \dots, n_k) d'entiers satisfaisant à :

$$(4.6.3) \quad n_1 \geq 0, n_2 \geq 0, \dots, n_k \geq 0 \quad \text{et} \quad n_1 + n_2 + \dots + n_k = p.$$

Démonstration. — Le développement de $(z_1 + z_2 + \dots + z_k)^p$ est égal à $\sum z_{i_1} z_{i_2} \dots z_{i_p}$, où la somme est étendue à toutes les suites (i_1, i_2, \dots, i_p) dont les termes sont pris dans $[k]$. Il y a donc k^p telles suites, donc k^p monômes dans la sommation. Lorsqu'on réarrange les lettres d'un monôme $z_{i_1} z_{i_2} \dots z_{i_p}$ suivant les indices croissants, on obtient un monôme de la forme $z_1^{n_1} z_2^{n_2} \dots z_k^{n_k}$, où les n_i satisfont les relations (4.6.3). D'après la Proposition 4.6.1, le nombre de monômes $z_{i_1} z_{i_2} \dots z_{i_p}$ de la somme initiale égaux à $z_1^{n_1} z_2^{n_2} \dots z_k^{n_k}$, est précisément égal au coefficient multinomial $\binom{p}{n_1, n_2, \dots, n_k}$. \square

Remarque. — Le nombre de termes distincts dans la somme de l'identité (4.6.2) est égal au nombre de solutions de l'équation

$$n_1 + n_2 + \dots + n_k = p,$$

c'est-à-dire, d'après la Proposition 4.5.2, à $\binom{p+k-1}{p}$.

Remarque. — L'identité multinomiale intervient dans de nombreux calculs explicites. Son écriture effective, même pour de petites valeurs de k et p , remplirait quantités d'écrans ! Le nombre de termes $\binom{p+k-1}{p}$ croît évidemment très vite.

5. Le principe de réflexion. — Dans ce paragraphe, nous nous proposons de montrer, à propos du problème du scrutin, comment une interprétation géométrique des suites finies de nombres permet une évaluation aisée de certains cardinaux et de là de certaines probabilités d'événements.

Soient n un entier fixé ($n \geq 1$) et $\omega = (x_1, x_2, \dots, x_n)$ une suite finie, telle que chaque x_i soit égal à 1 ou à -1 . On note $p(\omega) = p$ (resp. $q(\omega) = q$) le nombre de $+1$ (resp. -1) dans la suite ω , de sorte que l'on a $p + q = n$.

On appelle *sommes partielles* associées à ω , les sommes $s_k = x_1 + x_2 + \dots + x_k$ pour $k = 1, 2, \dots, n$. Par convention, $s_0 = 0$. Naturellement

$$s_k - s_{k-1} = x_k = \pm 1 \quad (1 \leq k \leq n) \quad \text{et} \quad s_n = p - q.$$

On peut identifier ω à une *ligne polygonale* (ou *chemin polygonal*) tracée dans un plan euclidien muni de deux axes rectangulaires, l'axe horizontal des t et l'axe vertical des s , de la façon suivante : on joint par une ligne droite successivement les points du plan : $(0, 0)$, $(1, s_1)$, $(2, s_2)$, \dots , (n, s_n) . On dit que l'entier n est la *longueur* du chemin. Il y a, évidemment, 2^n chemins de longueur n joignant $(0, 0)$ à un point de coordonnées (n, s) ($-n \leq s \leq n$).

Les chemins ainsi formés ne sont constitués que de courts segments SO-NE et NO-SE. Notons \mathcal{C} l'ensemble de tous les chemins constitués par ces seuls segments. Un chemin ω de \mathcal{C} allant de $(0, 0)$ à (n, s) doit satisfaire les équations :

$$(5.1) \quad p(\omega) + q(\omega) = n; \quad p(\omega) - q(\omega) = s;$$

ou encore :

$$(5.2) \quad p(\omega) = \frac{n+s}{2} \quad \text{et} \quad q(\omega) = \frac{n-s}{2}.$$

Par conséquent, le nombre de chemins allant de $(0, 0)$ à (n, s) est égal à :

$$(5.3) \quad c_{n,s} = \binom{p+q}{q} = \binom{p+q}{p} \quad \text{avec } p = \frac{n+s}{2}.$$

On pose $c_{n,s} = 0$, lorsque $n+s$ et $n-s$ ne sont pas tous deux pairs.

Soit A et B deux points du plan de coordonnées $A = (a, \alpha)$ et $B = (b, \beta)$. On suppose $0 \leq a < b$, $\alpha \geq 1$, $\beta \geq 1$. Ces hypothèses étant satisfaites, on a le lemme suivant, dit du *principe de réflexion*.

LEMME 5.1. — *Le nombre de chemins de \mathcal{C} allant de A à B , qui touchent ou traversent l'axe horizontal est égal au nombre de chemins allant du point $A' = (a, -\alpha)$ à B .*

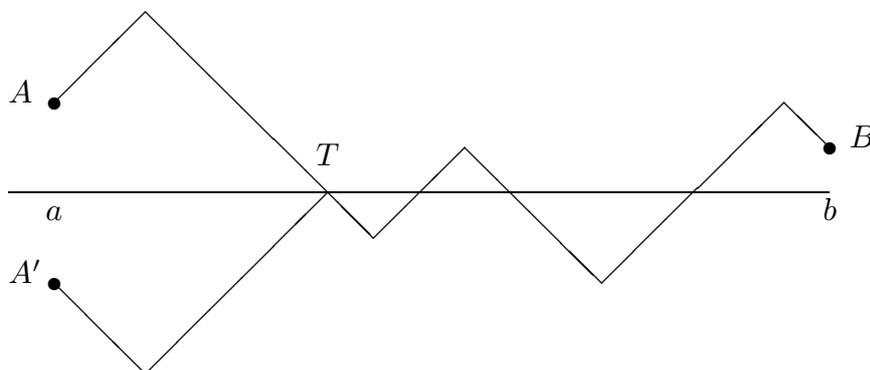


Fig. 2

Démonstration. — La démonstration est de nature purement géométrique (cf. Fig. 2). Soit ω un chemin touchant ou traversant l'axe horizontal. On désigne par T le point le plus à gauche où ω atteint l'axe horizontal. Soit ω_1 la portion du chemin ω allant de A à T et ω_2 la portion allant de T à B . On forme un nouveau chemin allant de A' à B en prenant le *symétrique* de ω_1 par rapport à l'axe horizontal et en lui adjoignant ω_2 . Soit ω' ce nouveau chemin. Il est clair que $\omega \mapsto \omega'$ est une bijection envoyant la première famille des chemins sur la seconde. \square

Analytiquement, le point T est le point d'abscisse t défini par

$$s_a > 0, s_{a+1} > 0, \dots, s_{t-1} > 0, s_t = 0,$$

et le nouveau chemin est défini par la séquence :

$$-s_a, -s_{a+1}, \dots, -s_{t-1}, s_t, s_{t+1}, \dots, s_b.$$

La plus célèbre application du principe de réflexion est le théorème du scrutin. Avant de l'établir, donnons encore un lemme de comptage, dans lequel on conserve les notations de (5.1), (5.2) et (5.3).

LEMME 5.2. — Soient $n \geq 1$ et $s \geq 1$. Le nombre de chemins allant de $(0, 0)$ à (n, s) et restant toujours strictement au-dessus de l'axe horizontal est égal à

$$\frac{s}{n} c_{n,s} = \frac{p-q}{p+q} \binom{p+q}{p}.$$

Démonstration. — Si un chemin ω partant de $(0, 0)$ reste toujours au-dessus de l'axe horizontal, il satisfait nécessairement $s_1 = 1$. De là, le nombre de tels chemins est aussi égal au nombre de chemins allant de $(1, 1)$ à (n, s) , ne touchant, ni traversant l'axe horizontal. D'après le précédent lemme et la formule (5.3), ce nombre est égal à :

$$\begin{aligned} c_{n-1,s-1} - c_{n-1,s+1} &= \binom{p+q-1}{p-1} - \binom{p+q-1}{p} \\ &= \left[\frac{p}{p+q} - \frac{q}{p+q} \right] \binom{p+q}{p} \\ &= \frac{p-q}{p+q} \binom{p+q}{p} = \frac{s}{n} c_{n,s}. \quad \square \end{aligned}$$

THÉORÈME 5.3 (du scrutin). — Dans un scrutin, il y a p bulletins pour le candidat P et q pour le candidat Q. On suppose $p > q$. Alors la probabilité pour que, durant tout le dépouillement, P soit toujours en tête est égale à $(p-q)/(p+q)$.

Démonstration. — Le problème revient à probabiliser l'ensemble de tous les chemins allant de $(0, 0)$ à $(p+q, p-q)$. Un tel chemin représente, en effet, un dépouillement (+1 si le vote est pour P, -1 si le vote est pour Q). En prenant l'équirépartition sur l'ensemble de ces chemins, il s'agit d'évaluer le nombre des chemins qui restent toujours strictement au-dessus de l'axe horizontal. C'est ce qui a été fait au lemme précédent. \square

COMPLÉMENTS ET EXERCICES

1. Une application de la formule de Poincaré. — Soit $n \geq 2$ et $n = p_1^{\alpha_1} \dots p_r^{\alpha_r}$ sa décomposition en facteurs premiers. Prenons $\Omega = \{1, 2, \dots, n\}$ et notons A_k la partie de Ω constituée par les entiers divisibles par le nombre premier p_k ($k = 1, \dots, r$). La réunion $A_1 \cup \dots \cup A_r$ est la partie de Ω constituée par les entiers qui sont divisibles par l'un au moins des premiers p_1, \dots, p_r et $(A_1 \cup \dots \cup A_r)^c$ est la partie de Ω constituée par les entiers qui ne sont divisibles par aucun des nombres premiers p_1, \dots, p_r , c'est-à-dire qui sont premiers avec n . Son cardinal est noté $\varphi(n)$. La fonction φ est appelée la *fonction arithmétique d'Euler*. On utilise la formule de Poincaré

pour démontrer la formule

$$\frac{\varphi(n)}{n} = \prod_{p|n} \left(1 - \frac{1}{p}\right),$$

où le produit au second membre est étendu à tous les facteurs premiers p de n qui divisent n . En effet, on a, d'après la formule de Poincaré :

$$|A_1 \cup \dots \cup A_r| = \sum_i |A_i| - \sum_{i < j} |A_i \cap A_j| + \dots + (-1)^{r-1} |A_1 \cap \dots \cap A_r|.$$

Or $|A_1 \cup \dots \cup A_r| = n - \varphi(n)$; d'autre part, si $d | n$, alors le nombre d'entiers de Ω qui sont divisibles par d est n/d . On a donc : $|A_i| = n/p_i$, $|A_i \cap A_j| = n/(p_i p_j)$, ..., $|A_i \cap \dots \cap A_r| = n/(p_1 \dots p_r)$. La formule de Poincaré s'écrit donc :

$$\begin{aligned} n - \varphi(n) &= \sum_i \frac{n}{p_i} - \sum_{i < j} \frac{n}{p_i p_j} + \dots + (-1)^{r-1} \frac{n}{p_1 \dots p_r}; \\ \frac{\varphi(n)}{n} &= 1 - \sum_i \frac{1}{p_i} + \sum_{i < j} \frac{1}{p_i p_j} + \dots + (-1)^r \frac{1}{p_1 \dots p_r} = \prod_{p|n} \left(1 - \frac{1}{p}\right). \quad \square \end{aligned}$$

2. *Le problème des rencontres.* — Une urne contient n boules numérotées de 1 à n . On les extrait successivement sans remise et après chaque tirage, on observe le numéro de la boule tirée. On décrit cette expérience au moyen du triplet $(\Omega, \mathfrak{A}, \mathbb{P})$, où Ω est l'ensemble des permutations de $\{1, \dots, n\}$, où \mathfrak{A} est $\mathfrak{P}(\Omega)$ et où \mathbb{P} est l'équirépartition.

On dit qu'il y a *rencontre au $i^{\text{ième}}$ tirage*, si la boule tirée porte le numéro i et l'on désigne par E_i l'évènement «il y a rencontre au $i^{\text{ième}}$ tirage»; E_i est la partie de Ω dont les éléments sont les permutations de $\{1, \dots, n\}$ dont la $i^{\text{ième}}$ place est occupée par le numéro i . De même, si (i_1, \dots, i_k) est une suite strictement croissante d'entiers compris entre 1 et n , l'intersection $E_{i_1} \cap \dots \cap E_{i_k}$ est la partie de Ω dont les éléments sont les permutations dont les $i_1^{\text{ième}}$, ..., $i_k^{\text{ième}}$ places sont occupées par les numéros i_1, \dots, i_k . On a donc, en vertu de l'équirépartition : $\mathbb{P}(E_i) = (n-1)!/n!$ ($i = 1, \dots, n$); $\mathbb{P}(E_{i_1} \cap E_{i_2}) = (n-2)!/n!$ ($1 \leq i_1 < i_2 \leq n$); $\mathbb{P}(E_1 \cap \dots \cap E_n) = 1/n!$.

a) Soit A l'évènement «il y a au moins une rencontre», c'est-à-dire $A = E_1 \cup \dots \cup E_n$. La formule de Poincaré donne :

$$\begin{aligned} \mathbb{P}(A) &= \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P}(E_{i_1} \cap \dots \cap E_{i_k}) \\ &= \sum_{k=1}^n (-1)^{k-1} \binom{n}{k} \frac{(n-k)!}{n!} = \sum_{k=1}^n \frac{(-1)^{k-1}}{k!}. \end{aligned}$$

Lorsque $n \rightarrow \infty$, cette quantité tend vers : $1 - e^{-1} \approx 0,63212$. On se convaincra que pour des valeurs modérées de n , en fait dès que $n = 7$, la probabilité $P(A)$ est déjà très proche de sa limite.

b) Soit B l'évènement : «il n'y a aucune rencontre». Alors $B = E_1^c \cap \dots \cap E_n^c = (E_1 \cup \dots \cup E_n)^c = A^c$. D'où

$$P(B) = 1 - P(A) = \sum_{k=0}^n \frac{(-1)^k}{k!}.$$

Lorsque $n \rightarrow \infty$, cette quantité tend vers $e^{-1} \approx 0,36788$.

c) Posons $d_n = n!P(B) = n! \sum_{k=0}^n (-1)^k/k!$. C'est évidemment un entier supérieur ou égal à 1 qui compte le nombre de permutations, parmi les $n!$ possibles, qui ne présentent aucune rencontre, c'est-à-dire telles qu'à aucun rang ne figure un numéro égal à ce rang. Le nombre d_n est appelé le *nombre de dérangements* de n objets.

Exemple 1. — Un facteur possède n lettres adressées à n destinataires deux à deux distincts. Alors d_n est le nombre de manières différentes dont il peut poster les lettres de telle façon qu'aucune n'arrive à destination.

Exemple 2. — Quel est le nombre de manières de disposer huit tours sur un échiquier de telle sorte qu'aucune tour ne puisse en attaquer une autre et que la diagonale blanche soit libre de tours? On voit que la solution est $d_8 = 14.833$.

d) Soit C l'évènement «il y a exactement une rencontre». On peut montrer que

$$P(C) = \sum_{k=0}^{n-1} \frac{(-1)^k}{k!}.$$

Cette quantité tend vers e^{-1} lorsque $n \rightarrow \infty$; d'autre part, $P(B) - P(C) = (-1)^n/n! \rightarrow 0$, de sorte que, pour n grand, la probabilité pour qu'il n'y ait aucune rencontre est pratiquement égale à celle pour qu'il y ait exactement une rencontre.

3. — Prenons $\Omega = \mathbb{N}^*$. On désigne par \mathcal{D} la classe des parties A de \mathbb{N}^* pour lesquelles la limite suivante existe :

$$d(A) = \lim_{n \rightarrow \infty} \frac{\text{card}(A \cap \{1, \dots, n\})}{n}.$$

Cette limite est alors appelée la *densité arithmétique* de A . On vérifie que $\Omega \in \mathcal{D}$, que \mathcal{D} est stable par passage au complémentaire et que \mathcal{D} est stable par réunion *disjointe finie*. Ainsi \mathcal{D} est une classe de Dynkin *faible*.

Remarque. — La classe \mathcal{D} n'est pas stable par intersection; ce n'est *pas* une algèbre. L'application $d : \mathcal{D} \rightarrow \mathbb{R}^+$ est simplement additive, mais n'est pas σ -additive sur \mathcal{D} .

4. — Une caisse contient dix paires de chaussures (en vrac), dont les miennes. On retire au hasard quatre chaussures. On veut calculer la probabilité pour que, parmi ces quatre chaussures, se trouve *ma paire*, ou encore la probabilité pour que, parmi ces quatre chaussures, se trouve *au moins une paire*.

Les dix chaussures *droites* (resp. *gauches*) sont numérotées : $(1, d)$, $(2, d)$, \dots , $(10, d)$ (resp. $(1, g)$, $(2, g)$, \dots , $(10, g)$). Soit X l'ensemble formé par ces vingt numéros.

a) Construire un triplet $(\Omega, \mathfrak{A}, P)$ qui décrit cette expérience. On choisira Ω de telle façon qu'il soit raisonnable de prendre pour P l'équipartition sur Ω .

b) Déterminer le sous-ensemble A_i de Ω qui correspond à l'évènement : « parmi les quatre chaussures retirées se trouve la $i^{\text{ième}}$ paire » ($i = 1, \dots, 10$).

c) Calculer $P(A_i)$ ($i = 1, \dots, 10$).

d) Soit (i_1, \dots, i_k) une suite strictement croissante d'entiers compris entre 1 et 10. Calculer $P(A_{i_1} \cap \dots \cap A_{i_k})$ pour $k = 2$, puis pour $k \geq 3$.

e) En déduire la probabilité pour que, parmi ces quatre chaussures, se trouve *au moins une paire*.

5. — On tire « au hasard » quatre cartes d'un jeu de cinquante-deux cartes. Quelle est la probabilité pour que, parmi ces quatre cartes, il y ait exactement deux rois? L'hypothèse « au hasard » amène à prendre l'équipartition sur un certain ensemble fondamental Ω . Préciser lequel.

6. *Jeu de « Passe-Dix »*. — On lance trois dés parfaits. Montrer que la probabilité pour que la somme des points amenés dépasse dix est égale à la probabilité pour que cette somme ne dépasse pas dix.

7. *Le « Paradoxe » du Chevalier de Méré*. — Ce personnage marquant de la cour de Louis XIV, qui « avait très bon esprit mais n'était pas géomètre » (cf. lettre de Pascal à Fermat du 29 juillet 1654) était un joueur impénitent, toujours à la recherche de règles cachées lui permettant de réaliser un avantage sur ses adversaires. Voici deux de ses « règles ».

a) « Il est avantageux de parier sur l'apparition d'au moins un “six” en lançant un dé quatre fois de suite. » Cette règle est bonne puisque la probabilité de l'évènement qui nous intéresse est :

$$1 - \left(\frac{5}{6}\right)^4 = 0,517747 > \frac{1}{2}.$$

La différence avec $\frac{1}{2}$ est faible, mais apte à fournir à long terme des gains assurés; le Chevalier devait jouer souvent!

b) « Il est avantageux de parier sur l'apparition d'au moins un “double-six” en lançant deux dés vingt-quatre fois de suite. » Cette règle est mauvaise,

puisque la probabilité de l'évènement cherché est :

$$1 - \left(\frac{35}{36}\right)^{24} = 0,491404 < \frac{1}{2}.$$

Le Chevalier était donc moins heureux avec cette règle qu'avec la précédente. En fait, il s'était laissé abuser par un soi-disant argument d'homothétie : en lançant un dé, il y a six issues, en lançant deux dés, il y en a $6^2 = 36$, soit six fois plus. Comme il est avantageux de parier sur l'apparition d'au moins un «six» en lançant un dé quatre fois de suite, il doit être avantageux de parier sur l'apparition d'au moins un «double-six» en lançant deux dés $4 \times 6 = 24$ fois de suite ! Paradoxe ! Noter que si le Chevalier avait parié sur l'apparition d'au moins un «double-six» en lançant deux dés vingt-cinq fois de suite, il aurait été avantagé, la probabilité de l'évènement désiré étant supérieure à $1/2$.

Dans les exercices suivants, on étudie le modèle des urnes et des boules. Il s'agit de répartir r boules dans n urnes et l'on suppose les *répartitions* équiprobables. Cependant, on obtient différentes situations suivant que l'on considère les boules et les urnes comme discernables ou indiscernables. Tous ces modèles sont effectivement utilisés en mécanique statistique.

8. *Boules et urnes discernables (modèle de Maxwell-Boltzmann)*

a) L'ensemble Ω de toutes les répartitions possibles est l'ensemble Ω de toutes les suites $(\omega_1, \dots, \omega_r)$, de longueur r , où ω_j désigne l'urne contenant la $j^{\text{ième}}$ boule ($j = 1, \dots, r$). Le cardinal de Ω est donc n^r . On prend comme probabilité sur Ω l'équirépartition P .

b) On désigne par A_i l'évènement «la $i^{\text{ième}}$ urne est vide» ($i = 1, \dots, n$). On a :

$$P(A_i) = \left(1 - \frac{1}{n}\right)^r \quad \text{pour tout } i = 1, \dots, n.$$

c) Soit B l'évènement «chaque urne contient au plus une boule». On a :

$$P(B) = \frac{n(n-1)\dots(n-r+1)}{n^r}.$$

Cette quantité est nulle si $r \geq n + 1$.

d) Soit C_{ik} l'évènement «la $i^{\text{ième}}$ urne contient exactement k boules» ($1 \leq i \leq n$; $0 \leq k \leq r$). Les k boules peuvent être choisies de $\binom{r}{k}$ façons et les $(r - k)$ boules restantes peuvent être placées dans les $(n - 1)$ urnes restantes de $(n - 1)^{r-k}$ façons. On a donc :

$$P(C_{ik}) = \frac{1}{n^r} \binom{r}{k} (n - 1)^{r-k} = \binom{r}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{r-k}.$$

e) Soit A l'évènement « parmi m urnes fixées à l'avance, aucune n'est vide » ($1 \leq m \leq n$). Pour calculer la probabilité de $A = A_1^c \cap \dots \cap A_m^c$ on utilise la formule de Poincaré :

$$\begin{aligned} P(A) &= \sum_{k=0}^m (-1)^k \sum_{1 \leq i_1 < \dots < i_k \leq m} P(A_{i_1} \cap \dots \cap A_{i_k}) \\ &= \sum_{k=0}^m (-1)^k \binom{m}{k} \left(1 - \frac{k}{n}\right)^r. \end{aligned}$$

f) Soit $C = C_{1,k_1} \cap C_{2,k_2} \cap \dots \cap C_{n,k_n}$. On a : $P(C) = \frac{1}{n^r} \frac{r!}{k_1! \dots k_n!}$, si $k_1 + \dots + k_n = r$ et $P(C) = 0$, sinon.

g) Dans Strasbourg *sept* accidents se produisent chaque *semaine*. En faisant des hypothèses convenables, calculer la probabilité pour qu'une semaine contienne un jour avec au moins deux accidents.

Réponse. — On a $r = 7$ boules (accidents) à répartir dans $n = 7$ urnes (les jours de la semaine). En désignant par A_i ($i = 1, \dots, 7$) l'évènement « le $i^{\text{ième}}$ jour est sans accident », l'évènement dont on cherche la probabilité est $A = A_1 \cup \dots \cup A_7$. La formule de Poincaré donne :

$$\begin{aligned} P(A) &= \sum_{k=1}^7 (-1)^{k-1} \binom{7}{k} P(A_1 \cap \dots \cap A_k) \\ &= \sum_{k=1}^7 (-1)^{k-1} \binom{7}{k} \left(1 - \frac{k}{7}\right)^7 \approx 0,99387. \end{aligned}$$

On a encore $P(A) = 1 - (7!/7^7)$. Un malheur n'arrive jamais seul!

9. Boules indiscernables et urnes discernables (Bose-Einstein).

a) Ici seul le *nombre* de boules dans les urnes intervient. On prend pour Ω l'ensemble des suites (x_1, \dots, x_n) qui sont solutions en nombre entiers positifs de l'équation : $x_1 + \dots + x_n = r$ (cf. Proposition 4.5.2). Le cardinal de Ω est donc $\binom{n+r-1}{r}$. Comme mesure de probabilité sur Ω on prend l'équirépartition.

b) Soit A_k l'évènement « une urne fixée contient exactement k boules ». On a : $P(A_k) = \binom{n+r-k-2}{r-k} / \binom{n+r-1}{r}$.

10. Boules indiscernables, urnes discernables et impossibilité d'avoir deux ou plus de deux boules dans une même urne (Fermi-Dirac).

a) L'ensemble Ω est l'ensemble des parties de r éléments (les r urnes non vides) prises d'un ensemble de n éléments (les n urnes au total). Le cardinal de Ω est $\binom{n}{r}$ (cf. Proposition 4.4.1). On prend l'équirépartition sur Ω .

b) Soit A_1 l'évènement « une urne fixée contient exactement une boule ». On a : $P(A_1) = \binom{n-1}{r-1} / \binom{n}{r} = r/n$.

CHAPITRE 5

VARIABLES ALÉATOIRES

Le modèle décrit dans les chapitres précédents n'est plus suffisant lorsque l'on veut mesurer des grandeurs dépendant du hasard ou décrire l'état d'un système aléatoire évoluant au cours du temps. On est amené à introduire des fonctions définies sur des espaces probabilisés. Les anciens les appelaient *variables aléatoires* et nous avons conservé l'appellation. Ce sont, en fait, de véritables fonctions, à valeurs réelles ou à valeurs dans \mathbb{R}^n . L'objet de ce chapitre est de donner la définition formelle de variable aléatoire, après avoir rappelé quelques points techniques sur les applications réciproques et fourni quelques résultats de base sur les fonctions mesurables.

1. Application réciproque. — L'application réciproque X^{-1} d'une application $X : E \rightarrow F$ est une application de l'ensemble des parties $\mathfrak{P}(F)$ dans l'ensemble $\mathfrak{P}(E)$, qui envoie toute partie B de F sur la partie $X^{-1}(B)$ de E formée de tous les éléments e tels que $X(e)$ appartient à B . L'ensemble $X^{-1}(B) = \{e \in E : X(e) \in B\}$ est appelé l'*image réciproque* de B .

Le fait essentiel, qui ne sera pas redémontré ici, est que l'application réciproque X^{-1} conserve les opérations élémentaires sur les ensembles. En d'autres termes, si B_n (avec ou sans indice) désigne une partie de F , on a les relations :

$$\begin{aligned} X^{-1}(\emptyset) &= \emptyset, & X^{-1}(F) &= E, & X^{-1}(B^c) &= (X^{-1}(B))^c, \\ X^{-1}\left(\bigcup_n B_n\right) &= \bigcup_n X^{-1}(B_n), & X^{-1}\left(\bigcap_n B_n\right) &= \bigcap_n X^{-1}(B_n). \end{aligned}$$

En particulier, l'image réciproque de la réunion de sous-ensembles *disjoints deux à deux* est la réunion de leurs images réciproques, disjointes deux à deux également. Avec la notation \sum , on a donc :

$$X^{-1}\left(\sum_n B_n\right) = \sum_n X^{-1}(B_n).$$

Comme on peut s'y attendre, l'application réciproque conserve les structures algébriques sur les ensembles, en particulier la structure de tribu; on a, en effet, la proposition suivante.

PROPOSITION 1.1. — *Soient (F, \mathfrak{F}) un espace mesurable et $X : E \rightarrow F$ une application. Alors la famille $X^{-1}(\mathfrak{F}) = \{X^{-1}(B) : B \in \mathfrak{F}\}$ est une tribu de E .*

Démonstration. — En effet, comme E est égal à $X^{-1}(F)$ et que F appartient à \mathfrak{F} , on a bien $E \in X^{-1}(\mathfrak{F})$. D'autre part, si B appartient à \mathfrak{F} , alors B^c est aussi dans \mathfrak{F} . Par suite, $(X^{-1}(B))^c = X^{-1}(B^c) \in X^{-1}(\mathfrak{F})$. Enfin, si (B_n) est une suite de parties appartenant à \mathfrak{F} , alors leur réunion y appartient aussi et l'on a : $\bigcup_n X^{-1}(B_n) = X^{-1}(\bigcup_n B_n) \in X^{-1}(\mathfrak{F})$. \square

La prochaine proposition dit que l'application réciproque commute avec l'opération de tribu engendrée.

PROPOSITION 1.2. — *Soient $X : E \rightarrow F$ une application et \mathcal{C} une classe de parties de F . Alors $\mathfrak{T}(X^{-1}(\mathcal{C})) = X^{-1}(\mathfrak{T}(\mathcal{C}))$, où $\mathfrak{T}(\mathcal{C})$ désigne la tribu engendrée par \mathcal{C} .*

Démonstration. — On a tout d'abord $\mathcal{C} \subset \mathfrak{T}(\mathcal{C})$, d'où $X^{-1}(\mathcal{C}) \subset X^{-1}(\mathfrak{T}(\mathcal{C}))$. Comme ce dernier ensemble est une tribu d'après la proposition précédente, on a l'inclusion : $\mathfrak{T}(X^{-1}(\mathcal{C})) \subset X^{-1}(\mathfrak{T}(\mathcal{C}))$.

Notons \mathfrak{F} la classe des parties B de F dont les images réciproques $X^{-1}(B)$ sont des éléments de $\mathfrak{T}(X^{-1}(\mathcal{C}))$. On a, par conséquent $X^{-1}(\mathfrak{F}) \subset \mathfrak{T}(X^{-1}(\mathcal{C}))$. Vérifions d'abord que \mathfrak{F} est une tribu de F :

- (i) On a $X^{-1}(F) = E \in \mathfrak{T}(X^{-1}(\mathcal{C}))$, d'où $F \in \mathfrak{F}$.
- (ii) Si (B_n) est une suite d'éléments de \mathfrak{F} , alors $X^{-1}(\bigcup_n B_n) = \bigcup_n X^{-1}(B_n) \in \mathfrak{T}(X^{-1}(\mathcal{C}))$, d'où $\bigcup_n B_n \in \mathfrak{F}$.
- (iii) Si B appartient à \mathfrak{F} , alors $X^{-1}(B^c) = (X^{-1}(B))^c$ appartient aussi à $\mathfrak{T}(X^{-1}(\mathcal{C}))$. D'où $B^c \in \mathfrak{F}$.

Notons ensuite que \mathfrak{F} contient \mathcal{C} , donc contient aussi $\mathfrak{T}(\mathcal{C})$. Par suite,

$$\mathfrak{T}(X^{-1}(\mathcal{C})) \supset X^{-1}(\mathfrak{F}) \supset X^{-1}\mathfrak{T}(\mathcal{C}).$$

Les deux tribus $\mathfrak{T}(X^{-1}(\mathcal{C}))$ et $X^{-1}\mathfrak{T}(\mathcal{C})$ sont donc identiques. \square

2. Fonctions mesurables

Définition. — Soient (E, \mathfrak{E}) et (F, \mathfrak{F}) deux espaces mesurables. On dit que $X : E \rightarrow F$ est une *fonction mesurable de (E, \mathfrak{E}) dans (F, \mathfrak{F})* , si $X^{-1}(\mathfrak{F})$ est une sous-tribu de \mathfrak{E} . Si $(F, \mathfrak{F}) = (\mathbb{R}, \mathcal{B}^1)$, on parle simplement de *fonction mesurable sur (E, \mathfrak{E})* .

La proposition suivante permet de vérifier la mesurabilité d'une application non nécessairement sur tous les éléments de la tribu \mathfrak{F} , mais sur les seuls éléments d'une classe qui engendre \mathfrak{F} .

PROPOSITION 2.1. — *Soient (E, \mathfrak{E}) et (F, \mathfrak{F}) deux espaces mesurables. Pour qu'une application $X : E \rightarrow F$ soit une application mesurable de (E, \mathfrak{E}) dans (F, \mathfrak{F}) , il suffit qu'il existe une classe \mathcal{C} de parties de \mathfrak{F} , qui engendre \mathfrak{F} et qui soit telle que $X^{-1}(\mathcal{C}) \subset \mathfrak{E}$.*

Démonstration. — En effet, si $X^{-1}(\mathcal{C})$ est contenue dans la tribu \mathfrak{E} , alors la tribu $\mathfrak{T}(X^{-1}(\mathcal{C}))$, égale à $X^{-1}(\mathfrak{T}(\mathcal{C}))$ d'après la Proposition 1.2, ou encore à $X^{-1}(\mathfrak{F})$, est aussi contenue dans \mathfrak{E} . L'application X est donc bien mesurable. \square

On a vu que la classe des intervalles ouverts (resp. des demi-droites, etc.) engendrait la tribu borélienne \mathcal{B}^1 de \mathbb{R} . Il résulte donc de la proposition précédente que pour vérifier qu'une fonction à valeurs réelles, définie sur un espace mesurable (E, \mathfrak{E}) , est mesurable, il suffit de montrer que pour tout couple de nombre réels (a, b) tels que $a < b$ on a $X^{-1}(]a, b[) \in \mathfrak{E}$ (resp. que pour tout nombre réel a l'image réciproque $X^{-1}(]-\infty, a])$ appartient à \mathfrak{E} , etc.).

La proposition suivante se démontre à l'aide de la Proposition 2.1. Sa démonstration n'est pas reproduite.

PROPOSITION 2.2. — *Soient (E, \mathfrak{E}) un espace mesurable, λ un nombre réel et X, Y (avec ou sans indice) des fonctions mesurables sur (E, \mathfrak{E}) . Alors $|X|, X + Y, \lambda X, X \cdot Y, \limsup_n X_n, \liminf_n X_n, \sup_n X_n, \inf_n X_n$ (à la condition que ces opérations conduisent à des fonctions à valeurs numériques finies) sont mesurables.*

Si, de plus, $X \neq 0$, alors $1/X$ est mesurable. Enfin, si X_n converge en tout point e de E vers un nombre fini, la fonction $X = \lim_n X_n$ est aussi mesurable.

On résume tout cet énoncé en disant que la mesurabilité est *stable sous toutes les opérations classiques de l'analyse*. Les applications constantes sont évidemment mesurables. Pour les fonctions continues, on a l'énoncé suivant.

PROPOSITION 2.3. — *Soient E un espace topologique et \mathfrak{E} la tribu engendrée par la classe des ouverts de E . Alors toute fonction continue $X : (E, \mathfrak{E}) \rightarrow (\mathbb{R}, \mathcal{B}^1)$ est mesurable.*

Démonstration. — Soit \mathfrak{D} la classe des ouverts de \mathbb{R} . Alors $X^{-1}(\mathfrak{D}) \subset \mathfrak{E}$. D'où $\mathfrak{T}(X^{-1}(\mathfrak{D})) = X^{-1}(\mathfrak{T}(\mathfrak{D})) \subset \mathfrak{E}$. Comme on a $\mathfrak{T}(\mathfrak{D}) = \mathcal{B}^1$, on en conclut que $X^{-1}(\mathcal{B}^1)$ est contenu dans \mathfrak{E} , c'est-à-dire que X est mesurable. \square

La prochaine proposition permet de comparer la mesurabilité d'une fonction à valeurs dans \mathbb{R}^n avec la mesurabilité des applications coordonnées de cette fonction.

PROPOSITION 2.4. — *Soient (E, \mathfrak{E}) un espace mesurable et $X = (X_1, \dots, X_n)$ une application de E dans \mathbb{R}^n . Alors X est une fonction mesurable de (E, \mathfrak{E}) dans $(\mathbb{R}^n, \mathcal{B}^n)$, si et seulement si, pour chaque $i = 1, \dots, n$, l'application coordonnée $X_i : E \rightarrow \mathbb{R}$ est une fonction mesurable de (E, \mathfrak{E}) dans $(\mathbb{R}, \mathcal{B}^1)$.*

Démonstration. — Soient (B_1, \dots, B_n) une suite de n boréliens de la droite réelle. Alors le produit cartésien $B = \prod_i B_i$ est un borélien de \mathbb{R}^n . De plus, $X^{-1}(B) = \bigcap_i X_i^{-1}(B_i)$. Si X est mesurable, choisissons un entier i dans $[1, n]$ et prenons $B_i = [a_i, b_i[$ et $B_j = \mathbb{R}$ pour tout $j \neq i$. Alors $X^{-1}(B) = X_i^{-1}([a_i, b_i[)$. D'où $X_i^{-1}([a_i, b_i[)$ appartient à \mathfrak{E} . Par suite, X_i est mesurable, d'après la Proposition 2.1.

Réciproquement, si tous les X_i sont mesurables, prenons $B_i = [a_i, b_i[$ ($i = 1, \dots, n$). L'ensemble B est un pavé de \mathbb{R}^n et $X^{-1}(B)$ appartient à \mathfrak{E} .

On a donc $X^{-1}(\mathcal{B}^n) \subset \mathfrak{E}$. La fonction X est donc elle-même mesurable, d'après la Proposition 2.1. \square

La prochaine proposition est donnée sans démonstration.

PROPOSITION 2.5. — *Soient (E, \mathfrak{E}) un espace mesurable, $X : (E, \mathfrak{E}) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$ et $f : (\mathbb{R}^n, \mathcal{B}^n) \rightarrow (\mathbb{R}, \mathcal{B}^1)$ deux applications mesurables. Alors $f \circ X$ est une fonction mesurable.*

La notion d'indicatrice d'évènement a été introduite au chap. 1, § 3.3. On prendra soin de ne pas confondre l'indicatrice I_A d'un évènement A appartenant à une tribu \mathfrak{A} d'un ensemble Ω et la *mesure singulière* ϵ_ω , qui est une fonction d'ensemble définie sur \mathfrak{A} . On a cependant la relation

$$I_A(\omega) = \epsilon_\omega(A), \quad \text{pour tout } \omega \in \Omega \text{ et } A \in \mathfrak{A}.$$

La proposition suivante est immédiate et donnée sans démonstration.

PROPOSITION 2.6. — *Soit (Ω, \mathfrak{A}) un espace mesurable. Alors l'indicatrice I_A d'un sous-ensemble A de Ω est une fonction mesurable si et seulement si A appartient à \mathfrak{A} .*

3. Variables aléatoires

Définition. — Un espace probabilisé $(\Omega, \mathfrak{A}, P)$ étant donné, on appelle *variable aléatoire réelle* (resp. *variable aléatoire à valeurs dans \mathbb{R}^n ou à n dimensions*) une fonction mesurable de (Ω, \mathfrak{A}) dans $(\mathbb{R}, \mathcal{B}^1)$ (resp. dans $(\mathbb{R}^n, \mathcal{B}^n)$).

Il résulte de ce qui précède qu'une variable aléatoire réelle (resp. à n dimensions) est une application de Ω dans \mathbb{R} (resp. dans \mathbb{R}^n) telle que pour tout a réel (resp. toute suite (a_1, \dots, a_n) de nombre réels) l'image réciproque $X^{-1}(] - \infty, a])$ (resp. $X^{-1}(] - \infty, a_1] \times \dots \times] - \infty, a_n])$) appartient à la tribu \mathfrak{A} .

Remarque. — Notons que dans la définition d'une variable aléatoire, la mesure de probabilité P figurant dans le triplet $(\Omega, \mathfrak{A}, P)$ ne joue aucun rôle; seule intervient la tribu \mathfrak{A} . La terminologie usuelle de variable *aléatoire* est donc, en toute rigueur, impropre. Cependant la variable «aléatoire» X définie sur le triplet permet de *probabiliser* l'espace d'arrivée $(\mathbb{R}, \mathcal{B}^1)$, comme il est décrit ci-après.

Notation probabiliste. — Supposons donnés un espace probabilisé $(\Omega, \mathfrak{A}, P)$ et un évènement A de la tribu \mathfrak{A} . La notation $P(A)$ se lit *probabilité de l'évènement A* ou *probabilité que l'évènement A se produise*.

De la même façon, si X est une variable aléatoire réelle définie sur cet espace et si B est un borélien de la droite, l'évènement $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$ appartient à \mathfrak{A} . On note cet évènement $\{X \in B\}$. Sa probabilité $P\{X \in B\}$ se lit *probabilité que X soit dans B* .

L'ensemble B peut prendre diverses formes. On écrira, par exemple, $P\{X = b\}$ au lieu de $P\{X \in \{b\}\}$ et $P\{X \leq b\}$ au lieu de $P\{X \in] - \infty, b])$.

Ces expressions sont lues *probabilité que X soit égal à b* et *probabilité que X soit inférieur ou égal à b* , respectivement.

De même, si (A_n) désigne une suite d'évènements de \mathfrak{A} , on écrit $P(A_1, A_2, \dots, A_n)$ pour $P(A_1 \cap A_2 \cap \dots \cap A_n)$. Cette expression se lit *probabilité que les évènements A_1, A_2, \dots, A_n se produisent en même temps*.

Enfin, si deux variables aléatoires X_1 et X_2 sont définies sur le même espace probabilisé $(\Omega, \mathfrak{A}, P)$, on écrit $P\{X_1 \leq b_1, X_2 \leq b_2\}$ pour $P(X_1^{-1}(] - \infty, b_1] \cap X_2^{-1}(] - \infty, b_2])$ et on lit *probabilité que X_1 soit inférieur ou égal à b_1 et que X_2 soit inférieur ou égal à b_2* .

4. Loi de probabilité d'une variable aléatoire. — La notion de variable aléatoire permet de *probabiliser* l'espace mesurable $(\mathbb{R}^n, \mathcal{B}^n)$; on a, en effet, la proposition suivante.

PROPOSITION 4.1. — *Soit $X : (\Omega, \mathfrak{A}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$ une variable aléatoire à n dimensions. Alors l'application $P_X : \mathcal{B}^n \rightarrow [0, 1]$ définie pour tout borélien B de \mathcal{B}^n par*

$$P_X(B) = P(X^{-1}(B)),$$

est une mesure de probabilité sur $(\mathbb{R}^n, \mathcal{B}^n)$.

La mesure de probabilité P_X ainsi définie est appelée *loi de probabilité de la variable aléatoire X* . Elle est encore notée $\mathcal{L}(X)$. On dit aussi que X *suit la loi de probabilité P_X* .

Démonstration. — Comme, pour tout sous-ensemble borélien B , l'image réciproque $X^{-1}(B)$ appartient à \mathfrak{A} , la précédente définition de P_X a bien un sens. D'autre part, $P_X(\mathbb{R}^n) = P(X^{-1}(\mathbb{R}^n)) = P(\Omega) = 1$. Enfin, si (B_n) est une suite de sous-ensembles boréliens deux à deux disjoints, on a les égalités : $P_X(\sum_n B_n) = P(X^{-1}(\sum_n B_n)) = P(\sum_n X^{-1}(B_n)) = \sum_n P(X^{-1}(B_n)) = \sum_n P_X(B_n)$. \square

Avec les mêmes notations que ci-dessus, on a donc

$$P\{X \in B\} = P(X^{-1}(B)) = P_X(B).$$

D'autre part, l'espace $(\mathbb{R}^n, \mathcal{B}^n, P_X)$ est un nouvel espace probabilisé. Il y a donc lieu de bien préciser la probabilité qu'on utilise lorsqu'on parle de probabilité d'un évènement.

Remarque 1. — La notion de loi de probabilité est fondamentale pour la raison suivante : à toute mesure de probabilité P' sur l'espace $(\mathbb{R}^n, \mathcal{B}^n)$, on peut associer un espace probabilisé $(\Omega, \mathfrak{A}, P)$ et une variable aléatoire $X : (\Omega, \mathfrak{A}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$, telle que la loi de probabilité P_X de X soit précisément P' . Il suffit de prendre, en effet, $\Omega = \mathbb{R}^n$, $\mathfrak{A} = \mathcal{B}^n$, $P = P'$ et pour X l'application identique de Ω .

On peut ainsi parler de variable aléatoire X ayant une loi de probabilité P_X , sans spécifier l'espace probabilisé $(\Omega, \mathfrak{A}, P)$ sur lequel X est défini. On ne

retient alors que l'espace probabilisé $(\mathbb{R}^n, \mathcal{B}^n, P_X)$, que l'on étudie *de façon autonome* en faisant abstraction du mécanisme qui a donné naissance à X . On est alors dans le domaine de la phénoménologie, puisque \mathbb{R}^n est l'ensemble des *valeurs observées* de X , dont l'information probabiliste sur X est concentrée sur la loi de probabilité P_X sur $(\mathbb{R}^n, \mathcal{B}^n)$.

Ce point de vue (appelé point de vue «en loi») est souvent adopté, principalement dans les exposés élémentaires. Il est parfaitement défendable, d'autant que pour le calcul de la plupart des valeurs typiques d'une variable aléatoire (espérance mathématique, variance, médiane, ...) il est suffisant de connaître l'espace probabilisé $(\mathbb{R}^n, \mathcal{B}^n, P_X)$.

Remarque 2. — La loi de probabilité P_X est l'image de la mesure P par l'application X . On la note aussi $X(P)$, notation qui a l'avantage suivant : si l'on considère la loi de $f \circ X$, la *transitivité de la mesure image* permet d'écrire : $(f \circ X)(P) = f(X(P))$.

5. Fonction de répartition d'une variable aléatoire réelle. — Soit $(\Omega, \mathfrak{A}, P)$ un espace probabilisé et $X : (\Omega, \mathfrak{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}^1)$ une variable aléatoire *réelle*. On appelle *fonction de répartition de X* la fonction qui à tout réel x associe le nombre $F(x)$ défini par :

$$F(x) = P\{X \leq x\} = P_X(]-\infty, x]).$$

PROPOSITION 5.1. — *La fonction de répartition F d'une variable aléatoire réelle X satisfait aux propriétés suivantes :*

- (i) $0 \leq F(x) \leq 1$;
- (ii) F est une fonction croissante (au sens large), continue à droite en tout point x de \mathbb{R} ;
- (iii) $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$.

Démonstration. — La propriété (i) est évidente. Prenons ensuite deux nombres réels x et x' tels que $x \leq x'$. Alors $] - \infty, x] \subset] - \infty, x']$. D'où $P_X(]-\infty, x]) \leq P_X(]-\infty, x'])$ et par conséquent $F(x) \leq F(x')$.

Soit maintenant une suite décroissante (ϵ_n) de nombres réels tendant vers 0. (En abrégé : $\epsilon_n \downarrow 0$.) Pour tout réel x on a :

$$\begin{aligned} P_X(]x, x + \epsilon_n]) &= P_X(]-\infty, x + \epsilon_n]) - P_X(]-\infty, x]) \\ &= F(x + \epsilon_n) - F(x). \end{aligned}$$

Or $]x, x + \epsilon_n] \downarrow \emptyset$. D'où $\lim_n P_X(]x, x + \epsilon_n]) = 0$, soit $F(x + 0) = F(x)$.

De même, $] - \infty, -n] \downarrow \emptyset$, d'où $\lim_{x \rightarrow -\infty} F(x) = \lim_n P_X(]-\infty, -n]) = P_X(\lim_n(]-\infty, -n]) = 0$.

Finalement, $] - \infty, n] \uparrow \mathbb{R}$. D'où $\lim_{x \rightarrow +\infty} F(x) = \lim_n P_X(]-\infty, n]) = P_X(\lim_n(]-\infty, n]) = P_X(\mathbb{R}) = 1$. \square

Remarque. — Il est souvent commode, pour une fonction réelle F d'une variable réelle, qui satisfait les propriétés (i), (ii) et (iii) de la Proposition 5.1, de dire qu'elle est une *fonction de répartition* tout court (sur la droite).

La Proposition 5.1 admet une réciproque, qui sera démontrée au chap. 10 (Théorème 3.1), que nous énonçons ici sous la forme suivante.

PROPOSITION 5.2. — *A toute fonction de répartition F correspond une et une seule mesure de probabilité P sur $(\mathbb{R}, \mathcal{B}^1)$, satisfaisant à $P(]-\infty, x]) = F(x)$ pour tout x réel.*

Remarque. — On peut aussi définir une fonction de répartition F associée à un vecteur aléatoire $X = (X_1, X_2, \dots, X_n)$ à n dimensions. Il suffit de poser :

$$F(x_1, x_2, \dots, x_n) = P\{X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n\}.$$

On obtient ainsi une fonction réelle, à valeurs dans $[0, 1]$, de n variables réelles. Les propriétés de cette fonction de répartition (conjointe) sont étudiées dans l'exercice 11.

6. La fonction de masse et les discontinuités de la fonction de répartition. — Soient $X : (\Omega, \mathfrak{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}^1)$ une variable aléatoire réelle et F sa fonction de répartition. On sait que pour tout $x \in \mathbb{R}$ on a $\{x\} \in \mathcal{B}^1$, de sorte que $P\{X = x\}$ est défini pour tout x .

Définition. — On appelle *fonction de masse de X* l'application $\pi(\cdot) : \mathbb{R} \rightarrow \mathbb{R}^+$ définie par $\pi(x) = P\{X = x\}$.

PROPOSITION 6.1. — *La fonction de masse $\pi(\cdot)$ est entièrement déterminée par la loi de probabilité, donc par la fonction de répartition F . Pour tout $x \in \mathbb{R}$, on a : $\pi(x) = F(x) - F(x - 0) = F(x + 0) - F(x - 0)$. Le nombre $\pi(x)$ est la hauteur du saut de F en x .*

Démonstration. — Soit $x \in \mathbb{R}$ et $(x_n)_{n \geq 1}$ une suite croissante de réels tels que pour tout $n \geq 1$ on ait $x_n < x$ et $x_n \uparrow x$. Pour tout $n \geq 1$ on a la relation

$$]-\infty, x] =]-\infty, x_n] \cup]x_n, x],$$

d'où, en prenant la probabilité P_X des deux côtés :

$$F(x) = F(x_n) + P_X(]x_n, x]).$$

Passant à la limite pour $n \rightarrow \infty$, on observe que :

(i) la suite $(F(x_n))_{n \geq 1}$ est croissante bornée, elle tend donc vers une limite notée $F(x - 0)$;

(ii) la suite d'ensembles $(]x_n, x])_{n \geq 1}$ est décroissante, sa limite est $\bigcap_{n \geq 1}]x_n, x] = \{x\}$; d'où $\lim_n P_X(]x_n, x]) = P_X(\{x\}) = P_X\{X = x\} = \pi(x)$. On obtient finalement : $F(x) = F(x - 0) + \pi(x)$. \square

PROPOSITION 6.2. — *Soient X une variable aléatoire, F sa fonction de répartition, $\pi(\cdot)$ sa fonction de masse. Posons :*

$$D_X = \{x \in \mathbb{R} : \pi(x) > 0\}.$$

- a) La fonction F est continue au point $x \in \mathbb{R}$, si et seulement si $\pi(x) = 0$.
 b) La fonction F est continue sur tout \mathbb{R} , si et seulement si $D_X = \emptyset$;
 on dit alors que la loi de probabilité de X est diffuse.
 c) L'ensemble D_X est fini ou dénombrable ; en d'autres termes, l'ensemble des points de discontinuité de F est fini ou dénombrable.

Démonstration. — Les propriétés a) et b) sont évidentes ; pour démontrer c) posons : $A_n = \{x \in \mathbb{R} : \pi(x) \geq 1/n\}$. Comme $\pi(x)$ représente la hauteur du saut de F au point x , et que F est croissante, de variation totale égale à 1, l'ensemble A_n contient au plus n éléments. En outre, $D_X = \bigcup_{n \geq 1} A_n$. Comme une réunion finie ou dénombrable d'ensembles finis ou dénombrables est encore finie ou dénombrable, la propriété est démontrée. \square

Définition. — Soient X une variable aléatoire réelle et $\pi(\cdot)$ sa fonction de masse. On a vu que D_X est fini ou dénombrable ; on peut donc effectuer la somme $\sum_{x \in D_X} \pi(x)$. Si cette somme est égale à 1, on dit que la variable aléatoire X est *discrète* et que D_X est son *support* ; nous désignerons par la suite ce support par S_X .

Remarque. — Le support d'une variable aléatoire discrète peut être *partout dense*, par exemple $S_X = \mathbb{Q}$. Dans ce cas, la fonction de répartition ne peut plus être représentée par une fonction en escalier élémentaire (cf. Exercice 7).

PROPOSITION 6.3. — Soient X une variable aléatoire, P_X sa loi de probabilité, F sa fonction de répartition et $\pi(\cdot)$ sa fonction de masse. Alors on a, pour $-\infty < a < b < +\infty$, les relations :

$$\begin{aligned} P_X(] - \infty, a]) &= F(a) ; \\ P_X(] - \infty, a[) &= F(a - 0) = F(a) - \pi(a) ; \\ P_X(]a, +\infty[) &= 1 - F(a) ; \\ P_X([a, +\infty[) &= 1 - F(a - 0) = 1 - F(a) + \pi(a) ; \\ P_X(]a, b]) &= F(b) - F(a) ; \\ P_X([a, b]) &= F(b) - F(a - 0) = F(b) - F(a) + \pi(a) ; \\ P_X(]a, b[) &= F(b - 0) - F(a) = F(b) - F(a) - \pi(a) ; \\ P_X([a, b[) &= F(b - 0) - F(a - 0) = (F(b) - F(a)) - (\pi(b) - \pi(a)). \end{aligned}$$

La vérification de toutes ces propriétés ne présente aucune difficulté.

7. Tribu engendrée par une variable aléatoire. — Soit X une variable aléatoire à n dimensions, définie sur un espace probabilisé $(\Omega, \mathfrak{A}, P)$. On appelle *tribu engendrée par X* la tribu notée $\mathfrak{T}(X)$, définie comme la plus petite tribu contenue dans \mathfrak{A} rendant X mesurable.

PROPOSITION 7.1. — On a $\mathfrak{T}(X) = X^{-1}(\mathcal{B}^n)$.

Démonstration. — Par définition même de la mesurabilité, une sous-tribu \mathfrak{T} de \mathfrak{A} rend X mesurable, si et seulement si l'on a $X^{-1}(\mathcal{B}^n) \subset \mathfrak{T}$. En particulier, on a $X^{-1}(\mathcal{B}^n) \subset \mathfrak{T}(X)$. L'inclusion inverse est aussi vraie, car $X^{-1}(\mathcal{B}^n)$ rend évidemment X mesurable. \square

COMPLÉMENTS ET EXERCICES

1. — Exprimer les indicatrices de $A_1 \cup A_2$, $A_1 \cap A_2$, $A_1 \setminus A_2$, $A_1 \Delta A_2$ (différence symétrique), $\limsup_n A_n$, $\liminf_n A_n$ en fonction des indicatrices de A_1 , A_2 , A_n , ($n \geq 1$).

2. — Montrer qu'une fonction réelle X définie sur un espace mesurable (Ω, \mathfrak{A}) est mesurable, si et seulement si l'ensemble $\{\omega \in \Omega : X(\omega) \leq r\}$ appartient à \mathfrak{A} pour tout r rationnel.

3. — Soit (X_n) une suite de fonctions réelles mesurables définies sur un espace mesurable (Ω, \mathfrak{A}) . Montrer que l'ensemble des éléments ω de Ω pour lesquels la suite $(X_n(\omega))$ est convergente est mesurable.

4. — Soit H_a la fonction de répartition de la mesure de probabilité singulière ϵ_a sur la droite \mathbb{R} . Donner l'expression de H_a («H» comme Heaviside).

5. — Soient a et b deux nombres réels. Déterminer la fonction de répartition de $Y = aX + b$.

6. — Si F est une fonction de répartition continue et si h est un nombre strictement positif, montrer que

$$\Phi_h(x) = \frac{1}{h} \int_x^{x+h} F(t) dt$$

est aussi une fonction de répartition.

7. — Soit $(a_n)_{n \geq 1}$ une numérotation de l'ensemble \mathbb{Q} des nombres rationnels. On considère une variable aléatoire discrète X dont le support S_X est \mathbb{Q} et dont la fonction de masse est : $\pi(x) = 0$, si $x \in \mathbb{R} \setminus \mathbb{Q}$ et $\pi(a_n) = 1/2^n$ pour $a_n \in \mathbb{Q}$ ($n \geq 1$). La fonction de répartition F de X est donnée par :

$$F(x) = P\{X \leq x\} = \sum_{a_n \leq x} \frac{1}{2^n} = \sum_{n \geq 1} \frac{1}{2^n} H_{a_n}(x).$$

Il en résulte que F a toutes les propriétés d'une fonction de répartition, et que ses seuls points de discontinuité sont les rationnels.

8. a) Soit X une variable aléatoire admettant une fonction de répartition F continue et strictement croissante. Montrer que la variable aléatoire $U = F \circ X$ admet pour fonction de répartition la fonction G définie par :

$$(E) \quad G(u) = \begin{cases} 0, & \text{si } u < 0; \\ u, & \text{si } 0 \leq u \leq 1; \\ 1, & \text{si } 1 < u. \end{cases}$$

b) Soit U une variable aléatoire dont la fonction de répartition G est donnée par (E). Soit d'autre part une fonction de répartition F dont on désigne par F^{-1} la fonction inverse généralisée (au sens de Paul Lévy) :

$$F^{-1}(x) = \sup\{y : F(y) \leq x\}.$$

Montrer que la variable aléatoire $X = F^{-1} \circ U$ admet F comme fonction de répartition.

9. — Soit X une variable aléatoire telle que X et $2X$ admettent la même fonction de répartition F . Déterminer F ; que peut-on dire de X ?

10. — Soit $X = (X_1, X_2)$ un vecteur aléatoire à deux dimensions qui prend toutes ses valeurs sur la première bissectrice. Montrer que sa fonction de répartition conjointe $F(x_1, x_2)$ est de la forme :

$$F(x_1, x_2) = h[\min(x_1, x_2)],$$

où $h(\cdot)$ est une fonction réelle. Réciproque ?

11. — Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire à n dimensions défini sur un espace probabilisé $(\Omega, \mathfrak{A}, P)$. On définit sa *fonction de répartition* comme étant l'application F qui envoie la suite (x_1, \dots, x_n) de nombres réels sur le nombre $P\{X_1 \leq x_1, \dots, X_n \leq x_n\}$. Démontrer les propositions suivantes :

- (i) F est une fonction croissante et continue à droite en chacun de ses arguments x_1, \dots, x_n .
- (ii) La limite $\lim F(x_1, \dots, x_n)$ est égale à 0 si l'une au moins des variables x_i tend vers $-\infty$ et elle est égale à 1 si toutes les variables x_i tendent vers $+\infty$.
- (iii) Si $G(x_1, \dots, x_n)$ est une fonction réelle de n variables réelles et si h est un nombre réel, l'accroissement de la fonction G , dû à l'accroissement h donné à la variable x_i , est défini par :

$$\Delta_h^{(i)} G(x_1, \dots, x_n) = G(x_1, \dots, x_i + h, \dots, x_n) - G(x_1, \dots, x_i, \dots, x_n).$$

Montrer que pour toute suite (h_1, \dots, h_n) de nombres réels positifs et toute suite (x_1, \dots, x_n) de nombres réels, on a l'inégalité :

$$\Delta_{h_1}^{(1)} \cdots \Delta_{h_n}^{(n)} F(x_1, \dots, x_n) \geq 0.$$

CHAPITRE 6

PROBABILITÉS CONDITIONNELLES. INDÉPENDANCE

Ce chapitre est consacré tout d'abord à l'étude des mesures de probabilité conditionnelles, un évènement A étant donné. C'est le cas dit élémentaire étudié dans tous les ouvrages d'initiation aux probabilités. On trouve également dans ce chapitre une étude de l'indépendance, non seulement des évènements, mais aussi des classes d'évènements. On y introduit aussi la notion importante de suites de variables aléatoires mutuellement indépendantes.

1. Probabilité conditionnelle. — Supposons que les faces paires d'un dé à six faces soient colorées en blanc et les faces impaires en noir. On lance ce dé; de loin on se rend compte qu'une face blanche est apparue; quelle est la probabilité que l'on ait obtenu un six? Chacun de nous répondra $1/3$, et non pas $1/6$. En effet, le fait d'avoir repéré la face blanche a modifié la pondération des évènements. On ne peut plus prendre l'équirépartition P sur l'ensemble $\Omega = \{1, 2, \dots, 6\}$; on se doit de donner à chaque nombre 2, 4, 6 la pondération $1/3$ et à chaque nombre 1, 3, 5 la pondération 0. Pour rendre compte de la nouvelle information A «il est apparu un nombre pair», il faut introduire une nouvelle pondération. Notons-la $P\{ \cdot | A \}$. Elle est définie par :

$$P\{\{i\} | A\} = \begin{cases} 1/3, & \text{si } i = 2, 4, 6; \\ 0, & \text{autrement.} \end{cases}$$

Puisque $P(A) = 1/2$, on a la relation :

$$P\{\{i\} | A\} = \frac{P(\{i\} \cap A)}{P(A)},$$

pour tout $i = 1, 2, \dots, 6$. L'exemple de ce jeu de dé suggère de prendre la définition ci-après pour la probabilité conditionnelle.

Dans tout le chapitre, on suppose donné un espace probabilisé $(\Omega, \mathfrak{A}, P)$ et les lettres A, B, C, \dots (avec ou sans indice) désignent des évènements (*i.e.* des éléments de la tribu \mathfrak{A}).

THÉORÈME ET DÉFINITION 1.1. — *Soit A un évènement de probabilité $P(A) > 0$. Alors la fonction d'ensembles $P\{ \cdot | A \}$, définie pour tout évènement B de \mathfrak{A} par*

$$(1.1) \quad P(B | A) = \frac{P(B \cap A)}{P(A)},$$

est une mesure de probabilité sur (Ω, \mathfrak{A}) .

On l'appelle *mesure de probabilité A-conditionnelle* ou *mesure de probabilité conditionnelle*, A étant donné.

Démonstration. — Tout d'abord, on a bien $P(\Omega | A) = \frac{P(\Omega \cap A)}{P(A)} = 1$.

D'autre part, si (B_n) désigne une suite d'évènements disjoints deux à deux, alors $(A \cap B_n)$ est aussi une suite d'évènements disjoints deux à deux. De là

$$\begin{aligned} P\left(\sum_n B_n | A\right) &= \frac{P(A \cap \sum_n B_n)}{P(A)} = \frac{P(\sum_n A \cap B_n)}{P(A)} \\ &= \frac{\sum_n P(A \cap B_n)}{P(A)} = \sum_n \frac{P(A \cap B_n)}{P(A)} = \sum_n P(B_n | A). \end{aligned}$$

Ainsi $P(\cdot | A)$ est bien une mesure de probabilité. \square

Notons que la mesure de probabilité A -conditionnelle *est portée par* A ; en d'autres termes, on a les relations :

$$AB = \emptyset \Rightarrow P(B | A) = 0 \quad \text{et} \quad B \supset A \Rightarrow P(B | A) = 1.$$

PROPOSITION 1.2 (Formule du double conditionnement). — *Soient A_1 et A_2 des évènements tels que la probabilité $P(A_1 A_2)$ soit strictement positive. Alors, pour tout évènement A_3 , on a la relation :*

$$P(A_1 A_2 A_3) = P(A_3 | A_1 A_2) P(A_2 | A_1) P(A_1).$$

Démonstration. — L'identité ci-dessus résulte immédiatement de la définition de la probabilité conditionnelle. On a, en effet, $P(A_1) > 0$, puisque $A_1 A_2$ est de probabilité strictement positive. D'où

$$P(A_1 A_2 A_3) = P(A_3 | A_1 A_2) P(A_1 A_2) = P(A_3 | A_1 A_2) P(A_2 | A_1) P(A_1). \quad \square$$

La formule précédente se généralise immédiatement au cas d'un nombre d'évènements supérieur à deux : *soient $n \geq 2$ et A_1, A_2, \dots, A_n une suite de n évènements tels que $P(A_1 A_2 \dots A_{n-1}) > 0$; on a l'identité :*

$$(1.2) \quad \begin{aligned} P(A_1 A_2 \dots A_n) \\ = P(A_n | A_1 \dots A_{n-1}) P(A_{n-1} | A_1 \dots A_{n-2}) \cdots P(A_2 | A_1) P(A_1). \end{aligned}$$

2. Systèmes complets d'évènements. — On dit qu'une suite (A_n) d'évènements est un *système complet*, si

- (i) $i \neq j \Rightarrow A_i A_j = \emptyset$ (les évènements A_n sont deux à deux incompatibles);
- (ii) $P(\sum_n A_n) = \sum_n P(A_n) = 1$ (presque sûrement l'un des évènements A_n se réalise).

En particulier, toute partition dénombrable de Ω en éléments de \mathfrak{A} forme un système complet. Dans la précédente définition, on impose seulement que la probabilité de l'évènement contraire à $\sum_n A_n$ soit nulle. Cet évènement n'est pas nécessairement impossible.

THÉORÈME 2.1 (Formule de Bayes). — Soit (A_n) un système complet d'évènements, tous de probabilité non nulle. Alors, pour tout évènement B , on a :

$$P(B) = \sum_n P(B | A_n)P(A_n) ;$$

si, de plus, $P(B) > 0$, on a, pour tout entier k , l'identité :

$$P(A_k | B) = \frac{P(B | A_k)P(A_k)}{\sum_n P(B | A_n)P(A_n)}.$$

Démonstration. — En posant $\Omega' = \sum_n A_n$, on a, pour tout évènement B , les relations : $B\Omega' = \sum_n BA_n$ et $P(B\Omega'^c) = 0$. D'où $P(B) = P(B\Omega') + P(B\Omega'^c) = P(\sum_n BA_n) = \sum_n P(BA_n) = \sum_n P(B | A_n)P(A_n)$. Enfin,

$$P(A_k | B) = \frac{P(BA_k)}{P(B)} = \frac{P(B | A_k)P(A_k)}{\sum_n P(B | A_n)P(A_n)}. \quad \square$$

Exemple (le problème du tricheur). — Un joueur joue à pile ou face, il parie sur *pile* et obtient *pile*. Quelle est la probabilité qu'il soit un tricheur ? Peut-on répondre à une telle question ?

Notons Ω l'ensemble de toutes les épreuves et soient p, f, h, t , les évènements «le résultat est *pile*», «le résultat est *face*», «le joueur est honnête», «le joueur est un tricheur», respectivement. Pour probabiliser l'ensemble $\{p, f, h, t\}$, on peut d'abord convenir que l'on a $P(p | h) = P(f | h) = 1/2$. On peut également convenir que $P(p | t)$ est égal à 1 (le tricheur obtient *pile* à la demande) et donc $P(f | t) = 0$. On peut poser enfin $P(t) = x$ ($0 \leq x \leq 1$). De la formule de Bayes, on déduit :

$$P(t | p) = \frac{P(p | t)P(t)}{P(p | t)P(t) + P(p | h)P(h)} = \frac{x}{x + (1/2)(1 - x)} = \frac{2x}{x + 1}.$$

On ne peut donc apporter une réponse numérique à ce problème que si l'on sait évaluer la proportion x de tricheurs dans la population. On obtient ainsi une réponse plus ou moins réconfortante, suivant l'opinion que l'on a de l'honnêteté de son prochain !

Exemple. — Un individu est tiré au hasard d'une population où l'on trouve une proportion de 10^{-4} de gens atteints du sida. On lui fait passer le test de détection de la maladie. Sachant que le test donne un résultat positif, quelle est la probabilité pour qu'il ait effectivement le sida ?

Considérons les évènements A_1 «l'individu est atteint du sida», puis A_2 «l'individu n'est pas atteint du sida», enfin B «le test de détection donne un résultat positif». Les données du problème fournissent $P(A_1) = 10^{-4}$, d'où $P(A_2) = 0,9999$. Or il faut encore connaître $P(B | A_1)$ et $P(B | A_2)$, c'est-à-dire les probabilités d'avoir un résultat positif lors de l'application du test si l'individu est atteint du sida, d'une part, et si l'individu ne l'est pas,

d'autre part. Ces probabilités peuvent être fournies par des expérimentations antérieures : prenons, par exemple, $P(B | A_1) = 0,99$, $P(B | A_2) = 0,001$ (les tests ne sont pas infallibles). On trouve alors :

$$P(A_1 | B) = \frac{10^{-4} \times 0,99}{10^{-4} \times 0,99 + 0,9999 \times 0,001} \approx 0,09.$$

On est surpris de la petitesse de cette probabilité. Elle est due au grand nombre d'erreurs de diagnostic provenant de la proportion *énorme* de gens non atteints du sida dans la population ($P(A_2) \gg P(A_1)$). Le dépistage d'une maladie coûte cher.

3. Systèmes de probabilités conditionnelles. — Dans beaucoup de situations, l'expérience fournit un système de probabilités conditionnelles sur des ensembles de suites d'épreuves et il s'agit de voir comment on peut en déduire une probabilité sur l'ensemble de toutes ces suites. On donne ici un résultat sur les suites *finies*. Au chap. 10, exercices 1-9, on étendra ce résultat au cas des suites infinies.

Supposons donnés un entier $n \geq 2$ et un ensemble fini ou dénombrable S . Formons l'ensemble (fondamental) $\Omega = S^n$ et pour tout entier $i = 1, 2, \dots, n$ définissons la *projection* $X_i : \Omega \rightarrow S$, comme étant la fonction qui à tout élément $\omega = (x_1, x_2, \dots, x_n)$ de Ω fait correspondre sa $i^{\text{ième}}$ coordonnée x_i , soit

$$X_i(\omega) = x_i.$$

Dans beaucoup de situations, la suite (x_1, x_2, \dots, x_n) est la suite des valeurs prises par un système aléatoire évoluant au cours du temps (ici discret). La variable aléatoire X_i est interprétée comme étant l'état du système à l'instant i .

Dans le théorème suivant, on suppose données une mesure de probabilité p_1 sur S et une suite de fonctions réelles positives q_2, \dots, q_n , respectivement définies sur S^2, \dots, S^n et satisfaisant, pour tout $i = 2, \dots, n$ et toute suite (x_1, \dots, x_{i-1}) de S^{i-1} , aux relations :

$$(3.1) \quad \sum_{x \in S} q_i(x_1, \dots, x_{i-1}, x) = 1.$$

THÉORÈME 3.1. — *La mesure de probabilité p_1 sur S et la famille (q_i) satisfaisant aux relations (3.1) étant données, il existe une et une seule mesure de probabilité P sur $(\Omega, \mathfrak{P}(\Omega))$ satisfaisant aux propriétés suivantes :*

- (i) $P\{X_1 = x_1\} = p_1(x_1)$, pour tout $x_1 \in S$;
- (ii) $P\{X_{i+1} = x_{i+1} | X_1 = x_1, \dots, X_i = x_i\} = q_{i+1}(x_1, \dots, x_i, x_{i+1})$, pour tout $i = 1, \dots, n-1$ et tout $(x_1, \dots, x_i, x_{i+1}) \in S^{i+1}$ tel que $P\{X_1 = x_1, \dots, X_i = x_i\} > 0$. On a alors, pour tout $(x_1, \dots, x_n) \in S^n$, l'identité :

$$(3.2) \quad P\{X_1 = x_1, \dots, X_n = x_n\} = q_n(x_1, \dots, x_n) \cdots q_2(x_1, x_2) p_1(x_1).$$

Démonstration. — Vérifions d'abord que si une telle mesure P existe, elle satisfait nécessairement la relation (3.2). Considérons, en effet, un élément $\omega = (x_1, \dots, x_n)$ de Ω . Si $p_1(x_1) = 0$, alors $P\{X_1 = x_1\} = 0$, ce qui entraîne $P\{X_1 = x_1, \dots, X_n = x_n\} = 0$. La relation (3.2) est donc vérifiée. Maintenant, si $p_1(x_1) > 0$, notons x_{n+1} un élément fixé de S et posons, par commodité, $q_{n+1}(x_1, \dots, x_n, x_{n+1}) = 0$. On peut alors désigner par $i + 1$ le plus petit entier tel que $2 \leq i + 1 \leq n + 1$ et $q_{i+1}(x_1, \dots, x_{i+1}) = 0$. On a alors, successivement,

$$\begin{aligned} P\{X_1 = x_1, X_2 = x_2\} &= P\{X_2 = x_2 \mid X_1 = x_1\}P\{X_1 = x_1\} \\ &= q_2(x_1, x_2)p_1(x_1) > 0, \\ &\dots = \dots \end{aligned}$$

$$\begin{aligned} P\{X_1 = x_1, \dots, X_i = x_i\} &= P\{X_i = x_i \mid X_1 = x_1, \dots, X_{i-1} = x_{i-1}\} \\ &\quad \times P\{X_1 = x_1, \dots, X_{i-1} = x_{i-1}\} \\ &= q_i(x_1, \dots, x_i) \\ &\quad \times P\{X_1 = x_1, \dots, X_{i-1} = x_{i-1}\} > 0. \end{aligned}$$

Par suite,

$$(3.3) \quad P\{X_1 = x_1, \dots, X_i = x_i\} = q_i(x_1, \dots, x_i) \cdots q_2(x_1, x_2)p_1(x_1).$$

Si $i + 1 = n + 1$, la relation (3.2) est vérifiée. Si $i + 1 \leq n$, on obtient :

$$\begin{aligned} P\{X_1 = x_1, \dots, X_{i+1} = x_{i+1}\} \\ &= P\{X_{i+1} = x_{i+1} \mid X_1 = x_1, \dots, X_i = x_i\}P\{X_1 = x_1, \dots, X_i = x_i\} \\ &= q_{i+1}(x_1, \dots, x_{i+1})P\{X_1 = x_1, \dots, X_i = x_i\} = 0. \end{aligned}$$

D'où

$$P\{X_1 = x_1, \dots, X_n = x_n\} = 0.$$

La relation (3.2) est donc encore vérifiée.

Vérifions, enfin, que la relation (3.2) définit bien une mesure de probabilité sur $(\Omega, \mathfrak{P}(\Omega))$ et que cette mesure satisfait les relations (i) et (ii). Remarquons, tout d'abord, que l'évènement $\{X_1 = x_1, \dots, X_n = x_n\}$ n'est autre que le singleton $\{(x_1, \dots, x_n)\}$ de Ω , auquel on attache une pondération par la formule (3.2). Soit i fixé tel que $1 \leq i \leq n$. En sommant les deux membres de la formule (3.2) successivement par rapport à x_n, \dots, x_{i+1} et en utilisant les propriétés (3.1), on obtient la formule (3.3). La relation (i) est donc, en particulier, vérifiée. Maintenant, si $P\{X_1 = x_1, \dots, X_i = x_i\} > 0$, la formule (3.3) implique immédiatement la relation (ii), par définition même de la probabilité conditionnelle. \square

4. Évènements indépendants. — Soient A et B deux évènements, tous deux de probabilité non nulle. En général, $P(A \mid B) = P(AB)/P(B)$ est différent de $P(A)$. Si $P(A \mid B) = P(A)$, on dit que A est *indépendant de* B .

Il est immédiat de vérifier que si A est indépendant de B , alors B est indépendant de A . On a donc recours plus volontiers à une expression qui reflète cette symétrie et l'on dit que A et B sont *mutuellement indépendants*. De façon générale, on utilise la définition suivante.

Définition. — Deux évènements A et B sont dits *indépendants* (pour la mesure de probabilité P), si l'on a :

$$P(AB) = P(A)P(B).$$

PROPOSITION 4.1. — Soient A, B, C (avec ou sans indices) des évènements.

- (i) Si A et B sont indépendants, alors A et B^c sont indépendants.
- (ii) Si A et B sont indépendants, si A et C sont indépendants et si de plus $C \supset B$, alors A et $C \setminus B$ sont indépendants.
- (iii) Tout évènement est indépendant de tout évènement de probabilité nulle ou de probabilité égale à 1.
- (iv) Si (A_n) est une suite d'évènements deux à deux incompatibles et si A est indépendant de A_n pour tout n , alors A est indépendant de l'union disjointe $\sum_n A_n$.

Démonstration. — Pour établir la propriété (i), on écrit simplement : $P(AB^c) = P(A \setminus AB) = P(A) - P(AB) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c)$. Pour la propriété (ii), on écrit : $P(A(C \setminus B)) = P(AC \setminus AB) = P(AC) - P(AB) = P(A)P(C) - P(A)P(B) = P(A)(P(C) - P(B)) = P(A)P(C \setminus B)$.

Pour établir (iii), considérons les évènements B et C , de probabilité $P(B) = 0$ et $P(C) = 1$. Pour tout évènement A , l'inclusion $AB \subset B$ entraîne $0 \leq P(AB) \leq P(B) = 0$, d'où $0 = P(AB) = P(A)P(B)$. Pour vérifier que A est indépendant de C , on remarque d'abord que A et C^c sont indépendants, puisque C^c est de probabilité nulle. De là, A et C sont indépendants d'après (ii). La propriété (iv) ne fait qu'utiliser la propriété de σ -additivité des probabilités. On a, en effet, $P(A \sum_n A_n) = P(\sum_n AA_n) = \sum_n P(AA_n) = \sum_n P(A)P(A_n) = P(A)P(\sum_n A_n)$. \square

Remarque. — Soit \mathcal{D}_A la classe des évènements qui sont indépendants d'un évènement donné A . On peut reformuler les propriétés ci-dessus en disant que \mathcal{D}_A est une classe d'évènements qui contient Ω et qui est stable par passage au complémentaire, par différence propre et par réunion dénombrable disjointe. Autrement dit, \mathcal{D}_A est un *système de Dynkin* (cf. chap. 2, § 3). On peut montrer que \mathcal{D}_A n'est en général pas stable par intersection, ce n'est en général pas une algèbre. (cf. Remarque 1 ci-dessous.)

Autres remarques

(i) Dans la Proposition 4.1, la première propriété est une conséquence de la deuxième et de la troisième (prendre $C = \Omega$).

(ii) Deux évènements ne peuvent être indépendants s'ils sont incompatibles, sauf si l'un d'eux a une probabilité nulle.

(iii) Les seuls évènements qui sont indépendants d'eux-mêmes sont les évènements de probabilité 0 ou 1.

On prolonge la notion d'indépendance de deux évènements au cas des suites d'évènements. À côté de l'indépendance de deux évènements de la suite, pris deux à deux, on peut aussi définir la notion d'évènements mutuellement indépendants :

Définition. — Soit (A_n) une suite finie ou infinie d'évènements. On dit que les évènements A_1, A_2, \dots sont *mutuellement indépendants*, si pour toute suite finie $A_{i_1}, A_{i_2}, \dots, A_{i_k}$ d'évènements distincts, on a :

$$P(A_{i_1}A_{i_2}\dots A_{i_k}) = P(A_{i_1})P(A_{i_2})\dots P(A_{i_k}).$$

Notons que si la suite (A_n) est finie et se compose de m ($m \geq 2$) évènements distincts, le nombre de conditions imposées dans la définition est égal à :

$$\binom{m}{2} + \binom{m}{3} + \dots + \binom{m}{m} = 2^m - m - 1.$$

Remarque 1. — L'exemple suivant montre que m évènements peuvent être indépendants deux à deux sans être indépendants mutuellement (on dit encore : dans leur ensemble). On lance deux dés et on désigne par A l'évènement «le premier dé amène un nombre pair», par B l'évènement «le second dé amène un nombre impair», par C l'évènement «les deux dés amènent des nombres de même parité».

On a $P(A) = P(B) = P(C) = 1/2$, puis $P(AB) = P(BC) = P(CA) = 1/4$, mais $P(ABC) = 0 \neq P(A)P(B)P(C)$. Cet exemple montre que A peut être indépendant de B et de C séparément, sans l'être de l'intersection $B \cap C$.

Remarque 2. — Prenons maintenant un exemple tiré de l'arithmétique montrant la différence entre indépendance mutuelle et indépendance deux à deux : on considère une urne contenant N boules numérotées de 1 à N . Une expérience consiste à tirer une boule au hasard et à noter son numéro. L'espace probabilisé associé est le triplet $(\Omega, \mathfrak{P}(\Omega), P)$, où $\Omega = \{1, \dots, N\}$ et où P est l'équirépartition sur Ω .

1) Pour chaque diviseur a de N on désigne par E_a l'évènement «la boule tirée porte un numéro divisible par a »; on a immédiatement $P(E_a) = 1/a$.

2) Soient a et b deux diviseurs de N ; on désigne par $[a, b]$ leur p.p.c.m.; c'est encore un diviseur de N . Il résulte alors de la relation $E_a \cap E_b = E_{[a, b]}$, que $P(E_a \cap E_b) = 1/[a, b]$.

On en déduit que les deux propriétés suivantes sont équivalentes :

a) $[a, b] = ab$, i.e. a et b sont premiers entre eux;

b) $P(E_a \cap E_b) = P(E_a)P(E_b)$, i.e., E_a et E_b sont indépendants.

3) Soient n un entier supérieur ou égal à 2 et a_1, \dots, a_n des diviseurs de N ; on désigne par $[a_1, \dots, a_n]$ leur p.p.c.m.; c'est encore un diviseur de N . Il résulte alors de la relation $E_{a_1} \cap \dots \cap E_{a_n} = E_{[a_1, \dots, a_n]}$ que

$P(E_{a_1} \cap \dots \cap E_{a_n}) = 1/[a_1, \dots, a_n]$. De même, on en déduit que les deux propriétés suivantes sont équivalentes :

a) $[a_1, \dots, a_n] = a_1 \cdots a_n$;

b) $P(E_{a_1} \cap \dots \cap E_{a_n}) = P(E_{a_1}) \dots P(E_{a_n})$.

Or on sait que la propriété a) est réalisée, si et seulement si les nombres sont premiers deux à deux. Pour toute partie $J \subset \{1, \dots, n\}$ on a donc $P(\bigcap_{j \in J} E_{a_j}) = \prod_{j \in J} P(E_{a_j})$. La propriété b) équivaut en fait à l'indépendance des évènements E_{a_1}, \dots, E_{a_n} dans leur ensemble.

4) Prenons $N = 12$, $n = 3$, $a_1 = 2$, $a_2 = 3$, $a_3 = 4$. On voit que E_2 et E_3 sont indépendants, ainsi que E_3 et E_4 , mais E_2 et E_4 ne sont *pas* indépendants. La relation d'indépendance n'est *pas* transitive.

5. Indépendance de classes d'évènements. — On prolonge la notion d'indépendance aux classes d'évènements de la façon suivante. Supposons donnée une suite (finie ou infinie) (\mathcal{C}_n) de classes d'évènements.

Définition. — On dit que \mathcal{C}_1 et \mathcal{C}_2 sont *indépendantes*, si quel que soit $A_1 \in \mathcal{C}_1$ et $A_2 \in \mathcal{C}_2$, les évènements A_1 et A_2 sont indépendants.

De même, la suite (\mathcal{C}_n) est une suite de classes *mutuellement indépendantes*, si pour toute sous-suite finie $\mathcal{C}_{i_1}, \dots, \mathcal{C}_{i_k}$ extraite de la suite (\mathcal{C}_n) et pour toute suite d'évènements $A_{i_1} \in \mathcal{C}_{i_1}, \dots, A_{i_k} \in \mathcal{C}_{i_k}$, on a :

$$P(A_{i_1} \dots A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k}).$$

La proposition suivante montre que pour vérifier que deux classes sont indépendantes, il suffit de le vérifier sur des sous-classes suffisamment stables. La notion de système de Dynkin joue alors un rôle essentiel.

PROPOSITION 5.1. — *Soient \mathcal{C}_1 et \mathcal{C}_2 deux classes d'évènements. On suppose qu'elles sont indépendantes et stables par intersection finie. Alors les tribus $\mathfrak{T}(\mathcal{C}_1)$ et $\mathfrak{T}(\mathcal{C}_2)$ engendrées par \mathcal{C}_1 et \mathcal{C}_2 sont indépendantes.*

Démonstration. — Soit \mathcal{E}_1 la classe des évènements qui sont indépendants de chaque évènement de \mathcal{C}_2 . On a observé plus haut que la classe \mathcal{D}_A de tous les évènements qui sont indépendants d'un évènement A formait un système de Dynkin. Comme \mathcal{E}_1 n'est autre que l'intersection $\bigcap \mathcal{D}_A$ ($A \in \mathcal{C}_2$), la classe \mathcal{E}_1 est aussi un système de Dynkin. Comme elle contient \mathcal{C}_1 , elle contient encore le système de Dynkin engendré $\mathfrak{D}(\mathcal{C}_1)$. De là, $\mathfrak{D}(\mathcal{C}_1)$ et \mathcal{C}_2 sont deux classes indépendantes.

De la même façon, on démontre que la classe \mathcal{E}_2 des évènements qui sont indépendants de chaque évènement de $\mathfrak{D}(\mathcal{C}_1)$ est aussi un système de Dynkin. Elle contient \mathcal{C}_2 , donc aussi $\mathfrak{D}(\mathcal{C}_2)$. Par suite, $\mathfrak{D}(\mathcal{C}_1)$ et $\mathfrak{D}(\mathcal{C}_2)$ sont indépendants.

Enfin, comme les classes \mathcal{C}_1 et \mathcal{C}_2 sont stables par intersection finie, les systèmes de Dynkin engendrés par celles-ci sont identiques aux tribus engendrées. D'où $\mathfrak{T}(\mathcal{C}_1) = \mathfrak{D}(\mathcal{C}_1)$ et $\mathfrak{T}(\mathcal{C}_2) = \mathfrak{D}(\mathcal{C}_2)$ sont indépendantes. \square

Comme une algèbre est stable par intersection finie, la Proposition 5.1 admet le corollaire suivant que nous énonçons sous forme de proposition, vu son importance.

PROPOSITION 5.2. — *Si \mathfrak{A}_1 et \mathfrak{A}_2 sont deux algèbres d'évènements, qui sont indépendantes, alors les tribus engendrées $\mathfrak{T}(\mathfrak{A}_1)$ et $\mathfrak{T}(\mathfrak{A}_2)$ sont aussi indépendantes.*

6. Variables aléatoires indépendantes. — On a vu précédemment la notion de tribu engendrée par une variable aléatoire. Si X est une variable aléatoire à n dimensions, définie sur un espace probabilisé $(\Omega, \mathfrak{A}, P)$, la tribu $\mathfrak{T}(X)$ engendrée par X n'est autre que la tribu $X^{-1}(\mathcal{B}^n)$. On peut donc prolonger la notion d'indépendance au cas des variables aléatoires de la façon suivante.

Définition. — Deux variables aléatoires (réelles ou à plusieurs dimensions) X et Y , définies sur un espace probabilisé $(\Omega, \mathfrak{A}, P)$, sont dites *indépendantes*, si les tribus $\mathfrak{T}(X)$ et $\mathfrak{T}(Y)$ sont indépendantes.

De façon plus explicite, soient X et Y deux variables aléatoires, respectivement à n et m dimensions, toutes deux définies sur un espace probabilisé $(\Omega, \mathfrak{A}, P)$. Elles sont *indépendantes*, si pour tout $A \in \mathcal{B}^n$ et tout $B \in \mathcal{B}^m$, on a :

$$P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\}.$$

Une des notions qui revient constamment dans la théorie des probabilités est la notion de suite de variables (mutuellement) indépendantes. La définition formelle en est la suivante :

Définition. — Soit (X_n) une suite de variables aléatoires définies sur un espace probabilisé $(\Omega, \mathfrak{A}, P)$. On dit que cette suite est une suite de *variables aléatoires indépendantes* (on dit encore *variables aléatoires mutuellement indépendantes*, pour éviter toute ambiguïté), si la suite des tribus engendrées $(\mathfrak{T}(X_n))$ est une suite de tribus mutuellement indépendantes.

La formulation utile de cette définition est la suivante : la suite (X_n) est une suite de *variables aléatoires indépendantes*, si pour toute suite finie extraite X_{i_1}, \dots, X_{i_k} et toute suite finie B_1, \dots, B_k de boréliens, on a :

$$P\{X_{i_1} \in B_1, \dots, X_{i_k} \in B_k\} = P\{X_{i_1} \in B_1\} \dots P\{X_{i_k} \in B_k\}.$$

La proposition suivante dit que pour vérifier que deux variables aléatoires réelles X et Y sont indépendantes, il suffit de se restreindre à des sous-classes d'ensembles, à savoir les demi-droites. Les probabilités $P\{X \in A\}$, lorsque $A =] - \infty, x]$, sont alors égales à $P\{X \leq x\}$, c'est-à-dire à $F(x)$, où F est la fonction de répartition de X . Il suffit alors de vérifier que la fonction de répartition du couple est égale au produit des fonctions de répartition de X et de Y .

PROPOSITION 6.1. — Soient $n \geq 2$ et X_1, \dots, X_n une suite de n variables aléatoires réelles, définies sur un espace probabilisé $(\Omega, \mathfrak{A}, \mathbb{P})$. Alors X_1, \dots, X_n sont mutuellement indépendantes, si et seulement si la fonction de répartition F du vecteur $X = (X_1, \dots, X_n)$ est égale au produit des fonctions de répartition F_1 de X_1, \dots, F_n de X_n , c'est-à-dire si pour toute suite (x_1, \dots, x_n) de \mathbb{R}^n , on a :

$$F(x_1, \dots, x_n) = F_1(x_1) \dots F_n(x_n).$$

Démonstration. — On donne la démonstration pour $n = 2$. Si X_1 et X_2 sont indépendantes, on a $\mathbb{P}\{X_1 \in B_1, X_2 \in B_2\} = \mathbb{P}\{X_1 \in B_1\}\mathbb{P}\{X_2 \in B_2\}$, pour tout couple de boréliens B_1, B_2 . En prenant $B_1 =]-\infty, x_1]$ et $B_2 =]-\infty, x_2]$, on obtient l'équation

$$(6.1) \quad F(x_1, x_2) = F_1(x_1)F_2(x_2).$$

Réciproquement, soit \mathcal{C}_i la classe des évènements $\{X_i \leq x_i\}$ ($i = 1, 2$). Cette classe est stable par intersection finie. L'équation (6.1) montre que \mathcal{C}_1 et \mathcal{C}_2 sont indépendantes. De là, leurs tribus engendrées sont indépendantes, mais celles-ci ne sont autres que les tribus $\mathfrak{T}(X_1)$ et $\mathfrak{T}(X_2)$. Les variables aléatoires X_1 et X_2 sont donc indépendantes. \square

La dernière proposition est fort utile lorsque l'on considère des transformations de variables aléatoires et que l'on veut s'assurer que les variables transformées sont toujours indépendantes.

PROPOSITION 6.2. — Soient $n \geq 2$ et X_1, \dots, X_n une suite de n variables aléatoires à m dimensions, définies sur un espace probabilisé $(\Omega, \mathfrak{A}, \mathbb{P})$ et supposées mutuellement indépendantes. Soient, d'autre part, $f_i : (\mathbb{R}^m, \mathcal{B}^m) \rightarrow (\mathbb{R}^p, \mathcal{B}^p)$ ($i = 1, \dots, n$) des fonctions mesurables. Alors $f_1 \circ X_1, \dots, f_n \circ X_n$ sont des variables aléatoires à p dimensions, mutuellement indépendantes.

Démonstration. — En effet,

$$\begin{aligned} \mathbb{P}\{f_1 \circ X_1 \in B_1, \dots, f_n \circ X_n \in B_n\} &= \mathbb{P}\{X_1 \in f_1^{-1}(B_1), \dots, X_n \in f_n^{-1}(B_n)\} \\ &= \mathbb{P}\{X_1 \in f_1^{-1}(B_1)\} \dots \mathbb{P}\{X_n \in f_n^{-1}(B_n)\} \\ &= \mathbb{P}\{f_1 \circ X_1 \in B_1\} \dots \mathbb{P}\{f_n \circ X_n \in B_n\}. \quad \square \end{aligned}$$

COMPLÉMENTS ET EXERCICES

1. — Soient A, B deux évènements. Montrer que si A et B sont indépendants, alors les tribus engendrées $\mathfrak{T}(A)$ et $\mathfrak{T}(B)$ sont indépendantes.

2. a) Si \mathcal{C}_1 et \mathcal{C}_2 sont deux classes indépendantes d'évènements, alors les classes monotones engendrées $\mathfrak{M}(\mathcal{C}_1)$ et $\mathfrak{M}(\mathcal{C}_2)$ sont indépendantes.

b) Si \mathfrak{A}_1 et \mathfrak{A}_2 sont deux algèbres d'évènements, qui sont indépendantes, alors les tribus engendrées $\mathfrak{T}(\mathfrak{A}_1)$ et $\mathfrak{T}(\mathfrak{A}_2)$ sont aussi indépendantes.

3. — On lance un dé parfait et l'on considère les deux évènements :

A «le point amené est divisible par 2» ;

B «le point amené est divisible par 3».

Montrer que les évènements A et B sont indépendants.

4. a) Montrer que si A et B sont deux évènements indépendants tels que A entraîne B , alors on a : ou $P(B) = 1$, ou $P(A) = 0$.

b) Montrer que si A est indépendant de lui-même, alors $P(A) = 0$ ou 1 .

c) Montrer que si $P(A) = 0$ ou 1 , alors A est indépendant de tout évènement.

d) (J.-P. Dion) La relation d'indépendance n'est *pas* transitive : il suffit de prendre deux évènements A, B indépendants avec $0 < P(A) < 1$. Alors A est indépendant de B et B de A , mais A n'est pas indépendant de A .

5. — Supposons A indépendant de $B \cap C$ et de $B \cup C$, puis B indépendant de $C \cap A$ et enfin C indépendant de $A \cap B$. Supposons, en outre, $P(A), P(B), P(C)$ strictement positifs. Alors A, B, C sont mutuellement indépendants.

6. — Montrer que le cas suivant peut se présenter : A est indépendant de $B \cap C$ et de $B \cup C$, mais ni de B , ni de C .

7. — Soient A, B, C trois évènements tels que A et B sont indépendants conditionnellement à C et à C^c et que A et C sont indépendants. Montrer que A et B sont alors indépendants.

On montre, de même : soit (X, Y, Z) un triplet de variables aléatoires, tel que X et Y sont indépendants conditionnellement à Z et que X et Z sont indépendants. Alors X et Y sont indépendants.

8. — Dans les deux exemples suivants, il convient, avant de calculer les probabilités conditionnelles demandées, de construire un triplet décrivant l'expérience.

a) Un père de famille déclare avoir deux enfants. Calculer la probabilité pour que les deux soient des garçons, sachant que

α) l'un au moins est un garçon; β) l'aîné est un garçon.

b) On choisit au hasard un enfant dans une famille de deux enfants. Sachant que l'enfant choisi est un garçon, quelle est la probabilité pour que les deux enfants de la famille soient des garçons?

9. — Trouver une condition nécessaire et suffisante pour qu'une variable aléatoire X soit indépendante d'elle-même.

10. — Soient X_1, X_2 deux variables aléatoires indépendantes ayant la loi de probabilité commune $\frac{1}{2}(\varepsilon_{-1} + \varepsilon_{+1})$. Est-ce que les trois variables $X_1, X_2, X_3 = X_1 X_2$ sont mutuellement indépendantes ou indépendantes deux à deux?

11. — Soit (X_1, \dots, X_n) un système de n variables aléatoires mutuellement indépendantes dont les fonctions de répartition sont respectivement F_1, \dots, F_n . Déterminer les fonctions de répartition de $Y = \max(X_1, \dots, X_n)$ et de $Z = \min(X_1, \dots, X_n)$.

12. — On désigne par $P_r(k)$ ($r \geq 1$) la probabilité pour qu'un central téléphonique reçoive k appels en r minutes. On suppose que le nombre d'appels reçus dans deux intervalles de temps disjoints sont deux variables aléatoires indépendantes.

a) Calculer en fonction de $P_1(k)$ ($k \geq 0$) la probabilité pour que le central reçoive s appels en deux minutes.

b) Si $P_1(k) = e^{-a} \frac{a^k}{k!}$ ($a > 0; k \in \mathbb{N}$), calculer $P_r(k)$ pour tout $r \geq 1$.

13. *Tirages avec et sans remise.* — Une urne contient $r + s$ boules dont r blanches et s noires ($r, s \geq 1$). On procède à n tirages successifs ($n \geq 1$), étant entendu qu'après chaque tirage on remet (resp. on ne remet pas) la boule tirée dans l'urne. On désigne par A_k ($k = 1, \dots, n$) l'évènement «on amène une boule blanche au k^e tirage» et l'on introduit les variables aléatoires : $X_k = I_{A_k}$ ($k = 1, \dots, n$) et $S_n = X_1 + \dots + X_n$ (nombre de blanches amenées au cours des n tirages).

Tirage avec remise; modèle binomial. — On prend pour Ω l'ensemble des éléments $\omega = A_1^{\varepsilon_1} \cap \dots \cap A_n^{\varepsilon_n}$, où $A^\varepsilon = A$, si $\varepsilon = 1$ et $A^\varepsilon = A^c$, si $\varepsilon = 0$, et pour mesure de probabilité P sur Ω celle définie par :

$$P(\{\omega\}) = P(A_1^{\varepsilon_1}) \dots P(A_n^{\varepsilon_n}), \quad \text{où } P(A_1) = \dots = P(A_n) = p.$$

Alors

a) Les variables aléatoires X_1, \dots, X_n sont mutuellement indépendantes, identiquement distribuées, la loi de probabilité de X_k étant :

$$P\{X_k = 1\} = p, \quad P\{X_k = 0\} = 1 - p, \quad k = 1, \dots, n.$$

b) La loi de probabilité de S_n est donnée par :

$$P\{S_n = i\} = \binom{n}{i} p^i (1 - p)^{n-i} \quad 0 \leq i \leq n.$$

La variable aléatoire S_n suit une *loi binomiale*, d'où le nom du modèle.

Tirage sans remise ; modèle hypergéométrique. — Dans ce cas, le procédé est exhaustif, l'urne est vide après le $(r + s)^e$ tirage, et $1 \leq n \leq r + s$. Étudions le $(r + s)$ -uplet (X_1, \dots, X_{r+s}) . A cet effet, nous prenons pour Ω l'ensemble des éléments $\omega = A_1^{\varepsilon_1} \cap \dots \cap A_{r+s}^{\varepsilon_{r+s}}$, où $(\varepsilon_1, \dots, \varepsilon_{r+s})$ est une suite contenant exactement r chiffres 1 et s chiffres 0. Les boules blanches étant indiscernables (de même que les noires), on a $\text{card } \Omega = \binom{r+s}{r}$. Nous prenons ensuite pour P l'équirépartition sur Ω . Alors

a) Les variables aléatoires X_1, \dots, X_{r+s} ne sont *pas* mutuellement indépendantes (on a, par exemple, $X_1 + \dots + X_{r+s} = r$), mais elles sont identiquement distribuées, la loi de probabilité de X_k étant donnée par :

$$P\{X_k = 1\} = \frac{r}{r+s} = p, \quad P\{X_k = 0\} = 1 - p, \quad k = 1, \dots, r+s.$$

b) Soit $1 \leq n \leq r + s$. La loi de probabilité de $S_n = X_1 + \dots + X_n$ est donnée par

$$P\{S_n = i\} = \begin{cases} \frac{\binom{r}{i} \binom{s}{n-i}}{\binom{r+s}{n}} & \text{si } \max(0, n-s) \leq i \leq \min(n, r); \\ 0, & \text{sinon.} \end{cases}$$

La variable aléatoire S_n suit une *loi hypergéométrique*, d'où le nom du modèle.

14. — Reprendre l'exercice 13, avec les mêmes notations. Calculer l'expression $P\{X_k = 1 \mid S_n = i\}$, $k \leq n$, dans le cas du tirage avec remise, puis sans remise.

15. *Généralisation de l'exercice 13 (modèle multinomial).* — On suppose que l'urne contient r_1 boules de couleur C_1, \dots, r_k boules de couleur C_k , les couleurs C_1, \dots, C_k étant distinctes, et l'on fait la même expérience que dans l'exercice 13 (n extractions suivies de n remises). On pose : $r_1 + \dots + r_k = m$, $p_i = r_i/m$ ($1 \leq i \leq k$). On désigne par A_{ij} l'évènement « on amène une boule de couleur C_i au j^e tirage » ($1 \leq i \leq k, 1 \leq j \leq n$). Enfin, on introduit les variables aléatoires :

$$X_{ij} = I_{A_{ij}} \quad (1 \leq i \leq k, 1 \leq j \leq n); \quad X_i = \sum_{j=1}^n X_{ij} \quad (1 \leq i \leq k).$$

La variable aléatoire X_i représente le nombre de boules de couleur C_i tirées au cours de n tirages. Montrer que l'on peut construire un triplet $(\Omega, \mathfrak{F}(\Omega), P)$ tel que :

a) Les variables aléatoires $X_{i_1,1}, \dots, X_{i_n,n}$ sont mutuellement indépendantes pour toute suite $(i_1, \dots, i_n) \in \{1, \dots, k\}^n$. En outre

$$P\{X_{ij} = 1\} = p_i \quad (1 \leq i \leq k, 1 \leq j \leq n).$$

b) La loi de probabilité du vecteur aléatoire à k dimensions $X = (X_1, \dots, X_k)$ est donnée par :

$$P\{X_1 = n_1, \dots, X_k = n_k\} = \binom{n}{n_1, \dots, n_k} p_1^{n_1} \dots p_k^{n_k}.$$

C'est la loi *multinomiale*.

16. — Trois personnes A, B, C sont placées « au hasard » sur une ligne droite. On considère les deux événements :

E « B est à la droite de A » ;

F « C est à la droite de A ».

Les événements E et F sont-ils indépendants, si l'on prend l'équirépartition sur l'ensemble fondamental ?

17. — Soit Ω l'ensemble des huit issues résultant de trois jets consécutifs d'une pièce de monnaie. On considère les deux événements :

A « la première pièce amène pile » ;

B « pile est amené au moins deux fois ».

a) Les événements A et B sont-ils indépendants, si l'on munit Ω de l'équirépartition ?

b) Existe-t-il une mesure de probabilité P sur Ω telle que A et B soient P -indépendants ?

18 (E. Kosmanek). — Soient $(\Omega, \mathfrak{A}, P)$ un espace probabilisé et A, B deux événements de \mathfrak{A} . Alors $|P(A \cap B) - P(A)P(B)| \leq \frac{1}{4}$.

On peut donner plusieurs démonstrations de cette inégalité qui découlent en fait de l'inégalité de Schwarz. Une démonstration directe consiste à considérer les « atomes » $A \cap B, A \cap B^c, A^c \cap B, A^c \cap B^c$. En notant $\alpha, \beta, \gamma, \delta$, leurs probabilités respectives, on a la relation $\alpha + \beta + \gamma + \delta = 1$. Posant $e(A, B) = P(A \cap B) - P(A)P(B)$, on en déduit : $e(A, B) = \alpha - (\alpha + \beta)(\alpha + \gamma) = \alpha(1 - \alpha - \beta - \gamma) - \beta\gamma = \alpha\delta - \beta\gamma$, d'où $e(A, B) \leq \alpha\delta \leq \frac{1}{4}$ (puisque $\alpha, \delta \geq 0, \alpha + \delta \leq 1$) et $e(A, B) \geq -\beta\gamma \geq -\frac{1}{4}$ (puisque $\beta, \gamma \geq 0, \beta + \gamma \leq 1$).

On remarque que l'identité $\alpha\delta - \beta\gamma = 0$ fournit une condition nécessaire et suffisante pour que les événements A et B soient indépendants.

19. — On dispose d'un dé parfait. Imaginer une expérience comportant douze événements disjoints équiprobables.

VARIABLES ALÉATOIRES DISCRÈTES. LOIS USUELLES

Le but de ce chapitre est de passer en revue les principales lois de probabilité discrètes, loi binomiale, loi hypergéométrique, loi géométrique, loi de Poisson, et de donner le champ d'application de ces lois. Les problèmes des « boîtes d'allumettes de Banach », de la « poissonisation » et du « paradoxe de l'inspection » sont traités dans les exercices 3, 8 et 9.

1. Variables aléatoires discrètes. — La notion de mesure de probabilité discrète a été introduite au chap. 4, § 1. Par ailleurs, on a noté au chap. 5, § 4, Remarque 1, que si P est une mesure de probabilité sur l'espace $(\mathbb{R}^n, \mathcal{B}^n)$, par exemple discrète, on peut toujours trouver une variable aléatoire qui admette P comme loi de probabilité. Il est donc naturel d'introduire la définition suivante.

Définition. — Une variable aléatoire X , à valeurs dans \mathbb{R}^n , est dite *discrète*, si la loi de probabilité P_X est une mesure de probabilité discrète sur $(\mathbb{R}^n, \mathcal{B}^n)$.

Soit $P_X = \sum_k \alpha_k \varepsilon_{x_k}$ la loi de probabilité d'une telle variable aléatoire. Comme la tribu borélienne \mathcal{B}^n contient tous les ensembles réduits à un seul élément $\{x\}$ ($x \in \mathbb{R}^n$), on peut écrire :

$$(1.1) \quad P_X\{\{x\}\} = \begin{cases} \alpha_k, & \text{si } x = x_k; \\ 0, & \text{si } x \notin \{x_1, x_2, \dots\}. \end{cases}$$

En particulier, si une telle variable aléatoire X est définie sur l'espace probabilisé $(\Omega, \mathfrak{A}, P)$, on peut écrire, pour tout $x \in \mathbb{R}^n$,

$$P_X\{\{x\}\} = P\{X \in \{x\}\} = P\{X = x\},$$

qui se lit « probabilité que X soit égal à x ». En récrivant les formules (1.1), on obtient ainsi :

$$P\{X = x_k\} = \alpha_k \quad \text{et} \quad P\{X = x\} = 0 \text{ si } x \notin \{x_1, x_2, \dots\}.$$

De même, pour tout $B \in \mathcal{B}^n$, on a $P\{X \in B\} = \sum P\{X = x_k\}$, où la sommation est étendue à tous les x_k pour lesquels $x_k \in B$.

Si x_0 est un élément de \mathbb{R}^n et si la loi de probabilité de la variable aléatoire X est *singulière* et égale à ε_{x_0} , on a :

$$P\{X = x_0\} = 1 \quad \text{et} \quad P\{X = x\} = 0 \text{ si } x \neq x_0.$$

On dit alors que X est *P-presque sûrement constante* (égale à x_0). Réciproquement, la loi de probabilité d'une fonction constante est singulière.

De même si $X : (\Omega, \mathfrak{A}, P) \rightarrow (\mathbb{R}^p, \mathcal{B}^p)$ est une variable aléatoire telle que $X(\Omega)$ est *au plus dénombrable*, alors X est une variable aléatoire discrète. En particulier, lorsque $p = 1$ et que $X(\Omega)$ est fini, on peut écrire

$$X = \sum_{k=1}^n x_k I_{A_k},$$

où $X(\Omega) = \{x_1, x_2, \dots, x_n\}$ et $A_k = X^{-1}(\{x_k\}) = \{X = x_k\}$ ($1 \leq k \leq n$). Les ensembles A_k ($1 \leq k \leq n$) appartiennent à \mathfrak{A} et sont disjoints deux à deux. Une telle variable aléatoire est dite *simple* ou *étagée*. Sa loi de probabilité P_X est donnée par :

$$P_X = \sum_{k=1}^n P(A_k) \varepsilon_{x_k}.$$

Une variable aléatoire simple est discrète, elle ne peut prendre qu'un nombre *fini* de valeurs (les x_k). En revanche, chacune de ces valeurs peut être prise en une infinité non dénombrable de points $\omega \in \Omega$. En effet, des ensembles A_k peuvent admettre la puissance du continu.

Dans les paragraphes suivants, nous passons en revue quelques mesures de probabilité discrètes rencontrées dans les applications. Pour des raisons de commodité, elles sont définies sur l'espace $(\mathbb{R}, \mathcal{B}^1)$.

2. La loi binomiale. — Soit p un nombre réel tel que $0 \leq p \leq 1$ et n un nombre entier positif. La mesure de probabilité $B(n, p)$ définie sur $(\mathbb{R}, \mathcal{B}^1)$ par

$$B(n, p) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} \varepsilon_k,$$

où $q = 1 - p$, est appelée *loi binomiale* de paramètres (n, p) . La loi $B(1, p) = q\varepsilon_0 + p\varepsilon_1$ est appelée *loi de Bernoulli* de paramètre p .

Définition. — Une variable aléatoire réelle X ayant une loi de probabilité binomiale $B(n, p)$ est dite *binomiale* de paramètres (n, p) .

Par exemple, dans le *tirage avec remise*, si la proportion de boules blanches dans l'urne est égale à p et si le nombre de tirages est égal à n , alors la variable aléatoire X « nombre de boules blanches tirées » est une variable aléatoire binomiale de paramètres (n, p) . De même, si A est un évènement d'un espace probabilisé $(\Omega, \mathfrak{A}, P)$, la variable aléatoire I_A est une variable de Bernoulli de paramètre $p = P(A)$.

3. La loi hypergéométrique. — Soient n, N, M trois entiers positifs tels que $n \leq N$, $M < N$. La mesure de probabilité $H(n, N, M)$ définie par

$$H(n, N, M) = \sum_k \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \varepsilon_k,$$

où $\max\{0, n - (N - M)\} \leq k \leq \min\{n, M\}$, est appelée *loi hypergéométrique*. Le fait que

$$\sum_k \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} = 1$$

se déduit directement de l'identité $(1+z)^M(1+z)^{N-M} = (1+z)^N$, en calculant le coefficient de z^n dans les deux membres. On appelle de même *variable aléatoire hypergéométrique* une variable aléatoire dont la loi de probabilité est hypergéométrique.

3.1. *Exemple (le tirage sans remise)*. — Une urne contient M boules blanches et $N - M$ boules noires ($M < N$). On tire, sans les remettre dans l'urne, n boules successivement ($n \leq N$). Le nombre X de boules blanches amenées parmi ces n boules est une variable aléatoire hypergéométrique de paramètres (n, N, M) . On a par exemple, pour $\max\{0, n - (N - M)\} \leq k \leq \min\{n, M\}$ la probabilité :

$$P\{X = k\} = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}.$$

Il est bon de noter, bien que cela ne soit pas apparent, que cette expression est symétrique en M et n .

Remarque. — Lorsque, n et k restant fixés ($0 \leq k \leq n$), on fait tendre M et $N - M$ vers l'infini de telle sorte que $M/N \rightarrow p \in]0, 1[$, un calcul de routine montre que

$$\frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \rightarrow \binom{n}{k} p^k (1-p)^{n-k}.$$

On dit que, dans les conditions indiquées, la loi hypergéométrique $H(n, N, M)$ converge vers la loi binomiale $B(n, p)$ (cf. chap. 16, § 6).

3.2. *Application : comment compter les poissons dans un étang*. — Un étang contient un nombre *inconnu* $N \geq 1$ de poissons. Pour déterminer N on effectue une première pêche qui amène $r \geq 1$ poissons que l'on marque, puis que l'on relâche. On effectue ensuite une deuxième pêche qui amène $n \geq 1$ poissons, parmi lesquels on trouve $k \geq 0$ poissons marqués. On se propose d'estimer N à partir de k .

a) La probabilité pour que, l'étang contenant N poissons, la deuxième pêche ait amené $k \geq 0$ poissons marqués est

$$(3.2.1) \quad p(k, N) = \frac{\binom{r}{k} \binom{N-r}{n-k}}{\binom{N}{n}}.$$

(N est inconnu; r, n, k sont connus, ce sont des observations.)

b) On a pêché en tout $r + (n - k)$ poissons différents. Par conséquent, $N \geq r + (n - k)$: c'est tout ce qu'on peut affirmer avec certitude. Il est parfaitement possible que l'étang ne contienne que $r + (n - k)$ poissons, mais cet évènement est hautement improbable.

c) Pour estimer N on applique le principe du maximum de vraisemblance, c'est-à-dire, les nombres r , n , k étant fixés, on essaie de déterminer N de telle sorte que l'expression (3.2.1) soit maximale. La valeur \hat{N} (si elle existe) qui réalise ce maximum est appelée l'*estimation de N par le maximum de vraisemblance*.

d) Montrons que si $k \geq 1$, il existe une valeur \hat{N} et une seule qui maximise (3.2.1) et cette valeur est l'entier le plus proche de nr/k .

En effet, formons le rapport

$$\frac{p(k, N)}{p(k, N-1)} = \frac{\binom{N-r}{n-k} \binom{N-1}{n}}{\binom{N-r-1}{n-k} \binom{N}{n}} = \frac{(N-r)(N-n)}{(N-r-n+k)N}.$$

Ce rapport est supérieur à 1 ou inférieur à 1, selon que $Nk < nr$ ou $Nk > nr$. Ceci montre que lorsque N croît, la suite de terme général $p(k, N)$ commence par croître, puis elle décroît et atteint son maximum lorsque N est égal à l'entier le plus proche de nr/k .

Supposons $r = n = k$; alors $\hat{N} = r$. En d'autres termes, si la deuxième pêche amène exactement autant de poissons que la première et si ces poissons sont tous marqués, alors l'estimation par le maximum de vraisemblance du nombre de poissons de l'étang est égal au nombre minimum de poissons. \square

Application numérique. — Prenons $r = n = 1000$, $k = 100$. Le nombre minimum des poissons dans l'étang est $r + (n - k) = 1.900$. L'estimation de N par le maximum de vraisemblance est $\hat{N} = \frac{1000 \times 1000}{100} = 10.000$.

e) *Cas où $k = 0$* : On a vu que l'estimation de N par le maximum de vraisemblance est possible lorsque $k \geq 1$. Dans le cas $k = 0$, c'est-à-dire, le cas où, lors de la deuxième pêche, on n'a amené aucun poisson marqué sur les n poissons pêchés, on peut inférer que le nombre N de poissons dans l'étang est *très grand*. Le calcul conforte cette intuition. En effet,

$$\frac{p(0, N)}{p(0, N-1)} = \frac{(N-r)(N-n)}{(N-r-n)N} > 1.$$

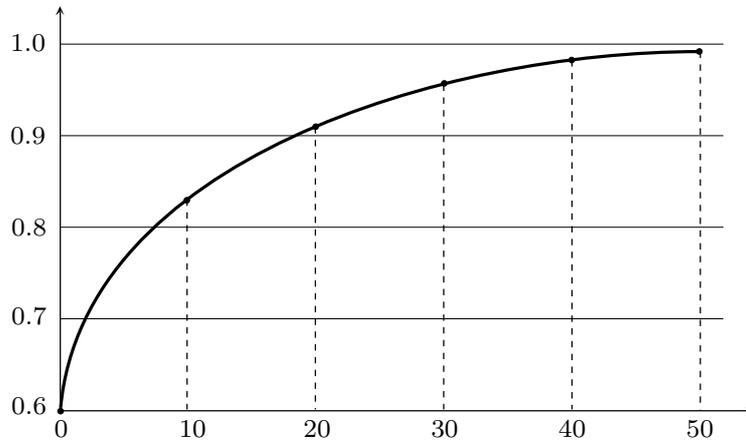
La suite de terme général $p(0, N)$ est strictement *croissante*; il n'existe donc pas de valeur de N qui maximise $p(0, N)$; mais on constate que $p(0, N)$ est d'autant plus grand que N est grand.

3.3. Loi hypergéométrique et décisions judiciaires. — Supposons que parmi 500 magistrats, conseillers à la Cour d'Appel, il y en ait $r = 200$ qui déclarent avoir des affinités avec les partis de gauche (appelons-les les magistrats «de gauche») et $s = 300$ qui pensent pencher politiquement vers les partis de droite (appelons-les les magistrats «de droite»). On choisit «au hasard» $n = 2p + 1$ magistrats parmi ces 500 magistrats pour former un tribunal. Quelle est la probabilité pour que ce tribunal ait une majorité de droite ?

Notons tout d'abord que la proportion de magistrats de droite dans l'ensemble de tous les magistrats est de $300/500 = 60\%$. On doit avoir, ensuite, $1 \leq 2p + 1 \leq 500$, soit $0 \leq p \leq 249$. Désignons par S_{2p+1} le nombre de magistrats de gauche figurant dans le tribunal. La probabilité cherchée est $P_{2p+1} = P\{S_{2p+1} \leq p\}$, soit,

$$P_{2p+1} = \sum_{k=0}^p P\{S_{2p+1} = k\} = \sum_{k=0}^p \frac{\binom{200}{k} \binom{300}{2p+1-k}}{\binom{500}{2p+1}}.$$

(On pose $\binom{n}{k} = 0$, si les entiers n, k ne vérifient pas $0 \leq k \leq n$.)



Courbe représentative de la probabilité P_{2p+1} en fonction de p .

p	0	1	2	3	4	5	8	12	28	249
P_{2p+1}	0,6	0,648	0,683	0,711	0,735	0,756	0,805	0,852	0,948	1

Fig. 1

L'étude des premières valeurs de P_{2p+1} montre que l'application $p \mapsto P_{2p+1}$ (cf. Fig. 1) croît très rapidement vers 1. Autrement dit, dès que l'effectif du tribunal augmente, il est quasiment certain que tribunal aura une majorité à droite. On ne peut donc jamais dans la composition d'un tribunal respecter l'équilibre politique de l'ensemble des magistrats!

4. La loi géométrique. — Soit p un nombre réel compris entre 0 et 1 et $q = 1 - p$. La mesure de probabilité P définie sur $(\mathbb{R}, \mathcal{B}^1)$ par

$$P = \sum_{k=1}^{\infty} pq^{k-1} \varepsilon_k$$

est appelée *loi géométrique de paramètre p* , on la notera $G(p)$. On définit bien là une mesure de probabilité, puisque $\sum_{k=1}^{\infty} pq^{k-1} = (p/(1 - q)) = 1$. On introduit souvent la *fonction de survie*

$$r(n) = P\{X > n\} = \sum_{k \geq n+1} pq^{k-1} = q^n.$$

On appelle quelquefois loi géométrique de paramètre p la loi définie par

$$P = \sum_{k=0}^{\infty} pq^k \varepsilon_k.$$

Il conviendra dans chaque cas de s'assurer de quelle loi géométrique il s'agit.

Exemple. — Un joueur procède à une suite de parties indépendantes de « pile » ou « face » et décide de s'arrêter de jouer dès que, *pour la première fois*, il aura amené « pile ». On s'intéresse au nombre X de parties qu'il lui faudra jouer pour réaliser son objectif. Pour définir X , on commence par introduire l'ensemble Ω de toutes suites *infinies* d'issues possibles : ainsi un élément $\omega \in \Omega$ est une suite $(\delta_1, \delta_2, \dots)$ de 0 et de 1, en convenant du fait que le terme général δ_k vaut 0 ou 1, suivant qu'à la $k^{\text{ième}}$ partie « face » ou « pile » apparaît.

Si l'on veut faire apparaître X comme une variable aléatoire définie sur Ω , l'évènement $\{X = k\}$ « le joueur s'arrête de jouer à la $k^{\text{ième}}$ partie » doit être probabilisable pour tout k fini. Il en est de même de l'évènement $\{X = \infty\}$ « le joueur joue indéfiniment » ! Pour tout $k = 1, 2, \dots$, notons A_k l'ensemble des suites $\omega = (\delta_1, \delta_2, \dots)$ telles que $\delta_k = 1$. Avec la convention prise, A_k est l'évènement « "pile" apparaît à la $k^{\text{ième}}$ partie ». Par conséquent, pour tout k fini, on a $\{X = k\} = A_1^c \dots A_{k-1}^c A_k$; de plus, $\{X = \infty\} = \lim_k A_1^c \dots A_k^c$. Ainsi X est une variable aléatoire définie sur (Ω, \mathfrak{A}) , si on prend pour \mathfrak{A} la tribu engendrée par les A_k .

Enfin, pour rendre compte de l'indépendance des parties et du fait que la probabilité d'amener « pile » en un coup est p ($0 \leq p \leq 1$), il faut prouver qu'il existe une mesure de probabilité P sur (Ω, \mathfrak{A}) telle que pour toute suite finie (i_1, i_2, \dots, i_k) d'entiers *distincts*, on ait $P(A_{i_1} A_{i_2} \dots A_{i_k}) = p^k$. Ce résultat sera établi au cours des Exercices 1-7 du chap. 10.

L'espace probabilisé $(\Omega, \mathfrak{A}, P)$ étant ainsi construit, on voit alors que X est définie sur cet espace et prend ses valeurs dans $(\overline{\mathbb{R}}, \overline{\mathcal{B}}^1)$, en désignant par $\overline{\mathbb{R}}$ la droite *achevée* et par $\overline{\mathcal{B}}^1$ la tribu engendrée par $\mathcal{B}^1 \cup \{+\infty\} \cup \{-\infty\}$. On peut alors calculer la loi de probabilité de X , soit, en posant $q = 1 - p$,

$$\begin{aligned} P\{X = 1\} &= P(A_1) = p; \\ P\{X = k\} &= P(A_1^c \dots A_{k-1}^c A_k) = q^{k-1} p, \quad \text{pour } k \geq 2; \\ P\{X = \infty\} &= P(\lim_k A_1^c \dots A_k^c) = \lim_k P(A_1^c \dots A_k^c) = \lim_k q^k = 0. \end{aligned}$$

Si l'on néglige la valeur $+\infty$, on voit que X suit la *loi géométrique* de paramètre p .

5. La loi de Poisson. — Soit λ un nombre réel strictement positif. La mesure de probabilité π_λ définie sur $(\mathbb{R}, \mathcal{B}^1)$ par

$$(5.1) \quad \pi_\lambda = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \varepsilon_k,$$

est appelée *loi de Poisson de paramètre* λ . (On la note souvent aussi $\mathcal{P}(\lambda)$.) Une variable aléatoire suivant la loi π_λ est dite *de Poisson de paramètre* λ .

LEMME 5.1. — *Pour tout entier* $k \geq 1$ *fixé, on a :*

$$\lim_n \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Le calcul de cette limite est banal et n'est pas reproduit ici. Le lemme dit essentiellement que si $B(n, p)$ est une loi binomiale telle que les deux paramètres n et p sont liés par la relation $np = \lambda > 0$, alors pour n grand la probabilité $\binom{n}{k} p^k q^{n-k}$ (pour qu'une variable binomiale de paramètres (n, p) prenne la valeur k) est approximativement égal à $e^{-\lambda} \lambda^k / k!$, qui est la probabilité pour qu'une variable de Poisson de paramètre λ prenne la valeur k . On dit encore que la loi de Poisson est la loi des événements « rares ».

Remarque. — D'un point de vue *pratique*, on est conduit à introduire la loi de Poisson dans les problèmes du type suivant. Supposons que l'on fasse des prélèvements de n unités dans une population ne comportant que deux sortes d'individus A et B en proportion p et q ($p + q = 1$). Si n est grand et p voisin de 0 de sorte que np soit, disons, compris entre 1 et 10, on peut admettre que le nombre d'individus de l'espèce A dans un prélèvement est approximativement une variable aléatoire de Poisson de paramètre $\lambda = np$.

Soit X le nombre d'individus prélevés de type A . La variable aléatoire X est *théoriquement* une variable binomiale de paramètres (n, p) . Avec l'approximation de la loi de Poisson, la probabilité pour que X prenne la valeur k n'est pas nulle, même pour $k > n$, mais dans ce dernier cas, elle est très faible si les conditions précédentes sont bien réalisées.

Remarque. — On peut trouver une majoration de l'erreur que l'on fait en approximant la loi binomiale par la loi de Poisson, c'est-à-dire une majoration et une minoration pour $\binom{n}{k} (\lambda/n)^k (1 - (\lambda/n))^{n-k} / (e^{-\lambda} (\lambda^k / k!))$.

Exemple. — Dans une solution liquide contenant des particules en suspension, disons des bactéries A_1 et d'autres particules A_2 , le nombre de bactéries est faible devant le nombre total des particules. D'autre part, même dans un volume réduit de solution, le nombre de particules est grand. Pour déterminer la loi de probabilité de la variable X « nombre de bactéries dans un volume élémentaire fixé », il faut, même en utilisant la loi de Poisson, connaître la proportion p de bactéries dans la solution. Pour *estimer* celle-ci, selon les canons de la statistique, on prélève une goutte de liquide et on l'étale sur un hématimètre. Celui-ci a un grand nombre de cases (de l'ordre de quatre cents). Si la solution est homogène, le nombre X_i de bactéries dans la i^e case ($i = 1, 2, \dots, 400$) est une variable aléatoire de Poisson de paramètre λ . Si on réalise une épreuve ω , c'est-à-dire une telle expérimentation, on obtiendra 400 observations $X_1(\omega), X_2(\omega), \dots, X_{400}(\omega)$ indépendantes (homogénéité du liquide). Le paramètre p est alors estimé par $\sum_{i=1}^{400} X_i(\omega) / 400$.

Tables. — On a construit autrefois des tables numériques (volumineuses) des lois binomiales et de Poisson. Au siècle de l'ordinateur, ces tables ont beaucoup moins d'utilité. Pour la loi binomiale, il fallait prévoir des tables à trois entrées (n, p, r) . On y trouvait la valeur de :

$$P\{X > r\} = \sum_{k=r+1}^n \binom{n}{k} p^k (1-p)^{n-k}.$$

Pour la loi de Poisson, les tables étaient à deux entrées (λ, c) . On y trouvait la valeur de : $P\{Y > c\} = \sum_{k=c+1}^{\infty} e^{-\lambda} \lambda^k / k!$ Si l'on utilise des tables numériques de fonctions eulériennes Γ (gamma) et B (bêta), il est bon de noter les relations

$$\sum_{k=r+1}^n \binom{n}{k} p^k (1-p)^{n-k} = \frac{B(r+1, n-r, p)}{B(r+1, n-r)},$$

où, pour $\alpha > 0$, $\beta > 0$ et $0 \leq x \leq 1$, on pose

$$B(\alpha, \beta, x) = \int_0^x t^{\alpha-1} (1-t)^{\beta-1} dt, \quad B(\alpha, \beta, 1) = B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)},$$

et les relations

$$\sum_{k=r+1}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = \frac{\Gamma(r+1, \lambda)}{\Gamma(r+1)},$$

où, pour $x > 0$ et $z > 0$, on pose

$$\Gamma(z, x) = \int_0^x t^{z-1} e^{-t} dt, \quad \Gamma(z) = \Gamma(z, +\infty) = \int_0^{+\infty} t^{z-1} e^{-t} dt.$$

Les relations ci-dessus s'obtiennent par de simples intégrations par parties.

COMPLÉMENTS ET EXERCICES

1. *La loi binomiale négative dite de Pascal.* — Considérons une suite de répétitions indépendantes d'une alternative à deux résultats possibles : A avec probabilité p et B avec probabilité $q = 1 - p$. Désignons par A_k l'évènement « la k^e répétition amène A » et introduisons les variables aléatoires :

$$X_k = I_{A_k} \quad (k \geq 1); \quad S_n = X_1 + \cdots + X_n \quad (n \geq 1).$$

On prend pour Ω l'ensemble des suites ω dont les termes appartiennent à $\{A, B\}$ et pour P l'(unique) mesure de probabilité sur Ω telle que

$$P\{X_k = 1\} = p \quad (k \geq 1)$$

et qui rend les variables X_1, X_2, \dots indépendantes.

On s'intéresse au nombre minimum T_r de répétitions qu'il faut effectuer pour amener r fois A ($r \geq 1$). Autrement dit,

$$T_r = \inf\{n : S_n = r\}.$$

On voit que le support de T_r est $\{r, r+1, \dots\}$. Cherchons sa loi de probabilité. Pour tout $n \geq r$ on a $\{T_r = n\} = \{S_{n-1} = r-1, X_n = 1\}$, d'où, en vertu de l'indépendance de S_{n-1} et de X_n :

$$P\{T_r = n\} = P\{S_{n-1} = r-1\}P\{X_n = 1\} = \binom{n-1}{r-1} p^{r-1} q^{n-r} p.$$

En faisant le changement d'indices $n = r + k$ ($k \geq 0$), il vient :

$$P\{T_r = r + k\} = \binom{r+k-1}{r-1} p^r q^k \quad (k \geq 0).$$

Remarque 1. — Comme $\binom{r+k-1}{r-1} = \binom{r+k-1}{k} = \frac{(r+k-1)!}{(r-1)!k!}$, on a immédiatement en vertu de l'identité binomiale la relation

$$\sum_{k \geq 0} P\{T_r = r + k\} = \sum_{k \geq 0} \frac{(r+k-1)!}{(r-1)!k!} p^r q^k = p^r (1-q)^{-r} = 1,$$

de sorte que $P\{T_r < +\infty\} = 1$. On peut encore écrire :

$$P\{T_r = r + k\} = \binom{-r}{k} p^r (-q)^k.$$

Cette dernière forme de la probabilité a fait donner à la loi de T_r le nom de *loi binomiale négative*.

Remarque 2. — Pour $r = 1$ on retrouve la loi géométrique $\sum_{k \geq 1} pq^{k-1} \varepsilon_k$.

Remarque 3. — La variable aléatoire T_r ($r \geq 1$) que nous venons de définir a pour support $\{r, r+1, \dots\}$. La translattée $X_r = T_r - r$ aura pour support $\{0, 1, \dots\}$; elle peut être interprétée comme le nombre d'échecs précédant la $r^{\text{ième}}$ apparition de A ; sa loi est donnée par $P\{X_r = k\} = P\{T_r = r + k\}$ ($k \geq 0$). Pour $r = 1$ on retrouve la loi géométrique $\sum_{k \geq 0} pq^k \varepsilon_k$.

2. — Une machine-outil produit à la chaîne des objets manufacturés et l'on sait qu'en période de marche normale la probabilité pour qu'un objet soit défectueux (resp. non défectueux) est p (resp. $q = 1 - p$). On se propose de vérifier la machine. A cet effet, on a besoin de la variable aléatoire T_r « nombre

minimum de prélèvements successifs qu'il faut effectuer pour amener r objets défectueux ». Calculer la loi de T_r .

3. *Le problème des boîtes d'allumettes de Banach.* — Un fumeur a dans chacune de ses poches droite et gauche une boîte d'allumettes (de contenance N chacune). Lorsqu'il désire une allumette, il choisit au hasard une de ses poches (chacune avec probabilité $\frac{1}{2}$). On considère le moment où, *pour la première fois*, le fumeur s'aperçoit que l'une des boîtes est vide. A ce moment, l'autre boîte peut contenir un nombre quelconque d'allumettes. On désigne par u_r la probabilité pour qu'elle en contienne r .

- Calculer u_r ($r = 0, 1, \dots, N$).
- Calculer la probabilité v_r pour qu'au moment où la première boîte est vidée de sa dernière allumette (mais non encore *reconnue* vide) l'autre boîte contienne exactement r allumettes.
- Calculer la probabilité v pour que la première boîte vidée ne soit pas la première boîte *reconnue* vide.
- Établir, pour tout entier $m \geq 0$ et tout réel a , l'identité :

$$\sum_{k=0}^n \binom{a-k}{N-1} = \binom{a+1}{N} - \binom{a-n}{N}.$$

- Déduire de c) et d) que $v = \binom{2N}{N} 2^{-(2N+1)}$.

4. — On pose $b(k; n, p) = \binom{n}{k} p^k q^{n-k}$ ($0 \leq k \leq n$, $0 < p < 1$, $q = 1 - p$). Montrer que l'application $k \mapsto b(k; n, p)$ est d'abord croissante, puis décroissante et qu'elle atteint sa valeur maximum pour $k = m$, où m est l'(unique) entier vérifiant : $(n+1)p - 1 < m \leq (n+1)p$. Si $m = (n+1)p$, la valeur maximum est atteinte pour $k = m$ et $k = m - 1$.

5. — Déterminer le maximum de la suite $p(k, \lambda) = e^{-\lambda} \frac{\lambda^k}{k!}$ pour $k \geq 0$ et $\lambda > 0$.

6. — On choisit 500 personnes « au hasard ». Quelle est la probabilité de l'évènement « trois personnes exactement, parmi les 500, ont leur anniversaire le 1^{er} mars » ?

7. (Caractérisation d'une loi de Poisson). — Soit X une variable aléatoire à valeurs dans \mathbb{N} , telle que pour tout n dans \mathbb{N} on ait $p_n = \mathbb{P}\{X = n\} > 0$. Montrer que pour tout $\lambda > 0$, les deux propriétés sont équivalentes :

- X suit la loi de Poisson de paramètre λ ;
- Pour tout $n \geq 1$ on a $\frac{p_n}{p_{n-1}} = \frac{\lambda}{n}$.

8. — Soit X une variable aléatoire de loi géométrique $\sum_{k \geq 1} pq^{k-1} \varepsilon_k$ ($0 < p < 1$). Calculer $\mathbb{E}[1/X]$.

9 (Poissonisation ; la solution proposée requiert des techniques de fonctions génératrices développées au chap. 9). — Considérons une suite de parties indépendantes de « pile » ou « face ». La probabilité d'apparition de « pile » en une partie est p ; on pose $q = 1 - p$. Désignons par (I_k) ($k \geq 1$) la suite correspondante des variables aléatoires indicatrices de « pile ».

1) Soit n un nombre entier tel que $n \geq 1$. Posons $N_1 = \sum_{k=1}^n I_k$, $N_2 = \sum_{k=1}^n (1 - I_k)$. Il est clair que $N_1 + N_2 = n$, que N_1 et N_2 ne sont pas indépendantes et que $\mathcal{L}(N_1) = B(n, p)$, $\mathcal{L}(N_2) = B(n, q)$.

2) Soit N une variable aléatoire à valeurs dans $\{0, 1, \dots\}$, indépendante de la suite (I_k) ($k \geq 1$). Posons $N_1 = \sum_{k=1}^N I_k$, $N_2 = \sum_{k=1}^N (1 - I_k)$. Il est clair que $N_1 + N_2 = N$. Montrer que

a) si N suit la loi de Poisson $\mathcal{P}(\lambda)$ de paramètre $\lambda > 0$, les variables aléatoires N_1 et N_2 sont indépendantes et $\mathcal{L}(N_1) = \mathcal{P}(\lambda p)$, $\mathcal{L}(N_2) = \mathcal{P}(\lambda q)$;

b) si N_1 et N_2 sont indépendantes, alors N suit une loi de Poisson.

10 (Le paradoxe de l'inspection¹). — On considère une suite de parties indépendantes de « pile » ou « face », la probabilité d'amener « pile » en une partie étant p ($0 < p < 1$) ; on pose $q = 1 - p$. Cette expérience peut être modélisée au moyen de l'espace probabilisé $(\Omega, \mathfrak{A}, P)$, où

Ω est l'ensemble des suites $\omega = (\varepsilon_1, \varepsilon_2, \dots)$, avec $\varepsilon_i \in \{0, 1\}$;

\mathfrak{A} est la tribu engendrée par les sous-ensembles

$$A_{i_1, \dots, i_n}(a_1, \dots, a_n) = \{\omega : \varepsilon_{i_1}(\omega) = a_1, \dots, \varepsilon_{i_n}(\omega) = a_n\},$$

avec $a_1, \dots, a_n \in \{0, 1\}$, $1 \leq i_1 < \dots < i_n$, $n \geq 1$;

P est la mesure de probabilité sur (Ω, \mathfrak{A}) telle que

$$P(A_{i_1, \dots, i_n}(a_1, \dots, a_n)) = p^{a_1 + \dots + a_n} q^{n - (a_1 + \dots + a_n)}.$$

On admettra l'existence d'un tel espace probabilisé. On pourra interpréter cette expérience comme un processus de comptage dans lequel l'apparition de « pile » est assimilée à l'arrivée d'un « événement » (panne d'une machine, passage d'un autobus, ...). Les notions introduites ci-dessous prendront tout leur sens dans une telle interprétation. Introduisons les notations :

$N_n = \sum_{k=1}^n \varepsilon_k$: le nombre de « piles » amenés jusqu'à la date n ($n = 1, 2, \dots$).

(On pose $N_0 = 0$.)

T_i ($i \geq 1$) : l'instant d'arrivée du $i^{\text{ième}}$ « pile ». (On pose $T_0 = 0$.)

On notera que :

$$\begin{aligned} \{T_1 = n\} &= \{\varepsilon_1 = \dots = \varepsilon_{n-1} = 0; \varepsilon_n = 1\} \\ &= \{N_1 = \dots = N_{n-1} = 0; N_n = 1\}; \\ \{T_i = n\} &= \{N_{n-1} = i - 1; N_n = i\} \quad (i \geq 2). \end{aligned}$$

¹ Nous devons à Anatole Joffe cette idée de retranscrire au cas discret, comme il est fait ici, ce paradoxe de l'inspection bien connu pour les processus de Poisson.

- 1) Déterminer la loi de T_1 .
- 2) On pose : $\tau_1 = T_1, \tau_2 = T_2 - T_1, \dots, \tau_n = T_n - T_{n-1}$. Montrer que τ_1, \dots, τ_n sont des variables aléatoires indépendantes, de même loi que T_1 .
- 3) Montrer que la loi conjointe de (T_1, \dots, T_n) est donnée par

$$P\{T_1 = t_1, \dots, T_n = t_n\} = \begin{cases} \left(\frac{p}{q}\right)^n q^{t_n}, & \text{si } 0 < t_1 < \dots < t_n; \\ 0, & \text{sinon.} \end{cases}$$

- 4) Étant donné qu'à la date m on a observé n « piles », calculer la loi conditionnelle des positions de ces n « piles ». De façon précise, posons

$$A = \{N_m = n\} \quad (0 \leq n \leq m);$$

$$B = \{T_1 = t_1, \dots, T_n = t_n\} \quad (1 \leq t_1 < \dots < t_n \leq m).$$

Calculer $P(B|A)$. On observera que le résultat ne dépend pas de p et peut s'interpréter comme le choix au hasard de n parmi m points.

- 5) On se place à la date $n \geq 1$. Alors

T_{N_n} est l'instant d'arrivée du *dernier* « pile » (on pose $T_0 = 0$);
 $T_{N_{n+1}}$ est l'instant d'arrivée du *prochain* « pile »;
 $\tau_{N_{n+1}} = T_{N_{n+1}} - T_{N_n}$ est la longueur de l'intervalle entre deux « piles » qui recouvre n .

Posons $U_n = n - T_{N_n}, V_n = T_{N_{n+1}} - n$. On notera que les valeurs possibles de U_n sont $0, 1, \dots, n$, et celles de V_n sont $1, 2, \dots$.

- a) Montrer que U_n et V_n sont des variables aléatoires indépendantes et déterminer leurs lois.

- b) Calculer $\lim_{n \rightarrow \infty} P\{U_n = i\}$ ($i = 0, 1, \dots$).

Remarque. — Comme $\tau_{N_{n+1}} = U_n + V_n$ et que $\mathcal{L}(V_n) = \mathcal{L}(\tau_1)$, il résulte de 5) a) que la longueur $\tau_{N_{n+1}}$ de l'intervalle entre deux « piles » qui recouvre n a tendance à être plus grande que τ_1 ; c'est le « paradoxe » de l'inspection : un inspecteur qui arrive à la date n dans l'intention d'estimer la distance qui sépare deux apparitions consécutives de « pile » enregistre un nombre en général trop grand. Pour de grandes valeurs de n , cette distance a pour loi la loi de $\tau_1 + \tau'_1 - 1$, où τ_1, τ'_1 sont deux variables aléatoires indépendantes, de même loi que τ_1 .

11. — L'exercice 3 du *problème des boîtes d'allumettes de Banach* fournit une démonstration probabiliste de l'identité $\sum_{r=0}^N \binom{2N-r}{N} (1/2)^{2N-r} = 1$, que l'on peut établir en utilisant l'*identité de Gauss* (cf. Bailey,² p. 11), qui donne une évaluation de la fonction hypergéométrique en $x = \frac{1}{2}$, à savoir :

$${}_2F_1\left(\begin{matrix} a, b \\ \frac{1}{2}(a+b+1) \end{matrix}; \frac{1}{2}\right) = \frac{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2} + \frac{1}{2}a + \frac{1}{2}b)}{\Gamma(\frac{1}{2} + \frac{1}{2}a)\Gamma(\frac{1}{2} + \frac{1}{2}b)}.$$

² Bailey (W.N.). — *Generalized Hypergeometric Series*. — Cambridge, Cambridge University Press, 1935.

CHAPITRE 8

ESPÉRANCE MATHÉMATIQUE, VALEURS TYPIQUES

Dans ce chapitre nous introduisons la notion d'espérance mathématique pour une variable aléatoire réelle *discrète*. Un tel chapitre serait donc inutile si l'on donnait tout d'abord la théorie de l'intégration des variables aléatoires quelconques, théorie qui repose sur l'étude de l'espace probabilisé $(\Omega, \mathfrak{A}, P)$. Au contraire, l'étude de l'espérance des variables aléatoires discrètes est faite directement sur l'espace probabilisé (image) $(\mathbb{R}, \mathcal{B}^1, P_X)$. Le rapport entre les deux théories est obtenu à l'aide du théorème dit du transfert. Une version discrète de ce théorème est donnée dans ce chapitre.

1. Transformation des variables aléatoires

PROPOSITION 1.1. — Soient X une variable aléatoire, discrète, à n dimensions, de loi

$$P_X = \sum_k \alpha_k \varepsilon_{x_k}$$

et g une fonction à valeurs dans \mathbb{R}^p , mesurable, définie sur $(\mathbb{R}^n, \mathcal{B}^n)$. Alors $g \circ X$ est une variable aléatoire à p dimensions, discrète, de loi

$$P_{g \circ X} = \sum_k \alpha_k \varepsilon_{g(x_k)}.$$

De plus, en écrivant la chaîne $(\Omega, \mathfrak{A}, P) \xrightarrow{X} (\mathbb{R}^n, \mathcal{B}^n, P_X) \xrightarrow{g} (\mathbb{R}^p, \mathcal{B}^p)$, on a, pour tout $z \in \mathbb{R}^p$

$$(1.1) \quad P_{g \circ X}\{z\} = P_X\{g = z\} = P\{g \circ X = z\}.$$

Démonstration. — La variable aléatoire $g \circ X$ est évidemment à valeurs dans \mathbb{R}^p . D'autre part, $g(X(\omega)) = z$ si et seulement si $X(\omega) \in g^{-1}(z)$. Par conséquent, $P_{g \circ X}\{z\} = P\{g \circ X = z\} = P\{X \in g^{-1}(z)\} = P_X(g^{-1}(z)) = P_X\{g = z\} = P_X\{x : g(x) = z\} = \sum_k \{\alpha_k : g(x_k) = z\}$. \square

COROLLAIRE 1.2. — Si $T = (X, Y)$ est une variable aléatoire discrète à deux dimensions de loi

$$P_T = \sum_{i,j} p(x_i, y_j) \varepsilon_{(x_i, y_j)},$$

où $\{(x_i, y_j) : (i, j) \in I \times J\}$ est une suite finie ou dénombrable d'éléments de \mathbb{R}^2 , alors X et Y sont discrètes, respectivement de lois

$$P_X = \sum_{i \in I} \left(\sum_{j \in J} p(x_i, y_j) \right) \varepsilon_{x_i} \quad \text{et} \quad P_Y = \sum_{j \in J} \left(\sum_{i \in I} p(x_i, y_j) \right) \varepsilon_{y_j}.$$

Les lois P_X et P_Y sont appelées les lois marginales (en X, Y) associées à la loi (conjointe) P_T .

Démonstration. — En effet, il suffit de remarquer que les projections $\pi_1 : (x, y) \mapsto x$ et $\pi_2 : (x, y) \mapsto y$ sont des applications mesurables de \mathbb{R}^2 dans \mathbb{R} et que l'on a : $X = \pi_1 \circ T$ et $Y = \pi_2 \circ T$. \square

COROLLAIRE 1.3. — Avec les mêmes notations que ci-dessus, la loi de $X + Y$ est donnée par :

$$P_{X+Y} = \sum_{i,j} p(x_i, y_j) \varepsilon_{(x_i+y_j)}.$$

Démonstration. — En effet, on a : $X + Y = g \circ T$ avec $g(x, y) = x + y$. \square

Le Corollaire 1.2 entraîne que la loi de T détermine complètement les lois de X et de Y . La réciproque n'est pas vraie, car si X et Y sont des variables aléatoires réelles, définies sur le même espace probabilisé $(\Omega, \mathfrak{A}, P)$ et de lois

$$(1.2) \quad P_X = \sum_{i \in I} P\{X = x_i\} \varepsilon_{x_i} \quad \text{et} \quad P_Y = \sum_{j \in J} P\{Y = y_j\} \varepsilon_{y_j},$$

il n'est pas possible, en général, d'en déduire la loi de $T = (X, Y)$, car à cet effet il faut connaître les quantités $p(x_i, y_j) = P\{X = x_i, Y = y_j\}$ pour tout $(i, j) \in I \times J$.

2. Indépendance. — Supposons que X et Y soient des variables aléatoires réelles définies sur le même espace probabilisé $(\Omega, \mathfrak{A}, P)$ et dont les lois P_X et P_Y sont données par les formules (1.2). On peut alors déterminer la loi du couple (X, Y) comme indiqué dans le corollaire de la proposition suivante.

PROPOSITION 2.1. — Les variables aléatoires réelles X et Y sont indépendantes, si et seulement si, pour tout $i \in I$ et tout $j \in J$ on a :

$$(2.1) \quad P\{X = x_i, Y = y_j\} = P\{X = x_i\} P\{Y = y_j\}.$$

Démonstration. — En effet, les variables X et Y sont indépendantes, si et seulement si $P\{X \in A, Y \in B\} = P\{X \in A\} P\{Y \in B\}$ pour toute paire de boréliens A, B . En prenant $A = \{x_i\}$ et $B = \{y_j\}$, on retrouve bien (2.1).

Réciproquement, supposons (2.1) satisfait pour tout $i \in I$ et $j \in J$ et prenons deux boréliens A et B . On obtient :

$$\begin{aligned} \mathbb{P}\{X \in A, Y \in B\} &= \sum \{ \mathbb{P}\{X = x_i, Y = y_j\} : x_i \in A, y_j \in B \} \\ &= \sum \{ \mathbb{P}\{X = x_i\} \mathbb{P}\{Y = y_j\} : x_i \in A, y_j \in B \} \\ &= \sum \{ \mathbb{P}\{X = x_i\} : x_i \in A \} \sum \{ \mathbb{P}\{Y = y_j\} : y_j \in B \} \\ &= \mathbb{P}\{X \in A\} \mathbb{P}\{Y \in B\}. \end{aligned}$$

D'où X et Y sont indépendants. \square

COROLLAIRE 2.2. — *Si X et Y sont indépendantes, la loi de $T = (X, Y)$ est déterminée dès que l'on connaît les lois de X et de Y .*

Le corollaire découle immédiatement des relations (2.1).

3. Convolution des lois de probabilité discrètes

Définition. — Soient $\mathbb{P} = \sum_{i \in I} \alpha_i \varepsilon_{x_i}$ et $\mathbb{Q} = \sum_{j \in J} \beta_j \varepsilon_{y_j}$ deux lois de probabilité discrètes. Le *produit de convolution* de \mathbb{P} par \mathbb{Q} est la mesure de probabilité, notée $\mathbb{P} * \mathbb{Q}$, définie par :

$$(3.1) \quad \mathbb{P} * \mathbb{Q} = \sum_{(i,j) \in I \times J} \alpha_i \beta_j \varepsilon_{(x_i + y_j)}.$$

Que $\mathbb{P} * \mathbb{Q}$ soit une mesure de probabilité découle des propriétés élémentaires sur les séries absolument convergentes. Ces mêmes propriétés entraînent aussi que le produit de convolution est *commutatif* et *associatif*.

Les lois binomiales et de Poisson sont conservées par le produit de convolution. En d'autres termes, on a la propriété suivante.

PROPOSITION 3.1. — *Soient $B(n, p)$ la loi binomiale de paramètres (n, p) ($0 \leq p \leq 1, n \geq 0$) et π_λ la loi de Poisson de paramètre λ ($\lambda > 0$). Alors*

$$\begin{aligned} B(n, p) * B(m, p) &= B(n + m, p) \quad (n, m \in \mathbb{N}); \\ \pi_\lambda * \pi_\nu &= \pi_{\lambda + \nu} \quad (\lambda > 0, \nu > 0). \end{aligned}$$

Démonstration. — On a

$$\begin{aligned} B(n, p) * B(m, p) &= \sum_{i=0}^n \sum_{j=0}^m \binom{n}{i} \binom{m}{j} p^{i+j} q^{n+m-i-j} \varepsilon_{i+j} \\ &= \sum_{k=0}^{n+m} \gamma_k p^k q^{n+m-k} \varepsilon_k, \end{aligned}$$

où, pour $k = 0, 1, \dots, n + m$, on a posé : $\gamma_k = \sum_{i=0}^k \binom{n}{i} \binom{m}{k-i}$.

Cette expression est égale à $\binom{n+m}{k}$ d'après l'identité binomiale. Ceci prouve la première relation.

Pour établir la seconde, on écrit

$$\pi_\lambda * \pi_\nu = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} e^{-(\lambda+\mu)} \frac{\lambda^i}{i!} \frac{\mu^j}{j!} \varepsilon_{i+j} = e^{-(\lambda+\mu)} \sum_{k=0}^{\infty} \gamma_k \varepsilon_k,$$

où, pour $k = 0, 1, \dots$, on a posé :

$$\gamma_k = \sum_{i=0}^k \frac{\lambda^i}{i!} \frac{\mu^{k-i}}{(k-i)!}.$$

Cette expression est égale à $(\lambda + \mu)^k/k!$ \square

PROPOSITION 3.2. — *Si X et Y sont des variables aléatoires réelles, discrètes, définies sur un même espace probabilisé, indépendantes et respectivement de lois P_X et P_Y , alors la loi de probabilité de la variable aléatoire $X + Y$ est le produit de convolution de P_X par P_Y . Autrement dit,*

$$P_{X+Y} = P_X * P_Y.$$

Cette proposition est une conséquence immédiate du Corollaire 1.3 et de la Proposition 2.1.

4. Espérance mathématique. — De même qu'en mécanique on introduit la notion de *barycentre* de points matériels, on parle en calcul des probabilités de valeur moyenne ou d'*espérance mathématique* d'une variable aléatoire réelle X . Chaque valeur prise par X est affectée d'une masse égale à la probabilité pour que X soit égal à cette valeur et l'espérance mathématique, notée $\mathbb{E}[X]$, est le barycentre de ces valeurs affectées de telles masses. Cet énoncé est suffisant pour traiter le cas des variables aléatoires discrètes.

Définition. — L'*espérance mathématique* d'une variable aléatoire réelle discrète X de loi $P_X = \sum_i \alpha_i \varepsilon_{x_i}$ est définie par

$$\mathbb{E}[X] = \sum_i \alpha_i x_i,$$

à condition que la série du second membre *converge absolument*. Dans ce cas, on dit que X a une espérance mathématique *finie*. Si la série $\sum_i \alpha_i |x_i|$ diverge, on dit que X n'a pas d'espérance mathématique finie.

Soit $\sum_j \beta_j \varepsilon_{y_j}$ une expression pour la mesure de probabilité P_X , où tous les y_j sont *distincts*. Pour tout j le nombre β_j est donc la somme de tous les α_i tels que $x_i = y_j$. Si la série $\sum \alpha_i x_i$ converge absolument, la série $\sum_j y_j \beta_j$ est elle-même absolument convergente et donc ne dépend pas de la numérotation retenue pour les couples (β_j, y_j) . De plus, on a

$$\sum_i x_i \alpha_i = \sum_j y_j \sum_{i; x_i=y_j} \alpha_i = \sum_j y_j \beta_j,$$

par la propriété d'associativité généralisée. Par conséquent, l'espérance mathématique de X ne dépend ni de l'expression retenue pour P_X , ni de la numérotation des couples (α_i, x_i) retenue pour la somme $\sum_i x_i \alpha_i$. Cette propriété de *commutativité complète* justifie l'interprétation de l'espérance mathématique comme *barycentre*.

Le théorème du transfert que nous énonçons maintenant montre la souplesse de la définition de la notion d'espérance mathématique. On considère ici un espace probabilisé $(\Omega, \mathfrak{A}, P)$, où Ω est *au plus dénombrable*, puis X une variable aléatoire réelle définie sur cet espace. L'image de Ω par X est elle-même *au plus dénombrable*, soit $X(\Omega) = \{x_n : n \in \mathbb{N}\}$. Notons P_X la loi de probabilité de X .

THÉORÈME 4.1 (théorème du transfert). — *On a l'identité*

$$\sum_{\omega \in \Omega} X(\omega)P(\{\omega\}) = \sum_n x_n P_X(\{x_n\}),$$

pourvu que l'une des séries intervenant dans les deux membres soit absolument convergente (l'autre l'est alors également). Si ceci est le cas, la valeur commune des deux membres est appelée l'espérance mathématique de X .

Démonstration. — Posons $A_n = X^{-1}(\{x_n\})$; la classe $\{A_n\}$ est une partition de Ω et l'on a, formellement

$$\sum_{\omega \in \Omega} X(\omega)P(\{\omega\}) = \sum_n \sum_{\omega \in A_n} X(\omega)P(\{\omega\});$$

d'où, puisque pour tout $\omega \in A_n$ on a $X(\omega) = x_n$

$$\begin{aligned} \sum_{\omega \in \Omega} X(\omega)P(\{\omega\}) &= \sum_n x_n \sum_{\omega \in A_n} P(\{\omega\}) \\ &= \sum_n x_n P(A_n) = \sum_n x_n P_X(\{x_n\}). \end{aligned}$$

Les calculs formels sont valables dès que l'une des séries des deux membres est absolument convergente. \square

Avant de donner les premières propriétés de l'espérance mathématique, introduisons la notion de *propriété vraie presque sûrement*.

Définition. — Soit $(\Omega, \mathfrak{A}, P)$ un espace probabilisé et \mathcal{P} une propriété susceptible ou non d'être vérifiée par tout $\omega \in \Omega$. On dit que \mathcal{P} est *vraie presque sûrement* (p.s.), s'il existe $A \in \mathfrak{A}$ tel que $P(A) = 0$ et \mathcal{P} est vraie pour tous les $\omega \in A^c$.

Dans cette définition, on n'impose pas que l'ensemble A' des $\omega \in \Omega$ qui n'ont pas la propriété \mathcal{P} soit de probabilité nulle, car A' n'appartient pas nécessairement à \mathfrak{A} . En fait, on a $A' \subset A$, $A \in \mathfrak{A}$, $P(A) = 0$ et \mathcal{P} vraie dans A^c (mais \mathcal{P} est aussi vraie dans $A \setminus A'$).

THÉORÈME 4.2. — *Soient X et Y deux variables aléatoires discrètes définies sur un espace probabilisé $(\Omega, \mathfrak{A}, P)$. On a les propriétés :*

- (D1) $\mathbb{E}[X]$ finie si et seulement si $\mathbb{E}[|X|]$ finie ;
- (D2) $|X| \leq Y$ et $\mathbb{E}[Y]$ finie entraînent $\mathbb{E}[X]$ finie ;
- (D3) $-\infty < a \leq X \leq b < +\infty \implies a \leq \mathbb{E}[X] \leq b$;

- (D4) $X = a$ p.s. $\implies \mathbb{E}[X] = a$;
 (D5) $\mathbb{E}[X]$ finie $\implies |\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.

Démonstration. — La propriété (D1) résulte de la définition même de l'espérance mathématique.

Pour démontrer (D2), on reprend les notations du Corollaire 1.2. Dans la chaîne $\Omega \xrightarrow{T} T(\Omega) \xrightarrow{\pi_2} Y(\Omega)$, l'ensemble $T(\Omega)$ est au plus dénombrable. D'autre part, pour tout y_j , on a, d'après la formule (1.1), $\mathbb{P}_Y\{y_j\} = \mathbb{P}_{\pi_2 \circ T}\{y_j\} = \mathbb{P}_T\{\pi_2 = y_j\}$. Posons $\mathbb{Q} = \mathbb{P}_T$, de sorte que \mathbb{Q} est une mesure de probabilité sur l'ensemble $T(\Omega)$, portée par les couples (x_i, y_j) . En notant \mathbb{Q}_{π_2} la loi de probabilité de la variable aléatoire π_2 , définie sur l'espace probabilisé $(T(\Omega), \mathfrak{P}(T(\Omega)), \mathbb{Q})$, on a $\mathbb{P}_Y\{y_j\} = \mathbb{Q}\{\pi_2 = y_j\} = \mathbb{Q}_{\pi_2}\{y_j\}$. Par le théorème du transfert qu'on applique à ce dernier espace probabilisé et à la variable π_2 , on en déduit

$$\begin{aligned} \mathbb{E}[Y] &= \sum_j y_j \mathbb{P}_Y\{y_j\} = \sum_j y_j \mathbb{Q}_{\pi_2}\{y_j\} \\ &= \sum_{(x_i, y_j) \in T(\Omega)} \pi_2(x_i, y_j) \mathbb{Q}\{(x_i, y_j)\} \\ &= \sum_{(x_i, y_j) \in T(\Omega)} y_j \mathbb{Q}\{(x_i, y_j)\}. \end{aligned}$$

Or $|X| \leq Y$ entraîne l'implication $(x_i, y_j) \in T(\Omega) \implies |x_i| \leq y_j$, soit

$$\begin{aligned} \mathbb{E}[Y] &\geq \sum_{(x_i, y_j) \in T(\Omega)} |x_i| \mathbb{Q}\{(x_i, y_j)\} \\ &\geq \left| \sum_{(x_i, y_j) \in T(\Omega)} x_i \mathbb{Q}\{(x_i, y_j)\} \right| \\ &\geq |\mathbb{E}[X]|, \end{aligned}$$

en appliquant cette fois le théorème du transfert à la variable $X = \pi_1 \circ T$.

Pour démontrer la propriété (D3), on écrit :

$$\mathbb{P}\{X = x_k\}a \leq \mathbb{P}\{X = x_k\}x_k \leq \mathbb{P}\{X = x_k\}b;$$

d'où

$$a = \sum_k \mathbb{P}\{X = x_k\}a \leq \sum_k \mathbb{P}\{X = x_k\}x_k \leq \sum_k \mathbb{P}\{X = x_k\}b = b.$$

Pour (D4), on note que si $X = a$ p.s., alors la loi de X est ε_a et l'on a bien $\mathbb{E}[X] = a$.

Enfin pour (D5), on a simplement :

$$|\mathbb{E}[X]| = \left| \sum_k \mathbb{P}\{X = x_k\}x_k \right| \leq \sum_k \mathbb{P}\{X = x_k\}|x_k| = \mathbb{E}[|X|]. \quad \square$$

Nous consignons dans le théorème suivant les propriétés usuelles de l'espérance mathématique.

THÉORÈME 4.3. — Soient X et Y deux variables aléatoires, discrètes, définies sur un espace probabilisé $(\Omega, \mathfrak{A}, \mathbb{P})$. Si $\mathbb{E}[|X|] < \infty$ et $\mathbb{E}[|Y|] < \infty$, on a les propriétés :

A. *Linéarité*

- (A1) $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$;
 (A2) $\mathbb{E}[\lambda X] = \lambda \mathbb{E}[X]$ ($\lambda \in \mathbb{R}$).

B. *Monotonie*

- (B1) $X \geq 0 \implies \mathbb{E}[X] \geq 0$;
 (B2) $X \geq Y \implies \mathbb{E}[X] \geq \mathbb{E}[Y]$;
 (B3) $X = Y$ p.s. $\implies \mathbb{E}[X] = \mathbb{E}[Y]$.

C. *Indépendance.* — Si X et Y sont indépendantes, alors $\mathbb{E}[XY]$ est finie et l'on a $\mathbb{E}[XY] = \mathbb{E}[X] \mathbb{E}[Y]$.

Démonstration. — Notons $\sum_i \mathbb{P}\{X = x_i\} \varepsilon_{x_i}$ et $\sum_j \mathbb{P}\{Y = y_j\} \varepsilon_{y_j}$ les lois respectives de X et de Y .

Pour démontrer (A1), on fait appel à la loi conjointe de X et Y . On a :

$$\sum_j \mathbb{P}\{X = x_i, Y = y_j\} |x_i| = \mathbb{P}\{X = x_i\} |x_i|,$$

d'où

$$\sum_i \left(\sum_j \mathbb{P}\{X = x_i, Y = y_j\} \right) |x_i| = \sum_i \mathbb{P}\{X = x_i\} |x_i| = \mathbb{E}[|X|] < +\infty.$$

On montre de même :

$$\sum_j \left(\sum_i \mathbb{P}\{X = x_i, Y = y_j\} \right) |y_j| = \mathbb{E}[|Y|] < +\infty.$$

Il en résulte que la série double $\sum_{i,j} \mathbb{P}\{X = x_i, Y = y_j\} (x_i + y_j)$ est absolument convergente; le calcul formel ci-dessous est donc valable :

$$\begin{aligned} & \sum_{i,j} \mathbb{P}\{X = x_i, Y = y_j\} (x_i + y_j) \\ &= \sum_i \left(\sum_j \mathbb{P}\{X = x_i, Y = y_j\} \right) x_i + \sum_j \left(\sum_i \mathbb{P}\{X = x_i, Y = y_j\} \right) y_j; \end{aligned}$$

c'est-à-dire

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

La propriété (A2) est banale à vérifier. Pour (B1), si l'on a $X \geq 0$, alors chacun des x_i est positif et $\mathbb{E}[X] = \sum_i \mathbb{P}\{X = x_i\} x_i \geq 0$. Soit $\sum_k \mathbb{P}\{Z = z_k\} \varepsilon_{z_k}$ la loi de $Z = X - Y$. Si $Z \geq 0$, alors $\mathbb{E}[Z] = \mathbb{E}[X] - \mathbb{E}[Y] \geq 0$. D'où (B2). Enfin, pour (B3), si $Z = 0$ p.s., alors $\mathbb{P}\{Z = 0\} = 1$. Par suite $\mathbb{P}\{Z = z\} = 0$ pour tout $z \neq 0$. D'où $\mathbb{E}[Z] = \sum_k \mathbb{P}\{Z = z_k\} z_k = 0$.

Pour démontrer (C), on pose $XY = g \circ T$ avec $T = (X, Y)$ et $g(x, y) = xy$. On part de la loi du couple T . D'après la Proposition 1.1, la loi du produit XY s'écrit à l'aide de la loi de T comme suit

$$\mathbb{P}_{XY} = \sum_{i,j} \mathbb{P}\{X = x_i, Y = y_j\} \varepsilon_{x_i y_j} = \sum_{i,j} \mathbb{P}\{X = x_i\} \mathbb{P}\{Y = y_j\} \varepsilon_{x_i y_j},$$

puisque X et Y sont indépendantes. Par suite,

$$\begin{aligned}\mathbb{E}[XY] &= \sum_{i,j} \mathbb{P}\{X = x_i\} \mathbb{P}\{Y = y_j\} x_i y_j \\ &= \left(\sum_i \mathbb{P}\{X = x_i\} x_i \right) \left(\sum_j \mathbb{P}\{Y = y_j\} y_j \right) = \mathbb{E}[X] \mathbb{E}[Y]. \quad \square\end{aligned}$$

5. Moments. — L'espérance mathématique d'une variable aléatoire X ne dépend que de la loi de X et indique la valeur *moyenne* autour de laquelle X prend ses valeurs. On introduit d'autres caractéristiques de la loi de X qui rendent compte de la *dispersion* de cette loi, par exemple les *moments*. Débutons par un lemme qui permet de comparer les moments de différents ordres.

LEMME 5.1. — Soient r et r' deux nombres réels tels que $0 < r' < r$ et X une variable aléatoire réelle. Si $\mathbb{E}[|X|^r]$ est fini, alors $\mathbb{E}[|X|^{r'}]$ est aussi fini.

Démonstration. — En effet, pour tout $a > 0$, on a l'inégalité $a^{r'} \leq 1 + a^r$, car pour $a \geq 1$, on peut écrire $a^r = a^{r'} a^{r-r'} \geq a^{r'}$ et pour $a < 1$ on a naturellement $a^{r'} < 1$.

Appliquons cette inégalité à $|X(\omega)|$. Il vient $|X(\omega)|^{r'} \leq 1 + |X(\omega)|^r$ pour tout $\omega \in \Omega$. Or $\mathbb{E}[1 + |X|^r] = 1 + \mathbb{E}[|X|^r]$ existe et est finie par hypothèse. De la propriété (D2) ci-dessus résulte alors que $\mathbb{E}[|X|^{r'}]$ est aussi fini. \square

Définition. — Soit X une variable aléatoire réelle, discrète, de loi $\mathbb{P}_X = \sum_{i \in I} \alpha_i \varepsilon_{x_i}$. Soit a un nombre réel et r un nombre réel. Si $\mathbb{E}[|X - a|^r]$ est fini, alors le *moment d'ordre r de X centré en a* est défini par :

$${}_a m_r = \mathbb{E}[(X - a)^r] = \sum_{i \in I} \alpha_i (x_i - a)^r.$$

Le *moment d'ordre r* (centré en 0) est défini par :

$$m_r = \mathbb{E}[X^r].$$

De même, si $\mathbb{E}[|X|]$ est fini, ainsi que $\mathbb{E}[|X - \mathbb{E}[X]|^r]$, le *moment centré d'ordre r* (à la moyenne) est défini par :

$$\mu_r = \mathbb{E}[(X - \mathbb{E}[X])^r].$$

Si $r = 1$, alors $m_1 = \mathbb{E}[X]$ et $\mu_1 = 0$. Pour $r = 2$, le *moment centré μ_2* est encore appelé *variance* de X et noté

$$\text{Var } X = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

La racine carrée positive de $\text{Var } X$ est désignée par $\sigma(X)$. C'est l'*écart-type* de X . Les variables aléatoires $(X - \mathbb{E}[X])$ et $(X - \mathbb{E}[X])/\sigma(X)$ sont respectivement appelées *centrée* et *centrée réduite* (dans ce dernier cas, on suppose $\sigma(X) > 0$).

Il résulte du précédent lemme que toute variable aléatoire qui a un moment d'ordre deux fini a une espérance mathématique finie.

PROPOSITION 5.2. — *Une variable aléatoire réelle X a un moment d'ordre deux $\mathbb{E}[X^2]$ fini, si et seulement si son espérance mathématique $\mathbb{E}[X]$ et sa variance $\text{Var } X$ existent et sont finies, et l'on a :*

$$(5.1) \quad \text{Var } X = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

Démonstration. — Si X a un moment d'ordre deux fini, elle a aussi une espérance mathématique, de sorte qu'on peut former le développement :

$$(X - \mathbb{E}[X])^2 = X^2 - 2X \mathbb{E}[X] + (\mathbb{E}[X])^2,$$

dont l'espérance mathématique, qui n'est autre que $\text{Var } X$, est donnée par $\mathbb{E}[X^2] - (\mathbb{E}[X])^2$, en vertu des propriétés de linéarité (A1) et (A2).

Réciproquement, supposons que $\mathbb{E}[|X|]$ et $\text{Var } X$ soient finies. En écrivant

$$X^2 = (X - \mathbb{E}[X] + \mathbb{E}[X])^2 = (X - \mathbb{E}[X])^2 + (\mathbb{E}[X])^2 + 2 \mathbb{E}[X] (X - \mathbb{E}[X]),$$

on remarque que tous les termes du second membre ont une espérance mathématique finie. De là, les propriétés de linéarité de l'espérance mathématique entraînent que $\mathbb{E}[X^2]$ est fini. Comme, de plus, on a

$$\mathbb{E}[\mathbb{E}[X] (X - \mathbb{E}[X])] = \mathbb{E}[X] \mathbb{E}[X - \mathbb{E}[X]] = 0,$$

on retrouve bien, une nouvelle fois la formule (5.1). \square

PROPOSITION 5.3. — *Soit X une variable aléatoire vérifiant $\mathbb{E}[X^2] < \infty$. Alors pour tout nombre réel a , on a l'inégalité :*

$$\mathbb{E}[(X - a)^2] \geq \mathbb{E}[(X - \mathbb{E}[X])^2] = \sigma^2.$$

En d'autres termes, c'est par rapport à l'espérance mathématique que le moment quadratique est minimum et la valeur du minimum est la variance. Si donc on prend l'espérance mathématique comme caractéristique de position, il convient de prendre la variance comme caractéristique de dispersion.

Démonstration. — Posons $g(a) = \mathbb{E}[(X - a)^2]$ et $\mu = \mathbb{E}[X]$. Alors

$$\begin{aligned} g(a) &= \mathbb{E}[\left((X - \mu) + (\mu - a)\right)^2] \\ &= \mathbb{E}[(X - \mu)^2] + 2(\mu - a) \mathbb{E}[X - \mu] + (\mu - a)^2 \\ &= \sigma^2 + (\mu - a)^2. \quad \square \end{aligned}$$

Définition. — Soient r un entier tel que $r \geq 1$ et X une variable aléatoire. Si $\mathbb{E}[|X|^r]$ est fini, on définit le *moment factoriel d'ordre r* par :

$$\mathbb{E}[X(X - 1) \dots (X - r + 1)].$$

Ces moments sont surtout utilisés dans le cas où la variable aléatoire X prend ses valeurs dans \mathbb{N} .

Définition. — Soient r un nombre réel et X une variable aléatoire. Si $\mathbb{E}[|X|^r] < +\infty$, on définit le *moment absolu d'ordre r* (centré en 0) par $\mathbb{E}[|X|^r]$. Dans le cas où en outre $r \neq 0$, on définit l'*écart d'ordre r* (par rapport à 0) par :

$$e_r = [\mathbb{E}[|X|^r]]^{1/r}.$$

On voit que e_2 est l'écart-type, si X est centré.

6. Covariance. — Soit $T = (X, Y)$ un couple de variables aléatoires réelles, de loi

$$P_T = \sum_{i,j} P\{X = x_i, Y = y_j\} \varepsilon_{(x_i, y_j)}.$$

La variable aléatoire XY a une espérance mathématique égale à

$$\mathbb{E}[XY] = \sum_{i,j} P\{X = x_i, Y = y_j\} x_i y_j,$$

à condition que la série du second membre converge absolument. Comme $|x_i y_j| \leq (x_i^2 + y_j^2)/2$, alors $\mathbb{E}[XY]$ existe si X et Y ont des moments du second ordre finis. Dans ce cas les espérances mathématiques existent et sont finies, leurs variances également. Par suite $(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])$ a aussi une espérance mathématique finie. La définition suivante a donc un sens.

Définition. — Soit (X, Y) un couple de variables aléatoires de loi conjointe donnée. Si X et Y ont des moments du second ordre finis, la *covariance* de X et de Y est définie par :

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Si $\text{Cov}(X, Y) = 0$, on dit que X et Y sont *non corrélées*.

Il résulte immédiatement de cette définition et de la propriété C (Indépendance), que si X et Y sont *indépendantes*, leur covariance $\text{Cov}(X, Y)$ est nulle. La réciproque n'est pas exacte : deux variables aléatoires peuvent être non corrélées sans être indépendantes.

Exemple. — Soit X la variable aléatoire de loi $P_X = \frac{1}{3}(\varepsilon_{-1} + \varepsilon_0 + \varepsilon_1)$. Posons $Y = X^2$. La loi du couple $T = (X, Y)$ est donnée par :

$$P_T = \frac{1}{3}(\varepsilon_{(-1,1)} + \varepsilon_{(0,0)} + \varepsilon_{(1,1)}).$$

On a $\mathbb{E}[X] = 0$ et $\mathbb{E}[XY] = 0$. D'où $\text{Cov}(X, Y) = 0$. Pourtant $Y = X^2$.

PROPOSITION 6.1. — *Si (X_1, X_2, \dots, X_n) est un système de n variables aléatoires ayant chacune un moment du second ordre fini, alors*

$$(6.1) \quad \text{Var} \sum_{i=1}^n X_k = \sum_{i=1}^n \text{Var} X_k + 2 \sum_{1 \leq j < k \leq n} \text{Cov}(X_j, X_k).$$

Si les variables sont mutuellement indépendantes (ou même seulement non corrélées deux à deux), alors

$$\operatorname{Var} \sum_{i=1}^n X_k = \sum_{i=1}^n \operatorname{Var} X_k.$$

Démonstration. — On peut, sans nuire à la généralité, supposer les variables X_1, X_2, \dots, X_n centrées. Or on a :

$$\left(\sum_k X_k \right)^2 = \sum_k X_k^2 + 2 \sum_{1 \leq j < k \leq n} X_j X_k.$$

En prenant l'espérance mathématique des deux membres, on obtient (6.1). Enfin, si X_1, \dots, X_n sont non corrélées deux à deux, les covariances $\operatorname{Cov}(X_j, X_k)$ sont nulles pour $1 \leq j < k \leq n$. On obtient bien la seconde formule. \square

On vérifie que $\operatorname{Cov}(aX + b, cY + d) = ac \operatorname{Cov}(X, Y)$, c'est-à-dire que la covariance est invariante par changement d'origine sur les axes $0x$ et $0y$, mais non par changement d'échelle. Ceci peut être un inconvénient dans les applications statistiques qu'il conviendra de corriger ; ce sera l'objet du prochain paragraphe.

7. Le coefficient de corrélation linéaire

Définition. — Soit (X, Y) un couple de variables aléatoires réelles telles que $\mathbb{E}[X^2] < \infty$ et $\mathbb{E}[Y^2] < \infty$. Supposons, d'autre part, que $\sigma(X)\sigma(Y) > 0$. On appelle *coefficient de corrélation* (linéaire) du couple (X, Y) le nombre

$$r(X, Y) = \frac{\operatorname{Cov}(X, Y)}{\sigma(X)\sigma(Y)} = \mathbb{E} \left[\left[\frac{X - \mathbb{E}[X]}{\sigma(X)} \right] \left[\frac{Y - \mathbb{E}[Y]}{\sigma(Y)} \right] \right].$$

On vérifie que si $ac \neq 0$, on a $r(aX + b, cY + d) = \operatorname{sg}(ac) r(X, Y)$. En prenant $a > 0, c > 0$, on voit que le *coefficient de corrélation linéaire est invariant à la fois par changement d'origine et par changement d'échelle sur les axes $0x$ et $0y$* . On tirera profit de cette propriété dans les calculs relatifs à $r(X, Y)$, en supposant les variables marginales X et Y centrées, réduites.

PROPRIÉTÉ 7.1. — $|r(X, Y)| \leq 1$.

Démonstration. — Supposons X et Y centrées réduites ; on a, pour tout λ

$$0 \leq \mathbb{E}[(X + \lambda Y)^2] = \mathbb{E}[X^2] + 2\lambda \mathbb{E}[XY] + \lambda^2 \mathbb{E}[Y^2] = 1 + 2\lambda r + \lambda^2.$$

On a un trinôme du second degré en λ , qui est positif ; son discriminant est donc négatif ou nul, c'est-à-dire $r^2 \leq 1$. \square

PROPRIÉTÉ 7.2. — Si $r(X, Y) = \pm 1$, alors X et Y sont liées par une relation fonctionnelle linéaire (mieux, affine). [D'où le nom de *coefficient de corrélation linéaire* donné à r .]

Démonstration. — Faisons-la pour $r = 1$. On suppose X et Y centrées réduites; on a alors pour tout λ

$$0 \leq \mathbb{E}[(X + \lambda Y)^2] = 1 + 2\lambda + \lambda^2 = (\lambda + 1)^2.$$

Pour $\lambda = -1$, on a $\mathbb{E}[(X - Y)^2] = 0$, c'est-à-dire $Y = X$ p.s. Pour $r = -1$, on trouve $Y = -X$ p.s. \square

Si les variables X, Y ne sont *pas* centrées réduites, la relation linéaire qui les lie est :

$$\frac{Y - \mathbb{E}[Y]}{\sigma(Y)} = \pm \frac{X - \mathbb{E}[X]}{\sigma(X)} \quad \text{p.s.}$$

8. L'inégalité de Tchebychev. — Il s'agit d'une inégalité extrêmement utilisée dans le calcul des probabilités, en particulier dans l'étude de la convergence en probabilité.

PROPOSITION 8.1. — Soit $r > 0$ un nombre réel et X une variable aléatoire réelle définie sur un espace probabilisé $(\Omega, \mathfrak{A}, \mathbb{P})$. Si $\mathbb{E}[|X|^r]$ est fini, on a, pour tout nombre réel $t > 0$, l'inégalité :

$$\mathbb{P}\{|X| \geq t\} \leq \frac{\mathbb{E}[|X|^r]}{t^r};$$

ou encore, de façon équivalente, en désignant par e_r l'écart d'ordre r , on a, pour tout nombre réel $t > 0$, l'inégalité :

$$\mathbb{P}\{|X| \geq te_r\} \leq \frac{1}{t^r}.$$

Démonstration. — En effet, soient $t, r > 0$. On a

$$\{|X| \geq t\} \Leftrightarrow \{|X|^r \geq t^r\}.$$

D'où

$$t^r I_{\{|X| \geq t\}} = t^r I_{\{|X|^r \geq t^r\}} \leq |X|^r,$$

et le résultat en prenant l'espérance mathématique. \square

Pour $r = 1, 2$, on obtient l'inégalité de Markov, de Tchebychev (ou de Bienaymé), respectivement. La version la plus utilisée de l'inégalité de Tchebychev est celle qui s'applique à la variable aléatoire centrée $(X - \mathbb{E}[X])$.

COROLLAIRE 8.2. — Si $\mathbb{E}[X^2] < +\infty$, on a pour tout $t > 0$ l'inégalité

$$\mathbb{P}\{|X - \mathbb{E}[X]| \geq t\} \leq \frac{\text{Var } X}{t^2}.$$

Remarque 1. — Posons $\mu = \mathbb{E}[X]$, $\sigma^2 = \text{Var } X$; alors pour tout $t > 0$ on a

$$\mathbb{P}\{|X - \mu| \geq t\} \leq \frac{\sigma^2}{t^2}; \quad \text{ou encore} \quad \mathbb{P}\{|X - \mu| \geq t\sigma\} \leq \frac{1}{t^2};$$

soit

$$\mathbb{P}\{|X - \mu| < t\sigma\} \geq 1 - \frac{1}{t^2}; \quad \text{et} \quad \mathbb{P}\{X \in]\mu - t\sigma, \mu + t\sigma[\} \geq 1 - \frac{1}{t^2}.$$

En particulier, pour $t = 2$ et $t = 3$, on a, respectivement :

$$(*) \quad \mathbb{P}\{X \in]\mu - 2\sigma, \mu + 2\sigma[\} \geq 1 - \frac{1}{4} = 0,75;$$

$$(**) \quad \mathbb{P}\{X \in]\mu - 3\sigma, \mu + 3\sigma[\} \geq 1 - \frac{1}{9} \approx 0,88.$$

On voit clairement le rôle de l'écart-type.

Remarque 2. — L'inégalité de Tchebychev est *universelle* (c'est-à-dire elle est valable pour toute variable aléatoire admettant un moment quadratique). En revanche, elle est *grossière*. On s'en convaincra en observant que si X est une variable aléatoire normale, de loi $\mathcal{N}(\mu, \sigma)$ (voir chap. 14, § 3), on a :

$$\mathbb{P}\{X \in]\mu - 2\sigma, \mu + 2\sigma[\} \approx 0,95; \quad [\text{r\egle des « } 2\sigma \text{ »}]$$

$$\mathbb{P}\{X \in]\mu - 3\sigma, \mu + 3\sigma[\} \approx 0,997.$$

On remarque la pauvreté des minoration (*) et (**).

9. Les inégalités relatives aux moments dans le cas fini. — Soit X une variable aléatoire discrète finie, à valeurs strictement positives. Pour fixer les idées, supposons que sa loi de probabilité P_X soit donnée par

$$P_X = \sum_{k=1}^l \alpha_k \varepsilon_{x_k},$$

où $\alpha_1, \dots, \alpha_l \geq 0$, $\sum_{k=1}^l \alpha_k = 1$ et $0 < x_1 < \dots < x_l < +\infty$. Nous observons que

a) pour tout nombre réel r le moment (absolu) d'ordre r existe et est égal à

$$m_r = \mathbb{E}[X^r] = \sum_{k=1}^l \alpha_k x_k^r;$$

b) pour tout nombre réel $r \neq 0$ l'écart d'ordre r existe et est égal à

$$e_r = (m_r)^{1/r}.$$

PROPOSITION ET DÉFINITION 9.1. — Lorsque r tend vers 0, l'écart e_r tend vers une limite finie, notée e_0 et l'on a :

$$e_0 = \prod_{k=1}^l x_k^{\alpha_k}.$$

Le nombre e_0 est appelé la *moyenne géométrique* de X .

Démonstration. — On a :

$$\begin{aligned} \text{Log } e_r &= \frac{1}{r} \text{Log} \left(\sum_{k=1}^l \alpha_k x_k^r \right) = \frac{1}{r} \text{Log} \left(\sum_{k=1}^l \alpha_k \exp(r \text{Log } x_k) \right) \\ &= \frac{1}{r} \text{Log} \left(\sum_{k=1}^l \alpha_k (1 + r \text{Log } x_k + o(r)) \right) \\ &= \frac{1}{r} \text{Log} \left(1 + r \sum_{k=1}^l \alpha_k \text{Log } x_k + o(r) \right). \end{aligned}$$

D'où $\text{Log } e_r$ tend vers $\sum_{k=1}^l \alpha_k \text{Log } x_k$, lorsque r tend vers 0. \square

THÉORÈME 9.2. — *Considérons l'application $r \mapsto e_r$ de \mathbb{R} dans \mathbb{R}^+ définie comme suit :*

$$e_r = \begin{cases} (m_r)^{1/r} & (\text{l'écart d'ordre } r), & \text{si } r \neq 0; \\ \prod_{k=1}^l x_k^{\alpha_k} & (\text{la moyenne géométrique}), & \text{si } r = 0. \end{cases}$$

Alors cette application est croissante sur \mathbb{R} .

Démonstration.

a) La fonction $r \mapsto \text{Log } m_r$ ($r \in \mathbb{R}$) est *convexe*. En effet, pour $r, s \in \mathbb{R}$, on a, d'après l'inégalité de Schwarz,

$$\begin{aligned} \sum_{k=1}^l \alpha_k x_k^{(r+s)/2} &\leq \left(\sum_{k=1}^l \alpha_k x_k^r \right)^{1/2} \left(\sum_{k=1}^l \alpha_k x_k^s \right)^{1/2} \\ m_{(r+s)/2} &\leq (m_r m_s)^{1/2} \\ \text{Log } m_{(r+s)/2} &\leq \frac{1}{2} (\text{Log } m_r + \text{Log } m_s), \end{aligned}$$

d'où le résultat puisque la fonction $r \mapsto \text{Log } m_r$ est *continue*.

b) La fonction $r \mapsto e_r$ ($r \in \mathbb{R} \setminus \{0\}$) est croissante dans $] -\infty, 0[$ et dans $]0, +\infty[$. En effet, d'après a) la courbe représentative de la fonction $r \mapsto \text{Log } m_r$ ($r \in \mathbb{R}$) est convexe et *passé par l'origine* (on a $m_0 = 1$). Considérons, pour $r \neq 0$, l'expression $\text{Log } e_r = \frac{1}{r} \text{Log } m_r = \frac{1}{r} (\text{Log } m_r - \text{Log } m_0)$ qui est la pente de la droite joignant l'origine au point $(r, \text{Log } m_r)$. Il résulte de a) que la fonction $r \mapsto \text{Log } e_r$ est croissante dans $] -\infty, 0[$ et dans $]0, +\infty[$. Il en est donc de même de la fonction $r \mapsto e_r$.

c) Le nombre $e_0 = \prod_{k=1}^l x_k^{\alpha_k}$ est le prolongement par continuité de la fonction $r \mapsto e_r$ ($r \in \mathbb{R} \setminus \{0\}$) au point $r = 0$. Le théorème en résulte. \square

Remarque. — L'application ci-dessus de \mathbb{R} dans \mathbb{R}^+ peut être prolongée en une application de $\overline{\mathbb{R}}$ dans \mathbb{R}^+ ; on a en effet :

$$\begin{aligned} \lim_{r \rightarrow -\infty} e_r &= \min_{k=1}^l x_k = x_1 \quad (= e_{-\infty}); \\ \lim_{r \rightarrow +\infty} e_r &= \max_{k=1}^l x_k = x_l \quad (= e_{+\infty}). \end{aligned}$$

Cas particulier 1. — Pour $r = n \in \mathbb{N}^*$ le théorème 9.2 donne $e_n \leq e_{n+1}$; c'est l'*inégalité de Liapounov*. Pour $n = 1$, il vient $e_1 \leq e_2$, c'est-à-dire $\mathbb{E}[|X|] \leq \sqrt{\mathbb{E}[X^2]}$. En prenant pour X la variable aléatoire centrée $(X - \mu)$ ($\mu = \mathbb{E}[X]$), on a $\mathbb{E}[|X - \mu|] \leq \sqrt{\mathbb{E}[(X - \mu)^2]}$, c'est-à-dire l'écart absolu par rapport à μ est majoré par l'écart-type.

Cas particulier 2. — Il résulte du Théorème 9.2 que l'on a $e_{-1} \leq e_0 \leq e_1$, où

$$e_{-1} = (\mathbb{E}[X^{-1}])^{-1} = \left(\sum_{k=1}^l \frac{\alpha_k}{x_k} \right)^{-1} \text{ est la « moyenne harmonique »;}$$

$$e_0 = \prod_{k=1}^l x_k^{\alpha_k} \text{ est la « moyenne géométrique »;}$$

$$e_1 = \mathbb{E}[X] = \sum_{k=1}^l \alpha_k x_k \text{ est la « moyenne arithmétique ».}$$

On retrouve les inégalités classiques entre ces moyennes.

Cas particulier 3. — On a vu que pour tout couple (r, s) de nombres réels on a $m_{(r+s)/2} \leq (m_r m_s)^{1/2}$. En prenant $r = 2n$, $s = 2n + 2$ ($n \in \mathbb{N}$), on a $m_{2n+1} \leq (m_{2n} m_{2n+2})^{1/2}$, inégalité qui permet de majorer tout moment d'ordre *impair* par des moments d'ordre *pair*.

10. Médiane. Écart moyen minimum. — Nous introduisons à présent une nouvelle caractéristique, appelée *médiane*, qui a l'avantage, sur l'espérance mathématique, d'exister pour toute variable aléatoire.

Définition. — Soit X une variable aléatoire réelle. On appelle *médiane* de X tout nombre M vérifiant :

$$P\{X \leq M\} \geq \frac{1}{2}, \quad P\{X \geq M\} \geq \frac{1}{2}.$$

Remarque 1. — Il résulte immédiatement de la définition que, si M est une médiane de X , on a :

$$P\{X \leq M\} \geq \frac{1}{2} \geq P\{X > M\} \quad \text{et} \quad P\{X \geq M\} \geq \frac{1}{2} \geq P\{X < M\}.$$

Remarque 2. — Toute variable aléatoire X admet au moins une médiane, elle peut en admettre plusieurs, qui jouent toutes le même rôle. Si la fonction de répartition F de X est *continue* et *strictement croissante*, alors X admet une médiane *et une seule*, M , et l'on a $F(M) = \frac{1}{2}$.

THÉORÈME 10.1. — *Soit X une variable aléatoire vérifiant $\mathbb{E}[|X|] < +\infty$. Si M est une médiane de X , alors pour tout nombre réel a on a l'inégalité :*

$$\mathbb{E}[|X - a|] \geq \mathbb{E}[|X - M|].$$

Démonstration. — Nous ferons la démonstration dans le cas où X est une variable discrète de loi $P_X = \sum_k \alpha_k \varepsilon_{x_k}$. Lorsque l'espérance mathématique

d'une variable aléatoire quelconque sera définie (*cf.* chap. 11), on pourra constater que la démonstration dans le cas général est tout à fait analogue. Supposons tout d'abord $M < a$. On partage \mathbb{R} en les trois intervalles disjoints $] - \infty, M]$, $]M, a]$, $]a, +\infty[$ et l'on peut écrire :

$$\begin{aligned} \mathbb{E}[|X - a|] - \mathbb{E}[|X - M|] &= \sum_k (|x_k - a| - |x_k - M|)\alpha_k \\ &= \sum_{x_k \in]-\infty, M]} (a - M)\alpha_k + \sum_{x_k \in]M, a]} (a + M - 2x_k)\alpha_k + \sum_{x_k \in]a, +\infty[} (M - a)\alpha_k. \end{aligned}$$

Notons A , B et C les trois sommes successives apparaissant dans la ligne précédente. On a :

$$\begin{aligned} A &= (a - M) \mathbb{P}\{X \leq M\}; \\ B &\geq \sum_{x_k \in]M, a]} (M - a)x_k = (M - a) \mathbb{P}\{M < X \leq a\}; \\ C &= (M - a) \mathbb{P}\{X > a\}; \end{aligned}$$

d'où, en définitive :

$$\mathbb{E}[|X - a|] - \mathbb{E}[|X - M|] \geq (a - M)(\mathbb{P}\{X \leq M\} - \mathbb{P}\{X > M\}).$$

Or, M étant une médiane, le second membre est positif. La démonstration se fait de façon analogue dans le cas où $M > a$. \square

Remarque. — Soit X une variable aléatoire vérifiant $\mathbb{E}[|X|] < +\infty$. Alors l'expression $\mathbb{E}[|X - M|]$ prend la même valeur pour toute médiane M de X : en effet, soient M_1 , M_2 deux médianes de X , avec $M_1 \neq M_2$; en prenant $a = M_1$, $M = M_2$, puis $a = M_2$, $M = M_1$ dans l'inégalité du Théorème 10.1, on voit que $\mathbb{E}[|X - M_1|] = \mathbb{E}[|X - M_2|]$. Cette circonstance justifie la définition suivante.

Définition. — Soit X une variable aléatoire réelle vérifiant $\mathbb{E}[|X|] < +\infty$. Alors $\mathbb{E}[|X - M|]$ prend la même valeur pour toute médiane M de X ; cette valeur commune est appelée *écart moyen minimum* ou *écart médian* de X .

Le Théorème 10.1 fournit alors un analogue du Théorème 5.3; il montre que si l'on adopte une médiane comme caractéristique de position, il convient de lui associer l'écart moyen minimum comme caractéristique de dispersion.

PROPOSITION 10.2. — *L'écart médian est majoré par l'écart-type.*

Démonstration. — Il résulte du Théorème 9.2 que $\mathbb{E}[|X - \mu|] \leq \sigma$. Or, M étant une médiane, on a, en appliquant le Théorème 10.1 pour $a = \mu$, l'inégalité $\mathbb{E}[|X - M|] \leq \mathbb{E}[|X - \mu|]$. D'où le résultat. \square

COMPLÉMENTS ET EXERCICES

1. — Déterminer l'espérance mathématique (resp. la variance) d'une variable aléatoire binomiale, de Poisson.

2. — Un concierge a n clefs dont une seule ouvre une porte. Il les essaie l'une après l'autre en éliminant après chaque essai la clef qui n'a pas convenu. Trouver le nombre moyen d'essais nécessaires pour trouver la bonne clef.

3. — Un processus de Bernoulli de paramètre p est une suite de variables aléatoires (X_n) ($n = 1, 2, \dots$) indépendantes et ne prenant chacune que deux valeurs (par exemple 1 et 0) avec la probabilité p et $q = 1 - p$. Il est commode d'interpréter X_n comme le résultat (succès ou échec) au $n^{\text{ième}}$ essai d'une même expérience que l'on répète indéfiniment de telle sorte que les conditions soient toujours les mêmes et que les résultats des différents essais soient sans influence mutuelle.

a) Montrer que la loi de la variable aléatoire $S_n = X_1 + \dots + X_n$ (nombre de succès aux n premiers essais) est la loi binomiale $B(n, p)$. Retrouver sans calcul l'espérance mathématique et la variance de cette loi.

b) Soit L le plus grand entier tel que $X_1 = X_2 = \dots = X_L$ et M le plus grand entier tel que $X_{L+1} = X_{L+2} = \dots = X_{L+M}$. Trouver les lois des variables aléatoires L et M , leurs espérances mathématiques et leurs variances. Montrer que les lois de L et M coïncident si et seulement si $p = 1/2$.

c) (E. Kosmanek) Montrer que $\mathbb{E}[L] \geq \mathbb{E}[M] = 2$, $\text{Var } L \geq \text{Var } M \geq 2$, $\text{Cov}(L, M) = -(p - q)^2 / (pq)$ et $-1/2 \leq r(L, M) \leq 0$.

d) (E. Kosmanek) Montrer que pour tout $n \geq 1$ on a

$$\lim_{l \rightarrow \infty} \text{P}\{M = n \mid L = l\} = \begin{cases} p^{n-1}q, & \text{si } p < 1/2; \\ q^{n-1}p, & \text{si } p > 1/2; \\ 1/2^n, & \text{si } p = 1/2. \end{cases}$$

e) Soit T le nombre d'échecs précédant le premier succès, i.e. le plus petit entier T tel que $X_{T+1} = 1$. Montrer que $\text{P}_T = \sum_{k \geq 0} pq^k \varepsilon_k$ (loi géométrique modifiée) et calculer $\mathbb{E}[T]$.

f) Plus généralement, si r est un entier au moins égal à 1, soit T_r le nombre d'échecs précédant le $r^{\text{ième}}$ succès. Montrer que

$$\text{P}\{T_r = k\} = \binom{r+k-1}{k} p^r q^k = \binom{-r}{k} p^r (-q)^k$$

(loi binomiale négative) et que $\mathbb{E}[T_r] = rq/p$.

4. — On reprend l'exercice 2 en supposant que le concierge remet après chaque échec la clef essayée dans le trousseau. On a donc un processus de Bernoulli avec $p = 1/n$. Calculer dans ce cas l'espérance mathématique du nombre d'essais pour trouver la bonne clef.

5. — Si X est une variable aléatoire prenant les valeurs x_k et A un évènement de probabilité non nulle, on pose :

$$\mathbb{E}[X | A] = \sum_k x_k P\{X = x_k | A\}.$$

Soit (B_n) ($n = 1, 2, \dots$) un système complet d'évènements. Montrer que

$$\mathbb{E}[X] = \sum_n P(B_n) \mathbb{E}[X | B_n].$$

6. — Soient (X_n) ($n = 1, 2, \dots$) une suite de variables aléatoires de même loi et N une variable aléatoire à valeurs entières telles que les termes de la suite N, X_1, X_2, \dots soient mutuellement indépendants. On pose $S_N = X_1 + \dots + X_N$. En utilisant l'exercice précédent et la proposition 6.2 du chapitre 6, établir la formule de Wald : $\mathbb{E}[S_N] = \mathbb{E}[N] \mathbb{E}[X_1]$.

7. — Soit (Z_n) ($n = 1, 2, \dots$) une suite de variables aléatoires ne prenant que deux valeurs, disons 0 et 1. Montrer que si les évènements $\{Z_n = 0\}$ ($n = 1, 2, \dots$) sont indépendants, les variables aléatoires Z_n sont mutuellement indépendantes.

8. — Un joueur possédant a pièces de monnaie différenciées joue une suite de parties, chaque partie consistant à lancer toutes les pièces. On se propose de calculer le nombre moyen de pièces ayant amené au moins une fois pile au cours des n premières parties, ainsi que le nombre minimum moyen de parties qu'il faut jouer pour que chaque pièce amène au moins une fois pile.

Pour $n = 1, 2, \dots$, considérons la variable aléatoire ξ_i^n , égale à 1 ou 0 suivant qu'à la $n^{\text{ième}}$ partie la $i^{\text{ième}}$ pièce a amené pile ou face. On fait l'hypothèse que les variables aléatoires ξ_i^n ($i = 1, 2, \dots, a; n = 1, 2, \dots$) sont indépendantes et ont la même loi : $\frac{1}{2}(\varepsilon_0 + \varepsilon_1)$.

Appelons Y_n le nombre de pièces ayant amené pile pour la première fois à la $n^{\text{ième}}$ partie et X_n le nombre de pièces ayant amené au moins une fois pile au cours des n premières parties. On a les relations $X_n = Y_1 + \dots + Y_n$ et $Y_n = \sum_{i \in A_n} \xi_i^n$, où A_n est l'ensemble des i tels que $\xi_i^1 = \dots = \xi_i^{n-1} = 0$.

a) Montrer que $\text{card } A_n = a - X_{n-1}$. En déduire à l'aide de l'exercice 6 la relation $\mathbb{E}[X_n] = \frac{1}{2} \mathbb{E}[X_{n-1}] + (a/2)$. Calculer $\mathbb{E}[X_n]$.

b) Pour n fixé et $1 \leq i \leq a$ appelons Z_i la variable prenant la valeur 1 ou 0 suivant qu'au cours des n premières parties la $i^{\text{ième}}$ pièce a donné au moins une fois pile ou non, soit $Z_i = \sup_{1 \leq k \leq n} \xi_i^k$. On a aussi $X_n = Z_1 + \dots + Z_a$.

Montrer que les Z_i sont indépendantes (pour n fixé). Trouver leur loi et en

déduire que : $P\{X_n = k\} = \binom{a}{k} \left(1 - \frac{1}{2^n}\right)^k \left(\frac{1}{2^n}\right)^{a-k}$. Retrouver la valeur de $\mathbb{E}[X_n]$ et calculer $\text{Var } X_n$.

9. — A l'entrée d'un restaurant, n personnes donnent leurs chapeaux à la consigne. Après le repas, elles trouvent leurs chapeaux complètement mélangés et chacune prend un chapeau au hasard. On désigne par X_k ($k = 1, 2, \dots, n$) la variable aléatoire qui prend la valeur 1, si la $k^{\text{ième}}$ personne récupère son chapeau et à 0 sinon. On pose $S_n = X_1 + \dots + X_n$, qui représente le nombre de personnes qui ont récupéré leur chapeau.

- a) Construire un espace probabilisé décrivant cette expérience.
- b) Calculer $\mathbb{E}[S_n]$ et $\text{Var } S_n$.
- c) Montrer que la probabilité pour que S_n soit au moins égal à 11 est au plus égale à 0,01 et ceci quel que soit $n \geq 11$.

10. — Soit (X, Y, Z) un triplet de variables aléatoires vérifiant $X + Y + Z = 1$. On suppose que l'on a : $\text{Var } X \leq \text{Var } Y \leq \text{Var } Z < +\infty$. Montrer que

- a) la variable Z est en corrélation négative avec X ainsi qu'avec Y ;
- b) l'on a $\text{Cov}(X, Y) \geq 0$ si et seulement si $\text{Var } X + \text{Var } Y \leq \text{Var } Z$;
- c) l'on a : $|\text{Cov}(X, Z)| \leq |\text{Cov}(Y, Z)|$.

11. — Une variable aléatoire X , de loi inconnue, a une espérance mathématique $\mu = 10$ et un écart-type $\sigma = 5$. Montrer que pour tout $n \geq 50$ la probabilité de l'évènement $\{10 - n < X < 10 + n\}$ est au moins égale à 0,99.

12. — Soit X une variable aléatoire. Montrer que si $\mathbb{E}[|X|] = 0$, alors $X = 0$ p.s. On a la même conclusion en supposant $\mathbb{E}[X^2] = 0$.

13. — Soient a, b deux nombres réels strictement positifs. On pose :

$$A = \frac{a+b}{2}, \quad G = \sqrt{ab}, \quad H = \left[\frac{1}{2} \left(\frac{1}{a} + \frac{1}{b} \right) \right]^{-1}.$$

Montrer que $H \leq G \leq A$ et que $G = \sqrt{AH}$. (G est la moyenne géométrique de A et de H .)

14. — Soit X une variable aléatoire à valeurs positives telle que $\mathbb{E}[X] < +\infty$, $\mathbb{E}[1/X] < +\infty$. Montrer que $\mathbb{E}[X] \mathbb{E}[1/X] \geq 1$.

15. — Soient X une variable aléatoire et r un nombre réel strictement positif tels que $\mathbb{E}[|X|^r] < +\infty$. Montrer que $\text{P}\{|X| \geq n\} = o(1/n^r)$, lorsque n tend vers $+\infty$.

16. — Soit (X_1, X_2, Y_1, Y_2) un système de quatre variables aléatoires admettant un moment du second ordre. Montrer que si (X_1, X_2) est indépendant de (Y_1, Y_2) , alors $\text{Cov}(X_1 + Y_1, X_2 + Y_2) = \text{Cov}(X_1, X_2) + \text{Cov}(Y_1, Y_2)$.

17. — Soit (X, Y) un couple de variables aléatoires indicatrices sur un espace probabilisé $(\Omega, \mathfrak{A}, \text{P})$, c'est-à-dire $X = I_A$, $Y = I_B$, où $A, B \in \mathfrak{A}$. Montrer que X et Y sont indépendantes si et seulement si elles sont non corrélées.

18. — Soit (X, Y) un couple de variables aléatoires tel que $\text{Var } X = \text{Var } Y < +\infty$. Montrer qu'alors les variables aléatoires $X + Y$ et $X - Y$ sont non corrélées.

19 (L'espérance comme approximation d'un paramètre). — Une urne contient des boules numérotées de 1 à N . On effectue n tirages (avec remise) et l'on désigne par X le *plus grand numéro amené*. On peut considérer X comme une variable aléatoire à valeurs dans $\{1, \dots, N\}$, dont la fonction de répartition et l'espérance sont données par

$$P\{X \leq k\} = P\{\text{les } n \text{ nombres amenés sont } \leq k\} = \left(\frac{k}{N}\right)^n \quad (k \in \{1, \dots, N\}).$$

$$\mathbb{E}[X] = \sum_{k=0}^{N-1} P\{X > k\} = \sum_{k=0}^{N-1} (1 - P\{X \leq k\}) = N - \frac{1}{N^n} \sum_{k=0}^{N-1} k^n.$$

Or $\sum_{k=0}^{N-1} k^n \sim N^{n+1}/(n+1)$, d'où $\mathbb{E}[X] \sim (n/(n+1))N$. On voit que pour de grandes valeurs de n l'espérance $\mathbb{E}[X]$ fournit une bonne approximation du nombre N de boules de l'urne. (En pratique, pour estimer N , on prendra plutôt X , le plus grand numéro amené, que $\mathbb{E}[X]$.)

20. — Soient $(\Omega, \mathfrak{A}, P)$ un espace probabilisé, A, B deux éléments de \mathfrak{A} et I_A, I_B les fonctions indicatrices de A, B .

a) On a $\text{Cov}(I_A, I_B) = P(A \cap B) - P(A)P(B)$.

α) $\text{Cov}(I_A, I_B) = 0$ si et seulement si A et B sont indépendants.

β) $\text{Cov}(I_{A^c}, I_B) = -\text{Cov}(I_A, I_B)$ (noter que $I_{A^c} = 1 - I_A$).

b) On a $\sigma^2(I_A) = \text{Var}(I_A) = P(A)(1 - P(A))$, d'où $\text{Var}(I_{A^c}) = \text{Var}(I_A)$.

c) Si $0 < P(A), P(B) < 1$, on peut définir le coefficient de corrélation linéaire du couple (I_A, I_B) (voir § 7). Alors

α) $r(I_{A^c}, I_B) = -r(I_A, I_B)$;

β) $r(I_A, I_B) = 1$ si et seulement si $B = A$ et $r(I_A, I_B) = -1$ si et seulement si $B = A^c$.

Cet exercice ne contenant que de simples vérifications est donné sans démonstration.

21. — Soit X une variable aléatoire de Bernoulli : $q\varepsilon_0 + p\varepsilon_1$ ($p, q \geq 0$; $p + q = 1$).

a) Si $p \neq q$, alors X admet une médiane et une seule M , égale à 0 si $p < q$, égale à 1 si $p > q$.

b) Si $p = q = \frac{1}{2}$, tout nombre de l'intervalle $[0, 1]$ est une médiane de X .

CHAPITRE 9
FONCTIONS GÉNÉRATRICES

Dans ce chapitre, on étudie plus particulièrement les mesures de probabilité discrètes *portées par les entiers positifs* et les variables aléatoires ayant pour loi ces mesures. Comme il sera expliqué, on peut ainsi envoyer bijectivement ces mesures sur des séries de puissances et utiliser ensuite l'analyse classique des séries pour calculer les valeurs caractéristiques de ces lois, comme l'espérance mathématique et les moments.

1. Définitions. — On désigne par \mathfrak{M} l'ensemble des mesures de probabilité de la forme

$$P = \sum_{i=0}^{\infty} \alpha_i \varepsilon_i$$

et par \mathfrak{M}' l'ensemble des variables aléatoires réelles (discrètes) définies sur un espace probabilisé $(\Omega, \mathfrak{A}, P)$ et dont la loi de probabilité appartient à \mathfrak{M} . Suivant la définition donnée au chapitre précédent, le *produit de convolution* de P par la mesure de probabilité $Q = \sum_{j=0}^{\infty} \beta_j \varepsilon_j$ est la mesure de probabilité

$$(1.1) \quad P * Q = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \alpha_i \beta_j \varepsilon_{(i+j)}.$$

On peut encore écrire :

$$P * Q = \sum_{k=0}^{\infty} \gamma_k \varepsilon_k, \quad \text{avec} \quad \gamma_k = \sum_{i=0}^k \alpha_i \beta_{k-i} \quad (k \geq 0).$$

PROPOSITION 1.1. — *Soient P, Q, R, \dots des mesures de probabilité de la classe \mathfrak{M} ; on a les propriétés :*

- (i) $P * Q \in \mathfrak{M}$;
- (ii) $P * Q = Q * P$;
- (iii) $P * (Q * R) = (P * Q) * R$;
- (iv) $P * \varepsilon_0 = \varepsilon_0 * P = P$;
- (v) $\varepsilon_n * \varepsilon_m = \varepsilon_{(n+m)}$ pour $n \geq 0$ et $m \geq 0$.

Ces propriétés résultent immédiatement de la définition du produit de convolution qui prend la forme (1.1) pour les mesures de probabilité appartenant à \mathfrak{M} .

D'après (iii) on peut poser pour tout $P \in \mathfrak{M}$

$$P^{1*} = P \quad \text{et} \quad P^{n*} = P * P^{(n-1)*} \quad \text{pour } n \geq 2.$$

En particulier

$$\varepsilon_1^{n*} = \varepsilon_n \quad \text{pour tout } n \geq 1.$$

La proposition suivante est alors une conséquence de cette dernière propriété et de la Proposition 3.2 du chapitre 8.

PROPOSITION 1.2. — *Soit P une mesure de probabilité appartenant à \mathfrak{M} . La somme S_n de n ($n \geq 1$) variables aléatoires réelles, discrètes, mutuellement indépendantes et de même loi égale à P a pour loi de probabilité P^{n*} .*

Désignons par $\mathfrak{M}(s)$ l'ensemble des séries de puissances à une indéterminée s dont les coefficients sont réels, positifs et de somme 1; soit, par conséquent,

$$\mathfrak{M}(s) = \left\{ \sum_{i=0}^{\infty} \alpha_i s^i : \alpha_i \geq 0, \sum_{i=0}^{\infty} \alpha_i = 1 \right\}.$$

De nouveau, la proposition suivante est une conséquence immédiate de la définition du produit de convolution pour les éléments de \mathfrak{M} , en se rappelant que dans $\mathfrak{M}(s)$ le produit de deux séries de puissances respecte les règles usuelles de distributivité pour le produit de deux sommes et le calcul sur les puissances : $s^i s^j = s^{i+j}$.

PROPOSITION 1.3. — *L'application $P \mapsto G_P(s)$, qui, à la mesure de probabilité $P = \sum_{i=0}^{\infty} \alpha_i \varepsilon_i$, de la classe \mathfrak{M} fait correspondre la série de puissances*

$$(1.2) \quad G_P(s) = \sum_{i=0}^{\infty} \alpha_i s^i,$$

est une bijection de \mathfrak{M} sur $\mathfrak{M}(s)$ satisfaisant à

$$(1.3) \quad G_{P*Q}(s) = G_P(s) G_Q(s),$$

pour tout P, Q appartenant à la classe \mathfrak{M} .

La série de puissances $G_P(s)$, donnée en (1.2), est appelée *fonction génératrice* de la mesure de probabilité P. Si X est une variable aléatoire dont la loi P appartient à \mathfrak{M} , on note aussi $G_X(s)$ la fonction génératrice de la loi P et par abus de langage on parle de *fonction génératrice de la variable aléatoire X*.

Remarque. — La fonction génératrice $G_P(s) = \sum_{i=0}^{\infty} \alpha_i s^i$ converge absolument et sa somme $G_P(s)$ est une fonction continue de s dans l'intervalle $[-1, +1]$ pour s réel (resp. dans le disque $|s| \leq 1$ pour s complexe). Elle admet, d'autre part, des dérivées de tous les ordres que l'on obtient comme sommes des séries dérivées, pour tout s tel que $|s| < 1$. On pourra ainsi faire appel aux techniques de dérivation et d'intégration des séries de puissances pour obtenir des propriétés sur les valeurs typiques comme l'espérance mathématique ou les moments.

THÉORÈME 1.4 (théorème d'unicité). — *La fonction génératrice d'une variable aléatoire (à valeurs entières positives) détermine la loi de cette variable. En d'autres termes, si deux variables aléatoires (à valeurs entières positives) admettent même fonction génératrice, alors elles ont même loi.*

Démonstration. — Soit X une variables aléatoire à valeurs dans \mathbb{N} . Posons $p_k = P\{X = k\}$ ($k \geq 0$) et désignons par $G(s)$ la fonction génératrice de X , soit

$$(1.4) \quad G(s) = \sum_{k \geq 0} p_k s^k.$$

Nous allons montrer que l'on peut déterminer la suite (p_k) ($k \geq 0$) à partir de la fonction G . En effet, on a, en se servant du fait que l'on peut dériver (1.4) terme à terme dans l'intervalle $] - 1, +1[$:

$$\begin{aligned} G(0) &= p_0; \\ G'(s) &= \sum_{k \geq 0} k p_k s^{k-1} = \sum_{k \geq 1} k p_k s^{k-1}, \quad G'(0) = p_1; \\ G''(s) &= \sum_{k \geq 0} k(k-1) p_k s^{k-2} = \sum_{k \geq 2} k(k-1) p_k s^{k-2}, \quad G''(0) = 2p_2; \\ &\dots \quad \dots \\ G^{(n)}(s) &= \sum_{k \geq 0} k(k-1) \dots (k-n+1) p_k s^{k-n} \\ &= \sum_{k \geq n} k(k-1) \dots (k-n+1) p_k s^{k-n}, \quad G^{(n)}(0) = n! p_n \quad (n \geq 0). \end{aligned}$$

On a donc, pour tout $n \geq 0$, l'identité $G^{(n)}(0) = n! p_n$; d'où il résulte que la connaissance de G détermine celle de la loi (p_k) ($k \geq 0$) de X . \square

2. Propriétés. — Donnons tout d'abord une première propriété montrant que l'espérance mathématique d'une variable aléatoire de la famille \mathfrak{M}' peut aussi se calculer à l'aide de la série de terme général $P\{X > i\}$ ($i \geq 0$).

PROPOSITION 2.1. — *Soit X une variable aléatoire de la classe \mathfrak{M}' . Alors l'identité*

$$(2.1) \quad \sum_{i=1}^{\infty} i P\{X = i\} = \sum_{i=0}^{\infty} P\{X > i\},$$

est valable dans la demi-droite achevée $[0, +\infty]$. Si l'une des séries figurant dans cette identité est convergente, il en est de même de l'autre et la valeur commune de leurs sommes est l'espérance mathématique $\mathbb{E}[X]$.

Démonstration. — Pour chaque $i \geq 0$ posons $\alpha_i = P\{X = i\}$. On a, dans $[0, +\infty]$,

$$(2.2) \quad \sum_{i \geq 0} P\{X > i\} = \sum_{i \geq 1} P\{X \geq i\} = \sum_{i \geq 1} \left(\sum_{j \geq i} \alpha_j \right).$$

Le dernier membre est une série double *itérée*, à termes *positifs*. En échangeant l'ordre des sommations, ce qui est licite en vertu du théorème de Fubini, il vient, toujours dans $[0, +\infty]$

$$\sum_{i \geq 1} \left(\sum_{j \geq i} \alpha_j \right) = \sum_{j \geq 1} \left(\sum_{1 \leq i \leq j} \alpha_j \right) = \sum_{j \geq 1} j \alpha_j. \quad \square$$

Remarque. — On peut démontrer la Proposition 2.1 comme suit : toutes les égalités étant dans $[0, +\infty]$, on a $X = \sum_{i=0}^{\infty} I_{\{X > i\}}$, d'où, en prenant l'espérance mathématique et en intervertissant les signes \mathbb{E} et \sum (ce qui est licite, toutes les quantités étant positives), il vient :

$$\mathbb{E}[X] = \sum_{i=0}^{\infty} \mathbb{E}[I_{\{X > i\}}] = \sum_{i=0}^{\infty} \mathbb{P}\{X > i\}.$$

La précédente propriété suggère l'introduction d'une seconde fonction génératrice associée à X et définie par :

$$(2.3) \quad H_X(s) = \sum_{i=0}^{\infty} \mathbb{P}\{X > i\} s^i.$$

Si l'on considère s comme une variable réelle, cette série est convergente dans l'intervalle ouvert $] -1, +1[$ et la relation fonctionnelle entre $G_X(s)$ et $H_X(s)$ est donnée dans la proposition suivante.

PROPOSITION 2.2. — *Pour $|s| < 1$, on a :*

$$H_X(s) = \frac{1 - G_X(s)}{1 - s}.$$

Démonstration. — Reprenons les notations de la démonstration de la Proposition 2.1. Pour $i \geq 1$ le coefficient de s^i dans le produit $(1 - s)H_X(s)$ est égal à $\beta_i - \beta_{i-1}$, c'est-à-dire $-\alpha_i$ et le coefficient de s^0 vaut $\beta_0 = 1 - \alpha_0$. On a donc bien : $(1 - s)H_X(s) = 1 - G_X(s)$. \square

Si l'on connaît l'expression de $G_X(s)$ ou de $H_X(s)$, on peut déterminer l'espérance mathématique de X et de ses moments — du moins, dans certaines conditions — comme on va le voir.

PROPOSITION 2.3. — *La fonction génératrice $G_X(s)$ admet une dérivée à gauche $G'_X(1)$ en $s = 1$, si et seulement si $\mathbb{E}[X]$ existe et est fini, et l'on a :*

$$(2.4) \quad \mathbb{E}[X] = G'_X(1).$$

En outre, la fonction $H_X(s)$ admet une limite à gauche $H_X(1)$ en $s = 1$, si et seulement si $\mathbb{E}[X]$ existe et est fini, et l'on a :

$$(2.5) \quad \mathbb{E}[X] = H_X(1).$$

Pour démontrer cette proposition, il est bon de faire appel au lemme d'Abel, qu'on peut énoncer de la façon suivante.

LEMME (Abel)

1) Si la série $\sum_i \alpha_i$ ($i \geq 0$) est convergente de somme α , alors

$$\lim_{s \rightarrow 1-0} \sum_{i=0}^{\infty} \alpha_i s^i = \sum_{i=0}^{\infty} \alpha_i = \alpha.$$

2) Si les α_i sont positifs et si $\lim_{s \rightarrow 1-0} \sum_{i=0}^{\infty} \alpha_i s^i = \alpha \leq +\infty$, alors

$$\sum_{i=0}^{\infty} \alpha_i = \alpha.$$

Démonstration

1) Il s'agit de montrer que $\lim_{s \rightarrow 1-0} \left| \sum_{i=0}^{\infty} \alpha_i (s^i - 1) \right| = 0$. Comme la série de terme général α_i converge, à tout $\varepsilon > 0$ on peut associer $N(\varepsilon)$ tel que pour tout $N' \geq N$ on ait $\left| \sum_{N \leq i \leq N'} \alpha_i \right| \leq \varepsilon/4$. Choisissons un tel N ; il vient

$$\left| \sum_{i=0}^{\infty} \alpha_i (s^i - 1) \right| \leq \left| \sum_{i=0}^N \alpha_i (s^i - 1) \right| + \left| \sum_{i=N+1}^{\infty} \alpha_i (s^i - 1) \right|.$$

Or pour tout $s \in [0, 1[$ on a

$$\left| \sum_{i=0}^N \alpha_i (s^i - 1) \right| \leq MN |s^N - 1| \text{ où } M = \max_{0 \leq i \leq N} |\alpha_i| < +\infty,$$

de sorte que, pour s suffisamment voisin de 1, on a : $\left| \sum_{i=0}^N \alpha_i (s^i - 1) \right| < \frac{\varepsilon}{2}$.

Pour majorer le second terme du membre de droite, on effectue la sommation par parties (précisément due à Abel). En posant $A_i = \sum_{k \geq i} \alpha_k$, il vient :

$$\begin{aligned} \left| \sum_{i=N+1}^{\infty} \alpha_i (s^i - 1) \right| &= \left| \sum_{i=N+1}^{\infty} (A_i - A_{i+1})(s^i - 1) \right| \\ &= \left| A_{N+1}(s^{N+1} - 1) + \sum_{i=N+2}^{\infty} A_i (s^i - s^{i-1}) \right| \\ &\leq \frac{\varepsilon}{4} |s^{N+1} - 1| + \frac{\varepsilon}{4} s^{N+1} < \frac{\varepsilon}{2}. \end{aligned}$$

On obtient finalement $\left| \sum_{i=0}^{\infty} \alpha_i (s^i - 1) \right| < \varepsilon$ pourvu que s soit suffisamment voisin de 1.

2) Puisque $\sum_{i=0}^{\infty} \alpha_i s^i \leq \sum_{i=0}^{\infty} \alpha_i$ pour $0 < s < 1$, le cas $\alpha = +\infty$ est évident. Supposons α fini. D'après l'hypothèse, on a $\sum_{i=0}^{\infty} \alpha_i s^i < \alpha < +\infty$ pour

$0 < s < 1$; on en déduit, pour tout $n \geq 1$, l'inégalité $\sum_{i=0}^n \alpha_i \leq \alpha$. Comme $\sum_{i=0}^n \alpha_i$ est une fonction de n bornée et croissante, elle tend vers une limite, soit α' . On peut alors appliquer la première partie du Lemme et en déduire $\alpha' = \alpha$. \square

La démonstration de la Proposition 2.3 se présente alors ainsi : si $\mathbb{E}[X]$ est finie, alors la série $\sum_i i\alpha_i$ a une somme finie. Pour $|s| < 1$ on peut dériver terme à terme la série $\sum_i \alpha_i s^i$, pour obtenir $G'_X(s) = \sum_{i=1}^{\infty} i\alpha_i s^{i-1}$. La première partie du lemme d'Abel entraîne alors $\lim_{s \rightarrow 1-0} G'_X(s) = \sum_{i=1}^{\infty} i\alpha_i = \mathbb{E}[X]$. Si $\lim_{s \rightarrow 1-0} \sum_{i=1}^{\infty} i\alpha_i s^{i-1} = \lim_{s \rightarrow 1-0} G'_X(s) = \alpha$, la seconde partie du lemme d'Abel entraîne que la somme $\sum_{i=0}^{\infty} i\alpha_i$ est égale à α , fini ou infini. La relation (2.4) est donc satisfaite.

Pour $|s| < 1$, on a $H_X(s) = \frac{1 - G_X(s)}{1 - s} = \frac{G_X(1) - G_X(s)}{1 - s} = G'_X(\sigma)$, pour $s \leq \sigma \leq 1$. Comme $H_X(s)$ et $G'_X(s)$ sont monotones, elles ont les mêmes limites en 1, finie ou infinie. \square

Les propositions suivantes se prouvent de façon analogue. Nous ne reproduisons pas leur démonstration.

PROPOSITION 2.4. — *La fonction $G_X(s)$ admet une dérivée à gauche $G_X^{(r)}(1)$ d'ordre r (r entier strictement positif) en $s = 1$, si et seulement si le moment factoriel d'ordre r , à savoir $\mathbb{E}[X(X-1)\dots(X-r+1)]$, existe et est fini. On a alors :*

$$(2.6) \quad \mathbb{E}[X(X-1)\dots(X-r+1)] = G_X^{(r)}(1);$$

en particulier, pour $r = 2$, on a :

$$(2.7) \quad \mathbb{E}[X(X-1)] = G_X''(1) = 2H'_X(1);$$

et par conséquent :

$$(2.8) \quad \begin{aligned} \text{Var } X &= G_X''(1) + G'_X(1) - (G'_X(1))^2 \\ &= 2H'_X(1) + H_X(1) - (H_X(1))^2. \end{aligned}$$

PROPOSITION 2.5. — *Supposons que la fonction $G_X(s)$ puisse être développée en série de Taylor au voisinage de $s = 1$, ou, ce qui revient au même, que la fonction $G_X(1+u)$ admette un tel développement au voisinage de $u = 0$. Alors le moment factoriel d'ordre $r \geq 1$ apparaît comme coefficient de $u^r/r!$ dans ce développement, c'est-à-dire,*

$$G_X(1+u) = 1 + \sum_{r \geq 1} \mathbb{E}[X(X-1)\dots(X-r+1)] \frac{u^r}{r!}.$$

3. Somme de variables aléatoires. — On examine d'abord le cas où le nombre de termes de cette somme est fixé, puis celui où ce nombre est lui-même aléatoire.

PROPOSITION 3.1. — *Si X et Y sont des variables aléatoires indépendantes, alors*

$$(3.1) \quad G_{X+Y}(s) = G_X(s) G_Y(s).$$

Démonstration. — En effet, si P et Q sont les lois de probabilité de X et de Y , alors la loi de $X + Y$ est $P * Q$. La proposition résulte alors de la Proposition 1.3. \square

COROLLAIRE. — *Si X_1, X_2, \dots, X_n sont des variables aléatoires mutuellement indépendantes et de même loi, dont la fonction génératrice est $G(s)$, alors la fonction génératrice de $S_n = X_1 + X_2 + \dots + X_n$ est donnée par :*

$$(3.2) \quad G_{S_n}(s) = (G(s))^n.$$

Nous étudions maintenant le cas où l'on prend un nombre *aléatoire* de variables aléatoires. Soit (X_n) une suite de variables aléatoires mutuellement indépendantes, de même loi $P_X \in \mathfrak{M}$, dont la fonction génératrice est $G_X(s)$ et toutes définies sur un même espace probabilisé $(\Omega, \mathfrak{A}, P)$. Soit de plus N une variable aléatoire, définie sur le même espace, indépendante des X_n , de loi $P_N \in \mathfrak{M}$ et dont la fonction génératrice est $G_N(s)$. On pose $S_0 = 0$, $S_n = X_1 + \dots + X_n$ ($n \geq 1$) et l'on considère la variable aléatoire :

$$S_N : \omega \longrightarrow S_{N(\omega)}(\omega) = X_1(\omega) + \dots + X_{N(\omega)}(\omega).$$

En fait, pour définir S_N , il faut introduire un produit (infini) d'espaces probabilisés, mais seul le calcul de la fonction génératrice de S_N nous importe ici.

Pour tout entier $j \geq 0$, on peut écrire

$$\{S_N = j\} = \sum_{n=0}^{\infty} \{S_N = j, N = n\} = \sum_{n=0}^{\infty} \{X_1 + \dots + X_n = j, N = n\}.$$

Ceci montre, en particulier, que S_N est une variable aléatoire, puisque le second membre est une réunion dénombrable d'évènements. Comme les variables X_n sont indépendantes de N , chaque variable S_n est indépendante de N . On en déduit :

$$\begin{aligned} P\{S_N = j\} &= \sum_{n=0}^{\infty} P\{S_N = j, N = n\} \\ &= \sum_{n=0}^{\infty} P\{S_n = j, N = n\} = \sum_{n=0}^{\infty} P\{S_n = j\} P\{N = n\}. \end{aligned}$$

D'où

$$G_{S_N}(s) = \sum_{j=0}^{\infty} P\{S_N = j\} s^j = \sum_{j=0}^{\infty} \left(\sum_{n=0}^{\infty} P\{S_n = j\} P\{N = n\} \right) s^j.$$

En échangeant l'ordre des sommations, on obtient :

$$\begin{aligned} G_{S_N}(s) &= \sum_{n=0}^{\infty} \mathbb{P}\{N = n\} \left(\sum_{j=0}^{\infty} \mathbb{P}\{S_n = j\} s^j \right) \\ &= \sum_{n=0}^{\infty} \mathbb{P}\{N = n\} (G_X(s))^n \quad [\text{d'après le corollaire précédent}] \\ &= G_N(G_X(s)). \end{aligned}$$

On a démontré ainsi la proposition suivante.

PROPOSITION 3.2. — *La fonction génératrice de la somme $S_N = X_1 + \dots + X_N$, où N est une variable aléatoire à valeurs entières, indépendante de la suite (X_n) ($n \geq 1$), est la fonction composée :*

$$(3.3) \quad G_{S_N}(s) = G_N(G_X(s)) = G_N \circ G_X(s).$$

4. Le théorème de continuité

THÉORÈME 4.1. — *Donnons-nous une suite $(p_{n,k}, k \geq 0)$ ($n \geq 0$) de lois de probabilité sur \mathbb{N} et une suite $(\alpha_k, k \geq 0)$ de nombres positifs vérifiant $\sum_{k \geq 0} \alpha_k \leq 1$. Posons :*

$$G_n(u) = \sum_{k \geq 0} p_{n,k} u^k \quad (n \geq 0) \quad \text{et} \quad G(u) = \sum_{k \geq 0} \alpha_k u^k.$$

Alors les deux propriétés suivantes sont équivalentes :

- a) Pour tout $k \geq 0$ on a : $\lim_{n \rightarrow \infty} p_{n,k} = \alpha_k$;
- b) Pour tout $u \in]0, 1[$ on a : $\lim_{n \rightarrow \infty} G_n(u) = G(u)$.

Démonstration

a) \Rightarrow b) : soit $u \in]0, 1[$; pour tout $\varepsilon > 0$ il existe un entier $N(\varepsilon)$ tel que $\sum_{k > N} u^k < \varepsilon$. De là

$$|G_n(u) - G(u)| \leq \sum_{k \geq 0} |p_{n,k} - \alpha_k| |u|^k \leq \sum_{k=0}^N |p_{n,k} - \alpha_k| + \sum_{k > N} u^k.$$

D'où $|G_n(u) - G(u)| < \sum_{k=0}^N |p_{n,k} - \alpha_k| + \varepsilon$. En maintenant N fixé et en faisant tendre n vers l'infini, on obtient le résultat, le nombre $\varepsilon > 0$ étant arbitraire.

b) \Rightarrow a) : le procédé diagonal classique montre que de toute suite $(P_n) = ((p_{n,k}, k \geq 1))$ ($n \geq 1$) de lois de probabilité sur \mathbb{N} on peut extraire une sous-suite $(P_{n'})$ convergente, c'est-à-dire, telle que pour tout $k \geq 0$ la limite $\lim_{n' \rightarrow \infty} p_{n',k}$ existe.

Si (P_n) admet deux sous-suites convergentes $(P_{n'})$ et $(P_{n''})$, on aura, en vertu de la partie a) \Rightarrow b) du théorème :

$$\lim_{n' \rightarrow \infty} G_{n'}(u) = G(u), \quad \lim_{n'' \rightarrow \infty} G_{n''}(u) = G(u).$$

Ainsi les limites de deux sous-suites convergentes ont même fonction génératrice; comme la fonction génératrice d'une suite détermine cette suite, toutes les sous-suites convergentes convergent vers la même limite. Donc pour tout $k \geq 0$ la limite $\lim_{n \rightarrow \infty} p_{n,k}$ existe. Appelons-la α_k . La suite $(\alpha_k, k \geq 0)$ a alors pour fonction génératrice $G(u)$. \square

Dans le cas où la limite $(\alpha_k, k \geq 0)$ est elle-même une loi de probabilité sur \mathbb{N} , on peut énoncer le résultat suivant.

THÉORÈME 4.2. — *Donnons-nous une suite (X_n) ($n \geq 1$) de variables aléatoires à valeurs dans \mathbb{N} et une variable aléatoire X à valeurs dans \mathbb{N} . Notons $(p_{n,k}, k \geq 0)$ la loi de X_n et $G_n(u) = \mathbb{E}[u^{X_n}]$ sa fonction génératrice, enfin $(\alpha_k, k \geq 0)$ la loi de X et $G(u) = \mathbb{E}[u^X]$ sa fonction génératrice. Alors les deux propriétés suivantes sont équivalentes :*

- a) *Pour tout $k \geq 0$ on a : $\lim_{n \rightarrow \infty} p_{n,k} = \alpha_k$ (X_n tend vers X en loi)¹ ;*
- b) *Pour tout $u \in]0, 1[$ on a : $\lim_{n \rightarrow \infty} G_n(u) = G(u)$.*

COMPLÉMENTS ET EXERCICES

1. — Déterminer l'expression de la fonction génératrice des lois binomiale $B(n, p)$, de Poisson π_λ , géométrique $\sum_{k \geq 1} pq^{k-1} \varepsilon_k$.

2. — Soit X une variable aléatoire de loi géométrique $\sum_{k \geq 1} q^{k-1} p \varepsilon_k$ ($0 < p < 1, q = 1 - p$). Montrer que tous les moments factoriels de X existent; les calculer; étudier le cas particulier $p = \frac{1}{2}$.

3. — Retrouver, en utilisant la technique des fonctions génératrices, les identités de convolution :

$$B(n, p) * B(m, p) = B(n + m, p); \quad \pi_\lambda * \pi_\mu = \pi_{\lambda + \mu}.$$

4. — Toujours au moyen de la technique des fonctions génératrices, évaluer $\mathbb{E}[X]$, $\text{Var } X$, lorsque X suit respectivement la loi $B(n, p)$ et la loi π_λ .

5. — On pose : $F(a, b, c; s) = {}_2F_1\left(\begin{matrix} a, b \\ c \end{matrix}; s\right) = \sum_{n \geq 0} \frac{(a)_n (b)_n}{(c)_n} \frac{s^n}{n!}$.

- a) Vérifier que l'on a $F'(a, b, c; s) = \frac{ab}{c} F(a + 1, b + 1, c + 1; s)$.

¹ cf. Théorème 6.1 du chapitre 16.

b) La loi hypergéométrique $H(n, N, M) = \sum_k \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \varepsilon_k$, où k varie dans l'intervalle $[\max\{0, n - (N - M)\}, \min\{n, M\}]$ a pour fonction génératrice

$$G_H(s) = \begin{cases} \frac{(-N+M)_n}{(-N)_n} {}_2F_1 \left(\begin{matrix} -M, -n \\ N-M-n+1 \end{matrix}; s \right), & \text{si } n \leq N-M; \\ \frac{(n-N+M+1)_{N-n}}{(n+1)_{N-n}} s^{n-N+M} {}_2F_1 \left(\begin{matrix} -N+n, -N+M \\ n-N+M+1 \end{matrix}; s \right), & \text{si } n \geq N-M+1. \end{cases}$$

si $n \geq N - M + 1$. La loi *hypergéométrique* doit son nom à cette propriété. Noter que $G_H(s)$ est symétrique en n et M .

c) En déduire, tout en utilisant l'identité de Chu-Vandermonde, que l'espérance mathématique d'une variable hypergéométrique (égale à $G'_H(1)$) vaut nM/N .

d) Vérifier que la dérivée seconde de $G_H(s)$ en $s = 1$ est égale à

$$\frac{M(M-1)n(n-1)}{N(N-1)}.$$

En tirer que la variance d'une variable hypergéométrique est donnée par

$$\text{Var } X = \frac{nM(N-M)}{N^2} \left[1 - \frac{n-1}{N-1} \right].$$

6. — Soit $G(s)$ la fonction génératrice d'une variable aléatoire $X \in \mathfrak{M}'$. Trouver la fonction génératrice des variables $X + b$ et aX , où a, b sont des entiers positifs.

7. — En désignant par $G(s)$ la fonction génératrice d'une variable aléatoire X de la classe \mathfrak{M}' , donner, en fonction de $G(s)$, l'expression de $J(s) = \sum_{n \geq 0} u_n s^n$, lorsque u_n est respectivement égal à :

- a) $P\{X \leq n\}$; b) $P\{X < n\}$; c) $P\{X \geq n\}$; d) $P\{X > n + 1\}$;
e) $P\{X = 2n\}$ pour tout $n \geq 0$.

8. — Un sac contient une boule blanche et deux boules rouges. On répète une infinité de fois l'opération qui consiste à tirer une boule, la remettre dans le sac si elle est blanche, l'éliminer si elle est rouge. On appelle X_n la variable aléatoire prenant la valeur 0 ou 1 suivant qu'à la $n^{\text{ième}}$ opération on a tiré une boule rouge ou blanche et l'on pose : $R_n = \{X_n = 0\}$ et $B_n = \{X_n = 1\}$.

a) Soit T_1 l'instant où l'on a tiré la première boule rouge (T_1 est ainsi le plus petit entier $m \geq 1$ tel que $X_m = 0$). Calculer $P\{T_1 = m\}$ ($m \geq 1$) et la fonction génératrice de T_1 . En déduire que $\sum_{m \geq 1} P\{T_1 = m\} = 1$ et les valeurs de l'espérance mathématique et de la variance de T_1 .

b) Soit T_2 l'instant où l'on a tiré la deuxième boule rouge. Calculer $P\{T_1 = m, T_2 = n\}$ pour $1 \leq m < n$.

c) En déduire $P\{T_2 = n\}$. Calculer la fonction génératrice de T_2 . Montrer que presque sûrement T_2 est fini. Calculer $\mathbb{E}[T_2]$ et $\text{Var } T_2$.

d) A l'aide de ce qui précède, calculer $P\{X_n = 0\}$ et en déduire la loi de X_n .

9. — On garde les mêmes hypothèses que dans la Proposition 3.2. Montrer que si X_1 et N ont une espérance et une variance finies, il en est de même pour S_N et que

$$\mathbb{E}[S_N] = \mathbb{E}[N] \mathbb{E}[X_1] \quad \text{et} \quad \text{Var } S_N = \mathbb{E}[N] \text{Var } X_1 + \text{Var } N (\mathbb{E}[X_1])^2.$$

10. — Dans une réaction nucléaire, une particule élémentaire provoque l'apparition de X_1 particules de même nature, dites de première génération. La $i^{\text{ième}}$ particule de la première génération engendre, indépendamment des autres, ξ_i^1 particules ($i = 1, 2, \dots, X_1$); le nombre de particules de la deuxième génération est donc $X_2 = \xi_1^1 + \dots + \xi_{X_1}^1$. Les variables aléatoires X_n et ξ_i^n sont définies par récurrence de la même façon : la taille de la $n^{\text{ième}}$ génération est X_n et ξ_i^n le nombre de descendants de la $i^{\text{ième}}$ particule de la $n^{\text{ième}}$ génération.

Pour $n \geq 1$ on a donc la relation : $X_{n+1} = \xi_1^n + \dots + \xi_{X_n}^n$. On fait l'hypothèse que pour tout $n \geq 1$ les variables aléatoires $X_n, \xi_1^n, \dots, \xi_{X_n}^n$ sont indépendantes et que les ξ_i^n ont toutes même loi que X_1 . On désigne par $G(s)$ la fonction génératrice de X_1 .

a) Soit $G_n(s)$ la fonction génératrice de X_n . Montrer que

$$G_{n+1}(s) = G_n(G(s)) = G(G_n(s)), \quad \text{pour } n \geq 1.$$

b) Montrer que la fonction $G(s)$ est croissante et convexe dans l'intervalle $[0, 1]$.

c) On pose $x_n = P\{X_n = 0\} = G_n(0)$; montrer que la suite (x_n) ($n \geq 1$) est croissante et que sa limite x est la plus petite solution comprise entre 0 et 1 de l'équation :

$$(*) \quad G(\xi) = \xi.$$

d) Soit $\mu = \mathbb{E}[X_1]$ le nombre moyen de descendants d'une particule. On suppose $G(s) \neq s$; à l'aide de b) montrer que :

(i) si $\mu \leq 1$, la seule racine de l'équation (*) comprise entre 0 et 1 est $\xi = 1$; donc $x = 1$;

(ii) si $\mu > 1$, l'équation (*) admet une solution unique ξ telle que $0 \leq \xi < 1$; on a donc : $x = \xi$.

e) Interpréter x et le résultat précédent.

f) Soit $\sigma^2 = \text{Var } X_1$. En établissant des formules de récurrence, calculer $\mathbb{E}[X_n]$ et $\text{Var } X_n$ en fonction de μ et σ^2 .

g) Calculer $G_n(s)$ dans le cas où $P_{X_1} = q\varepsilon_0 + p\varepsilon_1$ ($p + q = 1$).

11. — On reprend les notations de l'exercice 8 du chapitre précédent. Soit T le plus petit entier tel que $X_T = a$. Calculer $t_n = P\{T > n\}$ ($n \geq 0$). Trouver l'expression de la fonction génératrice $H(s) = \sum_{n \geq 0} t_n s^n$. En déduire $\mathbb{E}[T]$.

12. — Il est souvent possible de calculer explicitement les termes d'une suite (u_n) ($n \geq 0$) si sa fonction génératrice $U(s) = \sum_{n \geq 0} u_n s^n$ a une forme analytique particulière, par exemple si c'est une fraction rationnelle. Soit $U(s) = P(s)/Q(s)$ une fraction rationnelle irréductible telle que $Q(0) \neq 0$. On appelle s_1, \dots, s_m les racines (réelles ou complexes) de $Q(s)$ et r_1, \dots, r_m leur ordre de multiplicité. On suppose d'abord que le degré de $P(s)$ est strictement inférieur à celui de $Q(s)$. La décomposition de $U(s)$ en éléments simples a donc la forme :

$$U(s) = \sum_{1 \leq i \leq m} \sum_{1 \leq j \leq r_i} \frac{a_{ij}}{(s - s_i)^j}.$$

a) Montrer que $a_{ir_i} = r_i! P(s_i)/Q^{(r_i)}(s_i)$.

b) Montrer que pour $|s| < \inf_{1 \leq i \leq m} |s_i|$ la fonction $U(s)$ se développe en série entière

$$U(s) = \sum_{n \geq 0} u_n s^n, \quad \text{avec} \quad u_n = \sum_{1 \leq i \leq m} \sum_{1 \leq j \leq r_i} a_{ij} (-1)^j \frac{(j)_n}{n!} s_i^{-n-j},$$

où l'on a posé $(j)_n = j(j+1) \dots (j+n-1)$. On obtient ainsi une expression *exacte* pour u_n .

c) On suppose maintenant qu'il existe une racine et *une seule*, s_1 , qui soit strictement plus petite en module que les autres racines, i.e., $|s_1| < |s_i|$ pour $i = 2, \dots, m$. Montrer qu'on a alors

$$u_n \sim a_{1r_1} (-1)^{r_1} \frac{(r_1)_n}{n!} s_1^{-n-r_1},$$

lorsque n tend vers l'infini.

d) Montrer que le résultat de c) subsiste lorsque le degré de $P(s)$ est supérieur ou égal à celui de $Q(s)$.

13. — On effectue une suite de parties de pile ou face. Soit u_n ($n \geq 1$) la probabilité de ne pas avoir trois fois « face » à la suite au cours des n premières parties.

a) On a évidemment $u_1 = u_2 = 1$ et l'on pose $u_0 = 1$. Montrer que pour $n \geq 3$ on a la relation de récurrence :

$$u_n = \frac{1}{2}u_{n-1} + \frac{1}{4}u_{n-2} + \frac{1}{8}u_{n-3}.$$

b) On pose $U(s) = \sum_{n \geq 0} u_n s^n$. En déduire l'identité :

$$U(s) = \frac{2s^2 + 4s + 8}{8 - 4s - 2s^2 - s^3}.$$

c) On montrera que le dénominateur $Q(s) = 8 - 4s - 2s^2 - s^3$ a une racine strictement positive $s_1 = 1,087\dots$ et deux racines complexes dont le module est strictement plus grand que s_1 . (En effet, pour $|s| < s_1$, on a $|4s + 2s^2 + s^3| < 4s_1 + 2s_1^2 + s_1^3 = 8$ et l'on a la même inégalité pour $|s| = s_1$, $s \neq s_1$.)

d) Utiliser la technique du précédent exercice pour évaluer u_n .

14. — Cet exercice commenté, qui n'utilise que des techniques de ce chapitre, est donné sans solution. On y retrouve un cas très particulier du théorème dit des renouvellements (*cf.* Feller (*op. cit.*), chap. 13).

On considère une lampe électrique dont la durée de fonctionnement T est une variable aléatoire à valeurs entières. Les probabilités d'extinction $f_k = P\{T = k\}$ ($k = 1, 2, \dots$) sont données et vérifient $\sum_{k \geq 1} f_k = 1$. A l'instant initial $t = 0$, la lampe est neuve. Dès qu'elle s'éteint, on la remplace par une lampe neuve du même type; et ainsi de suite... On définit une suite de variables aléatoires (X_n) ($n = 1, 2, \dots$) comme suit : $X_n = 1$ ou 0 , suivant qu'à l'instant n il faut procéder à un remplacement ou non. Par hypothèse, on a donc : $P\{X_1 = 1\} = f_1$, puis $P\{X_1 = \dots = X_{n-1} = 0, X_n = 1\} = f_n$ pour $n \geq 2$ et enfin $P\{X_{k+1} = \dots = X_{n-1} = 0, X_n = 1 \mid X_k = 1\} = f_{n-k}$ pour $1 \leq k \leq n - 1$.

a) En posant $u_n = P\{X_n = 1\}$ pour $n \geq 1$ et aussi $u_0 = 1, f_0 = 0$, on a pour $n \geq 1$ l'équation de convolution

$$u_n = \sum_{0 \leq k \leq n} f_k u_{n-k}.$$

En effet, si l'évènement $\{X_n = 1\}$ est réalisé, ou bien, on n'a jamais changé de lampe avant l'instant n et l'évènement $\{X_1 = \dots = X_{n-1} = 0, X_n = 1\}$ est réalisé, ou bien, pour un certain k tel que $1 \leq k \leq n - 1$, l'évènement $\{X_k = 1, X_{k+1} = \dots = X_{n-1} = 0, X_n = 1\}$ est réalisé. La probabilité de ce dernier évènement est égal à

$$P\{X_{k+1} = \dots = X_{n-1} = 0, X_n = 1 \mid X_k = 1\} P\{X_k = 1\} = f_{n-k} u_k.$$

D'où, $u_n = P\{X_n = 1\} = f_n + \sum_{1 \leq k \leq n-1} f_{n-k} u_k = \sum_{0 \leq k \leq n} f_k u_{n-k}.$

b) Pour $|s| < 1$, on pose $F(s) = \sum_{k \geq 0} f_k s^k$ et $U(s) = \sum_{n \geq 0} u_n s^n$.

La précédente équation de convolution entraîne évidemment l'identité : $U(s)(1 - F(s)) = 1$. On suppose maintenant que f_k est nul pour k assez grand, de sorte que $F(s)$ est un polynôme et donc que $Q(s) = 1 - F(s)$ n'a

pas de racine dont le module est strictement plus petit que 1. On suppose, de plus, que 1 est la seule racine de module 1 de $Q(s)$. Comme le polynôme $F(s)$ n'a que des coefficients positifs de somme 1, le nombre 1 est racine *simple* de $Q(s)$. En utilisant les techniques de l'exercice 12, on peut donc déduire que $\lim_n u_n = 1/\mu$, où $\mu = \sum_{k \geq 1} k f_k$ est la durée moyenne de fonctionnement d'une lampe.

15. — Peut-on piper deux dés à six faces de sorte que la somme des points soit équirépartie sur $\{2, \dots, 12\}$?

16. — On lance un dé parfait n fois de suite. Montrer que la probabilité pour que la somme des points amenés soit égale à k est $\alpha_k/6^n$, où α_k est le coefficient de s^k dans le polynôme $(s + s^2 + \dots + s^6)^n$. Calculer cette probabilité.

17. — Calculer le moment factoriel d'ordre r ($r \geq 1$) d'une variable aléatoire de Poisson de paramètre λ ($\lambda > 0$).

18. — Soit (X_1, X_2, \dots, X_r) ($r \geq 1$) un système de r variables aléatoires indépendantes identiquement distribuées selon la loi géométrique de paramètre p ($0 \leq p \leq 1$). On pose $S_r = X_1 + \dots + X_r$ et l'on désigne par $\Pi(r, p)$ sa loi de probabilité (c'est la loi de Pascal, ou binomiale négative).

- Déterminer la fonction génératrice de S_r .
- En déduire la loi de probabilité $\Pi(r, p)$ de S_r .
- Montrer que, pour tout couple d'entiers $r_1, r_2 \geq 1$ et tout p ($0 \leq p \leq 1$), on a :

$$\Pi(r_1, p) * \Pi(r_2, p) = \Pi(r_1 + r_2, p).$$

19. — Soit X une variable aléatoire de loi $k \sum_{n \geq 1} \frac{\theta^n}{n} \varepsilon_n$, où θ désigne un paramètre donné appartenant à $]0, 1[$ et k un paramètre strictement positif.

- Déterminer la valeur du paramètre k en fonction de θ .
- Déterminer la fonction génératrice $G(u) = \mathbb{E}[u^X]$; préciser son domaine de définition.
- Déduire de b) les valeurs de $\mathbb{E}[X]$ et de $\text{Var } X$.

20 (Le cueilleur de champignons)². — On désigne par N le nombre de champignons ramassés par un cueilleur durant une période fixée. On suppose que N est une variable aléatoire à valeurs dans $\{1, 2, \dots\}$ de fonction génératrice G . On suppose, de plus, que la probabilité pour qu'un champignon cueilli soit comestible est p . En faisant les hypothèses d'indépendance qui vont de soi, montrer que la probabilité pour que *tous* les champignons ramassés soient comestibles est $G(p)$.

² Cet exercice qui nous a été communiqué par Anatole Joffe fait partie du « folklore » des probabilistes traitant des fonctions génératrices.

CHAPITRE 10

MESURES DE STIELTJES-LEBESGUE.

INTÉGRATION DES VARIABLES ALÉATOIRES RÉELLES

Nous avons déjà remarqué à propos de la discussion sur la loi géométrique au chap. 7, § 4, que l'étude probabiliste de la première apparition de « pile » au jeu de « pile » ou « face » nous amenait à considérer l'ensemble de toutes les suites *infinies* d'issues possibles. Si l'on identifie l'ensemble des issues d'une partie au doubleton $\{1, 0\}$, on est ainsi conduit à considérer l'ensemble des suites *infinies* $\omega = (\delta_1, \delta_2, \dots)$, où le terme général δ_k vaut 1 ou 0. Or cet ensemble a la puissance du continu (il peut être mis en bijection avec l'ensemble des nombres réels). On verra, dans les exercices 1–7 du présent chapitre, comment probabiliser un tel ensemble, comment le munir d'une tribu, puis comment définir sur cet espace probabilisable une mesure de probabilité qui rende compte de notre intuition sur la pondération d'évènements du type « "pile" sort exactement quatre fois dans les quinze premières parties ». Il est étonnant que pour l'étude théorique d'un jeu de hasard aussi simple, il faille recourir à des résultats profonds sur la théorie de la mesure et en particulier au théorème de prolongement de Carathéodory (voir Théorème 1.3 ci-après).

Par ailleurs, pour parler valablement de lois de probabilité non discrètes sur la droite réelle, on ne peut éviter de parler de *mesure de Lebesgue* sur l'espace mesurable $(\mathbb{R}, \mathcal{B}^1)$. Nous avons donc été amenés à reproduire dans les chapitres 10 et 11 des éléments de la théorie de la mesure qui permettent de faire le lien avec ce qui a été dit jusqu'ici. Nous donnons d'abord quelques notions sur les *mesures* et en plusieurs points l'exposé apparaîtra comme une redite de ce qui avait été développé pour les lois de probabilité. Nous continuons par un exposé sur l'intégration des variables aléatoires par rapport à une mesure et rejetons dans le chapitre suivant les propriétés relevant de l'intégration par rapport à une mesure de probabilité, définissant ainsi la notion d'espérance mathématique dans son cadre général.

1. Mesures. — Soit (Ω, \mathfrak{A}) le couple formé par un ensemble non vide Ω et une tribu \mathfrak{A} sur cet ensemble. Un tel couple avait été appelé *espace probabilisable*. On préfère parler d'*espace mesurable*, lorsqu'on souhaite définir sur \mathfrak{A} non plus une mesure de probabilité, mais une *mesure*. Par ce dernier terme, on entend une fonction μ définie sur \mathfrak{A} à valeurs dans $[0, +\infty]$ satisfaisant aux axiomes suivants :

(1) $\mu(\emptyset) = 0$;

(2) (axiome de σ -additivité) si (A_n) est une suite d'éléments de la tribu \mathfrak{A} , deux à deux disjoints, alors

$$(1.1) \quad \mu\left(\sum_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

Si A appartient à \mathfrak{A} , le nombre (fini ou non) $\mu(A)$ s'appelle la *mesure* de A . Le triplet $(\Omega, \mathfrak{A}, \mu)$ est appelé *espace mesuré*. La mesure μ est dite *finie* (ou *bornée*) si $\mu(\Omega)$ est fini. Par exemple, une mesure de probabilité est une mesure finie telle que $\mu(\Omega) = 1$. Nous verrons que certaines mesures, en particulier la mesure de Lebesgue sur la droite \mathbb{R} , ne sont pas finies, mais qu'il existe une suite d'ensembles, tous de mesure *finie*, dont la réunion est $\Omega = \mathbb{R}$.

Définition. — La mesure μ est dite σ -*finie*, s'il existe une suite (A_n) d'ensembles mesurables (i.e. appartenant à \mathfrak{A}) telle que $\bigcup_{n=1}^{\infty} A_n = \Omega$ et $\mu(A_n)$ est fini pour tout n . La mesure μ est dite *complète*, si tout sous-ensemble d'un ensemble A de \mathfrak{A} , de mesure nulle (i.e., $\mu(A) = 0$), appartient aussi à \mathfrak{A} .

Les propriétés suivantes ont été démontrées au chapitre 3 lorsque les mesures sous-jacentes étaient des mesures de probabilité. Leurs démonstrations sont quasiment identiques pour les mesures en général.

PROPOSITION 1.1. — Soient (A_n) une suite monotone d'ensembles de \mathfrak{A} et μ une mesure sur (Ω, \mathfrak{A}) . Si l'une des deux conditions suivantes est satisfaite

- (i) la suite (A_n) est croissante;
- (ii) la suite (A_n) est décroissante et il existe un entier m tel que $\mu(A_m)$ soit fini;

alors $\mu(\lim_n A_n) = \lim_n \mu(A_n)$.

PROPOSITION 1.2. — Soient (Ω, \mathfrak{A}) un espace mesurable et μ une fonction définie sur \mathfrak{A} à valeurs dans $[0, +\infty]$ ayant les deux propriétés suivantes

- (i) $\mu(\emptyset) = 0$;
- (ii) (additivité finie) pour tout A, B , disjoints, on a $\mu(A + B) = \mu(A) + \mu(B)$.

Si elle satisfait, en plus, l'une des deux conditions suivantes

- (iii) pour toute suite croissante (A_n) d'ensembles mesurables, on a $\lim_n \mu(A_n) = \mu(\lim_n A_n)$;
- (iii') μ est finie et $\lim_n \mu(A_n) = 0$ pour toute suite décroissante (A_n) d'ensembles mesurables tendant vers \emptyset ;

alors μ est une mesure sur (Ω, \mathfrak{A}) .

On définit également la notion de *mesure sur une algèbre* \mathfrak{A} . Il suffit de conserver l'axiome (1) et dans l'axiome (2) de supposer que (1.1) a lieu chaque

fois que la réunion $\sum_{n=1}^{\infty} A_n$ appartient à l'algèbre \mathfrak{A} .

Dans beaucoup de situations, on peut construire effectivement une mesure sur une algèbre et le problème est de savoir comment on peut prolonger cette mesure à tous les éléments appartenant à la tribu engendrée par cette algèbre. Le théorème de prolongement suivant, avec sa méthodologie due à Carathéodory, a la faveur des probabilistes.

THÉORÈME 1.3 (théorème de prolongement). — *Soit \mathfrak{A} une algèbre de parties d'un ensemble non vide Ω . Toute mesure μ sur \mathfrak{A} , σ -finie, se prolonge de façon unique en une mesure $\bar{\mu}$, σ -finie, sur la tribu $\mathfrak{T}(\mathfrak{A})$ engendrée par \mathfrak{A} .*

Nous esquissons seulement ici la démonstration dont l'idée remonte à Carathéodory.¹ On attribue tout d'abord à *tout* sous-ensemble A de Ω une *mesure extérieure* notée $\mu^*(A)$ définie de la façon suivante.

Pour tout $A \subset \Omega$ désignons par $H(A)$ l'ensemble de toutes les suites (A_n) d'éléments de l'algèbre \mathfrak{A} telles que A soit contenu dans la réunion $\bigcup_n A_n$ de ces A_n . L'ensemble $H(A)$ n'est pas vide, puisque $\Omega \in \mathfrak{A}$. On pose alors

$$\mu^*(A) = \inf \left\{ \sum_{n=1}^{\infty} \mu(A_n) ; (A_n) \in H(A) \right\}.$$

On peut vérifier que cette fonction μ^* a les propriétés suivantes :

- a) $\mu^*(A) \geq 0$, $\mu^*(\emptyset) = 0$;
- b) (monotonie) $A \subset B \Rightarrow \mu^*(A) \leq \mu^*(B)$;
- c) (sous-additivité) pour toute suite (A_n) de sous-ensembles de Ω , on a : $\mu^*(\bigcup_n A_n) \leq \sum_n \mu^*(A_n)$;
- d) $A \in \mathfrak{A} \Rightarrow \mu^*(A) = \mu(A)$.

On dit qu'un ensemble $A \subset \Omega$ est μ^* -mesurable si, pour tout $B \subset \Omega$, on a la relation

$$(1.2) \quad \mu^*(B) = \mu^*(AB) + \mu^*(A^c B).$$

Notons \mathfrak{A}^* la classe des sous-ensembles μ^* -mesurables de Ω . On peut montrer que

- e) \mathfrak{A}^* est une tribu; de plus, la restriction μ' de μ^* à la tribu \mathfrak{A}^* est aussi une mesure et elle est σ -additive.
- f) \mathfrak{A}^* contient \mathfrak{A} , donc $\mathfrak{T}(\mathfrak{A})$.

Des propriétés e) et f) on déduit que la restriction $\bar{\mu}$ de μ' sur $\mathfrak{T}(\mathfrak{A})$ satisfait les conditions du théorème de prolongement. L'*unicité* du prolongement peut aussi être démontrée. Dans le tableau suivant, nous schématisons la présente construction. La flèche montante symbolise le premier prolongement, les deux flèches descendantes les deux restrictions successives. Rappelons que l'on a : $\mathfrak{A} \subset \mathfrak{T}(\mathfrak{A}) \subset \mathfrak{A}^* \subset \mathfrak{P}(\Omega)$.

¹ Carathéodory (C.). — *Vorlesungen über reelle Funktionen*, 2. Auflage. — Leipzig, Teubner, 1927.

Fonction d'ensembles		définie sur
mesure extérieure	μ^*	$\mathfrak{P}(\Omega)$
mesure	μ'	\mathfrak{A}^* tribu des ensembles μ^* -mesurables
mesure	$\bar{\mu}$	
mesure	μ	$\mathfrak{T}(\mathfrak{A})$ tribu engendrée par \mathfrak{A}
		\mathfrak{A} algèbre

Donnons l'énoncé du théorème de prolongement lorsque μ est une mesure de probabilité sur \mathfrak{A} . Comme $\Omega \in \mathfrak{A}$ et $\mu(\Omega) = 1$, on a naturellement $\bar{\mu}(\Omega) = \mu(\Omega) = 1$ et ainsi $\bar{\mu}$ est une mesure de probabilité sur $\mathfrak{T}(\mathfrak{A})$.

THÉORÈME 1.4. — *Soit P une mesure de probabilité sur une algèbre \mathfrak{A} de parties d'un ensemble non vide Ω . Alors P se prolonge de façon unique en une mesure de probabilité \bar{P} sur la tribu $\mathfrak{T}(\mathfrak{A})$ engendrée par \mathfrak{A} .*

Soient μ une mesure sur un espace mesurable (Ω, \mathfrak{T}) et N un sous-ensemble de Ω . On dit que N est (μ) -négligeable, si N est un sous-ensemble d'un ensemble de mesure nulle appartenant à \mathfrak{T} . Pour tout $A \in \mathfrak{T}$ et tout ensemble N μ -négligeable, posons

$$\tilde{\mu}(A \cup N) = \mu(A).$$

On peut vérifier que la classe de tous les ensembles de la forme $A \cup N$ est une tribu \mathfrak{T}^μ contenant \mathfrak{T} et que $\tilde{\mu}$ est une mesure sur \mathfrak{T}^μ , qui prolonge la mesure μ . La tribu \mathfrak{T}^μ est appelée la *tribu complétée de \mathfrak{T} pour μ* et $\tilde{\mu}$ est la *complétée de μ* .

On peut montrer que dans le théorème de prolongement 1.3, la tribu complétée \mathfrak{T}^μ de la tribu $\mathfrak{T}(\mathfrak{A})$ pour $\bar{\mu}$ est contenue dans \mathfrak{A}^* , c'est-à-dire $\mathfrak{T}^\mu \subset \mathfrak{A}^*$. Ainsi le théorème de prolongement fournit nécessairement une mesure complète.

2. Mesures de Stieltjes-Lebesgue sur la droite. — Pour appliquer le théorème de prolongement, il faut déjà connaître une mesure sur une algèbre \mathfrak{A} . Le théorème suivant a pour but de construire une classe importante de telles mesures sur une algèbre qui engendre la tribu borélienne de la droite.

On suppose donnée une fonction réelle F , de variable réelle, ayant la propriété suivante : *F est une fonction croissante (au sens large), continue à droite en tout point x de \mathbb{R} .* On pose $\lim_{x \rightarrow -\infty} F(x) = F(-\infty)$ et $\lim_{x \rightarrow +\infty} F(x) = F(+\infty)$, ces deux nombres étant finis ou infinis. Si $F(-\infty) = 0$ et $F(+\infty) = 1$, on retrouve les propriétés d'une fonction de répartition.

Associons à F une fonction d'ensembles, notée $F\{\cdot\}$ définie sur la classe \mathcal{P}_0 des intervalles semi-ouverts de la forme $]a, b]$ ($-\infty < a \leq b < +\infty$), en posant

$$(1.3) \quad F\{]a, b]\} = F(b) - F(a).$$

La propriété suivante est immédiate.

PROPRIÉTÉ 2.1

- (i) $F\{\emptyset\} = 0$, $F\{]a, b]\} \geq 0$;
- (ii) $F\{]a, b]\} \downarrow 0$, lorsque $b \downarrow a$;
- (iii) $F\{\cdot\}$ est additive sur \mathcal{P}_0 : si $a \leq b \leq c$, alors
 $F\{]a, c]\} = F\{]a, b]\} + F\{]b, c]\}$;
- (iv) $F\{\cdot\}$ est monotone.

Montrons d'abord que $F\{\cdot\}$ est σ -additive sur \mathcal{P}_0 (prochaine Proposition) et ensuite qu'elle se prolonge de façon unique en une mesure sur l'algèbre \mathfrak{A} engendrée par \mathcal{P}_0 (Proposition suivante), qui n'est autre que la classe des réunions finies d'intervalles de la forme $] - \infty, a'$, $]a, b]$, $]a'', +\infty[$. Enfin, en utilisant le théorème de prolongement, on en déduit une mesure sur $(\mathbb{R}, \mathcal{B}^1)$.

PROPOSITION 2.2. — La fonction d'ensembles $F\{\cdot\}$ est σ -additive sur \mathcal{P}_0 .

Démonstration. — Soit $(U_i =]a_i, b_i])$ une suite d'intervalles de \mathcal{P}_0 , disjoints deux à deux, tels que $U = \sum_i U_i$ est encore un élément $]a, b]$ de \mathcal{P}_0 . Pour tout $n \geq 1$, on peut, si nécessaire, renuméroter les intervalles de la suite partielle (U_1, \dots, U_n) de sorte que $a \leq a_1 \leq b_1 \leq \dots \leq a_n \leq b_n \leq b$. Alors

$$\begin{aligned} \sum_{k=1}^n F\{]a_k, b_k]\} &\leq \sum_{k=1}^n F\{]a_k, b_k]\} + \sum_{k=1}^{n-1} F\{]b_k, a_{k+1}]\} = F\{]a_1, b_n]\} \\ &\leq F\{]a, b]\} = F\{U\}, \end{aligned}$$

soit $\sum_{k=1}^{\infty} F\{U_k\} \leq F\{U\}$.

Pour démontrer l'inégalité inverse, remarquons d'abord que le théorème est trivial si $a = b$. Si $a < b$, prenons $\varepsilon > 0$ tel que $\varepsilon < b - a$ et posons $V = [a + \varepsilon, b]$. Comme F est continue à droite, il existe, pour tout n , un nombre ε_n tel que $F(b_n + \varepsilon_n) - F(b_n) < \varepsilon/2^n$, soit $F\{]b_n, b_n + \varepsilon_n]\} < \varepsilon/2^n$.

Posons $V_n =]a_n, b_n + \varepsilon_n[$. On a $V_n \supset U_n$, d'où $\bigcup_n V_n \supset \bigcup_n U_n = U =]a, b] \supset [a + \varepsilon, b] = V$. Par le théorème de Borel-Lebesgue, il existe un entier n_0 tel que $\bigcup_{n=1}^{n_0} V_n \supset V$. En renumérotant les intervalles et éventuellement en en laissant tomber, on voit qu'il existe un entier m tel que $\bigcup_{n=1}^m V_n \supset V$, où les intervalles ouverts $V_n =]a_n, b_n + \varepsilon_n[$ ont la propriété suivante :

$$\begin{aligned} a_1 < a + \varepsilon, \quad a_2 < b_1 + \varepsilon_1, \quad a_3 < b_2 + \varepsilon_2, \quad \dots, \quad a_{k+1} < b_k + \varepsilon_k, \\ \dots, \quad a_m < b_{m-1} + \varepsilon_{m-1}, \quad b < b_m + \varepsilon_m. \end{aligned}$$

Il en résulte

$$\begin{aligned} F\{]a + \varepsilon, b]\} &\leq F\{]a_1, b_m + \varepsilon_m]\} \\ &\leq F\{]a_1, b_1 + \varepsilon_1]\} + F\{]a_2, b_2 + \varepsilon_2]\} + \dots + F\{]a_m, b_m + \varepsilon_m]\} \\ &\leq F\{]a_1, b_1]\} + F\{]a_2, b_2]\} + \dots + F\{]a_m, b_m]\} \\ &\quad + F\{]b_1, b_1 + \varepsilon_1]\} + F\{]b_2, b_2 + \varepsilon_2]\} + \dots + F\{]b_m, b_m + \varepsilon_m]\} \\ &\leq \sum_{k=1}^m F\{]a_k, b_k]\} + \sum_{k=1}^m \frac{\varepsilon}{2^k} \leq \sum_{k=1}^{\infty} F\{]a_k, b_k]\} + \varepsilon, \end{aligned}$$

soit encore

$$F(b) - F(a + \varepsilon) \leq \sum_{k=1}^{\infty} F\{]a_k, b_k]\} + \varepsilon,$$

et puisque F est continue à droite,

$$F(b) - F(a) = F\{]a, b]\} \leq \sum_{k=1}^{\infty} F\{]a_k, b_k]\},$$

en faisant tendre ε vers 0. \square

PROPOSITION 2.3. — *Il existe une mesure unique $\bar{F}\{\cdot\}$ sur \mathfrak{A} telle que si $U \in \mathcal{P}_0$, alors $\bar{F}\{U\} = F\{U\}$.*

Démonstration. — Remarquons d'abord que tout intervalle de la forme $] - \infty, a']$ ou $]a'', \infty[$ peut s'écrire comme réunion dénombrable d'intervalles appartenant à \mathcal{P}_0 et deux à deux disjoints. Pour prolonger $F\{\cdot\}$ à l'algèbre \mathfrak{A} , il suffit donc de montrer que si (U_i) est une suite dénombrable d'intervalles appartenant à \mathcal{P}_0 , de réunion A , alors $F\{A\}$ est complètement déterminée par

$$F\{A\} = \sum_i F\{U_i\}.$$

En effet, si l'on a aussi $A = \sum_j V_j$, où les V_j appartiennent à \mathcal{P}_0 et sont deux à deux disjoints, on peut écrire : $U_i = AU_i = \sum_j V_j U_i$ et $V_j = AV_j = \sum_i U_i V_j$. Or chaque $U_i V_j$ appartient à \mathcal{P}_0 , puisque \mathcal{P}_0 est stable par intersection finie. Comme $F\{\cdot\}$ est σ -additive sur \mathcal{P}_0 , il vient $\sum_i F\{U_i\} = \sum_i \sum_j F\{U_i V_j\} = \sum_j \sum_i F\{U_i V_j\} = \sum_j F\{V_j\}$. \square

En appliquant le théorème de prolongement, on obtient l'énoncé suivant.

THÉORÈME 2.4. — *Soit F une fonction réelle définie sur \mathbb{R} , croissante et continue à droite. Il existe une mesure unique $F\{\cdot\}$, définie sur la tribu borélienne \mathcal{B}^1 de \mathbb{R} telle que pour tout intervalle semi-ouvert borné $]a, b]$ on ait*

$$F\{]a, b]\} = F(b) - F(a).$$

La complétée $\tilde{F}\{\cdot\}$ de la mesure $F\{\cdot\}$, définie sur la tribu complétée \mathcal{B}^F pour F s'appelle la mesure de Lebesgue-Stieltjes induite par F .

Lorsque $F(x) = x$, la complétée que nous noterons λ^1 ou λ s'appelle la *mesure de Lebesgue sur la droite*. Elle fait correspondre à tout intervalle borné de la droite sa longueur. La tribu complétée s'appelle la *tribu des ensembles mesurables*. Notons que λ^1 n'est pas une mesure finie, puisque $\lambda^1\{\mathbb{R}\} = \lim_n \lambda^1\{]-n, n]\} = \lim_n 2n = +\infty$. En revanche, elle est σ -finie. Ce n'est évidemment pas une mesure de probabilité sur la droite.

Revenons au cas général d'une mesure de Stieltjes-Lebesgue $F\{\cdot\}$ induite par une fonction F . Les formules suivantes sont immédiates :

- (i) $F\{\{a\}\} = F(a) - F(a - 0)$;
- (ii) $F\{]a, b]\} = F(b) - F(a)$, $F\{]a, b[\} = F(b - 0) - F(a)$;
- (iii) $F\{[a, b[\} = F(b) - F(a - 0)$, $F\{[a, b]\} = F(b - 0) - F(a - 0)$.

3. Mesure de probabilité induite par une fonction de répartition. — On rappelle qu'une fonction de répartition sur la droite est une fonction réelle F , de variable réelle, croissante, continue à droite et telle que $F(-\infty) = 0$ et $F(+\infty) = 1$ (voir chap. 5, § 5). Le théorème précédent fournit immédiatement l'énoncé suivant.

THÉORÈME 3.1. — *Soit F une fonction de répartition sur la droite. Il existe une et une seule mesure de probabilité, notée $F\{\cdot\}$, sur la tribu borélienne \mathcal{B}^1 telle que $F\{]a, b]\} = F(b) - F(a)$ pour tout intervalle semi-ouvert borné $]a, b]$.*

Démonstration. — La seule propriété à vérifier est que la mesure $F\{\cdot\}$ est bien une mesure de probabilité. En effet, $] -n, +n] \uparrow \mathbb{R}$, d'où
 $\lim_n F\{] -n, +n]\} = \lim_n (F(n) - F(-n)) = 1. \quad \square$

Comme à toute mesure de probabilité P sur $(\mathbb{R}, \mathcal{B}^1)$, on peut associer une variable aléatoire réelle X admettant P comme loi de probabilité, on obtient le corollaire suivant.

COROLLAIRE. — *Toute fonction de répartition sur la droite est la fonction de répartition d'une variable aléatoire réelle.*

4. Mesures de Stieltjes-Lebesgue sur \mathbb{R}^n . — On peut faire la même construction que ci-dessus dans le cas de plusieurs dimensions. On part d'une fonction numérique $F(x_1, x_2, \dots, x_n)$ de n variables réelles x_1, x_2, \dots, x_n ayant les propriétés suivantes :

- (1) $F(x_1, x_2, \dots, x_n)$ est une fonction croissante, continue à droite, en chacune des variables;
- (2) Pour tout $h_k \geq 0$ et tout x_k réel ($k = 1, 2, \dots, n$), on a :

$$\Delta_{h_1}^{(1)} \Delta_{h_2}^{(2)} \dots \Delta_{h_n}^{(n)} F(x_1, x_2, \dots, x_n) \geq 0,$$

où naturellement $\Delta_{h_k}^{(k)} F$ désigne l'accroissement de F lorsqu'on augmente la $k^{\text{ième}}$ variable d'une valeur égale à h_k .

Soit I un pavé semi-ouvert de \mathbb{R}^n . Le pavé I consiste en les points (x_1, x_2, \dots, x_n) de \mathbb{R}^n satisfaisant les inégalités $a_k < x_k \leq b_k$ ($1 \leq k \leq n$). Posant $h_k = b_k - a_k$ ($1 \leq k \leq n$), on lui associe la mesure

$$(4.1) \quad F\{I\} = \Delta_{h_1}^{(1)} \Delta_{h_2}^{(2)} \dots \Delta_{h_n}^{(n)} F(a_1, a_2, \dots, a_n) \geq 0.$$

Suivant la même méthode que dans le cas d'une dimension, on peut montrer que $F\{\cdot\}$ peut être prolongée en une mesure sur les boréliens de \mathbb{R}^n , telle que (4.1) soit satisfait pour tout pavé semi-ouvert. Dans le cas spécial où $F(x_1, x_2, \dots, x_n) = x_1 x_2 \dots x_n$, on obtient par prolongement la *mesure de Lebesgue* λ^n sur $(\mathbb{R}^n, \mathcal{B}^n)$.

5. Variables aléatoires réelles. — Il est utile d'étendre légèrement la définition d'une variable aléatoire réelle et de la prendre à valeurs dans la *droite achevée*.

5.1. *La droite achevée.* — On appelle nombre *fini* tout nombre réel $x \in \mathbb{R}$ et nombre *infini* l'un des symboles $-\infty, +\infty$, vérifiant les propriétés suivantes :

- a) $-\infty < +\infty$;
- b) pour tout $x \in \mathbb{R}$,
$$\begin{cases} -\infty < x < +\infty; \\ \pm\infty = (\pm\infty) + x = x + (\pm\infty); \\ \frac{x}{\pm\infty} = 0; \end{cases}$$
- c) $x(\pm\infty) = (\pm\infty)x = \begin{cases} \pm\infty, & \text{si } 0 < x \leq +\infty; \\ 0, & \text{si } x = 0; \\ \mp\infty, & \text{si } -\infty \leq x < 0. \end{cases}$

On ne donnera pas de sens à la différence $+\infty - \infty$ et il conviendra donc de l'éviter. La *droite achevée* $\overline{\mathbb{R}}$ est l'ensemble de tous les nombres finis ou infinis. On écrit : $\mathbb{R} =]-\infty, +\infty[$, $\overline{\mathbb{R}} = [-\infty, +\infty]$.

5.2. *La tribu borélienne achevée.* — C'est la tribu $\overline{\mathcal{B}}$ engendrée par $\mathcal{B} \cup \{-\infty, +\infty\}$, où \mathcal{B} désigne la tribu borélienne sur \mathbb{R} . Un ensemble $A \subset \overline{\mathbb{R}}$ appartient à $\overline{\mathcal{B}}$ si et seulement si $A \cap \mathbb{R} \in \mathcal{B}$.

5.3. *Variable aléatoire réelle.* — Une *variable aléatoire réelle* définie sur un espace mesurable (Ω, \mathfrak{A}) est une application *mesurable* X de (Ω, \mathfrak{A}) dans $(\overline{\mathbb{R}}, \overline{\mathcal{B}})$; pour tout $B \in \overline{\mathcal{B}}$, elle vérifie donc la propriété $X^{-1}(B) \in \mathfrak{A}$.

Si $X(\Omega) \subset \mathbb{R}$, on dit que X est *finie*.

Si $X(\Omega) \subset [0, +\infty]$, on dit que X est *positive*.

Notation. — Soit X une variable aléatoire réelle. On note $X^+ = X \vee 0 = \sup(X, 0)$, $X^- = -X \wedge 0 = -\inf(X, 0)$, et l'on a : $X = X^+ - X^-$, $|X| = X^+ + X^-$.

PROPOSITION 5.3.1. — *Soient X, Y , deux variables aléatoires réelles définies sur un même espace mesurable (Ω, \mathfrak{A}) . Alors les applications $|X|$, $|Y|$, $X \pm Y$ (au cas où cette application est définie sur tout Ω), XY , $X \vee Y = \sup(X, Y)$, $X \wedge Y = \inf(X, Y)$ sont des variables aléatoires réelles.*

De même, si (X_n) ($n \geq 1$) est une suite de variables aléatoires réelles sur (Ω, \mathfrak{A}) , alors les applications $\sup_n X_n$, $\inf_n X_n$, $\limsup_n X_n$, $\liminf_n X_n$ sont des variables aléatoires.

C'est un cas particulier de la Proposition 2.2 du chap. 5.

PROPOSITION 5.3.2. — *Soient X une application de Ω dans \mathbb{R} et \mathfrak{A} une tribu sur Ω . Alors les deux propriétés suivantes sont équivalentes :*

- a) X est une variable aléatoire réelle ;
 b) X^+ et X^- sont des variables aléatoires réelles.

Démonstration. — L'implication a) \Rightarrow b) résulte de la Proposition 5.3.1, puisque 0 est une variable aléatoire réelle et b) \Rightarrow a) résulte de la même proposition, puisque $X = X^+ - X^-$. \square

5.4. *Variables aléatoires simples.* — Soit (A_1, \dots, A_n) une *partition* finie de Ω en éléments de \mathfrak{A} . Il résulte de la Proposition 5.3.1 que si (x_1, \dots, x_n) est un n -uplet de nombres réels, alors l'application $X = \sum_{k=1}^n x_k I_{A_k}$ de Ω dans \mathbb{R} est une variable aléatoire (à valeurs finies); d'où la définition suivante.

Définition. — Soient (A_1, \dots, A_n) une partition de Ω en éléments de \mathfrak{A} et (x_1, \dots, x_n) un n -uplet de nombres réels. L'application

$$X = \sum_{k=1}^n x_k I_{A_k}$$

de (Ω, \mathfrak{A}) dans $(\mathbb{R}, \mathcal{B})$ est appelée *variable aléatoire simple* ou *étagée*.

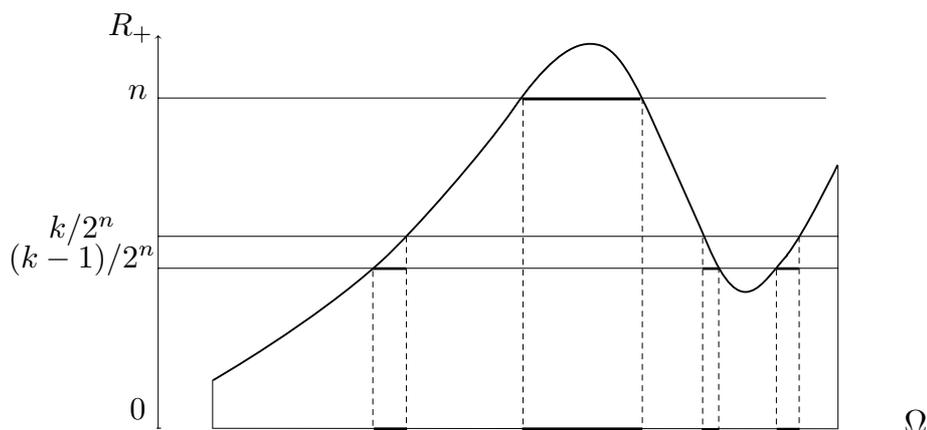
L'importance des variables simples apparaît clairement dans les lemmes suivants.

LEMME 5.4.1 (Lemme d'approximation). — *Soit X une application de (Ω, \mathfrak{A}) dans $\overline{\mathbb{R}}^+$. Une condition nécessaire et suffisante pour que X soit une variable aléatoire (positive) est qu'elle soit limite (au sens de la convergence simple) d'une suite croissante (X_n) ($n \geq 1$) de variables aléatoires simples, positives sur (Ω, \mathfrak{A}) .*

Démonstration. — Si $X = \lim_n X_n$, où X_n ($n \geq 1$) est une suite *croissante* de variables aléatoires *simples, positives*, alors, d'après la Proposition 5.3.1, la fonction X est une variable aléatoire (positive).

Réciproquement, soit X une variable aléatoire réelle *positive*; pour tout $n \geq 1$ et tout $\omega \in \Omega$ écrivons :

$$X_n(\omega) = \begin{cases} \frac{k-1}{2^n}, & \text{si } \frac{k-1}{2^n} \leq X(\omega) < \frac{k}{2^n} \quad (k = 1, \dots, n2^n); \\ n, & \text{si } X(\omega) \geq n. \end{cases}$$



On peut alors représenter X_n sous la forme :

$$(5.4.1) \quad X_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} I_{\left\{ \frac{k-1}{2^n} \leq X < \frac{k}{2^n} \right\}} + n I_{\{X \geq n\}}.$$

Pour tout $n \geq 1$ les ensembles $\{\frac{k-1}{2^n} \leq X < \frac{k}{2^n}\}$ ($k = 1, \dots, n2^n$) et $\{X \geq n\}$ sont disjoints et appartiennent à \mathfrak{A} . Il en résulte que, pour tout $n \geq 1$, la fonction X_n est une variable aléatoire *simple, positive* et il est clair que la suite (X_n) ($n \geq 1$) est *croissante*. Enfin, $X = \lim_n X_n = \sup_n X_n$, car, pour tout $\omega \in \Omega$, on a, ou bien $X(\omega) = +\infty$ et alors $X_n(\omega) = n$ pour tout $n \geq 1$, ou bien $X(\omega) < \infty$ et alors $0 \leq X(\omega) - X_n(\omega) \leq 1/2^n$ pour $n > X(\omega)$. \square

6. Intégrale d'une variable aléatoire réelle par rapport à une mesure. — Supposons que les variables aléatoires sont définies sur un *espace mesuré* $(\Omega, \mathfrak{A}, \mu)$. Au chapitre suivant, nous verrons les conséquences à apporter lorsque μ sera remplacée par une mesure de probabilité P .

Définition. — Soit $X = \sum_{k=1}^n x_k I_{A_k}$ une variable aléatoire *simple, positive* (les x_k sont des réels positifs et (A_1, \dots, A_n) est une partition de Ω en éléments de \mathfrak{A}). On appelle *intégrale de X par rapport à la mesure μ* le nombre positif, noté $\int X d\mu$, ou bien $\int_{\Omega} X d\mu$ et défini par :

$$(6.1) \quad \int X d\mu = \int_{\Omega} X d\mu = \sum_{k=1}^n x_k \mu(A_k).$$

Remarque. — Cette expression ne dépend pas de la combinaison linéaire particulière d'indicatrices qui représente X , comme on le vérifie facilement ; c'est un nombre qui ne dépend que de X .

La proposition suivante est facile à vérifier.

PROPOSITION 6.1 (Monotonie). — *Soient X, Y deux variables aléatoires simples, positives ; alors*

$$X \leq Y \implies \int X d\mu \leq \int Y d\mu.$$

La proposition suivante est plus difficile à vérifier.

PROPOSITION 6.2. — *Soient $(X_n), (Y_n)$ ($n \geq 1$) deux suites croissantes de variables aléatoires simples, positives ; alors*

$$\sup_n X_n = \sup_n Y_n \implies \sup_n \int X_n d\mu = \sup_n \int Y_n d\mu.$$

Définition. — Soit X une variable aléatoire réelle *positive*. D'après le Lemme 5.4.1, il existe une suite *croissante* (X_n) ($n \geq 1$) de variables aléatoires *simples, positives*, telle que, au sens de la convergence simple, on ait $X_n \uparrow X$ ($n \rightarrow \infty$). On appelle *intégrale de X par rapport à la mesure μ* le nombre fini ou infini

$$\int X d\mu = \lim_{n \rightarrow \infty} \int X_n d\mu.$$

Remarque 1. — La limite au dernier membre existe bien (comme élément de $[0, +\infty]$) en vertu de la Proposition 6.1. D'autre part, cette expression ne dépend pas de la suite particulière (X_n) ($n \geq 1$) dont la limite représente X ; c'est un nombre qui ne dépend que de X .

En effet, soient $(X_n), (Y_n)$ ($n \geq 1$) deux suites croissantes de variables aléatoires simples, positives, telles que $X = \lim_n X_n = \sup_n X_n$, $X = \lim_n Y_n = \sup_n Y_n$. Il résulte alors de la Proposition 6.2 que $\int X d\mu = \lim_n \int X_n d\mu = \sup_n \int X_n d\mu = \sup_n \int Y_n d\mu = \lim_n \int Y_n d\mu$. On pourrait également définir $\int X d\mu$ comme $\int X d\mu = \sup \int S d\mu$, où le « sup » est étendu à toutes les variables aléatoires simples positives S vérifiant $0 \leq S \leq X$.

Remarque 2. — On dit quelquefois que X est μ -intégrable, si $\int X d\mu < +\infty$; mais il est convenu de dire que $\int X d\mu$ existe même si $\int X d\mu = +\infty$.

Définition. — Soit X une variable aléatoire (à valeurs dans $\overline{\mathbb{R}}$); d'après la Proposition 5.3.2 les fonctions X^+ et X^- sont des variables aléatoires à valeurs dans $[0, +\infty]$, qui admettent donc des intégrales $\int X^+ d\mu$, $\int X^- d\mu$, éléments de $[0, +\infty]$. On dit que X est μ -intégrable, si $\int X^+ d\mu$ et $\int X^- d\mu$ sont toutes deux finies et le nombre réel

$$\int X d\mu = \int X^+ d\mu - \int X^- d\mu$$

est appelé l'intégrale de X par rapport à la mesure μ .

Remarque 3. — On a $|X| = X^+ + X^-$, d'où, en utilisant la propriété de linéarité de l'intégrale (établie dans le paragraphe 8 de ce chapitre) l'identité $\int |X| d\mu = \int X^+ d\mu + \int X^- d\mu$ (égalité dans $[0, +\infty]$). On en déduit que $\int X d\mu$ est μ -intégrable, si et seulement si $\int |X| d\mu < \infty$.

Remarque 4. — La différence $\int X^+ d\mu - \int X^- d\mu$ est encore définie si l'un au moins des deux termes $\int X^+ d\mu$, $\int X^- d\mu$ est fini. On pourra dans ce cas encore parler de l'intégrale de X par rapport à μ , soit $\int X d\mu = \int X^+ d\mu - \int X^- d\mu$.

7. Exemples

EXEMPLE 7.1. — Soit $(\Omega, \mathfrak{A}, P)$ un espace probabilisé avec $P = \varepsilon_{\omega_0}$, où ε_{ω_0} est la mesure (de Dirac) définie par la mesure unité au point $\omega_0 \in \Omega$. Soit, d'autre part, X une variable aléatoire sur cet espace. Si $X(\omega_0) \geq 0$, ou si $|X(\omega_0)| < +\infty$, alors $\int_{\Omega} X d\varepsilon_{\omega_0} = X(\omega_0)$.

Démonstration. — Supposons que $X = \sum_{k=1}^n x_k I_{A_k}$ soit une variable aléatoire simple, positive. Puisque (A_1, \dots, A_n) forme une partition de Ω , il existe un indice k_0 et un seul tel que $\omega_0 \in A_{k_0}$, d'où

$$\int X dP = \sum_{k=1}^n x_k P(A_k) = x_{k_0} P(A_{k_0}) = x_{k_0} = X(\omega_0).$$

Supposons que X soit une variable aléatoire *positive* et soit (X_n) ($n \geq 1$) une suite *croissante* de variables aléatoires *simples*, positives telle que $X = \lim_n X_n$. On a alors : $\int X dP = \lim_n \int X_n dP = \lim_n X_n(\omega_0) = X(\omega_0)$.

Supposons enfin que X soit une variable aléatoire réelle et que $|X(\omega_0)| < +\infty$. On a alors $\int X^+ dP = X^+(\omega_0) < +\infty$, $\int X^- dP = X^-(\omega_0) < +\infty$, d'où $\int X dP = \int X^+ dP - \int X^- dP = X^+(\omega_0) - X^-(\omega_0) = X(\omega_0)$. \square

EXEMPLE 7.2. — Soit $(\Omega, \mathfrak{A}, P)$ un espace probabilisé, où $P = \sum_i \alpha_i \varepsilon_{\omega_i}$ ($\alpha_i > 0$, $\sum_i \alpha_i = 1$) est une mesure de probabilité discrète sur (Ω, \mathfrak{A}) . Soit, d'autre part, X une variable aléatoire réelle sur cet espace. Si $X \geq 0$, ou si $\sum_i \alpha_i |X(\omega_i)| < \infty$, alors $\int X dP = \sum_i \alpha_i X(\omega_i)$.

Démonstration. — Supposons que $X = \sum_{k=1}^n x_k I_{A_k}$ soit une variable aléatoire *simple*, *positive*. Alors

$$\begin{aligned} \int X dP &= \sum_{k=1}^n x_k P(A_k) = \sum_{k=1}^n x_k \left(\sum_i \alpha_i \varepsilon_{\omega_i}(A_k) \right) \\ &= \sum_i \alpha_i \left(\sum_{k=1}^n x_k \varepsilon_{\omega_i}(A_k) \right) = \sum_i \alpha_i X(\omega_i). \end{aligned}$$

Supposons que X soit une variable aléatoire *positive* et soit (X_n) ($n \geq 1$) une suite *croissante* de variables aléatoires *simples* positives telle que $X = \lim_n X_n$. On a alors $\int X dP = \lim_n \int X_n dP = \lim_n (\sum_i \alpha_i X_n(\omega_i)) = \sum_i \alpha_i X(\omega_i)$.

Supposons enfin que $\sum_i \alpha_i |X(\omega_i)| < \infty$. On a alors $\int X^+ dP = \sum_i \alpha_i X^+(\omega_i) < \infty$, $\int X^- dP = \sum_i \alpha_i X^-(\omega_i) < \infty$, d'où

$$\int X dP = \int X^+ dP - \int X^- dP = \sum_i (X^+(\omega_i) - X^-(\omega_i)) = \sum_i \alpha_i X(\omega_i). \quad \square$$

8. Propriétés de l'intégrale

Définition. — Soient $(\Omega, \mathfrak{A}, \mu)$ un espace mesuré et \mathcal{P} une propriété dont la valeur de vérité dépend de $\omega \in \Omega$. On dit que \mathcal{P} est vraie μ -presque partout, s'il existe un $A \in \mathfrak{A}$ tel que $\mu(A) = 0$ et tel que la propriété \mathcal{P} est vraie pour tout $\omega \in A^c$.

Remarque. — Dans cette définition, on n'impose pas que l'ensemble A' des $\omega \in \Omega$ qui n'ont pas la propriété \mathcal{P} soit de mesure nulle, car A' n'appartient pas nécessairement à \mathfrak{A} . On a, en fait, $A' \subset A$, $A \in \mathfrak{A}$, $\mu(A) = 0$ et \mathcal{P} est vraie dans A^c (mais \mathcal{P} est aussi vraie dans $A \setminus A'$). (On pourrait exprimer ce fait en disant que l'ensemble A' des ω où \mathcal{P} n'est pas vérifiée est «négligeable».)

Dans ce paragraphe, les variables aléatoires X, Y qui interviennent sont définies sur un espace mesuré $(\Omega, \mathfrak{A}, \mu)$. Suivant notre convention, on dira que $\int X d\mu$ existe, si $X \geq 0$, ou si X est μ -intégrable. Enfin, si $\int X d\mu$ existe et si $A \in \mathfrak{A}$, on posera $\int_A X d\mu = \int_{\Omega} X I_A d\mu$.

PROPOSITION 8.1. — Si $\int X d\mu$ et $\int Y d\mu$ existent, alors les propriétés suivantes sont satisfaites :

A. *Linéarité*

$$(A1) \quad \int (X + Y) d\mu = \int X d\mu + \int Y d\mu ;$$

$$(A2) \quad \text{pour tout } \lambda \text{ réel, on a : } \int \lambda X d\mu = \lambda \int X d\mu ;$$

$$(A3) \quad \text{pour tous } A, B \in \mathfrak{A}, \text{ disjoints, on a : } \int_{A+B} X d\mu = \int_A X d\mu + \int_B X d\mu.$$

B. *Monotonie*

$$(B1) \quad X \geq 0 \implies \int X d\mu \geq 0 ;$$

$$(B2) \quad X \geq Y \implies \int X d\mu \geq \int Y d\mu ;$$

$$(B3) \quad X = Y \text{ } \mu\text{-presque partout} \implies \int X d\mu = \int Y d\mu.$$

C. *Intégrabilité*

$$(C1) \quad X \text{ } \mu\text{-intégrable} \iff |X| \text{ } \mu\text{-intégrable} ;$$

$$(C2) \quad X \text{ } \mu\text{-intégrable} \implies X \text{ presque partout fini.}$$

$$(C3) \quad |X| \leq Y \text{ et } Y \text{ } \mu\text{-intégrable} \implies X \text{ } \mu\text{-intégrable} ;$$

$$(C4) \quad X \text{ et } Y \text{ } \mu\text{-intégrables} \implies X + Y \text{ } \mu\text{-intégrable.}$$

D. *Majoration de l'intégrale*

(D1) Soient a et b deux nombres réels tels que pour tout ω d'un ensemble $A \in \mathfrak{A}$ on ait $a \leq X(\omega) \leq b$ et $\mu(A) < \infty$, alors $a\mu(A) \leq \int_A X d\mu \leq b\mu(A)$.

(D2) Si X est μ -intégrable, alors $|\int X d\mu| \leq \int |X| d\mu$.

9. Théorèmes de convergence. — Les trois théorèmes de convergence que nous énonçons maintenant sans démonstration sont fondamentaux. (cf. Revuz [12], pp. 61, 70 et 93.) Toutes les variables aléatoires réelles qui interviennent sont définies sur un même espace mesuré $(\Omega, \mathfrak{A}, \mu)$.

THÉORÈME 9.1 (Théorème de convergence monotone de Beppo Levi). Soit (X_n) ($n \geq 1$) une suite croissante de variables aléatoires positives. Cette suite converge au sens de la convergence simple vers une limite positive mesurable. On a alors

$$\int \lim_{n \rightarrow \infty} X_n d\mu = \lim_{n \rightarrow \infty} \int X_n d\mu,$$

cette égalité ayant lieu dans $[0, +\infty]$.

On peut donner de ce théorème la version suivante :

Soit (Y_n) ($n \geq 1$) une suite de variables aléatoires réelles positives. Alors

$$\int \sum_{n \geq 1} Y_n d\mu = \sum_{n \geq 1} \int Y_n d\mu,$$

cette égalité ayant lieu dans $[0, +\infty]$.

THÉORÈME 9.2 (Lemme de Fatou). — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires réelles positives. Alors

$$\int \liminf_{n \rightarrow \infty} X_n d\mu \leq \liminf_{n \rightarrow \infty} \int X_n d\mu,$$

cette inégalité ayant lieu dans $[0, +\infty]$.

COROLLAIRE. — Supposons, en outre, que

- a) $X_n \rightarrow X$ (presque partout);
 - b) il existe $M \in [0, +\infty[$ tel que pour tout $n \geq 1$ on ait $\int X_n d\mu \leq M$.
- Alors $\int X d\mu \leq M$.

THÉORÈME 9.3 (Théorème de convergence dominée de Lebesgue). — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires, μ -intégrables. On suppose que :

- a) $X_n \rightarrow X$ (presque partout);
- b) il existe une variable aléatoire Y , à valeurs positives, telle que $\int Y d\mu < \infty$ et que pour tout $n \geq 1$ on ait $|X_n| \leq Y$.

Alors X est μ -intégrable et l'on a :

$$\int \lim_{n \rightarrow \infty} X_n d\mu = \lim_{n \rightarrow \infty} \int X_n d\mu.$$

COMPLÉMENTS ET EXERCICES

Soient $S = \{1, 2, \dots, r\}$ ($r \geq 2$) un ensemble fini et Ω l'ensemble $S^{\mathbb{N}^*}$ de toutes les suites infinies $\omega = (x_1, x_2, \dots)$, où les x_i ($i = 1, 2, \dots$) appartiennent à S . Le but de la suite d'exercices 1–9 est de montrer comment on peut munir Ω d'une tribu d'évènements \mathfrak{T} , distincte de $\mathfrak{P}(\Omega)$, mais contenant tous les évènements dits *observables* ou qui ne font intervenir qu'un nombre *fini* d'instant (on précisera cette notion plus loin.) Ensuite, partant d'une classe de mesures de probabilité (p_n) ($n \geq 1$), où p_n est une mesure de probabilité sur S^n et supposant vérifiées certaines *relations de compatibilité* entre les p_n , le but sera de *probabiliser* l'espace (Ω, \mathfrak{T}) , c'est-à-dire de munir (Ω, \mathfrak{T}) d'une mesure de probabilité P . Lorsque $r = 2$, la construction proposée permet de «probabiliser les suites infinies du jeu de pile ou face». (Voir exercice 7.)

1 (L'algèbre des évènements observables). — Pour $n \geq 1$ désignons par $\pi_n : \Omega \rightarrow S^n$ la *projection* qui envoie chaque suite infinie $\omega = (x_1, x_2, \dots)$ de Ω sur la suite finie $\pi_n(\omega) = (x_1, x_2, \dots, x_n)$. Notons aussi $X_n : \Omega \rightarrow S$ la $n^{\text{ième}}$ *application coordonnée* définie par $X_n(\omega) = x_n$. Appelons enfin *n-cylindre* toute partie C de Ω de la forme $C = \{\pi_n \in A\} = \pi_n^{-1}(A)$, où A est une partie de S^n ($n \geq 1$) et notons \mathfrak{A}_n l'ensemble des *n-cylindres*.

- a) Pour $n \geq 1$, la classe \mathfrak{A}_n des *n-cylindres* est une tribu.
- b) La suite (\mathfrak{A}_n) ($n \geq 1$) est monotone croissante :
 $\mathfrak{A}_1 \subset \mathfrak{A}_2 \subset \dots \subset \mathfrak{A}_n \subset \mathfrak{A}_{n+1} \subset \dots$
- c) La classe $\mathfrak{A} = \lim_n \mathfrak{A}_n = \bigcup_n \mathfrak{A}_n$ est une algèbre et non pas une tribu.

2 (La tribu des événements observables). — Soit $\mathfrak{T} = \mathfrak{T}(\mathfrak{A})$ la tribu engendrée par \mathfrak{A} , dite *des événements observables*. C'est la plus petite tribu rendant tous les π_n (resp. X_n) ($n \geq 1$) mesurables.

3. — On considère, pour chaque $n \geq 1$, une fonction réelle p_n définie sur S^n , ayant les propriétés suivantes :

- (i) $p_n \geq 0$;
- (ii) $\sum_{x \in S} p_1(x) = 1$;
- (iii) $\sum_{x \in S} p_{n+1}(x_1, \dots, x_n, x) = p_n(x_1, \dots, x_n)$ pour toute suite (x_1, \dots, x_n) appartenant à S^n .

Si $C = \{\pi_n \in A\}$ est un cylindre, on pose $P(C) = P\{\pi_n \in A\} = \sum p_n(x_1, \dots, x_n)$, où la sommation se fait sur toutes les suites (x_1, \dots, x_n) appartenant à A . Montrer que cette formule ne dépend que de C (non pas de n , ni de A).

4. — Si (C_m) ($m \geq 1$) est une suite décroissante de cylindres non vides, leur intersection est non vide.

5. — L'application $P : \mathfrak{A} \rightarrow \mathbb{R}$ qui au cylindre $C = \{\pi_n \in A\}$ ($A \subset S^n$) fait correspondre le nombre $P(C) = \sum_A p_n(x_1, x_2, \dots, x_n)$, est une mesure de probabilité sur l'algèbre \mathfrak{A} .

6. — Soit S un ensemble fini. On suppose donné pour tout n une fonction $p_n : S^n \rightarrow \mathbb{R}$ satisfaisant les trois propriétés (i), (ii) et (iii) de l'exercice 3. Alors il existe sur (Ω, \mathfrak{T}) une et une seule mesure de probabilité P telle que pour tout $n \geq 1$ et tout $(x_1, \dots, x_n) \in S^n$, on ait :

$$P\{X_1 = x_1, \dots, X_n = x_n\} = p_n(x_1, \dots, x_n).$$

7 (Produit d'espaces probabilisés). — Supposons que l'ensemble fini S soit muni d'une mesure de probabilité p . Alors sur l'espace (Ω, \mathfrak{T}) il existe une et une seule mesure de probabilité P telle que

- (i) les projections X_n soient mutuellement indépendantes;
- (ii) pour toute partie U de S et tout n on ait : $P\{X_n \in U\} = p(U)$.

8 (Probabilités conditionnelles compatibles). — Soit p_1 une mesure de probabilité sur S et pour tout $n \geq 2$ soit $q_n : S^n \rightarrow \mathbb{R}_+$ une fonction satisfaisant pour tout $(x_1, \dots, x_{n-1}) \in S^{n-1}$ la relation $\sum_{x \in S} q_n(x_1, \dots, x_{n-1}, x) = 1$. Alors il existe une et une seule mesure de probabilité P sur (Ω, \mathfrak{T}) telle que $P\{X_1 = x_1\} = p_1(x_1)$ et telle que pour tout $n \geq 1$ et toute suite $(x_1, \dots, x_n) \in S^n$ on ait :

$$P\{X_n = x_n \mid X_{n-1} = x_{n-1}, \dots, X_1 = x_1\} = q_n(x_1, \dots, x_n).$$

9 (Chaînes de Markov homogènes). — Soient $(\Omega, \mathfrak{A}, P)$ un espace probabilisé et (X_n) une suite de variables aléatoires définies sur cet espace à valeurs

dans un même ensemble fini S . On dit que la suite (X_n) forme une *chaîne de Markov homogène* si les deux conditions suivantes sont remplies :

- (i) pour tout $n \geq 2$ et tout $(x_1, \dots, x_n) \in S^n$ on a :

$$P\{X_n = x_n \mid X_{n-1} = x_{n-1}, \dots, X_1 = x_1\} = P\{X_n = x_n \mid X_{n-1} = x_{n-1}\};$$
- (ii) pour tout $(x, y) \in S^2$ la probabilité $P\{X_n = y \mid X_{n-1} = x\}$ ne dépend pas de n . On note $p_{x,y}$ cette dernière probabilité.

Étant donnée une matrice stochastique $\mathcal{P} = (p_{x,y})$ ($(x, y) \in S^2$), c'est-à-dire une matrice dont les coefficients vérifient les propriétés $p_{x,y} \geq 0$ et pour tout $x \in S$, $\sum_{y \in S} p_{x,y} = 1$ et étant donné une mesure de probabilité (p_x) ($x \in S$) sur S , il existe une et une seule mesure de probabilité P sur (Ω, \mathfrak{F}) dont les projections $X_n : \Omega \rightarrow S$ forment une chaîne de Markov homogène telle que $P\{X_1 = x\} = p_x$ et $P\{X_n = x_n \mid X_{n-1} = x_{n-1}\} = p_{x_{n-1}, x_n}$.

CHAPITRE 11

ESPÉRANCE MATHÉMATIQUE. LOIS ABSOLUMENT CONTINUES

Dans ce chapitre, nous donnons la définition de l'espérance mathématique d'une variable aléatoire réelle par rapport à une mesure de probabilité quelconque et faisons le rapprochement avec la définition de l'espérance mathématique qui avait été donnée pour les variables aléatoires discrètes. La théorie de l'intégrale développée dans le chapitre précédent fournit tous les outils nécessaires. Comme nous le verrons, il y a essentiellement deux classes de variables aléatoires, les variables aléatoires discrètes et les variables aléatoires *absolument continues*. Il nous a paru naturel d'énoncer dans ce même chapitre les techniques de calcul de l'espérance mathématique lorsque les variables aléatoires appartiennent à cette seconde classe.

1. Espérance mathématique d'une variable aléatoire. — Soient $(\Omega, \mathfrak{A}, P)$ un espace probabilisé et X une variable aléatoire réelle définie sur cet espace. Suivant la théorie développée dans le chapitre précédent, on peut considérer l'intégrale de X par rapport à la *mesure de probabilité* P .

Définition. — Si X est positive ou si X est P -intégrable, on désigne par

$$(1.1) \quad \mathbb{E}[X] = \int X dP$$

l'*espérance mathématique* de X . Si l'on a $\mathbb{E}[X] \in \mathbb{R}$, on dit que X a une espérance mathématique *finie*.

Toutes les propriétés de l'intégrale énoncées dans le chapitre précédent, de linéarité, de monotonie, d'intégrabilité, de majoration, ainsi que les trois théorèmes de convergence restent valables. Il faut noter que les quatre premières séries de propriétés avaient déjà été données, presque mot pour mot, pour les variables aléatoires discrètes au chapitre 8.

Lorsque la mesure sous-jacente est une mesure de probabilité, on préfère parler de propriété vraie *P -presque sûrement*, plutôt que vraie *P -presque partout*. Dans les chapitres suivants, on fera, en particulier, une étude très approfondie de la convergence *presque sûre* des suites de variables aléatoires réelles.

Comme pour les variables aléatoires discrètes, on peut donc définir les notions de moment, de variance, ... Lorsque $\mathbb{E}[X] = 0$, on dit que X est *centrée*. Si $\mathbb{E}[X]$ est finie, on peut *centrer* la variable aléatoire en considérant $X - \mathbb{E}[X]$. Le *moment d'ordre k* ($k > 0$) est le nombre, s'il existe, $\mathbb{E}[X^k]$. Si $\mathbb{E}[X^2]$ est fini, la *variance* de X est définie par $\mathbb{E}[(X - \mathbb{E}[X])^2]$. Enfin, le *moment absolu d'ordre k* est le nombre $\mathbb{E}[|X|^k]$.

L'un des théorèmes fondamentaux sur l'intégration des variables aléatoires réelles et tout particulièrement sur leur calcul effectif est le théorème dit du transfert que nous énonçons maintenant et dont nous avons donné une version discrète au chap. 8, Théorème 4.1.

THÉORÈME 1.1 (théorème du transfert). — *Soient $(\Omega, \mathfrak{A}, P)$ un espace probabilisé, X une variable aléatoire définie sur cet espace, à valeurs réelles, de loi P_X et g une fonction mesurable réelle. On a donc le diagramme :*

$$\begin{array}{ccc} (\Omega, \mathfrak{A}, P) & \xrightarrow{X} & (\mathbb{R}, \mathcal{B}^1, P_X) \\ & \searrow g \circ X & \downarrow g \\ & & (\mathbb{R}, \mathcal{B}^1) \end{array}$$

Alors, si l'une des expressions $\int_{\Omega} (g \circ X) dP$, $\int_{\mathbb{R}} g dP_X$ existe, il en est de même de l'autre et l'on a :

$$(1.2) \quad \int_{\Omega} (g \circ X) dP = \int_{\mathbb{R}} g dP_X \left(= \int_{\mathbb{R}} g(x) dP_X(x) \right).$$

Les deux termes de cette identité sont deux représentations de $\mathbb{E}[g \circ X]$.

Le premier membre de l'identité (1.2) est l'intégrale de la variable aléatoire $g \circ X : (\Omega, \mathfrak{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}^1)$ par rapport à la mesure P ; le second membre est l'intégrale de la variable aléatoire $g : (\mathbb{R}, \mathcal{B}^1, P_X) \rightarrow (\mathbb{R}, \mathcal{B}^1)$ par rapport à la mesure (de Stieltjes-Lebesgue) P_X sur la droite réelle. Le théorème de transfert permet ainsi de *transférer* le calcul d'une intégrale sur un espace probabilisé quelconque $(\Omega, \mathfrak{A}, P)$ en le calcul d'une intégrale sur $(\mathbb{R}, \mathcal{B}^1, P_X)$.

Démonstration. — Pour démontrer le théorème de transfert, on le vérifie successivement pour les variables aléatoires simples positives, puis pour celles qui sont positives, puis pour les variables aléatoires quelconques. Soit $g = \sum_i x_i I_{A_i}$ une variable aléatoire simple positive. On a, pour $\omega \in \Omega$,

$$(g \circ X)(\omega) = g(X(\omega)) = \sum_i x_i I_{A_i}(X(\omega)) = \sum_i x_i I_{X^{-1}(A_i)}(\omega),$$

d'où

$$\int (g \circ X) dP = \sum_i x_i P(X^{-1}(A_i)) = \sum_i x_i P_X(A_i) = \int g dP_X.$$

Si maintenant (g_n) est une suite croissante de variables aléatoires simples positives telles que $g = \sup_n g_n$, on a aussi $g \circ X = \sup_n g_n \circ X$. De plus, $g_n \circ X$ est une variable aléatoire simple positive, puisque g_n l'est. D'où

$$\int (g \circ X) dP = \sup_n \int (g_n \circ X) dP = \sup_n \int g_n dP_X = \int g dP_X.$$

Enfin, si g est une variable aléatoire quelconque, on a $g^+ \circ X = (g \circ X)^+$ et $g^- \circ X = (g \circ X)^-$. Si $\int (g \circ X) dP$ est finie, alors les deux intégrales $\int (g \circ X)^+ dP$ et $\int (g \circ X)^- dP$ sont aussi finies. On a, de plus, $\int (g \circ X) dP = \int (g^+ \circ X) dP - \int (g^- \circ X) dP = \int g^+ dP_X - \int g^- dP_X = \int g dP_X$. D'où $\int g dP_X$ est finie et égale à $\int (g \circ X) dP$.

De même, si $\int g dP_X$ est finie, on démontre, par le même raisonnement, que $\int (g \circ X) dP$ est finie et lui est égale. \square

Exemple 1. — Prenons $g = \text{Id}_{\mathbb{R}}$, soit $g(x) = x$. Alors, si l'une des expressions $\int X dP$, $\int \text{Id}_{\mathbb{R}} dP_X = (\int x dP_X(x))$ existe, il en est de même de l'autre et l'on a :

$$(1.3) \quad \mathbb{E}[X] = \int X dP = \int x dP_X(x).$$

Exemple 2. — Soit $X : (\Omega, \mathfrak{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}^1)$ une variable aléatoire *discrète* de loi $P_X = \sum_i \alpha_i \varepsilon_{x_i}$ ($\alpha_i > 0$, $\sum_i \alpha_i = 1$). Si $X \geq 0$, ou si $\sum_i \alpha_i |x_i| < \infty$, on a toujours les relations (1.3). Or le dernier membre de (1.3) est l'intégrale de l'application identique de $(\mathbb{R}, \mathcal{B}^1, P_X)$ dans $(\mathbb{R}, \mathcal{B}^1)$. D'après l'Exemple 2 du paragraphe 7, chap. 10, elle vaut $\sum_i \alpha_i x_i$. On retrouve la définition élémentaire d'une variable discrète :

$$\mathbb{E}[X] = \sum_i \alpha_i x_i.$$

2. Mesures de probabilité produit et théorème de Fubini. — Dans ce paragraphe, nous donnons, sans démonstration, quelques résultats sur les produits de mesures de probabilité.

THÉORÈME 2.1. — Soient P_1 et P_2 deux mesures de probabilité sur $(\mathbb{R}, \mathcal{B}^1)$. Alors il existe une mesure de probabilité et une seule sur $(\mathbb{R}^2, \mathcal{B}^2)$, notée $P = P_1 \otimes P_2$ vérifiant pour tout $B_1, B_2 \in \mathcal{B}^1$ l'identité

$$P(B_1 \times B_2) = P_1(B_1)P_2(B_2).$$

La mesure P est appelée le *produit* des mesures P_1 et P_2 ; les mesures P_1 et P_2 sont appelées les mesures *marginales* de la mesure P . La mesure produit est d'une importance capitale dans l'étude des variables aléatoires réelles indépendantes, comme on le voit dans le théorème suivant.

THÉORÈME 2.2. — Soient X_1, X_2 deux variables aléatoires réelles sur un même espace probabilisé de lois respectives P_1, P_2 . Alors les deux propriétés suivantes sont équivalentes :

- a) la loi du couple (X_1, X_2) est la mesure de probabilité produit $P = P_1 \otimes P_2$;
- b) les variables aléatoires réelles X_1, X_2 sont indépendantes.

THÉORÈME 2.3 (Théorème de Fubini, cf. Revuz [12], p. 171). — Soient P_1, P_2 deux mesures de probabilité sur $(\mathbb{R}, \mathcal{B}^1)$, $P = P_1 \otimes P_2$ le produit des deux mesures et g une fonction mesurable $g : (\mathbb{R}^2, \mathcal{B}^2) \rightarrow (\mathbb{R}, \mathcal{B}^1)$ telle que $\int g dP$ existe (i.e. ou bien $g \geq 0$, ou bien $\int |g| dP < \infty$). Alors $h(x_2) = \int_{\mathbb{R}} g(x_1, x_2) dP_1(x_1)$ existe pour P_2 -presque tout x_2 ; de plus, $\int_{\mathbb{R}} h(x_2) dP_2(x_2)$ existe et est égal à $\int_{\mathbb{R}} g dP$; ou encore

$$(2.1) \quad \int_{\mathbb{R}^2} g(x_1, x_2) dP(x_1, x_2) = \int_{\mathbb{R}} dP_2(x_2) \left(\int_{\mathbb{R}} g(x_1, x_2) dP_1(x_1) \right).$$

Les mêmes affirmations sont vraies en échangeant les rôles des indices 1 et 2.

Dans l'identité (2.1) le membre de gauche est l'intégrale de la fonction g de deux variables réelles par rapport à la mesure de Stieltjes-Lebesgue $P = P_1 \otimes P_2$ sur $(\mathbb{R}^2, \mathcal{B}^2)$ (cf. chap. 10, § 4).

CAS PARTICULIER. — Si g est une fonction positive, alors l'identité (2.1) est toujours vraie (dans $[0, +\infty]$).

APPLICATION. — Soient X_1, X_2 deux variables aléatoires réelles indépendantes admettant des espérances mathématiques. Alors le produit $X_1 X_2$ admet une espérance mathématique et l'on a :

$$\mathbb{E}[X_1 X_2] = \mathbb{E}[X_1] \mathbb{E}[X_2].$$

Démonstration. — Il suffit d'appliquer le théorème de Fubini en prenant pour P_1, P_2 les lois de X_1, X_2 et pour $P = P_1 \otimes P_2$ la loi du couple (X_1, X_2) . \square

Certaines intégrales par rapport à des mesures de Stieltjes-Lebesgue se calculent directement comme des intégrales ordinaires. Nous nous proposons, dans le reste de ce chapitre, de donner les indications nécessaires pour arriver à de tels calculs.

3. Intégrale de Lebesgue. — On a appelé *mesure de Lebesgue* sur la droite réelle (cf. chap. 10, § 2) l'unique mesure, notée λ^1 ou λ , sur $(\mathbb{R}, \mathcal{B}^1)$, qui est telle que pour tout intervalle semi-ouvert $]a, b]$, on ait

$$(3.1) \quad \lambda\{]a, b]\} = b - a.$$

Comme l'application identique sur la droite qui induit la mesure de Lebesgue est évidemment continue, on a encore

$$\lambda\{[a, b]\} = \lambda\{[a, b[\} = \lambda\{]a, b[\} = b - a.$$

En fait, la mesure de Lebesgue est définie comme la complétée de λ . La précédente définition suffira pour les applications qui viennent.

Soit maintenant X une variable aléatoire réelle, définie sur $(\mathbb{R}, \mathcal{B}^1)$. On définit l'intégrale de Lebesgue de X comme l'intégrale de X par rapport à la mesure λ . On la note

$$(3.2) \quad \int X d\lambda \quad \text{ou} \quad \int X(x) dx.$$

L'intégrale de X sur un ensemble borélien A de la droite est définie comme l'intégrale de $I_A \cdot X$. On la note

$$\int_A X d\lambda = \int I_A \cdot X d\lambda = \int_A X(x) dx = \int I_A(x) X(x) dx.$$

En particulier, lorsque A est un intervalle borné $[a, b]$, $[a, b[$, $]a, b]$ ou $]a, b[$, l'intégrale de X sur A est notée

$$(3.3) \quad \int_a^b X d\lambda = \int_a^b X(x) dx,$$

toutes les intégrales de X sur chacun de ces intervalles, si elles existent, étant égales.

On reconnaît dans (3.3) l'expression formelle usuelle de l'intégrale de Riemann de la fonction X sur l'intervalle $[a, b]$. Nous consignons dans les deux propositions suivantes, données sans démonstration, des conditions permettant de comparer intégrale de Lebesgue et intégrale de Riemann.

PROPOSITION 3.1. — *Soit X une fonction réelle, mesurable, bornée, définie sur un intervalle borné $[a, b]$. Si X est intégrable au sens de Riemann, alors X est intégrable au sens de Lebesgue sur $[a, b]$ et l'intégrale de Riemann de X est égale à l'intégrale de Lebesgue de X sur $[a, b]$. (Pour les deux intégrales, on retrouve l'expression familière (3.3).)*

On rappelle que si X est une fonction réelle, définie sur la droite et intégrable au sens de Riemann sur tout intervalle borné $[a, b]$, son intégrale impropre de Riemann sur \mathbb{R} est définie par

$$(3.4) \quad \int_{-\infty}^{+\infty} X(x) dx = \lim_{\substack{a \rightarrow -\infty \\ b \rightarrow +\infty}} \int_a^b X(x) dx,$$

si la limite existe et est finie.

PROPOSITION 3.2. — *Soit X une fonction réelle, mesurable, définie sur la droite réelle. Si l'intégrale impropre de Riemann de $|X|$ existe et est finie, alors X est intégrable au sens de Lebesgue et l'intégrale de Lebesgue de X sur \mathbb{R} est égale à l'intégrale impropre de Riemann de X sur \mathbb{R} :*

$$\int_{\mathbb{R}} X(x) dx = \int_{-\infty}^{+\infty} X(x) dx.$$

4. Lois de probabilité absolument continues. — Les fonctions f réelles, positives, de variable réelle, qui sont intégrables au sens de Lebesgue et d'intégrale égale à 1 permettent de définir une classe importante de mesures de probabilité sur la droite, comme le montre la proposition suivante.

PROPOSITION 4.1. — *Soit f une fonction réelle positive, de variable réelle, intégrable au sens de Lebesgue et telle que*

$$\int_{\mathbb{R}} f d\lambda = \int_{-\infty}^{+\infty} f(x) dx = 1.$$

Alors la fonction $P : B \mapsto \int_B f d\lambda$ est une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}^1)$.

Démonstration. — D'abord $P \geq 0$ et $P(\mathbb{R}) = \int_{\mathbb{R}} f d\lambda = 1$. Ensuite $P(B) = \int_B f d\lambda \leq \int_{\mathbb{R}} f d\lambda = 1$.

D'autre part, considérons une suite (B_n) d'ensembles boréliens de la droite, disjoints deux à deux et de réunion B . On a alors $f.I_B = \sum_n f.I_{B_n}$, d'où $P(B) = \int f.I_B d\lambda = \int \sum_n f.I_{B_n} d\lambda = \sum_n \int f.I_{B_n} d\lambda = \sum_n P(B_n)$, d'après le théorème monotone de Beppo-Levi. Ainsi P est bien une mesure de probabilité. \square

Les mesures de probabilité auxquelles on peut faire correspondre une fonction f ayant les propriétés de la proposition précédente sont précisément les lois absolument continues, que l'on va maintenant étudier.

Étant donné une fonction de répartition F sur la droite, on sait qu'il existe une et une seule mesure de probabilité, notée $F\{\cdot\}$ sur $(\mathbb{R}, \mathcal{B}^1)$ telle que $F\{]a, b]\} = F(b) - F(a)$, pour tout intervalle borné $]a, b]$. La complétée de cette mesure F s'appelle la *mesure de Stieltjes-Lebesgue* induite par F . Si P est une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}^1)$, la complétée de P est identique à la mesure de Stieltjes-Lebesgue induite par la fonction $F : x \mapsto P\{]-\infty, x]\}$. Ainsi l'intégrale d'une variable aléatoire réelle g , de variable réelle, sera notée indifféremment

$$\int g dP \quad \text{ou} \quad \int g dF, \quad \text{ou encore} \quad \int g(x) dF(x),$$

et on l'appellera *intégrale de Stieltjes-Lebesgue de g par rapport à F* .

Définition. — On dit que F est *absolument continue*, s'il existe une fonction réelle f , de variable réelle, telle que

(i) $f \geq 0$;

(ii) f est intégrable au sens de Lebesgue sur \mathbb{R} et $\int f(x) dx = 1$;

(iii) $F\{B\} (= \int_B dF(x)) = \int_B f(x) dx (= \int_B f d\lambda)$ pour tout $B \in \mathcal{B}^1$.

La fonction f est appelée *densité* de la fonction de répartition F .

PROPOSITION 4.2. — *Soit F une fonction de répartition absolument continue, de densité f . Alors pour toute variable aléatoire $g : (\mathbb{R}, \mathcal{B}^1) \rightarrow (\mathbb{R}, \mathcal{B}^1)$, on a l'identité*

$$(4.1) \quad \int g(x) dF(x) \left(= \int g dF \right) = \int g(x) f(x) dx \left(= \int g f d\lambda \right),$$

dans le sens suivant : si l'une de ces intégrales existe, alors l'autre existe aussi et elles sont égales.

Démonstration. — Vérifions d'abord (4.1) pour les variables aléatoires simples positives. Soit $g = \sum_k x_k I_{A_k}$ une telle fonction. On a : $\int g dF = \sum_k x_k F\{A_k\} = \sum_k x_k \int_{A_k} f d\lambda = \sum_k x_k \int I_{A_k} f d\lambda = \int (\sum_k x_k I_{A_k}) f d\lambda = \int g f d\lambda$.

Si (g_n) est une suite croissante de variables aléatoires simples positives tendant vers g , alors $(g_n f)$ est une suite croissante de variables aléatoires positives tendant vers $g f$. D'où $\int g dF = \lim_n \int g_n dF = \lim_n \int g_n f d\lambda = \int g f d\lambda$, d'après le théorème de Beppo-Levi.

Enfin, si g est quelconque, on a : $\int g^+ dF = \int g^+ f d\lambda$ et $\int g^- dF = \int g^- f d\lambda$. L'identité (4.1) est encore vraie par définition de l'intégrale. \square

Remarque. — Si F est une fonction de répartition absolument continue, de densité f et si F est la fonction de répartition d'une variable aléatoire réelle X , on dira aussi que f est la *densité de probabilité de X* . On utilisera aussi les notations F_X , f_X , en homogénéité avec la notation P_X pour la loi de probabilité de X .

Dans l'identité (4.1), prenons pour F la loi de répartition F_X d'une variable aléatoire X de densité f_X . Puisque par convention d'écriture, on peut remplacer « dF_X » par « dP_X », l'identité (4.1) se récrit :

$$\int g(x) dP_X(x) = \int g(x) f_X(x) dx.$$

En prenant $g(x) = x$ pour tout x , on obtient la formule importante :

$$(4.2) \quad \mathbb{E}[X] = \int x dP_X(x) = \int x f_X(x) dx,$$

qui permet de calculer les espérances mathématiques des variables aléatoires réelles admettant des densités.

PROPOSITION 4.3. — *Si F est absolument continue, alors F est continue sur toute la droite et $F\{\{x\}\} = 0$ pour tout x . Si de plus F est la fonction de répartition d'une variable aléatoire réelle définie sur un espace probabilisé $(\Omega, \mathfrak{A}, P)$, alors $P\{X = x\} = 0$ pour tout x . [La variable aléatoire X est donc diffuse.]*

Démonstration. — Soit f la densité de F . On a $F\{\{x\}\} = F(x) - F(x-0)$ pour tout x . Or $F\{\{x\}\} = \int_x^x f(u) du = 0$, d'où $F(x-0) = F(x)$ et F est continue. Dans les hypothèses de l'énoncé, on a $P\{X = x\} = P_X\{x\} = F\{\{x\}\} = 0$. \square

L'ensemble $D_X = \{x : P\{X = x\} > 0\}$ a été introduit au chap. 5, § 6. Il résulte de la proposition précédente que D_X est vide, si X est absolument continue (voir aussi chap. 5, Proposition 6.2). On définit le *support* S_X d'une variable aléatoire réelle X comme étant l'ensemble D_X lui-même, lorsque X est discrète, et comme étant l'ensemble $S_X = \{x : f(x) > 0\}$, lorsque X est absolument continue de densité f .

PROPOSITION 4.4. — Si F est absolument continue, de densité f , alors $F'(x) = f(x)$ en tout point x où f est continue.

Démonstration. — Prenons $h \neq 0$ et supposons f continue au point x . Le nombre ε étant donné, on a, si $|t - x| < \eta$, l'encadrement $f(x) - \varepsilon < f(t) < f(x) + \varepsilon$. Or $(F(x+h) - F(x))/h = \int_x^{x+h} (f(t)/h) dt$, d'où pour $|h| < \eta$ l'encadrement $f(x) - \varepsilon < (F(x+h) - F(x))/h < f(x) + \varepsilon$, qui exprime bien que la dérivée $F'(x)$ de $F(x)$ est égale à $f(x)$. \square

Dans le cas particulier où X est à valeurs positives, on peut donner une représentation de l'espérance mathématique à l'aide de la fonction de survie $r(x)$ (cf. chap. 14, § 5. La loi exponentielle).

THÉORÈME 4.5. — Soit X une variable aléatoire à valeurs dans $[0, +\infty[$, absolument continue, de densité f , de fonction de survie $r(x) = P\{X > x\}$. Alors on a, dans la demi-droite achevée $[0, +\infty]$,

$$\int_0^{+\infty} r(x) dx = \int_0^{+\infty} x f(x) dx.$$

Démonstration. — En effet

$$\int_0^{+\infty} r(x) dx = \int_0^{+\infty} \left(\int_x^{+\infty} f(t) dt \right) dx = \int_0^{+\infty} \left(\int_0^{+\infty} f(t) I_{\{t > x\}}(t, x) dt \right) dx;$$

soit d'après le théorème de Fubini

$$\int_0^{+\infty} r(x) dx = \int_0^{+\infty} \left(\int_0^{+\infty} I_{\{t > x\}}(t, x) dx \right) f(t) dt = \int_0^{+\infty} t f(t) dt. \quad \square$$

Dans la démonstration précédente on a fait appel au théorème de Fubini pour les intégrales doubles, mais, en fait, il n'y a aucune difficulté ici puisque la fonction à intégrer est positive.

Remarque. — Si $X \geq 0$ et si $\mathbb{E}[X]$ est finie, alors $\mathbb{E}[X]$ admet la représentation :

$$(4.3) \quad \mathbb{E}[X] = \int_0^{+\infty} r(x) dx.$$

5. Les trois types de fonctions de répartition. — Soit X une variable aléatoire réelle de fonction de répartition F . On distingue tout d'abord deux cas :

a) F est « totalement » discontinue, auquel cas l'ensemble des points de discontinuité de F est fini ou dénombrable, et la variable aléatoire X est *discrète*.

b) F est continue, auquel cas la variable aléatoire X est *diffuse*.

Parmi les variables aléatoires diffuses, nous avons rencontré les variables aléatoires *absolument continues*. On montre que ces dernières ne constituent

pas le cas général des variables aléatoires diffuses : en d'autres termes, *il existe des variables aléatoires diffuses qui ne sont pas absolument continues*. Ces variables aléatoires sont appelées *singulières*; l'exemple type en est fourni par une variable aléatoire dont la fonction de répartition est celle de l'ensemble triadique de Cantor sur $[0, 1]$. Nous ne les étudierons pas ici et nous nous bornerons à énoncer le théorème suivant dû à Lebesgue.

THÉORÈME 5.1 (cf. Munroe [8], chap. 6). — *Soit F une fonction de répartition. Alors il existe trois fonctions de répartition F_1 discrète, F_2 absolument continue, F_3 singulière et trois nombres réels $\alpha_1, \alpha_2, \alpha_3$ vérifiant $\alpha_1, \alpha_2, \alpha_3 \geq 0, \alpha_1 + \alpha_2 + \alpha_3 = 1$ tels que F puisse être représentée sous la forme :*

$$F = \alpha_1 F_1 + \alpha_2 F_2 + \alpha_3 F_3.$$

En d'autres termes, toute fonction de répartition peut être représentée comme combinaison convexe de fonctions de répartition des trois types fondamentaux. Cette représentation est en outre unique, si tous les coefficients de la combinaison linéaire sont strictement positifs.

6. Convolution. — Le produit de convolution de deux lois discrètes a été étudié au chap. 8, § 3. Nous traitons ici le cas des mesures de probabilité quelconques sur la droite.

Définition. — Soient P et Q deux mesures de probabilité sur la droite, dont les fonctions de répartition correspondantes sont F et G. On appelle *produit de convolution de P et Q* (ou de F et G), la mesure de probabilité, notée $P * Q$, dont la fonction de répartition, notée $F * G$, est donnée par

$$(6.1) \quad H(z) = (F * G)(z) = \int F(z - y) dG(y) = \int G(z - x) dF(x).$$

On vérifie que H est bien une fonction de répartition. Que les deux termes de droite de (6.1) soient égaux résulte de la proposition suivante.

PROPOSITION 6.1. — *Soient X et Y deux variables aléatoires réelles, de lois P_X et P_Y , de fonctions de répartition F et G. Si X et Y sont indépendantes, la loi de la somme $X + Y$ est le produit de convolution de P_X et P_Y ; ou encore la loi P_{X+Y} de $X + Y$ est donné par*

$$(6.2) \quad P_{X+Y} = P_X * P_Y;$$

ou enfin la fonction de répartition H de $X + Y$ est donnée par

$$(6.3) \quad H(z) = (F * G)(z) = \int F(z - y) dG(y) = \int G(z - x) dF(x).$$

Démonstration. — D'après le Théorème 2.2, la loi du couple (X, Y) est la mesure de probabilité produit $P_X \otimes P_Y$. Supposons que X et Y sont définies sur le même espace probabilisé $(\Omega, \mathfrak{A}, P)$ (en toute rigueur, il

faudrait démontrer que c'est toujours possible ; admettons ce résultat ici) et notons g la fonction réelle de deux variables réelles : $g(x, y) = x + y$. En désignant par P_T la loi du couple $T = (X, Y)$, on a $X + Y = g \circ T$, puis $H(z) = P\{X + Y \leq z\} = P\{g \circ T \leq z\} = P_T\{g \leq z\}$. D'où

$$H(z) = \int_{\mathbb{R}^2} I_{\{g \leq z\}}(x, y) dP_T(x, y),$$

soit, d'après le Théorème de Fubini 2.3,

$$H(z) = \int_{\mathbb{R}} dP_Y(y) \int_{\mathbb{R}} I_{\{g \leq z\}}(x, y) dP_X(x).$$

Or $I_{\{g \leq z\}}(x, y) = 1$ si $x + y \leq z$ et 0 autrement. D'où

$$I_{\{g \leq z\}}(x, y) = I_{\{]-\infty, z-y]\}}(x)$$

et donc

$$\begin{aligned} H(z) &= \int_{\mathbb{R}} dP_Y(y) \int_{\mathbb{R}} I_{\{]-\infty, z-y]\}}(x) dP_X(x) \\ &= \int_{\mathbb{R}} F(z - y) dP_Y(y) = \int_{\mathbb{R}} F(z - y) dG(y). \end{aligned}$$

On obtient l'égalité avec le membre tout à droite de (6.3) en intégrant d'abord par rapport à y . \square

Cette proposition, et tout particulièrement la formule (6.2), entraîne que le produit de convolution est *commutatif* et *associatif* :

$$P * Q = Q * P, \quad P * (Q * R) = (P * Q) * R.$$

Lorsque les lois F et G sont absolument continues, on a le résultat suivant.

PROPOSITION 6.2. — Si F et G admettent des densités f et g , alors $H = F * G$ admet la densité h donnée par :

$$(6.4) \quad h(z) = \int_{\mathbb{R}} f(z - y) g(y) dy = \int_{\mathbb{R}} g(z - x) f(x) dx.$$

La densité h est appelée le produit de convolution des densités f , g , et est notée $f * g$.

Démonstration. — Par définition des densités de probabilité, on a :

$$\begin{aligned} H(z) &= \int_{\mathbb{R}} F(z - y) dG(y) = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z-y} f(x) dx \right) g(y) dy \\ &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^z f(x - y) dx \right) g(y) dy \\ &= \int_{-\infty}^z \left(\int_{-\infty}^{+\infty} f(x - y) g(y) dy \right) dx, \end{aligned}$$

par le théorème de Fubini appliqué aux intégrales doubles sur \mathbb{R}^2 . D'où $h(z) = \int f(z - y)g(y) dy$ est une densité pour $H(z)$. \square

COMPLÉMENTS ET EXERCICES

1. — Dans l'Application du Théorème 2.3, on a démontré que, lorsque deux variables aléatoires X_1 et X_2 sont indépendantes, l'espérance mathématique de leur produit est égale au produit de leurs espérances mathématiques. La démonstration donnée utilise le fait que l'on peut définir ces espérances par rapport à l'espace probabilisé $(\mathbb{R}^2, \mathcal{B}^2, P_1 \otimes P_2)$. On peut aussi faire la démonstration par rapport à l'espace probabilisé abstrait $(\Omega, \mathfrak{A}, P)$ sur lequel ces deux variables aléatoires sont définies. On a recours alors aux techniques de construction de l'intégrale développées au chapitre 10.

D'abord, $\mathbb{E}[X_1 X_2] = \mathbb{E}[X_1] \mathbb{E}[X_2]$ lorsque X_1 et X_2 sont des variables aléatoires simples positives, puisque cette relation a déjà été établie au chap. 8, Théorème 4.3 (C), pour les variables aléatoires discrètes.

a) Utiliser la décomposition (5.4.1) du chap. 10, appliquée à deux variables aléatoires positives, pour montrer que la précédente relation est encore vraie pour les variables aléatoires positives.

b) Conclure que la relation est vraie pour des variables aléatoires quelconques (admettant toutefois des espérances mathématiques finies!) en utilisant les représentations habituelles : $X_i = X_i^+ - X_i^-$ ($i = 1, 2$).

2. — Soient X une variable aléatoire absolument continue, f_X sa densité de probabilité, S_X son support (cf. le commentaire suivant la Proposition 4.3). Soit d'autre part $h = \mathbb{R} \rightarrow \mathbb{R}$ une fonction mesurable telle que l'ensemble $h(S_X)$ soit *fini ou dénombrable*. Montrer que la variable aléatoire $Y = h \circ X$ est alors discrète

a) de support $S_Y \subset h(S_X)$;

b) de fonction de masse : $\forall y \in S_Y, \quad \pi_Y(y) = \int_{h^{-1}(\{y\})} f_X(x) dx.$

3. — Soient X une variable aléatoire absolument continue, f_X sa densité de probabilité supposée continue, S_X son support. Soit d'autre part $h : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continûment dérivable et *strictement monotone*. Montrer que la variable aléatoire $Y = \varphi \circ X$ est alors absolument continue

a) de support $S_Y = f(S_X)$;

b) de densité de probabilité

$$f_Y(y) = \begin{cases} f_X(h^{-1}(y)) |(h^{-1}(y))'|, & \text{si } y \in S_Y; \\ 0, & \text{sinon.} \end{cases}$$

4. (Ph. Artzner). — Soit X une variable aléatoire *positive* admettant une densité f continûment dérivable et *strictement décroissante*. Désignons

par M la médiane (voir chap. 8, § 10) de X et par $\mathbb{E}[X]$ son espérance mathématique (qui peut être égale à $+\infty$). Alors $M \leq \mathbb{E}[X]$.

5. — Soient X_1, X_2, X_3 trois variables aléatoires indépendantes, chacune de loi uniforme sur $[0, 1]$. Calculer les densités de probabilité de $X_1 + X_2$, $X_1 + X_2 + X_3$, $X_1 - X_2$.

6. — Soit X une variable aléatoire de densité $f(x) = I_{[0,1]}(x)$. Alors la variable centrée $Y = X - \frac{1}{2}$ a pour densité $g(x) = f(y + \frac{1}{2}) = I_{[-1/2, 1/2]}(y)$; sa loi est la loi uniforme sur $[-1/2, 1/2]$. Soient ensuite X_1, X_2, X_3 trois variables aléatoires indépendantes, chacune de loi uniforme sur $[0, 1]$. Alors $Y_1 = X_1 - \frac{1}{2}$, $Y_2 = X_2 - \frac{1}{2}$, $Y_3 = X_3 - \frac{1}{2}$ sont trois variables aléatoires indépendantes, chacune de loi uniforme sur $[-1/2, 1/2]$. Déterminer les lois de $Y_1 + Y_2$, de $Y_1 + Y_2 + Y_3$ et de $Y_1 - Y_2$.

7. — Soient $n \geq 2$ et X_1, X_2, \dots, X_n une suite de n variables aléatoires indépendantes, chacune de loi uniforme sur $[0, 1]$. Soit f_n la densité de la somme $X_1 + X_2 + \dots + X_n$. Pour tout x réel, on pose $(x)^+ = \max(0, x)$. La densité f_n est donnée par :

$$f_n(x) = \begin{cases} \sum_{k=0}^{n-1} (-1)^k \binom{n}{k} \frac{((x-k)^+)^{n-1}}{(n-1)!}, & \text{si } 0 \leq x \leq n; \\ 0, & \text{sinon.} \end{cases}$$

Noter qu'un calcul direct de f_2 et f_3 a été demandé dans l'exercice 5.

8. — Soient $n \geq 1$ et X_1, X_2, \dots, X_n une suite de n variables aléatoires, indépendantes, chacune suivant la loi uniforme sur $[0, 1]$. On note f_n (resp. F_n) la densité (resp. la fonction de répartition) du produit $X = X_1 X_2 \dots X_n$. Alors

$$(a) \quad f_n(x) = \frac{(-\text{Log } x)^{n-1}}{(n-1)!} I_{[0,1]}(x);$$

$$(b) \quad F_n(x) = x \left(1 + \frac{(-\text{Log } x)}{1!} + \frac{(-\text{Log } x)^2}{2!} + \dots + \frac{(-\text{Log } x)^{n-1}}{(n-1)!} \right), \text{ si } x \in [0, 1] \text{ et } F_n(x) = 0 \text{ si } x < 0; F_n(x) = 1, \text{ si } x > 1.$$

On remarque que, lorsque n tend vers l'infini, on a $F_n(x) \rightarrow x e^{-\log x} = 1$, pour tout $x \in]0, 1[$, c'est-à-dire $X_1 X_2 \dots X_n \xrightarrow{\mathcal{L}} 0$ (cf. Chap. 16, § 1).

[Indication : pour (a) commencer par calculer f_2 au moyen de la loi du produit (p. 200). Une simple récurrence fournit la valeur f_n pour $n \geq 2$. Pour obtenir (b), effectuer une simple intégration.]

VECTEURS ALÉATOIRES; ESPÉRANCE CONDITIONNELLE. LOIS NORMALES

On trouve dans ce chapitre, d'abord une courte étude des vecteurs aléatoires à deux dimensions et de leurs lois de probabilité. Pour la sous-classe des vecteurs aléatoires dont la loi de probabilité est absolument continue, on donne ensuite un exposé sur l'espérance conditionnelle, avec la liste des règles de calcul. Le chapitre se termine par un traitement des lois normales à deux dimensions.

1. Définitions et premières propriétés. — Comme déjà défini au chap. 5, § 3, une *variable aléatoire à deux dimensions* (on dit encore *vecteur aléatoire à deux dimensions*) est une application mesurable $X : (\Omega, \mathfrak{A}, \mathbb{P}) \rightarrow (\mathbb{R}^2, \mathcal{B}^2)$. Considérons les deux projections canoniques $\pi_i : (x_1, x_2) \mapsto x_i$ ($i = 1, 2$) de \mathbb{R}^2 dans \mathbb{R} . Alors les applications coordonnées de $(\Omega, \mathfrak{A}, \mathbb{P})$ dans $(\mathbb{R}, \mathcal{B}^1)$ définies par $X_i = \pi_i \circ X$ ($i = 1, 2$) sont toutes deux mesurables. Ce sont donc deux variables aléatoires, appelées *variables aléatoires marginales*.

On écrit souvent $X = (X_1, X_2)$ ou $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ et l'on parle aussi de *couple* de variables aléatoires pour désigner X .

Soit $X : (\Omega, \mathfrak{A}, \mathbb{P}) \rightarrow (\mathbb{R}^2, \mathcal{B}^2)$ un vecteur aléatoire à deux dimensions. Au chap. 5, § 4, on a défini la *loi de probabilité du vecteur X* comme étant l'application, notée \mathbb{P}_X , qui envoie tout $B \in \mathcal{B}^2$ sur le nombre $\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B))$. D'après la Proposition 4.1 du chap. 5, l'application \mathbb{P}_X est une mesure de probabilité sur $(\mathbb{R}^2, \mathcal{B}^2)$ appelée *loi de probabilité de X* . On dira encore que c'est la *loi conjointe du couple $X = (X_1, X_2)$* .

Comme déjà vérifié pour les variables aléatoires discrètes (voir Corollaire 1.2 du chap. 8), la loi conjointe de $X = (X_1, X_2)$ *détermine* les lois des variables aléatoires marginales X_1, X_2 (lois marginales). Ceci fait l'objet de la prochaine proposition.

PROPOSITION 1.1. — *La loi conjointe \mathbb{P}_X de $X = (X_1, X_2)$ permet de déterminer les lois (dites marginales) $\mathbb{P}_{X_1}, \mathbb{P}_{X_2}$ des variables aléatoires marginales X_1, X_2 , de la façon suivante. On a pour tout $B \in \mathcal{B}^1$*

$$\mathbb{P}_{X_i}(B) = \mathbb{P}_X(\pi_i^{-1}(B)) \quad (i = 1, 2),$$

c'est-à-dire \mathbb{P}_{X_i} est l'image de \mathbb{P}_X par l'application π_i .

Démonstration. — On a $X_i = \pi_i \circ X$. Pour tout $B \in \mathcal{B}^1$ on en déduit : $\mathbb{P}_{X_i}(B) = \mathbb{P}(X_i^{-1}(B)) = \mathbb{P}(X^{-1} \circ \pi_i^{-1}(B)) = \mathbb{P}_X(\pi_i^{-1}(B))$. \square

Définition. — On appelle *fonction de répartition (conjointe)* de (la loi) de $X = (X_1, X_2)$ la fonction de deux variables réelles définie par

$$F(x_1, x_2) = P\{X_1 \leq x_1, X_2 \leq x_2\}.$$

On voit qu'elle peut s'exprimer au moyen de la loi conjointe de X par

$$F(x_1, x_2) = P_X([\!-\infty, x_1] \times [\!-\infty, x_2]).$$

Les fonctions de répartition conjointes des couples de variables aléatoires ne sont pas très utilisées, ne serait-ce qu'en raison du fait qu'il n'existe pas de relation d'ordre naturelle sur \mathbb{R}^2 ; aussi n'allons-nous pas en décrire les propriétés (certaines ont déjà été étudiés dans l'exercice 11 du chap. 5). Nous mentionnons toutefois les trois propositions suivantes.

PROPOSITION 1.2. — *La donnée de la fonction de répartition conjointe d'un couple $X = (X_1, X_2)$ équivaut à celle de la loi de X .*

PROPOSITION 1.3. — *La fonction de répartition conjointe F d'un couple $X = (X_1, X_2)$ permet de déterminer les fonctions de répartition F_1, F_2 des variables aléatoires marginales X_1, X_2 (à savoir les fonctions de répartition marginales) comme suit :*

$$\begin{aligned} F_1(x_1) &= P\{X_1 \leq x_1\} = \lim_{x_2 \rightarrow +\infty} F(x_1, x_2) = F(x_1, +\infty); \\ F_2(x_2) &= P\{X_2 \leq x_2\} = \lim_{x_1 \rightarrow +\infty} F(x_1, x_2) = F(+\infty, x_2). \end{aligned}$$

PROPOSITION 1.4. — *Les variables aléatoires marginales X_1, X_2 sont indépendantes, si et seulement si la fonction de répartition conjointe du couple (X_1, X_2) est égale au produit des fonctions de répartition marginales.*

Le théorème du transfert a déjà été énoncé au chap. 11, Théorème 1.1, pour les variables aléatoires réelles. Nous nous contentons d'en donner un nouvel énoncé pour les vecteurs aléatoires à deux dimensions. Sa démonstration est complètement analogue.

THÉORÈME 1.5 (théorème du transfert). — *Soient un vecteur aléatoire $X : (\Omega, \mathfrak{A}, P) \rightarrow (\mathbb{R}^2, \mathcal{B}^2)$ et P_X sa loi. Soit, d'autre part, $g : (\mathbb{R}^2, \mathcal{B}^2) \rightarrow (\mathbb{R}, \mathcal{B}^1)$ une fonction mesurable. Alors $g \circ X$ est une variable aléatoire. On a, de plus, l'identité*

$$\int_{\Omega} (g \circ X) dP = \int_{\mathbb{R}^2} g dP_X \left(= \int_{\mathbb{R}^2} g(x_1, x_2) dP_X(x_1, x_2) \right),$$

pourvu que l'un de ses deux membres soit défini en tant qu'intégrale abstraite, c'est-à-dire soit absolument convergent. Si c'est le cas, la valeur commune de ses deux membres est notée $\mathbb{E}[g \circ X]$ ou $\mathbb{E}[g(X_1, X_2)]$.

Notons que le membre de droite de cette identité est une intégrale sur $(\mathbb{R}^2, \mathcal{B}^2, P_X)$ et peut donc être calculé à partir de la loi P_X de X .

2. Loi de probabilité absolument continue, densité de probabilité

Comme pour les mesures de probabilité sur la droite, il existe une classe importante de mesures de probabilité sur $(\mathbb{R}^2, \mathcal{B}^2)$, que l'on peut définir à partir de fonctions réelles positives, de deux variables réelles, intégrables par rapport à la mesure de Lebesgue sur ce dernier espace (*cf.* chap. 10, § 5). Ce sont les mesures de probabilité *absolument continues*. Leur définition formelle est donnée ci-après en utilisant la terminologie des couples de variables aléatoires. Par commodité, nous désignerons désormais par (X, Y) un couple de variables aléatoires, au lieu de la notation (X_1, X_2) utilisée plus haut.

Définition. — Soit (X, Y) un couple de variables aléatoires et $P_{X,Y}$ sa loi conjointe. On dit que la loi $P_{X,Y}$ est *absolument continue* (par rapport à la mesure de Lebesgue sur $(\mathbb{R}^2, \mathcal{B}^2)$), s'il existe une fonction mesurable à valeurs positives $f : (\mathbb{R}^2, \mathcal{B}^2) \rightarrow (\mathbb{R}^+, \mathcal{B}^1)$ telle que pour tout $B \in \mathcal{B}^2$ on ait :

$$(2.1) \quad P_{X,Y}(B) = \int_B f(x, y) dx dy = \int_{\mathbb{R}^2} f(x, y) I_B(x, y) dx dy.$$

La fonction f est appelée *densité de probabilité conjointe* de (la loi) de (X, Y) . On la note aussi $f_{X,Y}(x, y)$.

PROPOSITION 2.1. — *Toute densité de probabilité conjointe f vérifie*

- a) $f \geq 0$;
- b) $\int_{\mathbb{R}^2} f(x, y) dx dy = 1$;
- c) *la fonction de répartition conjointe $F = F_{X,Y}$ peut être représentée par $F(x, y) = \int_{]-\infty, x] \times]-\infty, y]} f(u, v) du dv$;*
- d) *si f est continue au point (x_0, y_0) , alors $f(x_0, y_0) = \frac{\partial^2}{\partial x \partial y} F(x, y)$ au point $(x, y) = (x_0, y_0)$.*

PROPOSITION 2.2. — *Si le couple (X, Y) est absolument continu, alors ses variables aléatoires marginales sont absolument continues et sa densité conjointe $f(x, y) = f_{X,Y}(x, y)$ détermine les densités marginales $f_X(x)$, $f_Y(y)$ au moyen des formules :*

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y) dy, \quad f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y) dx.$$

PROPOSITION 2.3. — *Soit (X, Y) un couple de variables aléatoires, $f(x, y)$ sa densité conjointe et $f_X(x)$, $f_Y(y)$ les densités marginales de X , Y , respectivement. Alors les deux propriétés suivantes sont équivalentes :*

- a) *X et Y sont indépendantes ;*
- b) *pour (Lebesgue)-presque tout $(x, y) \in \mathbb{R}^2$, on a*

$$f_{X,Y}(x, y) = f_X(x)f_Y(y).$$

Démonstration

a) \Rightarrow b). Supposons (X, Y) indépendant. Alors, en désignant par $F = F_{X,Y}$ sa fonction de répartition conjointe et par F_X, F_Y ses fonctions de répartition marginales, il vient, pour tout $(x, y) \in \mathbb{R}^2$

$$F(x, y) = F_X(x)F_Y(y).$$

En prenant la dérivée mixte $\partial^2/\partial x \partial y$ des deux côtés (elle existe Lebesgue-presque partout), on obtient b).

b) \Rightarrow a). Supposons b) réalisée; alors, pour tout $(x, y) \in \mathbb{R}^2$, on a :

$$\begin{aligned} F(x, y) &= \int_{]-\infty, x] \times]-\infty, y]} f_X(u) f_Y(v) du dv \\ &= \int_{]-\infty, x]} f_X(u) du \int_{]-\infty, y]} f_Y(v) dv = F_X(x)F_Y(y). \end{aligned}$$

Il en résulte que X et Y sont indépendantes. \square

PROPOSITION 2.4. — Soit (X, Y) un couple de variables aléatoires absolument continu, de densité conjointe f et soit $g : (\mathbb{R}^2, \mathcal{B}^2) \rightarrow (\mathbb{R}, \mathcal{B}^1)$ une fonction mesurable. Alors $g \circ (X, Y)$ est une variable aléatoire dont l'espérance mathématique est donnée par

$$\mathbb{E}[g \circ (X, Y)] = \int_{\mathbb{R}^2} g(x, y) f(x, y) dx dy,$$

pourvu que l'intégrale au second membre soit absolument convergente.

3. Loi de probabilité conditionnelle, espérance mathématique conditionnelle, régression. — Quelle que soit la manière d'aborder la question, la définition de l'espérance mathématique conditionnelle reste un sujet difficile, si l'on veut en donner une définition avec toute la rigueur souhaitée. Nous allons traiter successivement les deux cas usuels, celui où le couple de variables aléatoires (X, Y) est discret et celui où (X, Y) est *absolument continu*. On peut évidemment introduire un formalisme qui permet d'englober ces deux cas — c'est ce que nous proposons dans l'exercice 1. De toute façon, il faut dans chaque cas trouver la formule explicite de la loi de probabilité conditionnelle ou de l'espérance conditionnelle.

(A) *Cas où le couple est discret.* — Supposons que (X, Y) soit un couple de variables aléatoires discrètes, prenant les valeurs (x_i, y_j) , où i (resp. j) varie dans un ensemble fini ou dénombrable I (resp. J). Posons $P\{X = x_i, Y = y_j\} = p_{ij}$, $P\{X = x_i\} = p_i$, $P\{Y = y_j\} = p_j$. On suppose que les x_i (resp. y_j) sont tous distincts et les probabilités p_i (resp. p_j) toutes strictement positives. Pour $i \in I$ fixé et tout $j \in J$ posons :

$$(3.1) \quad b_i(j) = P\{Y = y_j | X = x_i\} = \frac{P\{X = x_i, Y = y_j\}}{P\{X = x_i\}} = \frac{p_{ij}}{p_i}.$$

La mesure discrète $\sum_{j \in J} b_i(j) \varepsilon_{y_j}$, portée par les y_j ($j \in J$), est une mesure de probabilité. Il est naturel de l'appeler *loi de probabilité de Y conditionnelle à $\{X = x_i\}$* .

Supposons que Y ait une espérance mathématique finie, de sorte que la série de terme général $p_j y_j$ ($j \in J$) est absolument convergente. Par suite,

pour i fixé, la série de terme général $b_i(j)y_j$ ($j \in J$) est aussi absolument convergente. Il est naturel d'appeler la somme de cette série l'*espérance mathématique de Y conditionnelle à $\{X = x_i\}$* et de la noter $\mathbb{E}[Y | X = x_i]$. On pose donc

$$\mathbb{E}[Y | X = x_i] = \sum_{j \in J} b_i(j) y_j.$$

Si à chaque valeur $\mathbb{E}[Y | X = x_i]$ on associe la probabilité p_i ($i \in I$), on définit la loi de probabilité d'une certaine variable aléatoire, qu'on note $\mathbb{E}[Y | X]$. On l'appelle *espérance mathématique de Y en X* . Il faut bien noter que $\mathbb{E}[Y | X]$ est une variable aléatoire. Calculons son espérance mathématique. On a :

$$\begin{aligned} \mathbb{E}[\mathbb{E}[Y | X]] &= \sum_{i \in I} p_i \mathbb{E}[Y | X = x_i] = \sum_{i \in I} p_i \sum_{j \in J} b_i(j) y_j \\ &= \sum_{i \in I} p_i \sum_{j \in J} \frac{p_{ij}}{p_i} y_j = \sum_{j \in J} p_{.j} y_j = \mathbb{E}[Y]. \end{aligned}$$

On retrouvera cette formule dans le Théorème 3.3. De même, si on pose $a_j(i) = P\{X = x_i | Y = y_j\} = p_{ij}/p_{.j}$, on peut définir la *loi de X conditionnelle à $\{y = y_j\}$* comme étant $\sum_{i \in I} a_j(i) \varepsilon_{x_i}$, puis, si X est d'espérance mathématique finie, l'*espérance de X conditionnelle à $\{Y = y_j\}$* , à savoir $\mathbb{E}[X | Y = y_j] = \sum_{i \in I} a_j(i) x_i$, enfin $\mathbb{E}[X | Y]$ comme étant une variable aléatoire prenant la valeur $\mathbb{E}[X | Y = y_j]$ avec la probabilité $p_{.j}$ pour tout $j \in J$.

(B) *Cas où le couple est absolument continu.* — Il s'agit de trouver une formule analogue à la formule (3.1). La difficulté ici vient du fait que pour tout x réel on a $P\{X = x\} = 0$. Désignons par $f(x, y) = f_{X,Y}(x, y)$ la densité conjointe et par $f_X(x)$, $f_Y(y)$ ses densités marginales. Nous allons voir qu'en remplaçant les quantités $P\{X = x, Y = y\}$ et $P\{X = x\}$ par $f_{X,Y}(x, y)$ et $f_X(x)$, respectivement, on peut définir la loi de probabilité et l'espérance mathématiques conditionnelles de façon satisfaisante.

Définition. — Soit (X, Y) un couple de variables aléatoires absolument continu, soit $f_{X,Y}(x, y)$ sa densité conjointe et $f_X(x)$, $f_Y(y)$ ses densités marginales. Soient, de plus, $g_0(y)$ et $h_0(x)$ deux densités *arbitraires* sur $(\mathbb{R}, \mathcal{B}^1)$.

On appelle *densité de Y conditionnelle à $\{X = x\}$* , la fonction $f_{Y|X}(\cdot | x)$ de y définie par

$$f_{Y|X}(y | x) = \begin{cases} \frac{f_{X,Y}(x, y)}{f_X(x)}, & \text{si } x \text{ est tel que } f_X(x) > 0; \\ g_0(y), & \text{sinon.} \end{cases}$$

De même, on appelle *densité de X conditionnelle à $\{Y = y\}$* , la fonction $f_{X|Y}(\cdot | y)$ de x définie par

$$f_{X|Y}(x | y) = \begin{cases} \frac{f_{X,Y}(x, y)}{f_Y(y)}, & \text{si } y \text{ est tel que } f_Y(y) > 0; \\ h_0(x), & \text{sinon.} \end{cases}$$

Remarque 1. — Il résulte de cette définition que, pour presque tout x , on a, pour presque tout y ,

$$(3.2) \quad f_{X,Y}(x, y) = f_X(x)f_{Y|X}(y|x).$$

En effet, cette égalité est vraie par définition si $f_X(x) > 0$. Supposons $f_X(x_0) = 0$, c'est-à-dire $\int_{\mathbb{R}} f_{X,Y}(x_0, y) dy = 0$, alors $f_{X,Y}(x_0, y) = 0$ pour presque tout y , d'où $f_{X,Y}(x_0, y) = f_X(x_0)f_{Y|X}(y|x_0)$ pour presque tout y . On voit de même que, pour presque tout y , on a, pour presque tout x ,

$$(3.3) \quad f_{X,Y}(x, y) = f_Y(y)f_{X|Y}(x|y).$$

Remarque 2. — Il en résulte que

$$f_X(x) = \int f_{X|Y}(x|y) f_Y(y) dy, \quad f_Y(y) = \int f_{Y|X}(y|x) f_X(x) dx,$$

c'est-à-dire chacune des densités marginales est une combinaison convexe de densités conditionnelles.

PROPOSITION 3.1. — *Pour tout x la fonction $f_{Y|X}(\cdot|x)$ possède toutes les propriétés d'une densité de probabilité. De même, pour tout y , la fonction $f_{X|Y}(\cdot|y)$ possède toutes les propriétés d'une densité de probabilité.*

Les démonstrations sont immédiates.

PROPOSITION 3.2. — *Soit (X, Y) un couple de variables aléatoires absolument continu. Désignons par $f_X(x)$, $f_Y(y)$, $f_{Y|X}(\cdot|x)$, $f_{X|Y}(\cdot|y)$ les densités marginales et conditionnelles. Si le couple (X, Y) est indépendant, alors*

- 1) *Pour tout x tel que $f_X(x) > 0$, on a $f_{Y|X}(y|x) = f_Y(y)$.*
- 2) *Pour tout y tel que $f_Y(y) > 0$, on a $f_{X|Y}(x|y) = f_X(x)$.*

Les démonstrations sont immédiates.

Définition (Espérance mathématique conditionnelle). — Soit (X, Y) un couple de variables aléatoires absolument continu. On conserve toutes les notations précédentes pour les densités marginales et conditionnelles. En particulier, $f_{Y|X}(\cdot|x)$ désigne la densité de Y conditionnelle à $\{X = x\}$. Pour tout x réel, l'intégrale $\int_{\mathbb{R}} y f_{Y|X}(y|x) dy$, si elle est absolument convergente, peut s'interpréter comme l'espérance mathématique de Y par rapport à la loi de probabilité de densité $f_{Y|X}(\cdot|x)$. Si donc l'intégrale précédente est absolument convergente, on pose

$$(3.4) \quad \mathbb{E}[Y | X = x] = \int_{\mathbb{R}} y f_{Y|X}(y|x) dy$$

et on appelle cette intégrale l'*espérance mathématique de Y conditionnelle à $\{X = x\}$* .

Maintenant l'application $e : x \mapsto e(x) = \mathbb{E}[Y | X = x]$ est une fonction réelle d'une variable réelle. La fonction composée $e \circ X$ est une variable aléatoire réelle (définie sur $(\Omega, \mathfrak{A}, \mathbb{P})$). On la note $\mathbb{E}[Y | X]$ et on l'appelle l'*espérance mathématique conditionnelle de Y en X*. On définit, de façon analogue, l'espérance mathématique conditionnelle de X en Y .

Dans le théorème suivant, on est amené à considérer l'espérance mathématique de la variable aléatoire réelle $\mathbb{E}[Y | X]$. On notera que cette espérance mathématique est calculée non sur l'espace $(\Omega, \mathfrak{A}, \mathbb{P})$, mais sur l'espace $(\mathbb{R}, \mathcal{B}^1, \mathbb{P}_X)$, par rapport à la loi \mathbb{P}_X de X . Toutefois, dans des exposés plus avancés, la notion d'espérance mathématique conditionnelle est définie de façon naturelle sur l'espace $(\Omega, \mathfrak{A}, \mathbb{P})$.

THÉORÈME 3.3 (Théorème de l'espérance mathématique conditionnelle)
Soit (X, Y) un couple de variables aléatoires. On suppose $\mathbb{E}[|Y|] < +\infty$. Alors on a l'identité

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y | X]].$$

Démonstration. — On a formellement :

$$\begin{aligned} \mathbb{E}[Y] &= \int_{\mathbb{R}^2} y f_{X,Y}(x, y) dx dy = \int_{\mathbb{R}^2} y f_X(x) f_{Y|X}(y|x) dx dy \\ &= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} y f_{Y|X}(y|x) dy \right] f_X(x) dx = \int_{\mathbb{R}} \mathbb{E}[Y | X = x] f_X(x) dx \\ &= \mathbb{E}[\mathbb{E}[Y | X]]. \end{aligned}$$

De par l'hypothèse $\mathbb{E}[|Y|] < +\infty$, ces calculs formels sont licites. \square

Définition (Courbes de régression). — Soit (X, Y) un couple de variables aléatoires vérifiant $\mathbb{E}[|X|] < +\infty$, $\mathbb{E}[|Y|] < +\infty$. Le graphe de $x \mapsto \mathbb{E}[Y | X = x]$ est appelé la *courbe de régression de Y en X*. Le graphe de $y \mapsto \mathbb{E}[X | Y = y]$ est appelé la *courbe de régression de X en Y*.

Remarque. — Ces deux courbes sont en général distinctes. Si, par exemple, X et Y sont indépendantes, le graphe de $x \mapsto \mathbb{E}[Y | X = x]$ est une droite parallèle à $0x$ et celui de $y \mapsto \mathbb{E}[X | Y = y]$ est une droite parallèle à $0y$.

Les courbes de régression vérifient des propriétés de minimalité, utilisées en statistique, comme le montre le théorème suivant.

THÉORÈME 3.4. — Soit (X, Y) un couple de variables aléatoires vérifiant $\mathbb{E}[Y^2] < \infty$. La courbe de régression de Y en X vérifie la propriété de minimalité suivante : soit u une fonction réelle mesurable ; on suppose que l'expression

$$(3.5) \quad \mathbb{E}[(Y - u(X))^2]$$

est finie. Lorsque la fonction mesurable u varie, il en est de même de l'expression (3.5) et celle-ci passe par un minimum atteint pour la fonction $u(x) = \mathbb{E}[Y | X = x]$. La valeur de ce minimum est $\mathbb{E}[(Y - \mathbb{E}[Y | X])^2]$.

Démonstration. — Supposons toujours le couple (X, Y) absolument continu et utilisons les notations ci-dessus

$$\begin{aligned}\mathbb{E}[[Y - u(X)]^2] &= \int_{\mathbb{R}^2} [y - u(x)]^2 f_{X,Y}(x, y) dx dy \\ &= \int_{\mathbb{R}} f_X(x) \left[\int_{\mathbb{R}} [y - u(x)]^2 f_{Y|X}(y|x) dy \right] dx.\end{aligned}$$

Pour tout nombre réel x fixé, l'intégrale entre crochets au dernier membre est minimum pour $u(x) = \int_{\mathbb{R}} y f_{Y|X}(y|x) dy = \mathbb{E}[Y | X = x]$, d'après la Proposition 5.3 du chap. 8, énoncée pour les variables aléatoires discrètes, mais en fait valable pour les variables aléatoires quelconques. \square

4. Règles de calcul concernant les espérances conditionnelles

Nous regroupons dans ce paragraphe quelques règles de calcul sur les espérances conditionnelles. Les notations précédentes concernant toutes les espérances (conditionnelles ou non) de X et de Y sont conservées. On écrit indifféremment $\mathbb{E}[Y | X]$ ou $\mathbb{E}_X[Y]$. Les symboles g, h (avec ou sans indices) désignent des fonctions réelles mesurables dont les arguments sont évidents d'après l'écriture. Nous supposons enfin que toutes les espérances qui interviennent existent.

Soit tout d'abord h une fonction réelle mesurable de deux variables réelles. Le produit de composition $h \circ (X, Y)$ est une variable aléatoire réelle. On définit son *espérance mathématique conditionnelle* à $\{X = x\}$ par :

$$(4.1) \quad \mathbb{E}[h \circ (X, Y) | X = x] = \int_{\mathbb{R}} h(x, y) f_{Y|X}(y|x) dy.$$

Lorsque $h(x, y) = y$, on retrouve bien la définition (3.4).

THÉORÈME 4.1

1) *On a les identités :*

$$\mathbb{E}[\mathbb{E}[h \circ (X, Y) | X]] = \int \mathbb{E}[h \circ (X, Y) | X = x] f_X(x) dx = \mathbb{E}[h \circ (X, Y)].$$

En prenant, en particulier, $h(x, y) = g(y)$, on obtient la formule de l'espérance conditionnelle du Théorème 3.3 pour la variable $g \circ Y$:

$$\mathbb{E}[\mathbb{E}[g \circ Y | X]] = \int \mathbb{E}[g \circ Y | X = x] f_X(x) dx = \mathbb{E}[g \circ Y].$$

2) *Si X, Y sont indépendantes, alors $\mathbb{E}[g \circ Y | X] = \mathbb{E}[g \circ Y]$.*

3) *On a toujours : $\mathbb{E}[g \circ X | X] = g \circ X$.*

4) *Pour X et Y quelconques, on a*

$$\begin{aligned}\mathbb{E}_X[\mathbb{E}_X[Y]] &= \mathbb{E}_X[Y]; \\ \mathbb{E}[(g_1 \circ X)(g_2 \circ Y) | X] &= (g_1 \circ X) \mathbb{E}[g_2 \circ Y | X].\end{aligned}$$

Ainsi, dans le calcul de l'espérance mathématique conditionnelle par rapport à X , la fonction $g_1 \circ X$ se comporte comme une constante.

Démonstration. — La démonstration de 1) est tout à fait similaire à la démonstration du Théorème 3.3, en faisant usage de (4.1). Si l'on pose $e(x) = \mathbb{E}[h \circ (X, Y) | X = x]$, l'espérance mathématique conditionnelle de $h \circ (X, Y)$ en X , notée $\mathbb{E}[h \circ (X, Y) | X]$, est la fonction composée $e \circ \pi_1 \circ (X, Y) = e \circ X$. C'est encore une variable aléatoire définie sur $(\Omega, \mathfrak{A}, \mathbb{P})$. La première identité de 1) montre que l'espérance mathématique de cette variable aléatoire est calculée sur $(\mathbb{R}, \mathcal{B}^1, \mathbb{P}_X)$:

$$\begin{aligned} \mathbb{E}[\mathbb{E}[h \circ (X, Y) | X]] &= \mathbb{E}[e \circ X] = \int_{\mathbb{R}} e(x) f_X(x) dx \\ &= \int_{\mathbb{R}} \mathbb{E}[h \circ (X, Y) | X = x] f_X(x) dx. \\ &= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} h(x, y) f_{Y|X}(y|x) dy \right] f_X(x) dx \\ &= \int_{\mathbb{R}^2} h(x, y) f_{X,Y}(x, y) dx dy = \mathbb{E}[h \circ (X, Y)]. \end{aligned}$$

2) Si X et Y sont indépendantes, on a :

$$\mathbb{E}[g \circ Y | X = x] = \int_{\mathbb{R}} g(y) f_{Y|X}(y|x) dy = \int_{\mathbb{R}} g(y) f_Y(y) dy = \mathbb{E}[g \circ Y],$$

d'après la Proposition 3.2 et le fait qu'on peut calculer $\mathbb{E}[g \circ Y]$ sur l'espace probabilisé $(\mathbb{R}, \mathcal{B}^1, \mathbb{P}_Y)$. La quantité $e(x) = \mathbb{E}[g \circ Y | X = x]$ est alors une *constante* égale à $\mathbb{E}[g \circ Y]$. Il en est de même de $\mathbb{E}[g \circ Y | X]$ égale à $e \circ Y$. Ceci prouve la formule 2).

3) On note que pour tout x on a :

$$\mathbb{E}[g \circ X | X = x] = \int_{\mathbb{R}} g(x) f_Y(y) dy = g(x) \int_{\mathbb{R}} f_Y(y) dy = g(x),$$

qui vaut aussi $e(x)$ dans les notations précédentes. D'où $g = e$ et $\mathbb{E}[g \circ X | X] = e \circ X = g \circ X$.

4) La première identité résulte de 3) en prenant $g \circ X = \mathbb{E}[Y | X]$. Pour prouver la seconde identité, il est commode de poser $e_2(x) = \mathbb{E}[g_2 \circ Y | X = x]$, de sorte que $\mathbb{E}[g_2 \circ Y | X] = e_2 \circ X$ et $e(x) = \mathbb{E}[(g_1 \circ X)(g_2 \circ Y) | X = x]$, de sorte que $\mathbb{E}[(g_1 \circ X)(g_2 \circ Y) | X] = e \circ X$. On a :

$$\begin{aligned} e(x) &= \int_{\mathbb{R}} g_1(x) g_2(y) f_{Y|X}(y|x) dy = g_1(x) \int_{\mathbb{R}} g_2(y) f_{Y|X}(y|x)(y) dy \\ &= g_1(x) \mathbb{E}[g_2 \circ Y | X = x] = g_1(x) e_2(x). \end{aligned}$$

D'où $e \circ X = (g_1 \circ X)(e_2 \circ X)$. \square

Supposons que A soit un évènement dont l'indicatrice I_A s'exprime comme une fonction mesurable $h \circ (X, Y)$ du couple (X, Y) . Ainsi $A = \{X < Y\}$

est un tel évènement, puisqu'on peut écrire : $I_{\{X < Y\}} = h \circ (X, Y)$ avec $h(x, y) = I_{\{x < y\}}(x, y)$. On peut alors utiliser les formules précédentes pour une telle indicatrice. Se rappelant que l'on a $P(A) = \mathbb{E}[I_A]$, on trouve encore

$$P(A) = \mathbb{E}[I_A] = \mathbb{E}[\mathbb{E}[I_A | X]].$$

D'où, en posant les définitions

$$(4.2) \quad P\{A | X = x\} = \mathbb{E}[I_A | X = x],$$

$$(4.3) \quad P\{A | X\} = \mathbb{E}[I_A | X],$$

on trouve l'identité

$$(4.4) \quad P(A) = \int_{\mathbb{R}} P\{A | X = x\} f_X(x) dx.$$

La fonction $P\{A | X = x\}$ est la *probabilité de A conditionnelle à $\{X = x\}$* . La formule (4.4) est fréquemment utilisée dans l'évaluation de certaines probabilités, une fois que l'on sait calculer $P\{A | X = x\}$.

Exemple 1. — Soit (X, Y) un couple de variables aléatoires absolument continues. Évaluons d'abord la formule $P\{X < Y | X = x\}$, en utilisant les fonctions $h(x, y) = I_{\{x < y\}}(x, y)$ et $g(y) = I_{\{x < y\}}(y)$. On a :

$$\begin{aligned} P\{X < Y | X = x\} &= \mathbb{E}[I_{\{X < Y\}} | X = x] = \mathbb{E}[h \circ (X, Y) | X = x] \\ &= \int_{\mathbb{R}} h(x, y) f_{Y|X}(y|x) dy = \int_{\mathbb{R}} I_{\{x < y\}}(x, y) f_{Y|X}(y|x) dy \\ &= \int_{\mathbb{R}} I_{\{x < y\}}(y) f_{Y|X}(y|x) dy = \int_{\mathbb{R}} g(y) f_{Y|X}(y|x) dy \\ &= \mathbb{E}[g \circ Y | X = x] = \mathbb{E}[I_{\{x < Y\}} | X = x] = P\{x < Y | X = x\}. \end{aligned}$$

Si les variables X et Y sont indépendantes, on a $f_{Y|X}(y|x)(y) = f_Y(y)$, de sorte que

$$\begin{aligned} P\{X < Y | X = x\} &= P\{x < Y | X = x\} = \int_{\mathbb{R}} I_{\{x < y\}}(y) f_{Y|X}(y|x) dy \\ &= \int_{\mathbb{R}} I_{\{x < y\}}(y) f_Y(y) dy = P\{x < Y\}. \end{aligned}$$

Exemple 2. — Si X et Y sont deux variables aléatoires réelles, indépendantes, suivant chacune une loi exponentielle de paramètre λ et μ , respectivement, alors $P\{X < Y\} = \lambda/(\lambda + \mu)$.

En effet, la densité de X est donnée par $f_X(x) = \lambda e^{-\lambda x} I_{\{x \geq 0\}}$. D'après (4.3) et l'Exemple 1, on a :

$$\begin{aligned} P\{X < Y\} &= \int_0^{\infty} P\{X < Y | X = x\} f_X(x) dx \\ &= \int_0^{\infty} P\{Y > x | X = x\} f_X(x) dx = \int_0^{\infty} P\{Y > x\} f_X(x) dx \\ &= \int_0^{\infty} e^{-\mu x} \lambda e^{-\lambda x} dx = \lambda \int_0^{\infty} e^{-(\lambda + \mu)x} dx = \frac{\lambda}{\lambda + \mu}. \end{aligned}$$

5. La loi normale à deux dimensions. — La loi normale $\mathcal{N}(0, 1)$ sera étudiée au chap. 14, § 3. Pour traiter ce qui va suivre, on a simplement besoin de savoir que cette loi admet la densité $(1/\sqrt{2\pi})e^{-x^2/2}$ sur toute la droite. Si donc $M' = \begin{pmatrix} X'_1 \\ X'_2 \end{pmatrix}$ est un couple de variables aléatoires *indépendantes* de loi normale $\mathcal{N}(0, 1)$, il est absolument continu, d'après la Proposition 2.3 et sa densité est égale au produit des densités de X'_1 et de X'_2 . Par ailleurs, la fonction génératrice d'un couple de variables aléatoires sera étudiée au chap. 13, § 6, comme étant la fonction $\mathbb{E}[e^{u'_1 X'_1 + u'_2 X'_2}]$ des deux variables réelles u'_1 et u'_2 . En posant $u' = \begin{pmatrix} u'_1 \\ u'_2 \end{pmatrix}$ et $x' = \begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix}$, il est ainsi immédiat que M' admet une fonction génératrice et une densité conjointe données par

$$(5.1) \quad g_{M'}(u') = \mathbb{E}[e^{t u' M'}] = \exp\left(\frac{1}{2}(u_1'^2 + u_2'^2)\right) = \exp\left(\frac{1}{2} t u' u'\right);$$

$$(5.2) \quad f_{M'}(x') = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x_1'^2 + x_2'^2)\right) = \frac{1}{2\pi} \exp\left(-\frac{1}{2} t x' x'\right).$$

Associons à M' le couple $M = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ au moyen de la transformation $M = A M'$, où A est une matrice 2×2 , réelle. En utilisant des notations matricielles qui apparaissent évidentes dans ce contexte, on vérifie que :

$$a) \quad M \text{ est centré : } \mathbb{E}[M] = A \mathbb{E}[M'] = A \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix};$$

b) la *matrice des variances-covariances* $\mathbb{E}[(M - \mathbb{E}[M])^t (M - \mathbb{E}[M])]$ de M est égale à : $\mathbb{E}[M^t M] = A \mathbb{E}[M'^t M'] A = A I A = A^t A$.

PROPOSITION 5.1. — *Le couple M admet pour fonction génératrice*

$$(5.3) \quad g_M(u) = \mathbb{E}[e^{t u M}] = \exp\left(\frac{1}{2} t u (A^t A) u\right), \quad u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}.$$

Si, en outre, A est régulière, M admet une densité conjointe

$$(5.4) \quad f_M(x) = \frac{1}{2\pi} \exp\left(-\frac{1}{2} t x (A^t A)^{-1} x\right) |\det A^{-1}|, \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Démonstration. — Pour calculer $g_M(u)$ on fait simultanément les transformations $M = A M'$ et $t u' = t u A$. On voit que $t u M = t u' M'$, d'où $g_M(u) = \mathbb{E}[e^{t u M}] = \mathbb{E}[e^{t u' M'}] = \exp\left(\frac{1}{2} t u' u'\right) = \exp\left(\frac{1}{2} t u (A^t A) u\right)$. Dans le cas où $\det A \neq 0$, on fait appel à la formule du changement de variables du Théorème 2.1 du chap. 15, appliquée à (5.2). On utilise la transformation $G : x' \mapsto x = A x'$, de sorte que la bijection inverse est $H : x \mapsto x' = A^{-1} x$ et que le jacobien de celle-ci est $\det A^{-1}$. On obtient ainsi

$$f_M(x) = \frac{1}{2\pi} \exp\left(-\frac{1}{2} t x^t (A^{-1}) A^{-1} x\right) |\det A^{-1}|,$$

d'où la formule (5.4), puisque $t(A^{-1}) = (t A)^{-1}$ et $(t A)^{-1} A^{-1} = (A^t A)^{-1}$. \square

La matrice $A^t A$ qui figure dans (5.3) et (5.4) est la matrice des variances-covariances du vecteur aléatoire $M = AM'$. En fait, nous allons voir que toute matrice des variances-covariances Γ admet une représentation de la forme $A^t A$, c'est-à-dire est la matrice des variances-covariances d'un vecteur aléatoire de la forme $M = AM'$.

LEMME 5.2. — Soit Γ une matrice des variances-covariances 2×2 . Alors il existe une matrice 2×2 , réelle, A telle que $\Gamma = A^t A$.

Démonstration. — Par définition Γ est réelle, symétrique, définie-positive (puisque ${}^t u \Gamma u = \mathbb{E}[\|{}^t u M\|^2] \geq 0$); elle admet donc des valeurs propres λ_1, λ_2 positives ou nulles. Il existe donc une matrice 2×2 , réelle, orthogonale S telle que

$$S^{-1} \Gamma S = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix} \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix}.$$

On en déduit

$$\Gamma = S \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix} \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix} S^{-1},$$

d'où $\Gamma = A^t A$, en posant $A = S \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix}$. [Noter que $S^{-1} = {}^t S$, puisque S est orthogonale.] \square

En choisissant A conforme au Lemme 5.2, le couple M défini par $M = A M'$ a une loi dont la fonction génératrice et la densité conjointe, le cas échéant, sont données par (5.3) et (5.4). On voit que cette loi ne dépend que de Γ ; on la note $\mathcal{N}_2(0, \Gamma)$; elle fait l'objet de la définition suivante.

Définition. — Donnons-nous une matrice des variances-covariances $\Gamma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, ($\sigma_1 > 0, \sigma_2 > 0, |\rho| \leq 1$). On dit qu'un couple aléatoire $M = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ suit la loi normale centrée $\mathcal{N}_2(0, \Gamma)$, s'il admet pour fonction génératrice

$$(5.5) \quad g_M(u) = \exp\left(\frac{1}{2} {}^t u \Gamma u\right) = \exp\left(\frac{1}{2} (\sigma_1^2 u_1^2 + 2\rho\sigma_1\sigma_2 u_1 u_2 + \sigma_2^2 u_2^2)\right).$$

Si en outre $\det \Gamma \neq 0$, c'est-à-dire si $|\rho| < 1$, il admet une densité conjointe donnée par

$$(5.6) \quad f_M(x) = \frac{\sqrt{\det \Gamma^{-1}}}{2\pi} \exp\left(-\frac{1}{2} {}^t x \Gamma^{-1} x\right) \\ = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1 - \rho^2}} \exp\left(-\frac{1}{2(1 - \rho^2)} \left(\frac{x_1^2}{\sigma_1^2} - 2\rho \frac{x_1 x_2}{\sigma_1 \sigma_2} + \frac{x_2^2}{\sigma_2^2}\right)\right).$$

La loi $\mathcal{N}_2(0, \Gamma)$ est dite dégénérée ou non dégénérée, selon que $\det \Gamma = 0$ ou $\neq 0$. Seules les lois non dégénérées admettent des densités conjointes. La propriété suivante n'est qu'une redite.

PROPRIÉTÉ 5.3. — Le couple $M = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ est centré et sa matrice des variances-covariances est Γ ; en d'autres termes, les variables aléatoires marginales X_1 et X_2 sont centrées et leur coefficient de corrélation linéaire est ρ .

Les deux propriétés suivantes sont évidentes sur la forme de $g_M(u)$.

PROPRIÉTÉ 5.4. — Les variables aléatoires marginales X_1, X_2 sont des variables aléatoires normales, centrées, de variances respectives σ_1^2 et σ_2^2 .

PROPRIÉTÉ 5.5. — Une condition nécessaire et suffisante pour que les variables aléatoires marginales X_1, X_2 soient indépendantes est qu'elles soient non corrélées, c'est-à-dire que $\rho = 0$. Ainsi, dans le cas particulier des lois normales à deux dimensions, les propriétés d'indépendance et de non-corrélation de X_1, X_2 sont équivalentes.

THÉORÈME 5.6. — Soient $M = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ un couple de loi $\mathcal{N}_2(0, \Gamma)$ et A une matrice 2×2 réelle. Alors $M' = AM$ est un couple de loi $\mathcal{N}_2(0, \Gamma')$, où $\Gamma' = A\Gamma^tA$.

Démonstration. — Faisons simultanément les transformations $M' = AM$, ${}^t u = {}^t u' A$. On a ${}^t u M = {}^t u' M'$, d'où $g_{M'}(u') = \mathbb{E}[e^{{}^t u' M'}] = \mathbb{E}[e^{{}^t u M}] = \exp\left(\frac{1}{2} {}^t u \Gamma u\right) = \exp\left(\frac{1}{2} {}^t u' (A\Gamma^tA) u'\right)$. \square

COROLLAIRE 1. — Supposons que $\Gamma = I$, c'est-à-dire que M soit un couple de variables aléatoires indépendantes de loi $\mathcal{N}(0, 1)$. Une condition nécessaire et suffisante pour qu'il en soit de même de $M' = AM$ est que $A^tA = I$, c'est-à-dire que A soit une matrice orthogonale.

COROLLAIRE 2. — Si le couple (X_1, X_2) suit la loi normale $\mathcal{N}_2(0, \Gamma)$, alors $\left(\frac{X_1}{\sigma_1}, \frac{X_2}{\sigma_2} - \rho \frac{X_1}{\sigma_1}\right)$ est un couple de variables aléatoires indépendantes, normales, centrées, de variances respectives $1, 1 - \rho^2$.

Démonstration. — Définissons $M' = AM$ par

$$\begin{cases} X'_1 = \frac{X_1}{\sigma_1}; \\ X'_2 = \frac{X_2}{\sigma_2} - \rho \frac{X_1}{\sigma_1}; \end{cases} \quad \text{soit } A = \begin{pmatrix} \frac{1}{\sigma_1} & 0 \\ -\frac{\rho}{\sigma_1} & \frac{1}{\sigma_2} \end{pmatrix}.$$

Alors $\Gamma' = A\Gamma^tA = \begin{pmatrix} 1 & 0 \\ 0 & 1 - \rho^2 \end{pmatrix}$. \square

Remarques sur les lois dégénérées. — Considérons la loi $\mathcal{N}_2(0, \Gamma)$ avec $\det \Gamma = 0$, c'est-à-dire $\rho = \pm 1$. Elle n'admet pas de densité conjointe, mais admet une fonction génératrice obtenue en faisant $\rho = \pm 1$ dans (5.5), soit

$$(5.7) \quad g_M(u) = \exp\left(\frac{1}{2}(\sigma_1 u_1 \pm \sigma_2 u_2)^2\right).$$

On voit que c'est la fonction génératrice d'un couple (X_1, X_2) , où $X_1 = \sigma_1 U$, $X_2 = \pm \sigma_2 U$, avec U une variable aléatoire de loi $\mathcal{N}(0, 1)$. Ainsi une loi $\mathcal{N}_2(0, \Gamma)$ dégénérée admet pour support une droite de pente $\pm \sigma_2 / \sigma_1$ et passant par l'origine.

Remarque sur les lois non centrées. — Soient $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ ($\mu_1, \mu_2 \in \mathbb{R}$) et M' un point aléatoire de loi $\mathcal{N}_2(0, \Gamma)$. On note $\mathcal{N}_2(\mu, \Gamma)$ la loi de $M = M' + \mu$ et on l'appelle *loi normale à deux dimensions de point moyen μ et de matrice de variances-covariances Γ* . Sa fonction génératrice est donnée par

$$(5.8) \quad g(u) = \exp\left({}^t u \mu + \frac{1}{2} {}^t u \Gamma u\right) \\ = \exp\left(u_1 \mu_1 + u_2 \mu_2 + \frac{1}{2} (\sigma_1^2 u_1^2 + 2\rho \sigma_1 \sigma_2 u_1 u_2 + \sigma_2^2 u_2^2)\right).$$

Si en outre $|\rho| < 1$, elle admet une densité conjointe donnée par

$$(5.9) \quad f_M(x) = \frac{\sqrt{\det \Gamma^{-1}}}{2\pi} \exp\left(-\frac{1}{2} {}^t (x - \mu) \Gamma^{-1} (x - \mu)\right) \\ = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1 - \rho^2}} \exp\left(-\frac{1}{2(1 - \rho^2)} \left(\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 - 2\rho \left(\frac{x_1 - \mu_1}{\sigma_1}\right) \left(\frac{x_2 - \mu_2}{\sigma_2}\right) + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2\right)\right).$$

On voit que si $M = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ suit la loi $\mathcal{N}_2(\mu, \Gamma)$, alors le couple $\begin{pmatrix} U \\ V \end{pmatrix}$ avec $U = (X_1 - \mu_1) / \sigma_1$, $V = (X_2 - \mu_2) / \sigma_2$ suit la loi $\mathcal{N}_2(0, \gamma)$ avec $\gamma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$.

Le prochain théorème présente un double intérêt, car, d'une part, il fournit une manière alternative pour introduire la loi normale à deux dimensions, d'autre part, il permet, convenablement généralisé, de définir des lois normales dans des espaces plus généraux, comme par exemple les espaces de Banach.

THÉORÈME 5.7. — *Soit $M = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ un vecteur aléatoire centré. Les deux propriétés suivantes sont équivalentes :*

- a) *M suit une loi normale centrée à deux dimensions ;*
- b) *toute combinaison linéaire de X_1, X_2 suit une loi normale centrée à une dimension.*

Démonstration. — Notons $L_u = {}^t u M$, $u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$, une combinaison linéaire à coefficients réels de X_1, X_2 .

a) \Rightarrow b) : supposons $\mathcal{L}(M) = \mathcal{N}_2(0, \Gamma)$. La fonction génératrice de L_u est donnée, pour $v \in \mathbb{R}$, par $g(v) = \mathbb{E}[e^{v L_u}] = \mathbb{E}[e^{({}^t (v u) M)}] = \exp\left(\frac{1}{2} ({}^t u \Gamma u) v^2\right)$. Ceci montre, d'après chap. 14, § 3.2 c), que L_u est une variable aléatoire normale, centrée, de variance ${}^t u \Gamma u$.

b) \Rightarrow a) : par hypothèse, pour tout u dans \mathbb{R}^2 , la variable aléatoire L_u suit une loi normale centrée à une dimension ; elle admet donc une

fonction génératrice donnée, pour v réel, par $g(v) = \mathbb{E}[e^{vL_u}] = e^{Q(u)(v^2/2)}$, où $Q(u) = \text{Var } L_u = \text{Var}({}^t u M) = \mathbb{E}[({}^t u M)^2] = \mathbb{E}[({}^t u M)({}^t u M)]$. Comme ${}^t u M = {}^t M u$, on en déduit $Q(u) = {}^t u \mathbb{E}[M {}^t M] u = {}^t u \Gamma u$, où Γ est la matrice des variances-covariances de M .

En prenant $v = 1$, il vient $g(1) = \mathbb{E}[e^{L_u}] = \mathbb{E}[e^{{}^t u M}] = e^{(1/2) {}^t u \Gamma u}$, qui est l'expression pour la fonction génératrice de M . On voit donc que M suit la loi $\mathcal{N}_2(0, \Gamma)$. \square

Les notations utilisées étant matricielles, seules des modifications mineures sont à apporter à l'étude de vecteurs aléatoires normaux M à un nombre de dimension $n \geq 2$. En particulier, lorsque la matrice des variances-covariances Γ est régulière, la densité de M est donnée par la formule (5.9), lorsqu'on y remplace 2π par $(2\pi)^{n/2}$ au dénominateur de la fraction.

COMPLÉMENTS ET EXERCICES

1. Autre traitement des lois conditionnelles (X. Fernique). — Cette méthode permet de traiter *simultanément* le cas des variables aléatoires discrètes et celui des variables aléatoires absolument continues. Considérons un couple de variables aléatoires réelles (X, Y) défini sur un espace probabilisé $(\Omega, \mathfrak{A}, P)$. Notons μ, P_X, P_Y , la loi du couple, celle de X , celle de Y , respectivement.

Soit $(Q_y(A)) (y \in \mathbb{R}, A \in \mathcal{B}^1)$ une famille de nombres réels indexée par le couple $(y, A) \in \mathbb{R} \times \mathcal{B}^1$. On dit que cette famille est la *loi conditionnelle de X relativement à Y*, si les propriétés suivantes sont remplies :

- (1) pour tout nombre réel y l'application $Q_y : A \mapsto Q_y(A)$ est une *loi de probabilité sur \mathbb{R}* , donc, en particulier, une application de \mathcal{B}^1 dans $[0, 1]$;
- (2) pour tout borélien $A \in \mathcal{B}^1$ l'application $Q_{(\cdot)}(A) : y \mapsto Q_y(A)$ est *mesurable*, donc, en particulier, une application de $(\mathbb{R}, \mathcal{B}^1)$ dans $(\mathbb{R}, \mathcal{B}^1)$;
- (3) pour tout couple de boréliens A, B on a l'identité :

$$P\{X \in A, Y \in B\} = \mathbb{E}[Q_{(\cdot)}(A) \cdot I_{\{Y \in B\}}].$$

Noter que $Q_{(\cdot)}(A)$ est une variable aléatoire réelle définie sur $(\mathbb{R}, \mathcal{B}^1)$ et qu'elle est bornée, puisque $0 \leq Q_y(A) \leq 1$ pour tout y . Le produit $Q_{(\cdot)}(A) \cdot I_{\{Y \in B\}}$ est donc intégrable. L'espérance mathématique « \mathbb{E} » dans l'identité précédente est prise par rapport à la loi P_Y de Y .

a) Montrer que l'identité en (3) se réécrit :

$$\int_{x \in A, y \in B} d\mu(x, y) = \int_{y \in B} \left(\int_{x \in A} dQ_y(x) \right) dP_Y(y).$$

b) Supposons que Y soit une variable aléatoire discrète, de loi $P_Y = \sum_{j \in J} P\{Y = y_j\} \varepsilon_{y_j}$, où J est fini ou dénombrable, en supposant les y_j tous

distincts et les probabilités $P\{Y = y_j\}$ toutes *strictement positives*. Soit Q_0 une loi de probabilité arbitraire sur $(\mathbb{R}, \mathcal{B}^1)$ et définissons

$$Q_y(\cdot) = \begin{cases} Q_0, & \text{si } y \neq y_j \text{ pour tout } j; \\ P\{X \in \cdot | Y = y_j\}, & \text{si } y = y_j. \end{cases}$$

En particulier, pour $y = y_j$ et tout ensemble borélien A , on pose

$$Q_y(A) = P\{X \in A | Y = y_j\}.$$

Vérifier que la fonction $Q_{(\cdot)}$ satisfait les propriétés (1), (2) et (3).

c) Supposons le couple (X, Y) est absolument continu et reprenons les mêmes notations que dans les paragraphes 3 et 4. Pour tout nombre réel y notons Q_y la loi de probabilité sur \mathbb{R} , de densité $f_{X|Y}(x|y)$. Vérifier les trois propriétés (1), (2) et (3).

2. — Soit (X, Y) un couple de variables aléatoires de densité conjointe

$$f(x, y) = \begin{cases} e^{-(x+y)}, & \text{si } x, y \geq 0; \\ 0, & \text{sinon.} \end{cases}$$

- a) Déterminer les densités marginales de X, Y .
- b) Les variables X et Y sont-elles indépendantes ?

3. — Mêmes questions qu'en 2) pour un couple (X, Y) de densité conjointe :

$$f(x, y) = \begin{cases} 2e^{-(x+y)}, & \text{si } 0 \leq x \leq y; \\ 0, & \text{sinon.} \end{cases}$$

4. — Soit (X, Y) un couple de variables aléatoires de densité conjointe $f(x, y)$. Montrer que X, Y sont indépendantes, si et seulement si f se factorise en le produit $f(x, y) = g(x)h(y)$, d'une fonction de x et d'une fonction de y .

5. — Soit D le disque de centre 0 et de rayon $r > 0$

$$D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq r^2\}.$$

On désigne par (X, Y) un point aléatoire à valeurs dans D et dont la loi conjointe est la loi uniforme sur D , c'est-à-dire est donnée par :

$$f(x, y) = \begin{cases} \frac{1}{\pi r^2}, & \text{si } (x, y) \in D; \\ 0, & \text{sinon.} \end{cases}$$

- a) Déterminer les densités marginales de X, Y . Calculer $\mathbb{E}[X], \mathbb{E}[Y]$.
- b) Les variables X et Y sont-elles indépendantes ?
- c) Calculer $\text{Cov}(X, Y)$. Quelle conclusion peut-on tirer de b), c) ?

d) Déterminer la fonction de répartition $G(u)$ puis la densité de probabilité $g(u)$ de la variable aléatoire $U = X^2 + Y^2$.

e) Calculer $\mathbb{E}[U]$; en déduire $\mathbb{E}[X^2]$, $\mathbb{E}[Y^2]$, puis $\text{Var } X$, $\text{Var } Y$.

f) Déterminer la densité $f_{Y|X}(\cdot|x)$ de Y liée par $\{X = x\}$. Calculer $\mathbb{E}[Y^2 | X = x]$, puis $\mathbb{E}[X^2 + Y^2 | X = x]$ et $\mathbb{E}[X^2 + Y^2 | X]$.

g) On suppose à présent qu'un tireur tire sur une cible ayant la forme d'un disque D et que la loi du point d'impact (X, Y) sur la cible est la loi uniforme sur D . Au point (X, Y) on associe la variable aléatoire $L = \sqrt{X^2 + Y^2}$, qui est la distance de (X, Y) au centre de la cible. A un système de n tirs indépendants correspond alors un système de n points aléatoires indépendants, puis un système (L_1, \dots, L_n) de n variables aléatoires qui représentent les distances au centre de ces points. Ce système est formé de variables aléatoires indépendantes et identiquement distribuées. On demande de calculer $P\{\min(L_1, \dots, L_n) < a\}$, où a est un réel vérifiant $0 < a < r$. Interpréter cette probabilité.

6. — Soit $M = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ un vecteur aléatoire de loi $\mathcal{N}_2(0, \Gamma)$ avec $\Gamma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ et $|\rho| < 1$. Montrer que

a) la loi de X_2 liée par $\{X_1 = x_1\}$ est la loi $\mathcal{N}(\rho x_1, \sqrt{1 - \rho^2})$. Il en résulte

α) $\mathbb{E}[X_2 | X_1 = x_1] = \rho x_1$: l'espérance conditionnelle est linéaire en x_1 ; en d'autres termes, la courbe de régression de X_2 en X_1 est une droite passant par l'origine et de pente ρ .

β) $\text{Var}(X_2 | X_1 = x_1) = 1 - \rho^2$; elle est indépendante de x_1 .

b) $(X_1, X_2 - \mathbb{E}[X_2 | X_1])$ est un couple de variables aléatoires *indépendantes*, normales, centrées, de variances respectives 1, $1 - \rho^2$.

7. — Soit (X_1, X_2) un couple de variables aléatoires dont les lois marginales sont des lois normales $\mathcal{N}(0, 1)$. Il n'en résulte pas que (X_1, X_2) soit normal. Imaginer un exemple.

8. — Soit $M = (X, Y)$ un vecteur aléatoire absolument continu, de support \mathbb{R}^2 , dont les composantes X, Y sont indépendantes. Montrer que les deux propriétés suivantes sont équivalentes :

a) X et Y suivent toutes deux une même loi normale centrée;

b) la loi de M est isotrope, c'est-à-dire est invariante par rapport à toute rotation autour de l'origine.

9. — On lance un dé parfait; ceci étant fait, on lance une pièce de monnaie parfaite un nombre de fois égal au nombre de points amenés par le dé. On désigne par X le nombre de points amenés par le dé et par Y le nombre de "pile" amenés par la pièce de monnaie.

a) Déterminer la loi conjointe du couple (X, Y) .

b) Calculer $\mathbb{E}[Y]$.

10. (Calcul de l'espérance mathématique d'une variable géométrique) Soit une urne contenant r boules blanches et s boules noires ($r, s \geq 1$); on pose $p = r/(r + s)$. On procède à une suite de tirages *avec remise* et l'on désigne par N le nombre de tirages nécessaires pour amener pour la première fois une boule blanche (N est alors une variable aléatoire géométrique).

On introduit une variable aléatoire X qui prend la valeur 1 ou 0, suivant que la *première* boule tirée est blanche ou non. Calculer l'espérance mathématique de N en utilisant les formules sur l'espérance conditionnelle, lorsque l'on conditionne N par rapport à la variable X .

11. (Suite de l'exercice 10). — Soit une urne contenant r boules blanches et s boules noires ($r, s \geq 1$). On procède à une suite de tirages *sans remise* et l'on désigne par $N_{r,s}$ le nombre de tirages nécessaires pour amener pour la première fois une boule blanche. Pour calculer $\mathbb{E}[N_{r,s}]$, on peut procéder comme suit : soit X la variable aléatoire qui prend la valeur 1 ou 0 suivant que la *première* boule tirée est blanche ou noire. On calcule $\mathbb{E}[N_{r,s}]$ par la formule $\mathbb{E}[N_{r,s}] = \mathbb{E}[\mathbb{E}[N_{r,s} | X]]$.

a) Montrer que les nombres $a_{r,s} = \mathbb{E}[N_{r,s}]$ vérifient le système de récurrence : $a_{r,s} = 1 + (s/(r + s))a_{r,s-1}$ pour $r, s \geq 1$ et $a_{r,0} = 1$ pour tout $r \geq 1$.

b) Montrer que ce système admet une solution et une seule donnée par $a_{r,s} = \mathbb{E}[N_{r,s}] = (r + s + 1)/(r + 1)$.

12. — Soit (X, Y) un couple de variables aléatoires à valeurs dans \mathbb{N} et vérifiant $0 \leq Y \leq X$, $\mathbb{E}[X] < +\infty$. On suppose que la loi de Y , conditionnellement à X , est la loi uniforme sur $\{0, 1, \dots, X\}$.

1) Calculer $\mathbb{E}[X]$ en fonction de $\mathbb{E}[Y]$.

2) Montrer que les deux propriétés suivantes sont équivalentes :

a) le couple $(X - Y, Y)$ est indépendant ;

b) la variable aléatoire Y suit une loi géométrique : $P\{Y = n\} = q^n p$ ($n \geq 0$).

13. (Stéphanie et l'assurance-chômage). — Soit X le temps s'écoulant jusqu'à l'instant où un individu d'une population donnée se retrouve au chômage. On suppose que X suit une loi exponentielle de paramètre λ (voir chap. 14, § 5). La compagnie d'assurance qui assure cette population contre le chômage veut calculer le temps moyen de travail pour la sous-population qui sera au chômage entre le temps a et le temps b ($0 < a < b < +\infty$).

a) Soit $g(a, b)$ ce temps moyen. Comment faut-il faire le calcul de $g(a, b)$?

b) Déterminer la limite de $g(a, b)$ quand b tend vers l'infini. Aurait-on pu prévoir ce résultat sans calcul ?

c) Déterminer la limite de $g(a, a + \varepsilon)$, lorsque ε tend vers 0 par valeurs positives.

FONCTION GÉNÉRATRICE DES MOMENTS ; FONCTION CARACTÉRISTIQUE

Nous avons étudié au chapitre 9 la notion de fonction génératrice d'une variable aléatoire discrète à *valeurs dans* \mathbb{N} . Dans le cas des variables aléatoires quelconques, il convient de modifier légèrement la définition de cette notion. On introduit alors une fonction réelle qui, si elle est définie dans un voisinage de l'origine, contient toute l'information utile sur les moments de cette variable aléatoire. La fonction caractéristique est la version complexe de la fonction génératrice des moments.

1. Introduction. — Soit X une variable aléatoire réelle; associons-lui la fonction $g_X(u) = \mathbb{E}[e^{uX}]$ de la variable réelle u , définie sur l'ensemble des valeurs u pour lesquelles $\mathbb{E}[e^{uX}] < +\infty$. Les propriétés suivantes sont immédiates :

1) La fonction g_X est toujours définie pour $u = 0$ et $g_X(0) = 1$. Il existe des variables aléatoires dont la fonction g_X associée n'est définie que pour $u = 0$ (par exemple, la variable aléatoire de Cauchy).

2) Si X est *bornée*, alors g_X est définie et continue sur tout \mathbb{R} .

3) Si X est à *valeurs positives*, alors g_X est continue et bornée sur $] -\infty, 0]$. Dans ce cas, on fait souvent le changement de variable $u = -v$ et l'on obtient alors la *transformée de Laplace* de (la loi de) X :

$$L_X(v) = g_X(-v) = \mathbb{E}[e^{-vX}].$$

La fonction L_X ainsi définie est une fonction continue et bornée sur $[0, +\infty[$.

Définition. — Soit X une variable aléatoire réelle telle que la fonction

$$g_X(u) = \mathbb{E}[e^{uX}]$$

de la variable réelle u soit définie *dans un voisinage ouvert de l'origine*. Cette fonction est alors appelée *fonction génératrice des moments* de (la loi de) X .

La terminologie *fonction génératrice des moments* est bien appropriée. En effet, on montrera (Théorème 3.1) que si une variable aléatoire réelle X admet une fonction génératrice des moments g_X , alors

a) elle admet des moments de tous les ordres (entiers positifs);

b) la fonction g_X est développable en série entière dans un voisinage de l'origine et le moment d'ordre n de X apparaît comme le coefficient de $u^n/n!$ dans ce développement.

2. Propriétés élémentaires

PROPOSITION 2.1. — Soit g_X la fonction génératrice des moments d'une variable aléatoire X .

- (1) Pour a, b réels, on a : $g_{aX+b}(u) = e^{bu}g_X(au)$.
- (2) La fonction $g_X(-u)$ est aussi une fonction génératrice des moments.
- (3) Si la loi de X est symétrique, c'est-à-dire si $\mathcal{L}(X) = \mathcal{L}(-X)$, alors g_X est une fonction paire.
- (4) La fonction g_X est convexe.

Démonstration. — La propriété (1) est évidente. Pour établir la propriété (2), on note que la fonction $g_X(-u)$ est la fonction génératrice des moments de la variable aléatoire $-X$. La propriété (3) résulte de :

$$g_X(u) = \mathbb{E}[e^{uX}] = \mathbb{E}[e^{u(-X)}] = \mathbb{E}[e^{(-u)X}] = g_X(-u).$$

Enfin, la propriété (4) résulte du fait qu'une fonction génératrice est combinaison convexe de fonctions exponentielles. \square

THÉORÈME 2.2 (théorème d'unicité). — La fonction génératrice des moments d'une variable aléatoire détermine la loi de cette variable. En d'autres termes, si deux variables aléatoires admettent même fonction génératrice des moments, alors elles ont même loi.

Démonstration. — Elle n'est aisée que lorsque la variable aléatoire X prend ses valeurs dans \mathbb{N} . Dans ce cas, en effet, il est de coutume de remplacer la fonction génératrice des moments $g(u) = \mathbb{E}[e^{uX}]$ par la fonction génératrice $G(s) = \mathbb{E}[s^X]$, qui joue essentiellement le même rôle. Or on a pu établir de façon élémentaire que la fonction G détermine la loi de X (cf. Théorème 1.4 du chap. 9).

Dans le cas général, la démonstration du théorème d'unicité est plus complexe (cf. Billingsley.¹) \square

THÉORÈME 2.3. — Soit (X, Y) un couple de variables aléatoires indépendantes dont chacune admet une fonction génératrice des moments. Alors la somme $X + Y$ admet une fonction génératrice des moments et l'on a :

$$(2.2) \quad g_{X+Y} = g_X g_Y.$$

Démonstration. — On a : $g_{X+Y}(u) = \mathbb{E}[e^{u(X+Y)}] = \mathbb{E}[e^{uX}e^{uY}]$; comme X et Y sont indépendantes, les variables e^{uX} et e^{uY} le sont aussi. D'après l'Application du Théorème 2.2 du chap. 11, on en déduit : $g_{X+Y}(u) = \mathbb{E}[e^{uX}] \mathbb{E}[e^{uY}] = g_X(u) g_Y(u)$. \square

Remarque. — La propriété (2.2) n'est pas caractéristique de l'indépendance : il est possible de construire un couple (X, Y) de variables aléatoires

¹ Billingsley (P.). — *Probability and Measure*, troisième édition. — New York, Wiley, 1995, p. 390.

admettant des fonctions génératrices des moments, non indépendantes, et qui, néanmoins, vérifient (2.2), comme le montre l'exemple suivant.

Soit (X, Y) un couple de variables aléatoires dont la loi conjointe admet la densité :

$$f(x, y) = \begin{cases} 2, & \text{si } (x, y) \in E; \\ 0, & \text{sinon;} \end{cases}$$

où E est la partie hachurée de $[0, 1] \times [0, 1]$ représentée dans la Figure 1.

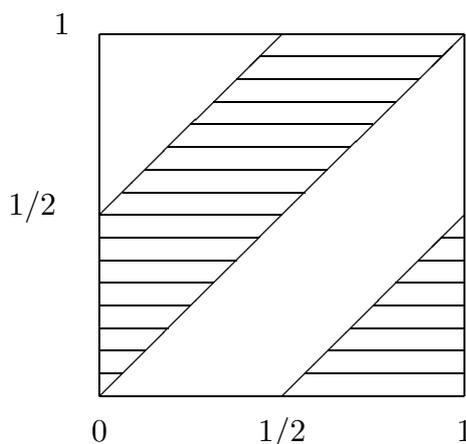


Fig. 1

Les variables aléatoires X et Y ne sont *pas indépendantes* : on a, par exemple, $P\{X \leq 1/4, Y \geq 3/4\} = 0$, alors que $P\{X \leq 1/4\} = P\{Y \geq 3/4\} = 1/4$. Prenons, par ailleurs, un couple (X^*, Y^*) de variables aléatoires *indépendantes*, uniformément réparties sur $[0, 1]$. On vérifie que $\mathcal{L}(X) = \mathcal{L}(X^*)$, $\mathcal{L}(Y) = \mathcal{L}(Y^*)$, $\mathcal{L}(X + Y) = \mathcal{L}(X^* + Y^*)$ (cf. Exercice 1). Comme toutes les variables aléatoires $X, Y, X + Y, X^*, Y^*, X^* + Y^*$ sont *bornées*, elles admettent des fonctions génératrices des moments et l'on a, pour tout nombre réel u les identités : $g_{X+Y}(u) = g_{X^*+Y^*}(u) = g_{X^*}(u) g_{Y^*}(u) = g_X(u) g_Y(u)$. Ainsi le couple (X, Y) de variables aléatoires non indépendantes vérifie aussi $g_{X+Y} = g_X g_Y$. \square

Le théorème 2.3 admet les corollaires suivants.

COROLLAIRE 1. — *Le produit de deux fonctions génératrices des moments est une fonction génératrice des moments.*

COROLLAIRE 2. — *Si g est une fonction génératrice des moments, il en est de même de $h(u) = g(u)g(-u)$. De façon précise, si X est une variable aléatoire admettant une fonction génératrice des moments $g(u)$ et si X' est une variable aléatoire indépendante de X et admettant même loi que X , alors la variable aléatoire $Y = X - X'$ admet une fonction génératrice des moments qui n'est autre que $h(u) = g(u)g(-u)$. La variable aléatoire Y est appelée la symétrisée de X .*

3. Moments. — Le théorème suivant justifie la terminologie de fonction génératrice des moments qui a été adoptée.

THÉORÈME 3.1. — Soit X une variable aléatoire admettant une fonction génératrice des moments g . Supposons que $g(u)$ soit définie pour tout u dans l'intervalle ouvert $]u_1, u_2[$, où $u_1 < 0 < u_2$. Alors

- 1) pour tout $k \geq 1$ on a $\mathbb{E}[|X|^k] < +\infty$;
- 2) pour tout $t \in]-s, +s[$, où $0 < s < u_0 = \min(-u_1, u_2)$, on a :

$$g(t) = 1 + \mathbb{E}[X] \frac{t}{1!} + \mathbb{E}[X^2] \frac{t^2}{2!} + \cdots + \mathbb{E}[X^n] \frac{t^n}{n!} + \cdots$$

- 3) pour tout $k \geq 1$, on a : $g^{(k)}(0) = \mathbb{E}[X^k]$.

Démonstration. — Nous faisons la démonstration dans le seul cas où X admet une densité f .

1) Il s'agit de montrer que pour tout $k \geq 1$, on a $\mathbb{E}[|X|^k] = \int_{\mathbb{R}} |x|^k f(x) dx < +\infty$. Soit $0 < s < u_0 = \min(-u_1, u_2)$ fixé. Le développement en série de $e^{s|x|}$ pour x réel permet d'écrire l'inégalité $s^k |x|^k / k! \leq e^{s|x|}$ valable pour tout $k \geq 1$, d'où encore $|x|^k \leq k! e^{s|x|} / s^k$. Il en résulte, pour tout $k \geq 1$, les relations :

$$\begin{aligned} \mathbb{E}[|X|^k] &\leq \frac{k!}{s^k} \int_{-\infty}^{+\infty} e^{s|x|} f(x) dx = \frac{k!}{s^k} \left(\int_0^{+\infty} e^{sx} f(x) dx + \int_{-\infty}^0 e^{-sx} f(x) dx \right) \\ &\leq \frac{k!}{s^k} \left(\int_{-\infty}^{+\infty} e^{sx} f(x) dx + \int_{-\infty}^{+\infty} e^{-sx} f(x) dx \right) = \frac{k!}{s^k} [g_X(s) + g_X(-s)]. \end{aligned}$$

La dernière quantité est finie puisque par hypothèse $g_X(s)$ et $g_X(-s)$ sont finies.

$$2) \text{ Posons : } g(t) = \int_{-\infty}^{+\infty} e^{tx} f(x) dx = \int_{-\infty}^{+\infty} \left(\sum_{k \geq 0} \frac{(tx)^k}{k!} \right) f(x) dx.$$

Il s'agit de montrer que l'on peut permuter les signes de sommation et d'intégration. Définissons :

$$h(x) = f(x) e^{tx} \quad \text{et} \quad h_n(x) = f(x) \sum_{k=0}^n \frac{(tx)^k}{k!} \quad (n = 1, 2, \dots).$$

On constate que pour tout nombre réel x on a $\lim_n h_n(x) = h(x)$. De plus, pour tout $n = 1, 2, \dots$ on a les majorations :

$$|h_n(x)| \leq f(x) \sum_{k=0}^n \frac{|tx|^k}{k!} \leq f(x) \sum_{k \geq 0} \frac{|tx|^k}{k!} = f(x) e^{|tx|} \leq f(x) e^{s|x|}.$$

D'autre part, par un raisonnement analogue à celui de 1), on a :

$$\int_{-\infty}^{+\infty} e^{s|x|} f(x) dx \leq g(s) + g(-s) < +\infty.$$

On se trouve dans les conditions d'application du théorème de convergence dominée et l'on obtient donc :

$$\begin{aligned} g(t) &= \int_{-\infty}^{+\infty} h(x) dx = \lim_n \int_{-\infty}^{+\infty} h_n(x) dx = \lim_n \sum_{k=0}^n \frac{t^k}{k!} \int_{-\infty}^{+\infty} x^k f(x) dx \\ &= \sum_{k \geq 0} \frac{t^k}{k!} \mathbb{E}[X^k]. \end{aligned}$$

3) On sait qu'une série entière peut être dérivée terme à terme dans son domaine de convergence. Il résulte donc de la partie 2) que pour tout t dans l'intervalle $] -s, s[$ et tout $k = 1, 2, \dots$ on a :

$$g^{(k)}(t) = \sum_{n \geq k} n(n-1) \dots (n-k+1) \mathbb{E}[X^n] \frac{t^{n-k}}{n!};$$

d'où $g^{(k)}(0) = \mathbb{E}[X^k]$. \square

Remarques. — On vient de voir que si une variable aléatoire admet une fonction génératrice des moments, elle admet des moments de tous les ordres entiers positifs. La réciproque est fautive : une variable aléatoire peut admettre des moments de tous les ordres entiers positifs, sans admettre de fonction génératrice des moments (au sens de la définition donnée dans le paragraphe 1), comme le montre l'exemple suivant.

Soit $X = e^Y$, où Y suit une loi normale centrée, réduite $\mathcal{N}(0, 1)$ (voir chap. 14, § 3 et 4). La loi de X est appelée *log-normale* $(0, 1)$.

a) D'abord X admet des moments de tous les ordres; en effet, on a :

$$\mathbb{E}[X^n] = \int_{-\infty}^{+\infty} e^{ny} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = e^{n^2/2} < +\infty.$$

b) D'autre part, $g(u) = \mathbb{E}[e^{uX}] = \int_{-\infty}^{+\infty} e^{ue^y} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$.

Cette intégrale est convergente si et seulement si $u \in] -\infty, 0]$. (Noter que pour $u > 0$ et y assez grand, on a : $ue^y > y^2/2$.) Or $] -\infty, 0]$ n'est pas un voisinage ouvert de $u = 0$, de sorte que g n'est pas une fonction génératrice des moments.

Ainsi X admet des moments de tous les ordres, alors qu'elle n'admet pas de fonction génératrice des moments. Ceci a pour conséquence que la suite des moments ne détermine pas la loi de probabilité : il existe d'autres lois, différentes de la loi log-normale, admettant même suite des moments.

Cas particulier. — Soit X une variable aléatoire à valeurs dans \mathbb{N} ; il est d'usage de lui associer la *fonction génératrice des moments factoriels* G :

$$G(u) = \mathbb{E}[u^X],$$

qui est liée à la fonction génératrice des moments g

$$g(u) = \mathbb{E}[e^{uX}],$$

par la relation : $g(u) = G(e^u)$. (Pour une étude des ces fonctions, voir chap. 9.) Le théorème suivant résulte alors du théorème 3.1.

THÉORÈME 3.1'. — Soient X une variable aléatoire à valeurs dans \mathbb{N} et G sa fonction génératrice des moments factoriels. Supposons que G soit définie dans un voisinage ouvert de $u = 1$. Alors

a) tous les moments d'ordre entier positif $\mathbb{E}[X^n]$ existent et sont finis ; il en résulte que tous les moments factoriels d'ordre entier strictement positif de X existent et sont finis ;

b) pour tout $n \geq 1$, on a $g^{(n)}(0) = \mathbb{E}[X^n]$.

Démonstration

a) Supposons que $G(u)$ soit définie dans un voisinage ouvert de 1 ; elle est alors définie dans un intervalle $U =]u_1, u_2[$ ($0 < u_1 < 1 < u_2$). Pour tout $u \in U$, posons $u = e^t$, $t = \text{Log } u$. Alors $g(t) = G(e^t)$ est définie dans l'intervalle ouvert $]t_1, t_2[$, où $t_1 < 0 < t_2$, $t_1 = \text{Log } u_1$, $t_2 = \text{Log } u_2$. Il résulte alors du Théorème 3.1 1) que tous les moments d'ordre entier positif de X existent et sont finis ; il en est de même de tous ses moments factoriels.

b) résulte du Théorème 3.1 3). \square

4. Fonctions caractéristiques. — A un stade plus avancé, on préfère travailler avec la *fonction caractéristique* d'une variable aléatoire plutôt qu'avec sa fonction génératrice des moments. La fonction caractéristique de X , définie pour tout t réel par $\varphi(t) = \mathbb{E}[e^{itX}]$, a l'énorme avantage d'exister pour tout nombre réel t et toute variable aléatoire X . En outre, la correspondance entre loi de probabilité d'une variable aléatoire et fonction caractéristique est *bijection*. En revanche, le maniement des fonctions caractéristiques est plus délicat, puisqu'il fait appel à la théorie des fonctions d'une variable complexe.

Définition. — Soit X une variable aléatoire réelle. On appelle *fonction caractéristique* de X (ou de la loi de X) la fonction de la variable réelle t définie par :

$$\varphi_X(t) = \mathbb{E}[e^{itX}].$$

a) Si X est une variable aléatoire de loi discrète $\sum_k p_k \varepsilon_{x_k}$, alors $\varphi(t) = \sum_k p_k e^{itx_k}$.

b) Si X est une variable absolument continue, de densité f , alors $\varphi(t) = \int_{\mathbb{R}} e^{itx} f(x) dx$.

THÉORÈME 4.1 (Propriétés élémentaires). — Désignons par φ la fonction caractéristique d'une variable aléatoire réelle X . Alors

- 1) φ est une fonction définie et continue pour tout nombre réel t ;
- 2) φ est bornée, et en fait, pour tout t réel, on a : $|\varphi(t)| \leq \varphi(0) = 1$;
- 3) φ est une fonction hermitique, c'est-à-dire pour tout nombre réel t on a : $\varphi(-t) = \overline{\varphi(t)}$; il en résulte que si une fonction caractéristique est réelle, alors elle est paire ;

4) pour tous a, b réels, on a, pour tout réel t , l'identité : $\varphi_{aX+b}(t) = e^{ibt} \varphi_X(at)$;

5) si la loi de X est symétrique, c'est-à-dire si $\mathcal{L}(-X) = \mathcal{L}(X)$, alors φ_X est une fonction réelle, donc paire ;

6) toute combinaison convexe de fonctions caractéristiques est une fonction caractéristique.

Démonstration. — Donnons seulement une indication pour la propriété 5). Il suffit d'écrire : $\varphi_X(t) = \varphi_{-X}(t) = \varphi_X(-t) = \overline{\varphi_X(t)}$. D'où φ_X est réelle, donc paire. \square

Le tableau des fonctions caractéristiques des lois usuelles est reproduit dans la table suivante. Les trois premières lois sont discrètes, de forme $\sum_k p_k \varepsilon_k$. On précise chaque fois l'ensemble des valeurs prises par l'entier k . Les autres sont absolument continues, de densité f . Une description complète de ces quatre dernières lois est faite au chapitre suivant, § 5, 6, 3 et 7, respectivement.

Loi	Expression analytique	Fonction caractéristique
Loi binomiale	$p_k = \binom{n}{k} p^k q^{n-k} \quad (0 \leq k \leq n)$	$(q + pe^{it})^n$
Loi de Poisson	$p_k = e^{-\lambda} \frac{\lambda^k}{k!} \quad (k \geq 0, \lambda > 0)$	$e^{\lambda(e^{it}-1)}$
Loi géométrique	$p_k = q^{k-1} p \quad (k \geq 1)$	$\frac{pe^{it}}{1 - qe^{it}}$
Loi exponentielle $\mathcal{E}(\lambda)$	$f(x) = \lambda e^{-\lambda x} I_{[0, \infty[}(x) \quad (\lambda > 0)$	$\frac{\lambda}{\lambda - it}$
Première loi de Laplace	$f(x) = \frac{1}{2} e^{- x } \quad (x \in \mathbb{R})$	$\frac{1}{1 + t^2}$
Loi normale $\mathcal{N}(0, 1)$	$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad (x \in \mathbb{R})$	$e^{-t^2/2}$
Loi de Cauchy $C(0, 1)$	$f(x) = \frac{1}{\pi} \frac{1}{1 + x^2} \quad (x \in \mathbb{R})$	$e^{- t }$

Le calcul concernant les trois premières lois est banal ; il n'en est pas de même des autres. Cependant, pour le calcul de la fonction caractéristique de loi normale $\mathcal{N}(0, 1)$, on peut éviter de recourir à la théorie des fonctions de variable complexe en s'y prenant comme suit.

La loi $\mathcal{N}(0, 1)$ d'une variable aléatoire normale est évidemment symétrique. Sa fonction caractéristique est donc égale à :

$$\varphi(t) = \mathbb{E}[e^{itX}] = \mathbb{E}[\cos tX] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \cos(tx) e^{-x^2/2} dx.$$

On est donc ramené au calcul classique d'une intégrale dépendant d'un paramètre. Par dérivation sous le signe somme, qu'il est facile de justifier, on obtient

$$\varphi'(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (-x \sin(tx)) e^{-x^2/2} dx;$$

soit en intégrant par parties,

$$\varphi'(t) = -t \varphi(t);$$

comme $\varphi(0) = 1$, on obtient bien par intégration :

$$\varphi(t) = e^{-t^2/2}.$$

Remarque. — *Formellement*, la fonction caractéristique $\varphi(t)$ se déduit de la fonction génératrice des moments $g(u)$ en remplaçant u par it . Dans de nombreux cas, comme par exemple les six premières lois du tableau précédent, l'expression $g(it)$ obtenue par substitution a bien un sens en tant que fonction à valeurs complexes et est bien égale à la fonction caractéristique calculée directement par sommation ou par intégration dans le plan complexe. On ne peut toutefois ériger cette substitution en règle générale, car, d'une part, $g(u)$ peut ne pas être définie en dehors de l'origine (c'est le cas, par exemple, de la loi de Cauchy), d'autre part, même si $g(u)$ est bien définie comme fonction de la variable *réelle* u , la fonction complexe $g(it)$ peut avoir plusieurs déterminations dans le plan complexe, par exemple, dès que l'expression algébrique de g contient des puissances de la variable complexe non entières. Il convient donc de faire chaque fois un calcul direct de $\varphi(t)$.

THÉORÈME 4.2 (théorème d'unicité). — *La fonction caractéristique d'une variable aléatoire détermine la loi de cette variable. En d'autres termes, si deux variables aléatoires admettent même fonction caractéristique, elles ont même loi.*

Ce théorème justifie la terminologie de fonction *caractéristique* : la fonction caractéristique *caractérise* la loi. Nous donnerons une démonstration de ce théorème à la fin de ce chapitre.

THÉORÈME 4.3. — *Soit (X, Y) un couple de variables aléatoires indépendantes. Alors, pour tout réel t , on a :*

$$(4.1) \quad \varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t).$$

Démonstration. — En effet, $\varphi_{X+Y}(t) = \mathbb{E}[e^{it(X+Y)}] = \mathbb{E}[e^{itX}e^{itY}]$, d'où, puisque X et Y sont indépendantes, donc aussi e^{itX} et e^{itY} :

$$\varphi_{X+Y}(t) = \mathbb{E}[e^{itX}] \mathbb{E}[e^{itY}] = \varphi_X(t)\varphi_Y(t). \quad \square$$

Remarque. — La propriété (4.1) n'est *pas* caractéristique de l'indépendance : il est possible de construire un couple de variables aléatoires *non*

indépendantes qui, néanmoins, vérifie (4.1). On peut donner ici un exemple particulièrement frappant (qui n'est *pas* valable pour les fonctions génératrices des moments).

Soit X une variable aléatoire de Cauchy $C(0, 1)$; sa fonction caractéristique est donnée par $\varphi_X(t) = e^{-|t|}$. Le couple (X, Y) , où $Y = X$ fournit l'exemple désiré : en effet,

$$\varphi_{X+Y}(t) = \varphi_{2X}(t) = e^{-2|t|} = (e^{-|t|})^2 = \varphi_X(t)\varphi_Y(t). \quad \square$$

COROLLAIRE. — *Le produit de deux fonctions caractéristiques est une fonction caractéristique.*

THÉORÈME 4.4. — *Si φ est une fonction caractéristique, il en est de même de $\bar{\varphi}$, $|\varphi|^2$, $\Re \varphi$. Si, en outre, φ est la fonction caractéristique d'une variable aléatoire absolument continue, il en est de même de $\bar{\varphi}$, $|\varphi|^2$.*

Démonstration. — Soit X une variable aléatoire admettant φ comme fonction caractéristique. La variable aléatoire $-X$ admet pour fonction caractéristique $\bar{\varphi}$. En outre, si X est absolument continue, il en est de même de $-X$.

Soit X' une variable aléatoire indépendante de X et admettant même loi que X . Alors la variable aléatoire $Y = X - X'$ admet pour fonction caractéristique : $\varphi_{X-X'}(t) = \varphi_X(t)\overline{\varphi_{X'}(t)} = \varphi_X(t)\overline{\varphi_X(t)} = |\varphi_X(t)|^2$. En outre, si X est absolument continue, il en est de même de Y . La variable aléatoire Y est dite la *symétrisée* de X .

On a $\Re \varphi = (\varphi + \bar{\varphi})/2$, où φ et $\bar{\varphi}$ sont des fonctions caractéristiques. Or le barycentre à coefficients positifs de fonctions caractéristiques est encore une fonction caractéristique. \square

Étudions à présent les rapports entre fonction caractéristique et moments.

THÉORÈME 4.5. — *Soit X une variable aléatoire de fonction caractéristique φ . Si $\mathbb{E}[|X|] < +\infty$, alors φ est continûment dérivable et $\varphi'(0) = i \mathbb{E}[X]$.*

Démonstration. — Nous faisons la démonstration dans le cas où X admet une densité f . Dans ce cas,

$$\varphi(t) = \mathbb{E}[e^{itX}] = \int_{\mathbb{R}} e^{itx} f(x) dx.$$

On constate que pour tout nombre réel t on a :

$$\left| \int_{\mathbb{R}} ix e^{itx} f(x) dx \right| \leq \int_{\mathbb{R}} |x| f(x) dx = \mathbb{E}[|X|] < +\infty.$$

On peut donc dériver sous le signe somme et écrire, pour tout nombre réel t , l'expression $\varphi'(t) = \int_{\mathbb{R}} ix e^{itx} f(x) dx$, d'où $\varphi'(0) = i \mathbb{E}[X]$. \square

Suivant la même méthode, on démontre le théorème suivant concernant les moments d'ordre supérieur.

THÉORÈME 4.6. — Soit X une variable aléatoire de fonction caractéristique φ . Supposons qu'il existe un entier $n \geq 1$ tel que $\mathbb{E}[|X|^n] < +\infty$. Alors

- 1) φ est continûment dérivable jusqu'à l'ordre n inclus ;
- 2) pour tout $k = 0, 1, \dots, n$ on a : $\varphi^{(k)}(0) = i^k \mathbb{E}[X^k]$;
- 3) φ peut être représentée par la formule de Taylor :

$$\varphi(t) = \sum_{k=0}^n \frac{(it)^k}{k!} \mathbb{E}[X^k] + o(|t|^n), \quad \text{lorsque } t \rightarrow 0.$$

Le théorème suivant est un outil technique qui nous servira au chapitre 18, § 4, pour la démonstration du théorème « central limit ».

THÉORÈME 4.7. — Soit X une variable aléatoire de L^2 , centrée. Posons $\text{Var } X = \sigma^2$ et désignons par φ sa fonction caractéristique. Alors

- a) pour tout t réel, la fonction $\varphi(t)$ admet la représentation

$$\varphi(t) = 1 - \frac{t^2 \sigma^2}{2} - t^2 \int_0^1 (1-u) \mathbb{E}[X^2 (e^{ituX} - 1)] du.$$

- b) pour tout t réel tel que $3t^2 \sigma^2 \leq 1$, la fonction $\text{Log } \varphi(t)$ existe et admet la représentation

$$\text{Log } \varphi(t) = -\frac{t^2}{2} \sigma^2 - t^2 \int_0^1 (1-u) \mathbb{E}[X^2 (e^{ituX} - 1)] du + 3\theta t^4 \sigma^4,$$

où θ est un nombre complexe vérifiant $|\theta| \leq 1$.

Démonstration. — Pour tout t et tout x réels, on a l'identité :

$$e^{itx} = 1 + itx - \frac{t^2 x^2}{2} - t^2 x^2 \int_0^1 (1-u) (e^{itux} - 1) du.$$

- a) En intégrant cette identité par rapport à μ , où μ est la loi de X , on obtient la première représentation a).

- b) La fonction $\varphi(t)$ est de la forme $1 - r$, où

$$r = \frac{t^2}{2} \sigma^2 + t^2 \int_0^1 (1-u) \mathbb{E}[X^2 (e^{ituX} - 1)] du.$$

Or $|r| \leq \frac{3}{2} t^2 \sigma^2$, de sorte que $|r| \leq \frac{1}{2}$ si $3t^2 \sigma^2 \leq 1$. D'autre part, pour tout nombre complexe tel que $|r| \leq \frac{1}{2}$, on a

$$\text{Log}(1-r) = -r \int_0^1 \frac{dx}{1-rx}, \quad \text{d'où} \quad \text{Log}(1-r) + r = -r^2 \int_0^1 \frac{x}{1-rx} dx.$$

Or l'intégrale au second membre peut être majorée en module par 1, puisque, pour tout x tel que $0 \leq x \leq 1$, on a $|1 - rx| \geq 1 - |rx| \geq 1 - x \geq \frac{1}{2}$. On a donc

$$|\operatorname{Log}(1 - r) + r| \leq |r|^2 \leq \left(\frac{3}{2}t^2\sigma^2\right)^2 \leq 3t^4\sigma^4,$$

d'où

$$\operatorname{Log}(1 - r) = -r + 3\theta t^4\sigma^4,$$

où θ est un nombre complexe vérifiant $|\theta| \leq 1$. Ceci établit b). \square

5. Seconde fonction caractéristique

Définition. — Soient X une variable aléatoire réelle, $\varphi(t)$ sa fonction caractéristique. On appelle *seconde fonction caractéristique* de (la loi) de X la fonction $\psi(t) = \operatorname{Log} \varphi(t)$ ($t \in \mathbb{R}$).

Comme la fonction caractéristique $\varphi(t)$ est une fonction à valeurs complexes, il faut préciser celle des déterminations du logarithme que l'on considère : puisque $\varphi(0) = 1$, on convient de prendre ψ continue et $\psi(0) = 0$.

THÉORÈME 5.1. — *Soit (X, Y) un couple de variables aléatoires indépendantes. Alors, pour tout t réel, on a :*

$$\psi_{X+Y}(t) = \psi_X(t) + \psi_Y(t).$$

La démonstration résulte immédiatement du Théorème 4.3.

Définition. — Soit X une variable aléatoire réelle et supposons que sa seconde fonction caractéristique ψ admette un développement de Taylor au voisinage de l'origine

$$\psi(t) = \sum_{n \geq 1} \kappa_n \frac{(it)^n}{n!}.$$

Le coefficient κ_n de $(it)^n/n!$ dans ce développement est appelé le $n^{\text{ième}}$ *cumulant de (la loi) de X* .

THÉORÈME 5.2. — *Soit (X, Y) un couple de variables aléatoires dont chacune vérifie les hypothèses de la définition précédente. Le cumulants d'ordre n de $X + Y$ est la somme des cumulants d'ordre n de X et du cumulants d'ordre n de Y (d'où la terminologie de « cumulants »).*

La démonstration est une conséquence directe du Théorème 5.1 et de la définition précédente.

PROPOSITION 5.3. — *Désignons par m_1, m_2, m_3 les moments d'ordre 1, 2, 3 de X et par $\mu_1 = 0, \mu_2, \mu_3$ ses moments centrés d'ordre 1, 2, 3. Alors*

$$\begin{aligned} \varphi(t) &= 1 + itm_1 + \frac{(it)^2}{2!}m_2 + \frac{(it)^3}{3!}m_3 + o(|t|^3) = 1 + \lambda(t) \\ \psi(t) &= \operatorname{Log}(1 + \lambda(t)) = itm_1 + \frac{(it)^2}{2!}\mu_2 + \frac{(it)^3}{3!}\mu_3 + o(|t|^3). \end{aligned}$$

La démonstration nécessite un simple calcul. Notons que les cumulants d'ordre 1, 2, 3 sont respectivement :

l'espérance mathématique $m_1 = \mathbb{E}[X]$;

le moment centré d'ordre deux, c'est-à-dire la variance

$$\mu_2 = \mathbb{E}[(X - m_1)^2] ;$$

le moment centré d'ordre trois $\mu_3 = \mathbb{E}[(X - m_1)^3]$.

Les propriétés d'additivité des coefficients κ s'étendent *aux moments centrés d'ordre 2 et 3*, mais pas au-delà ; le quatrième cumulant est par exemple $\kappa_4 = \mu_4 - 3\mu_2^2$.

Exemples

1) Soit X de loi $\mathcal{N}(\mu, \sigma)$. On a $\varphi(t) = \exp(it\mu - \frac{1}{2}t^2\sigma^2)$, d'où $\psi(t) = it\mu - \frac{1}{2}t^2\sigma^2$ et par conséquent $\kappa_1 = \mu$, $\kappa_2 = \sigma^2$ et $\kappa_n = 0$ pour tout $n \geq 3$.

2) Soit X de loi $\mathcal{P}(\lambda)$. On a $\varphi(t) = \exp(\lambda(e^{it} - 1))$, d'où $\psi(t) = t(e^{it} - 1) = \lambda \sum_{n \geq 1} (it)^n / n!$, d'où $\kappa_n = \lambda$ pour tout $n \geq 1$.

3) La loi de Cauchy n'a *pas* de cumulants.

6. Fonctions génératrice et caractéristique d'un vecteur aléatoire

Nous donnons ici quelques indications sur les fonctions génératrices et les fonctions caractéristiques des vecteurs aléatoires à deux dimensions.

Définition. — Soit (X, Y) un vecteur aléatoire tel que la fonction

$$g(u, v) = \mathbb{E}[e^{uX+vY}]$$

des deux variables réelles u, v soit définie dans un voisinage ouvert de $(0, 0)$. Cette fonction est appelée *fonction génératrice des moments de (la loi de) (X, Y)* .

Les deux théorèmes suivants sont les analogues des théorèmes 3.1 et 3.2 au cas de deux variables et se démontrent de la même façon.

THÉORÈME 6.1 (théorème d'unicité). — *La fonction génératrice des moments d'un vecteur aléatoire détermine la loi de ce vecteur.*

THÉORÈME 6.2. — *Soit (X, Y) un vecteur aléatoire admettant une fonction génératrice des moments $g(u, v)$ (donc définie dans un voisinage ouvert de $(0, 0)$). Alors*

1) *Pour tous entiers $k, l \geq 1$, on a $\mathbb{E}[|X|^k |Y|^l] < +\infty$;*

2) $\mathbb{E}[X^k Y^l] = \frac{\partial^{k+l}}{\partial^k u \partial^l v} g(u, v) \Big|_{(u, v) = (0, 0)}$.

THÉORÈME 6.3. — *Soit (X, Y) un vecteur aléatoire admettant une fonction génératrice des moments $g(u, v)$. Alors chacune des variables aléatoires marginales X, Y admet une fonction génératrice $g_1(u), g_2(v)$. On a, de plus : $g_1(u) = g(u, 0)$; $g_2(v) = g(0, v)$.*

La démonstration est immédiate.

Le théorème suivant donne une caractérisation de l'indépendance des variables aléatoires marginales X, Y en termes des fonctions génératrices des moments.

THÉORÈME 6.4. — *Soit (X, Y) un vecteur aléatoire admettant une fonction génératrice des moments $g(u, v)$ définie dans un voisinage ouvert V de $(0, 0)$. Alors les deux propriétés suivantes sont équivalentes.*

- 1) X et Y sont indépendantes ;
- 2) Pour tout $(u, v) \in V$ on a : $g(u, v) = g(u, 0)g(0, v)$.

Démonstration.

1) \Rightarrow 2) Supposons X, Y indépendantes ; alors e^{uX}, e^{vY} sont indépendantes et l'on a, pour tout $(u, v) \in V$ les relations : $g(u, v) = \mathbb{E}[e^{uX+vY}] = \mathbb{E}[e^{uX}]\mathbb{E}[e^{vY}] = g(u, 0)g(0, v)$.

2) \Rightarrow 1) Supposons que 2) soit vérifié. Désignons par $F(x, y)$ la fonction de répartition conjointe de (X, Y) et par $F_X(x)$ et $F_Y(y)$ ses fonctions de répartition marginales. Il vient

$$(6.1) \quad g(u, v) = \int_{\mathbb{R}^2} e^{ux+vy} dF(x, y)$$

$$g(u, 0)g(0, v) = \left(\int_{\mathbb{R}} e^{ux} dF_X(x) \right) \left(\int_{\mathbb{R}} e^{vy} dF_Y(y) \right)$$

$$(6.2) \quad = \int_{\mathbb{R}^2} e^{ux+vy} dF_X(x) dF_Y(y)$$

Comme (6.1) = (6.2) pour tout $(u, v) \in V$, il résulte alors du Théorème 2.2 (théorème d'unicité) que $F(x, y) = F_X(x)F_Y(y)$ pour tout couple (x, y) ; d'où l'indépendance de X, Y . \square

On peut également définir la fonction caractéristique d'un couple (X, Y) arbitraire par

$$\varphi(u, v) = \mathbb{E}[e^{i(uX+vY)}] \quad ((u, v) \in \mathbb{R}^2).$$

On peut établir des théorèmes analogues aux Théorèmes 6.1, 6.2, 6.3, 6.4. Nous nous bornons à énoncer le théorème suivant.

THÉORÈME 6.4'. — *Soit (X, Y) un couple de variables aléatoires réelles, dont la fonction caractéristique est notée $\varphi(u, v)$. Alors les deux propriétés suivantes sont équivalentes :*

- a) X et Y sont indépendantes.
- b) Pour tout $(u, v) \in \mathbb{R}^2$ on a : $\varphi(u, v) = \varphi(u, 0)\varphi(0, v)$.

7. Propriété fondamentale

THÉORÈME 7.1. — *La fonction caractéristique d'une mesure de probabilité détermine cette mesure.*

La démonstration de ce théorème repose sur le lemme suivant qui sera de nouveau utilisé dans le chapitre 16 sur les convergences stochastiques. Dans

ce lemme, on fait usage du produit de convolution $\mu * \nu$ de deux mesures, introduit au chap. 11, § 6 et aussi de la propriété « ancipitale » de certaines densités de probabilité.

Définition. — Soit g une densité de probabilité sur \mathbb{R} . On dit qu'elle a la propriété *ancipitale*, si, à un facteur près, elle est aussi la fonction caractéristique d'une densité de probabilité sur \mathbb{R} .

Si on se reporte au tableau des fonctions caractéristiques usuelles reproduit après le Théorème 4.1, on constate que la densité de probabilité de la loi normale $\mathcal{N}(0, 1)$ a la propriété ancipitale. Il en est de même de la densité de la loi $\mathcal{N}(0, a)$ ($a > 0$), puisque la densité de $\mathcal{N}(0, a)$ est aussi, à un facteur près, égale à la fonction caractéristique de la densité de $\mathcal{N}(0, 1/a)$

LEMME 7.2. — Soient μ une mesure de probabilité sur \mathbb{R} et g une densité de probabilité sur \mathbb{R} ayant la propriété ancipitale. Alors la mesure $\mu * g$ (c'est-à-dire la mesure $\mu * (g\lambda)$, où λ est la mesure de Lebesgue sur \mathbb{R}), admet une densité qui ne dépend de μ qu'à travers sa propre fonction caractéristique $\hat{\mu}$.

Démonstration. — En effet, par définition, il existe $c > 0$ et une densité f tels que $g(t) = c \int_{\mathbb{R}} e^{itx} f(x) dx$. Par ailleurs, la densité h de $\mu * g$ est donnée par :

$$h(u) = \int_{\mathbb{R}} g(u - v) d\mu(v) = c \int_{\mathbb{R}} \left[\int_{\mathbb{R}} e^{i(u-v)x} f(x) dx \right] d\mu(v) \quad (u \in \mathbb{R}).$$

L'intégrable double au dernier membre est absolument convergente : on peut donc appliquer le théorème de Fubini et l'on obtient, pour tout u réel :

$$(7.1) \quad h(u) = c \int_{\mathbb{R}} \left[\int_{\mathbb{R}} e^{i(u-v)x} f(x) d\mu(v) \right] dx = c \int_{\mathbb{R}} e^{iux} f(x) \hat{\mu}(-x) dx.$$

On constate que cette quantité ne dépend de μ qu'à travers $\hat{\mu}$. \square

Reprenons la démonstration du théorème 7.1.

1) Soit z réel tel que $\mu(\{z\}) = 0$. On a, en appliquant le théorème de Fubini,

$$\int_{-\infty}^z h(u) du = \int_{\mathbb{R}} \left[\int_{-\infty}^z g(u - v) du \right] d\mu(v) = \int_{\mathbb{R}} \left[\int_{-\infty}^{z-v} g(y) dy \right] d\mu(v),$$

c'est-à-dire en désignant par G la fonction de répartition de g :

$$(7.2) \quad \int_{-\infty}^z h(u) du = \int_{\mathbb{R}} G(z - v) d\mu(v).$$

2) Prenons pour g la densité d'une variable aléatoire X_a de loi $\mathcal{N}(0, a)$ ($a > 0$). Elle a la propriété ancipitale, comme remarqué plus haut. On peut

donc lui appliquer la théorie précédente pour tout $a > 0$. Or lorsque $a \downarrow 0$, la variable X_a tend en loi vers 0, ce que l'on vérifie directement en montrant que lorsque $a \downarrow 0$, on a : $G(x) \rightarrow \begin{cases} 1, & \text{si } x > 0; \\ 0, & \text{si } x < 0. \end{cases}$

3) Reprenons l'égalité (7.2) avec cette valeur de g pour $a > 0$ et faisons $a \downarrow 0$. La fonction G est positive, majorée (uniformément en $a > 0$) par 1 (qui est μ -intégrable) et l'on a, lorsque $a \downarrow 0$:

$$G(z - v) \rightarrow \begin{cases} 1, & \text{si } z - v > 0, v < z; \\ 0, & \text{si } z - v < 0, v > z; \end{cases}$$

c'est-à-dire

$$G(z - v) \rightarrow I_{]-\infty, z[}(v), \quad v \neq z.$$

Le théorème de convergence dominée s'applique et l'on a, pour $a \downarrow 0$

$$\int_{-\infty}^z h(u) du \rightarrow \int_{\mathbb{R}} I_{]-\infty, z[}(v) = \mu(]-\infty, z[).$$

Or pour tout $a > 0$ le premier membre ne dépend que de $\hat{\mu}$; il en est donc de même de sa limite pour $a \downarrow 0$. Il en résulte que $\hat{\mu}$ détermine la valeur $\mu(]-\infty, z[)$ de la fonction de répartition de μ en tout point z tel que $\mu(\{z\}) = 0$; l'ensemble de ces points est partout dense, $\hat{\mu}$ détermine la fonction de répartition de μ , donc μ . \square

COMPLÉMENTS ET EXERCICES

1. — Soient (X, Y) un couple de variables aléatoires réelles dont la loi conjointe est décrite dans la remarque suivant le Théorème 2.3 du présent chapitre et (X^*, Y^*) un couple de variables aléatoires réelles, indépendantes, uniformément réparties sur $[0, 1]$. Si Z est une variable aléatoire réelle, on notera F_Z sa fonction de répartition. Les propriétés ci-dessous se démontrent géométriquement :

- a) On a $F_{X^*} = F_X$ et $F_{Y^*} = F_Y$.
- b) Lorsque $1 \leq z \leq 2$, on a $F_{X^*+Y^*}(z) = 1 - F_{X^*+Y^*}(2 - z)$ et $F_{X+Y}(z) = 1 - F_{X+Y}(2 - z)$.
- c) Lorsque $0 \leq z \leq 1$, on a : $F_{X+Y}(z) = F_{X^*+Y^*}(z)$. [Distinguer les cas : $0 \leq z \leq 1/2$ et $1/2 \leq z \leq 1$.]
- d) De a), b) et c) découle $F_{X+Y}(z) = F_{X^*+Y^*}(z)$ pour tout z . Enfin,

$$F_{X^*+Y^*}(z) = \begin{cases} 0, & \text{si } z < 0; \\ z^2/2, & \text{si } 0 \leq z < 1; \\ 1 - (2 - z)^2/2, & \text{si } 1 \leq z < 2; \\ 1, & \text{si } z \geq 2. \end{cases}$$

2. — Soit X une variable aléatoire de loi $p\varepsilon_1 + q\varepsilon_0$ ($0 \leq p \leq 1$, $p + q = 1$). Calculer la fonction génératrice des moments de la variable centrée $X - p$, ainsi que ses moments d'ordre entier $k \geq 1$.

3. (A. Joffe). — Soit (X, Y) un couple de variables aléatoires, *indépendantes*, à valeurs positives, vérifiant $P\{X = 0\} = P\{Y = 0\} = 0$. Désignons par $g_1(u)$, $g_2(v)$ les transformées de Laplace de X , Y , respectivement définies pour $u \geq 0$ par :

$$g_1(u) = \mathbb{E}[e^{-uX}], \quad g_2(u) = \mathbb{E}[e^{-uY}] \quad (u \geq 0).$$

Alors on a les propriétés suivantes :

a) $\mathbb{E}\left[\frac{X}{X+Y}\right] = - \int_0^\infty g_1'(u)g_2(u) du.$

b) Si $g(u)$ désigne la transformée de Laplace d'une variable aléatoire X à valeurs positives vérifiant $P\{X = 0\} = 0$, alors $\lim_{u \rightarrow +\infty} g(u) = 0$.

4. (A. Joffe). — Soient X une variable aléatoire à valeurs *strictement positives* et $g(u) = \mathbb{E}[e^{-uX}]$ sa transformée de Laplace définie pour $u \geq 0$. Alors, pour tout $p > 0$, on a l'identité suivante valable dans $[0, +\infty[$:

$$\mathbb{E}\left[\frac{1}{X^p}\right] = \frac{1}{\Gamma(p)} \int_0^\infty g(u)u^{p-1} du.$$

En particulier, pour $p = \frac{1}{2}$,

$$\mathbb{E}\left[\frac{1}{\sqrt{X}}\right] = \frac{1}{\sqrt{\pi}} \int_0^\infty g(u)u^{-1/2} du.$$

Enfin, pour $p = 1$, on a :

$$\mathbb{E}\left[\frac{1}{X}\right] = \int_0^\infty g(u) du.$$

5. — Soit X une variable aléatoire admettant une fonction génératrice des moments $g(u)$. Considérons $h(u) = \text{Log } g(u)$. Montrer que $h'(0) = \mathbb{E}[X]$, $h''(0) = \text{Var } X$, $h'''(0) = \mathbb{E}[(X - \mathbb{E}[X])^3]$. On remarquera que pour $n > 3$ la quantité $h^{(n)}(0)$ n'est pas nécessairement égale au moment centré d'ordre n de X .

6. — On dit qu'une variable aléatoire réelle X , à valeurs dans $[1, +\infty[$, suit la loi de Pareto $\mathcal{P}(a, 1)$, avec $a > 0$, si elle admet une densité donnée par :

$$f(x) = \frac{a}{x^{a+1}} I_{[1, +\infty[}(x).$$

a) Les moments $\mathbb{E}[X^n]$ existent-ils pour toutes les valeurs de $n \geq 1$?

b) Pour quelles valeurs de u la fonction $g(u) = \mathbb{E}[e^{uX}]$ est-elle définie ? La variable aléatoire X admet-elle une fonction génératrice des moments ? Admet-elle une fonction caractéristique ?

7. — Calculer la fonction génératrice des moments de la variable aléatoire $|X|$, lorsque X suit une loi normale centrée réduite $\mathcal{N}(0, 1)$.

8. — Soit (X, Y) un couple de variables aléatoires indépendantes, suivant chacune une loi normale centrée réduite $\mathcal{N}(0, 1)$. Montrer que le produit XY admet une fonction génératrice des moments donnée par : $g(u) = \frac{1}{\sqrt{1-u^2}}$ ($|u| < 1$).

En remplaçant formellement u par it (t réel), on obtient la fonction caractéristique de XY , soit $\varphi(t) = 1/\sqrt{1+t^2}$. Sa densité est donnée par la formule d'inversion de Fourier

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \varphi(t) dt = \frac{1}{\pi} \int_0^{\infty} \frac{\cos tx}{\sqrt{1+t^2}} dt. = \frac{1}{\pi} K_0(x),$$

où $K_0(x)$ est la fonction de Bessel modifiée de genre 2 (cf. Abramowitz & Stegun³).

9. — Soit (X_1, X_2, X_3, X_4) un système de quatre variables aléatoires indépendantes suivant chacune une loi normale centrée réduite $\mathcal{N}(0, 1)$. Montrer que le déterminant $\Delta = \begin{vmatrix} X_1 & X_2 \\ X_3 & X_4 \end{vmatrix}$ admet une fonction génératrice donnée par : $g(u) = \frac{1}{1-u^2}$ ($|u| < 1$). C'est la fonction génératrice de la première loi de Laplace.

10. — Soient (φ_k) ($k \geq 0$) une suite de fonctions caractéristiques et (α_k) ($k \geq 0$) une suite de nombres positifs, de somme 1. Alors $\sum_{k \geq 0} \alpha_k \varphi_k$ est une fonction caractéristique. En prenant $\varphi_k = (\varphi)^k$, où φ est une fonction caractéristique, on voit que $\sum_{k \geq 0} \alpha_k (\varphi)^k$ est une fonction caractéristique.

a) Prenons $\alpha_k = e^{-\lambda} \lambda^k / k!$ ($\lambda > 0, k \in \mathbb{N}$). Si φ est une fonction caractéristique, alors, pour tout $\lambda > 0$, la fonction $\varphi_\lambda = e^{\lambda(\varphi-1)}$ est une fonction caractéristique (théorème de Finetti). Ainsi, en prenant $\lambda = 1$ et $\varphi(t) = 1/(1+t^2)$, on voit que $\exp(-t^2/(1+t^2))$ est une fonction caractéristique. Dans le cas particulier où $\varphi(t) = e^{it}$, déterminer la loi admettant φ_λ pour fonction caractéristique.

b) Prenons $\alpha_k = pq^k$, ($0 < p < 1, p+q = 1, k \in \mathbb{N}$). Montrer que si φ est une fonction caractéristique, alors, pour tout $\lambda > 1$, la fonction $\varphi_\lambda = (\lambda-1)/(\lambda-\varphi)$ est une fonction caractéristique. Dans le cas particulier où $\varphi(t) = e^{it}$, déterminer la loi admettant φ_λ pour fonction caractéristique.

11. — Soient (φ_λ) ($\lambda \in I$) une famille de fonctions caractéristiques indexée par un indice λ prenant ses valeurs dans un intervalle I non vide et f une

³ Abramowitz (Milton) & Stegun (Irene). — *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. — New York, Dover, 1968, section 9.6.21.

densité de probabilité sur I . Alors $\varphi(t) = \int_I \varphi_\lambda(t) f(\lambda) d\lambda$ est une fonction caractéristique.

a) Montrer que si φ est une fonction caractéristique, alors $\Phi(t) = (1/t) \int_0^t \varphi(u) du$ est également une fonction caractéristique (Khinchine).

b) Montrer que pour tout $\gamma > 0$ la fonction $\varphi_\gamma(t) = 1/(1+t^2)^\gamma$ est une fonction caractéristique. Pour $\gamma = 1$, c'est la fonction caractéristique de la première loi de Laplace; pour $\gamma = \frac{1}{2}$, c'est la fonction caractéristique de XY , où (X, Y) est un couple de variables aléatoires indépendantes suivant chacune la loi normale $\mathcal{N}(0, 1)$.

c) On rencontre en analyse la formule

$$e^{-|t|} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{1}{2}\left(\frac{t^2}{x^2} + x^2\right)\right) dx.$$

Ceci montre que $e^{-|t|}$ est une fonction caractéristique, puisque pour tout $x \neq 0$ la fonction $\exp\left(-\frac{1}{2}\frac{t^2}{x^2}\right)$ est une fonction caractéristique et que $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ ($x \in \mathbb{R}$) est une densité.

12. — Soient (X, Y) un couple de variables aléatoires et $Z = X + Y$. Montrer que si X est indépendante de Z et Y indépendante de Z , alors Z est presque sûrement constante.

13. — Étant donné un système de trois variables aléatoires indépendantes (X, X_1, X_2) , on lui associe le couple (Y_1, Y_2) , où $Y_1 = X + X_1$, $Y_2 = X + X_2$. Montrer que les variables aléatoires Y_1 et Y_2 sont indépendantes si et seulement si X est presque sûrement constante.

14. — Soit X une variable aléatoire admettant une fonction génératrice des moments $g(u)$. Montrer que l'on a l'inégalité de Chernoff

$$\forall x > 0 \quad \mathbb{P}\{X \geq x\} \leq \inf_{u \geq 0} e^{-ux} g(u).$$

15. — La fonction caractéristique φ d'une variable aléatoire exponentielle X de paramètre $\lambda > 0$ vérifie

$$(1) \quad |\varphi(t)|^2 = \Re \varphi(t).$$

Si $\varphi(t)$ vérifie (1), il en est de même de $\varphi(-t)$, la fonction caractéristique de la variable aléatoire $-X$. D'autre part, la seule fonction réelle qui vérifie (1) est $\varphi(t) = 1$, la fonction caractéristique de la constante 0. Existe-t-il d'autres lois de probabilité dont la fonction caractéristique vérifie (1). [Le problème reste ouvert.]

**LES PRINCIPALES LOIS DE PROBABILITÉ
(ABSOLUMENT CONTINUES)**

Nous décrivons dans ce chapitre les principales lois de probabilité admettant une densité. Pour chacune d'elles, nous donnons ses propriétés, ainsi que son champ d'application.

1. La loi uniforme sur $[0, 1]$

Définition. — Une variable aléatoire U à valeurs dans $[0, 1]$ est dite *uniformément répartie sur $[0, 1]$* si elle est absolument continue et admet pour densité :

$$f(x) = I_{[0,1]}(x).$$

La loi de densité f est appelée *loi uniforme sur $[0, 1]$* .

Les propriétés suivantes sont immédiates à établir :

$$\mathbb{E}[U] = \frac{1}{2}; \quad \text{Var } U = \frac{1}{12};$$

$$g(u) = \mathbb{E}[e^{uU}] = \frac{e^u - 1}{u} \quad (u \neq 0), \quad g(0) = 1; \quad \mathcal{L}(U) = \mathcal{L}(1 - U).$$

La loi uniforme a une grande importance dans la théorie de la simulation des lois de probabilité en raison du théorème suivant dû à P. Lévy.

THÉORÈME 1.1. — *Soit $F : \mathbb{R} \rightarrow [0, 1]$ une fonction croissante, continue à droite, et vérifiant $F(-\infty) = 0$, $F(+\infty) = 1$. On note $h :]0, 1[\rightarrow \mathbb{R}$ sa fonction inverse généralisée définie par :*

$$h(\omega) = \inf\{x : F(x) \geq \omega\} \quad (\omega \in]0, 1[).$$

Soit enfin U une variable aléatoire uniformément répartie sur $[0, 1]$. Alors la variable aléatoire $X = h \circ U$ admet F comme fonction de répartition.

Démonstration. — Il résulte de la définition que l'on a l'équivalence : $\omega \leq F(x) \Leftrightarrow h(\omega) \leq x$. On a alors pour tout réel x

$$\mathbb{P}\{X \leq x\} = \mathbb{P}\{h \circ U \leq x\} = \mathbb{P}\{U \leq F(x)\} = F(x).$$

Ceci montre que X admet F comme fonction de répartition. \square

2. La loi uniforme sur $[a, b]$

Définition. — Une variable aléatoire X à valeurs dans $[a, b]$ ($-\infty < a < b < +\infty$) est dite *uniformément répartie sur $[a, b]$* si elle est absolument continue et admet pour densité

$$f(x) = \frac{1}{b-a} I_{[a,b]}(x).$$

La loi de densité f est appelée *loi uniforme sur $[a, b]$* .

Les propriétés suivantes sont également faciles à établir :

$$\begin{aligned} \mathbb{E}[X] &= \frac{a+b}{2}; & \text{Var } X &= \frac{(a-b)^2}{12}; \\ g(u) = \mathbb{E}[e^{uX}] &= \frac{1}{b-a} \frac{e^{bu} - e^{au}}{u} \quad (u \neq 0), & g(0) &= 1. \end{aligned}$$

Si $a = -1$, $b = +1$, on a : $g(u) = \frac{\text{sh } u}{u}$ ($u \neq 0$), $g(0) = 1$.

De façon générale, si $a = -l$, $b = +l$ et $l > 0$, on a :

$$g(u) = \frac{\text{sh } lu}{lu} \quad (u \neq 0), \quad g(0) = 1.$$

3. La loi normale ou de Laplace-Gauss

3.1. La loi normale réduite

Définition. — Une variable aléatoire X à valeurs dans \mathbb{R} est dite *normale réduite* si elle est absolument continue et admet pour densité

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad (x \in \mathbb{R}).$$

La loi de densité f est appelée *normale réduite* et est notée $\mathcal{N}(0, 1)$.

Donnons une liste de propriétés de cette loi classique :

a) Le graphe (*cf.* Fig. 1) de la densité f a l'allure d'une courbe en cloche assez aplatie. Pour le voir, il suffit de remarquer que f est paire et que f admet un maximum pour $x = 0$, qui vaut $f(0) = 1/\sqrt{2\pi} \approx 0,399$. Enfin, $f''(x) = 0$ si et seulement si $x = \pm 1$.

b) On désigne par Φ la fonction de répartition de X :

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt \quad (x \in \mathbb{R}).$$

Son graphe a l'allure d'une courbe en S assez étalée et est symétrique par rapport au point $(0, 1/2)$ et la pente de la tangente en ce point est $1/\sqrt{2\pi}$.

c) On a : $\mathbb{E}[X] = 0$ et $\text{Var } X = 1$. C'est la raison pour laquelle la loi est appelée (centrée) *réduite* et notée $\mathcal{N}(0, 1)$.

d) On a : $g(u) = \mathbb{E}[e^{uX}] = e^{u^2/2}$ ($u \in \mathbb{R}$).

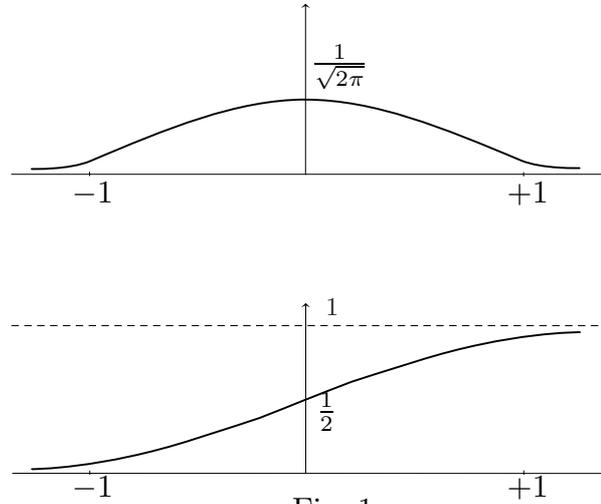


Fig. 1

Démonstration. — En effet,

$$g(u) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{ux} e^{-x^2/2} dx.$$

Or

$$ux - x^2/2 = -(x - u)^2/2 + u^2/2;$$

d'où

$$g(u) = e^{u^2/2} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(x-u)^2/2} dx;$$

en faisant le changement de variables $x - u = t$, on obtient :

$$g(u) = e^{u^2/2} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-t^2/2} dt = e^{u^2/2}. \quad \square$$

e) La fonction $g(u) = e^{u^2/2}$ est définie pour tout réel u et est développable en série entière pour tout réel u . Il en résulte que X admet des moments de tous les ordres. Le moment d'ordre n ($n \geq 0$) apparaît comme le coefficient de $u^n/n!$ dans le développement de g autour de 0. Or

$$g(u) = e^{u^2/2} = \sum_{n \geq 0} \frac{1}{n!} \left(\frac{u^2}{2}\right)^n = \sum_{n \geq 0} \frac{1}{n!} \frac{1}{2^n} u^{2n} = \sum_{n \geq 0} \frac{(2n)!}{n!} \frac{1}{2^n} \frac{u^{2n}}{(2n)!}.$$

D'où les valeurs des moments :

$$\begin{aligned} \mathbb{E}[X^{2n}] &= \frac{1}{2^n} \frac{(2n)!}{n!} = 1 \times 3 \times \cdots \times (2n-3) \times (2n-1) \quad (n \geq 1); \\ \mathbb{E}[X^{2n+1}] &= 0 \quad (n \geq 0). \end{aligned}$$

f) Pour tout $r > -1$, on a : $\mathbb{E}[|X|^r] = \frac{2^{r/2}}{\sqrt{\pi}} \Gamma\left(\frac{r+1}{2}\right).$

Démonstration. — On a :

$$\mathbb{E}[|X|^r] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} |x|^r e^{-x^2/2} dx = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} x^r e^{-x^2/2} dx;$$

d'où, en faisant le changement de variable $x^2/2 = u$

$$\mathbb{E}[|X|^r] = \frac{2^{r/2}}{\sqrt{\pi}} \int_0^{\infty} u^{(r-1)/2} e^{-u} du = \frac{2^{r/2}}{\sqrt{\pi}} \Gamma\left(\frac{r+1}{2}\right).$$

3.2. La loi normale générale

Définition. — Considérons une variable aléatoire Y de la loi $\mathcal{N}(0, 1)$ et deux nombres réels μ et $\sigma > 0$. La variable aléatoire $X = \mu + \sigma Y$ est appelée *normale de loi $\mathcal{N}(\mu, \sigma)$* (ou $\mathcal{N}_1(\mu, \sigma^2)$ dans les notations du chap. 12, § 5).

On voit que la loi de X dépend de deux paramètres μ et σ dont l'interprétation est aisée. En effet, $\mathbb{E}[X] = \mu$ et $\text{Var } X = \sigma^2 \text{Var } Y = \sigma^2$. De là $\sigma(X) = \sigma$.

Les propriétés de la loi normale générale sont décrites ci-après :

a) La densité de X est donnée par :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \quad (x \in \mathbb{R}).$$

Pour dessiner son graphe, on pourra observer que f est symétrique par rapport à $x = \mu$, que f admet un maximum pour $x = \mu$, qui vaut $f(\mu) = 1/(\sigma\sqrt{2\pi})$, enfin que $f''(x) = 0$ si et seulement si $x = \mu \pm \sigma$. Il est facile d'en conclure que ce graphe a l'allure d'une courbe en cloche symétrique par rapport à $x = \mu$, très pointue pour σ petit, très aplatie pour σ grand.

b) La fonction de répartition de X est donnée par :

$$F(x) = \text{P}\{X \leq x\} = \text{P}\left\{Y \leq \frac{x-\mu}{\sigma}\right\} = \Phi\left(\frac{x-\mu}{\sigma}\right) \quad (x \in \mathbb{R}).$$

Son graphe est symétrique par rapport au point $(\mu, 1/2)$ et la pente de la tangente en ce point est $1/(\sigma\sqrt{2\pi})$. Il en résulte que ce graphe a l'allure d'une courbe en S, très pentue pour σ petit, très étirée pour σ grand.

c) La fonction génératrice des moments est donnée par :

$$g(u) = \mathbb{E}[e^{uX}] = \exp\left(\mu u + \frac{\sigma^2 u^2}{2}\right) \quad (u \in \mathbb{R}).$$

d) Il résulte de la forme de la fonction génératrice des moments que si (X_1, X_2) est un couple de variables aléatoires indépendantes de lois respectives $\mathcal{N}(\mu_1, \sigma_1)$, $\mathcal{N}(\mu_2, \sigma_2)$, avec μ_1, μ_2 réels et $\sigma_1 > 0$, $\sigma_2 > 0$, alors la somme $X = X_1 + X_2$ a pour loi $\mathcal{N}(\mu, \sigma)$, où $\mu = \mu_1 + \mu_2$, $\sigma^2 = \sigma_1^2 + \sigma_2^2$.

Occurrences de la loi normale. — L'expérience montre qu'un grand nombre de caractères physiques, biométriques, ... suivent une loi normale.

Une des explications de ce phénomène est fournie par le théorème « central limit », dont il sera question au chapitre 18. Il y a pourtant des cas où la loi normale est contre-indiquée pour décrire un phénomène. En effet, on vérifie aisément que si X est une variable aléatoire normale, alors $Y = 1/X$ n'admet pas d'espérance mathématique (i.e. $\mathbb{E}[1/|X|] = +\infty$). Considérons, par exemple, la loi d'Ohm $I = V/R$ et supposons que la tension V soit non aléatoire et connue, mais que la résistance R suive une loi normale, ce qui, à première vue, semble raisonnable. Or ceci aurait pour conséquence que l'intensité du courant I serait une variable aléatoire dont l'espérance mathématique n'existe pas, ce qui est difficilement acceptable pour l'ingénieur. L'hypothèse que R suive une loi normale est donc irréaliste.

4. La loi Log-normale

Définition. — Une variable aléatoire X à valeurs dans $]0, +\infty[$ est dite suivre la loi Log-normale de paramètres (μ, σ) (μ réel, $\sigma > 0$) si $Y = \text{Log } X$ suit la loi $\mathcal{N}(\mu, \sigma)$.

PROPRIÉTÉS

a) La fonction de répartition de X est donnée par :

$$F(x) = \begin{cases} \Phi\left(\frac{\text{Log } x - \mu}{\sigma}\right), & \text{si } x > 0; \\ 0, & \text{sinon.} \end{cases}$$

b) La densité de X est donnée par :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{x} \exp\left(-\frac{1}{2}\left(\frac{\text{Log } x - \mu}{\sigma}\right)^2\right) I_{]0, +\infty[}(x).$$

c) On a : $\mathbb{E}[X] = e^{\mu + \sigma^2/2}$, $\text{Var } X = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1)$.

d) Si X suit la loi Log-normale de paramètres (μ, σ) , alors X^r ($r > 0$) suit la loi Log-normale de paramètres $(r\mu, r\sigma)$.

Remarque. — Il résulte de c) et de d) que pour tout $r > 0$ le moment $\mathbb{E}[X^r]$ est fini et que $\mathbb{E}[X^r] = e^{r\mu + (r^2\sigma^2)/2}$.

e) La fonction $g(u) = \mathbb{E}[e^{uX}]$ n'est pas définie pour $u > 0$; en d'autres termes, la loi Log-normale n'admet pas de fonction génératrice $g(u)$ qui soit définie dans un intervalle ouvert contenant $u = 0$. Néanmoins, X admet des moments entiers positifs de tous les ordres.

Démonstration

a) Pour tout $x > 0$, on a :

$$\begin{aligned} F(x) &= \mathbb{P}\{X \leq x\} = \mathbb{P}\{Y \leq \text{Log } x\} \\ &= \mathbb{P}\left\{\frac{Y - \mu}{\sigma} \leq \frac{\text{Log } x - \mu}{\sigma}\right\} = \Phi\left(\frac{\text{Log } x - \mu}{\sigma}\right). \end{aligned}$$

b) La densité s'obtient en dérivant $F(x)$.

c) Posons $X = e^{\sigma Y + \mu}$ où $\mathcal{L}(Y) = \mathcal{N}(0, 1)$. D'après le théorème de transfert, on obtient

$$\begin{aligned}\mathbb{E}[X] &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{\sigma x + \mu} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(x-\sigma)^2/2 + \sigma^2/2 + \mu} dx \\ &= e^{\mu + \sigma^2/2} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(x-\sigma)^2/2} dx,\end{aligned}$$

d'où en faisant le changement de variable $x - \sigma = t$,

$$\mathbb{E}[X] = e^{\mu + \sigma^2/2} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-t^2/2} dt = e^{\mu + \sigma^2/2}.$$

On obtient l'expression de la variance par un calcul analogue.

d) En effet $\text{Log } X^r = r \text{Log } X$. Or $\text{Log } X$ suit la loi $\mathcal{N}(\mu, \sigma)$; donc $r \text{Log } X$ suit la loi $\mathcal{N}(r\mu, r\sigma)$.

e) Évident. \square

Occurrence de la loi Log-normale. — La loi Log-normale a trouvé un domaine d'application inattendu : la linguistique. En effet, le nombre de mots par phrase (c'est-à-dire la longueur de la phrase mesurée en nombre de mots) suit approximativement une loi Log-normale.¹

5. La loi exponentielle

Définition. — Soit λ un nombre strictement positif; une variable aléatoire X à valeurs dans $]0, +\infty[$ est dite *exponentielle de paramètre λ* si elle est absolument continue et admet pour densité

$$f(x) = \lambda e^{-\lambda x} I_{]0, +\infty[}(x).$$

La loi de densité f est appelée *loi exponentielle de paramètre λ* ($\lambda > 0$) et est notée $\mathcal{E}(\lambda)$.

Voici quelques propriétés de la loi exponentielle.

a) La fonction de répartition de X est donnée par :

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{si } x > 0; \\ 0, & \text{sinon.} \end{cases}$$

Il est préférable de travailler avec la *fonction de survie* (ou *fonction de fiabilité*) définie par $r(x) = 1 - F(x) = P\{X > x\}$ et qui est donnée par :

$$r(x) = \begin{cases} e^{-\lambda x}, & \text{si } x > 0; \\ 1, & \text{sinon.} \end{cases}$$

b) On a : $\mathbb{E}[X] = 1/\lambda$, $\text{Var } X = 1/\lambda^2$, $\mathcal{L}(\lambda X) = \mathcal{E}(1)$.

c) La fonction génératrice des moments est donnée par :

$$g(u) = \mathbb{E}[e^{uX}] = \frac{\lambda}{\lambda - u} = \frac{1}{1 - \frac{u}{\lambda}} \quad (u \in]-\infty, \lambda[).$$

¹ Williams (C.B.). — Studies in the history of probability and statistics : a note on a early statistical study of literary style, *Biometrika*, t. **43**, 1956, p. 248–356.

Démonstration. — On a :

$$g(u) = \lambda \int_0^{\infty} e^{-x(\lambda-u)} dx.$$

L'intégrale au second membre est convergente si et seulement si $\lambda - u > 0$. Plaçons-nous dans ce cas et faisons le changement de variable $x(\lambda - u) = t$. Il vient :

$$g(u) = \lambda \int_0^{\infty} e^{-t} \frac{1}{\lambda - u} dt = \frac{\lambda}{\lambda - u}. \quad \square$$

d) La fonction $g(u) = 1/(1 - (u/\lambda))$ est définie dans l'intervalle $] -\infty, \lambda[$, ($\lambda > 0$), qui est un voisinage ouvert de l'origine; c'est donc la fonction génératrice des moments de X ; elle est développable en série entière pour tout u appartenant à l'intervalle $] -\lambda, +\lambda[$. La variable X admet des moments de tous les ordres et le moment d'ordre n ($n \geq 0$) apparaît comme le coefficient de $u^n/n!$ dans le développement de g autour de 0. Or

$$g(u) = \frac{1}{1 - \frac{u}{\lambda}} = \sum_{n \geq 0} \left(\frac{u}{\lambda}\right)^n = \sum_{n \geq 0} \frac{n!}{\lambda^n} \frac{u^n}{n!} \quad (u \in] -\lambda, +\lambda[);$$

d'où les valeurs des moments :

$$\mathbb{E}[X^n] = \frac{n!}{\lambda^n} \quad (n \geq 0).$$

e) Pour tout $r > -1$ on a : $\mathbb{E}[X^r] = \frac{\Gamma(r+1)}{\lambda^r}$.

Démonstration. — On a :

$$\mathbb{E}[X^r] = \lambda \int_0^{\infty} x^r e^{-\lambda x} dx,$$

d'où, en faisant le changement de variable $\lambda x = u$

$$\mathbb{E}[X^r] = \frac{1}{\lambda^r} \int_0^{\infty} u^r e^{-u} du = \frac{\Gamma(r+1)}{\lambda^r}.$$

Cette quantité est finie si et seulement si $r + 1 > 0$. \square

f) Si X suit la loi $\mathcal{E}(\lambda)$, alors la partie entière $Y = [X]$ suit la loi $P\{Y = n\} = pq^n$ ($n \geq 0$), où $q = e^{-\lambda}$, $p = 1 - q$; c'est une loi géométrique; la loi géométrique $\sum_{n \geq 0} pq^n \varepsilon_n$ est la version discrète de la loi exponentielle.

Démonstration. — Supposons que X suive la loi $\mathcal{E}(\lambda)$ ($\lambda > 0$). En posant $q = e^{-\lambda}$, on a, pour tout entier $n \geq 0$:

$$P\{[X] > n\} = P\{X \geq n + 1\} = e^{-\lambda(n+1)} = (e^{-\lambda})^{n+1} = q^{n+1}. \quad \square$$

g) Si X suit la loi $\mathcal{E}(\lambda)$, alors $U = e^{-\lambda X}$ suit la loi uniforme sur $]0, 1[$.

Démonstration. — Prenons u dans l'intervalle $]0, 1[$. Alors

$$\begin{aligned} P\{U \leq u\} &= P\{-\lambda X \leq \text{Log } u\} = P\left\{X > -\frac{\text{Log } u}{\lambda}\right\} \\ &= \exp\left(-\lambda\left(-\frac{\text{Log } u}{\lambda}\right)\right) = \exp \text{Log } u = u. \quad \square \end{aligned}$$

Remarque. — Il résulte de g) que si U suit une loi uniforme sur $]0, 1[$, alors $X = -(1/\lambda) \text{Log } U$, pour $\lambda > 0$, suit la loi $\mathcal{E}(\lambda)$. Cette propriété est utilisée pour *simuler* la loi $\mathcal{E}(\lambda)$.

h) *La propriété d'absence de mémoire.*

Définition. — Une variable aléatoire X , à valeurs positives, vérifiant pour tout $x \geq 0$ la propriété

$$r(x) = \text{P}\{X > x\} > 0,$$

est dite *sans mémoire*, si elle vérifie pour tous $x, y \geq 0$ l'identité

$$(5.1) \quad \text{P}\{X > x + y \mid X > y\} = \text{P}\{X > x\}.$$

Interprétons X comme la durée de vie d'un individu A . Cette propriété exprime que A *ne vieillit pas* : si A a vécu au moins y unités de temps, alors il vivra encore x unités de temps supplémentaires avec la même probabilité qu'un individu analogue à A qui viendrait de naître. La propriété (5.1) est donc une propriété de *non-vieillesse* ou *d'absence de mémoire* (l'individu A ne se souvient pas d'avoir vieilli). Il est remarquable qu'elle soit équivalente au fait que X suive une loi exponentielle.

THÉORÈME 5.1. — *Soit X une variable aléatoire absolument continue, à valeurs positives, vérifiant, pour tout $x \geq 0$, la propriété :*

$$r(x) = \text{P}\{X > x\} > 0.$$

Alors les deux propriétés suivantes sont équivalentes :

- (a) X suit une loi exponentielle.
- (b) X est « sans mémoire ».

Démonstration. — Remarquons que la propriété (b) est équivalente à la propriété

$$(b') \quad \forall x, y \geq 0 \quad r(x + y) = r(x)r(y).$$

Il suffit donc de démontrer que (a) \Leftrightarrow (b').

(a) \Rightarrow (b'). Supposons que X suive la loi $\mathcal{E}(\lambda)$ ($\lambda > 0$); alors, pour tout $x > 0$, on a $r(x) = e^{-\lambda x}$, qui vérifie (b').

(b') \Rightarrow (a). L'équation (b') est l'équation fonctionnelle de l'exponentielle; la fonction $r(x)$ étant *continue à droite*, sa solution est de la forme $r(x) = e^{\alpha x}$ ($\alpha \in \mathbb{R}$, $x > 0$) (cf. exercice 1). Or $r(x)$ doit être décroissante, d'où $\alpha = -\lambda$ ($\lambda \geq 0$). Comme le cas $\lambda = 0$ ne correspond pas à une vraie variable aléatoire, on a $\lambda > 0$. \square

Occurrences de la loi exponentielle

1) On a de bonnes raisons de supposer que la *durée de vie* X d'un appareil (d'un organisme, d'une ampoule, d'un atome radioactif, ...) suit une loi exponentielle. Prenons pour origine l'instant où cet appareil est mis en

route; la quantité $r(x) = P\{X > x\}$ ($x > 0$) peut alors être interprétée comme la probabilité pour que cet appareil ait encore été en fonctionnement à l'instant x (fonction de *survie*) ou comme la probabilité pour qu'on ait pu s'y fier jusqu'à cet instant (fonction de *fiabilité*). Il faut toutefois garder à l'esprit que si l'on suppose que X suit une loi exponentielle, on suppose de façon équivalente que X a la propriété de *non-vieillesse* (ou d'absence de mémoire). Une loi qui décrit la réalité d'une façon un peu plus fidèle est la loi gamma de paramètres (p, λ) avec $p > 1$ et $\lambda > 0$ (cf. Exercice 7).

2) Soit (X_1, X_2) un couple de variables aléatoires *indépendantes* dont chacune suit la loi $\mathcal{N}(0, 1)$. Alors la variable aléatoire $Y = X_1^2 + X_2^2$ (le « chi-deux » à deux degrés de liberté) suit la loi $\mathcal{E}(1/2)$.

6. La première loi de Laplace

Définition. — Une variable aléatoire X à valeurs réelles est dite suivre la *première loi de Laplace*, si elle est absolument continue et admet pour densité :

$$f(x) = \frac{1}{2}e^{-|x|} \quad (x \in \mathbb{R}).$$

Les propriétés de la première loi de Laplace sont les suivantes.

a) On a : $g(u) = \mathbb{E}[e^{uX}] = \frac{1}{1-u^2} \quad (-1 < u < +1)$.

Démonstration. — En effet, pour $-1 < u < +1$, on a :

$$\begin{aligned} g(u) &= \frac{1}{2} \int_{\mathbb{R}} e^{ux-|x|} dx = \frac{1}{2} \left[\int_0^{\infty} e^{-(1-u)x} dx + \int_{-\infty}^0 e^{(1+u)x} dx \right] \\ &= \frac{1}{2} \left[\int_0^{\infty} e^{-(1-u)x} dx + \int_0^{\infty} e^{-(1+u)x} dx \right] \\ &= \frac{1}{2} \left[\frac{1}{1-u} + \frac{1}{1+u} \right] = \frac{1}{1-u^2}. \quad \square \end{aligned}$$

b) La variable aléatoire X admet des moments de tous les ordres entiers positifs; le moment d'ordre $n \geq 0$ apparaît comme coefficient de $(u^n/n!)$ dans le développement de g autour de 0. De

$$g(u) = \frac{1}{1-u^2} = \sum_{n \geq 0} (2n)! \frac{u^{2n}}{(2n)!} \quad (-1 < u < +1),$$

on déduit

$$\mathbb{E}[X^{2n}] = (2n)!, \quad \mathbb{E}[X^{2n+1}] = 0; \quad (n \geq 0).$$

c) Comme $g(u) = \frac{1}{1-u^2} = \frac{1}{1-u} \frac{1}{1+u}$ ($-1 < u < +1$) et puisque $1/(1-u)$ est la fonction génératrice des moments d'une variable aléatoire Y de loi $\mathcal{E}(1)$, on voit que la *symétrisée* de Y suit la première loi de Laplace.

Occurrences de la première loi de Laplace. — Cette loi, qui a été proposée en premier lieu par Laplace² pour rendre compte des erreurs d'expérience, a été supplantée par la loi normale, appelée aussi *seconde loi de Laplace*. Il est intéressant, dans cette perspective, de comparer les densités de ces deux lois.

PROPOSITION 6.1. — *Posons :*

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad g(x) = \frac{1}{2} e^{-|x|} \quad (x \in \mathbb{R}).$$

On a alors pour tout réel x

$$f(x) \leq c g(x), \quad \text{avec } c = \sqrt{\frac{2e}{\pi}}.$$

Démonstration. — En effet,

$$\frac{1}{2}(|x| - 1)^2 = \frac{x^2}{2} + \frac{1}{2} - |x|, \quad \text{d'où } -\frac{x^2}{2} \leq \frac{1}{2} - |x|;$$

et par conséquent

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \leq \frac{1}{\sqrt{2\pi}} e^{((1/2) - |x|)} = \sqrt{\frac{2e}{\pi}} \frac{1}{2} e^{-|x|} = c g(x). \quad \square$$

7. La loi de Cauchy

Définition. — Une variable aléatoire à valeurs réelles est dite suivre la *loi de Cauchy* $C(0, 1)$ si elle est absolument continue et admet pour densité

$$f(x) = \frac{1}{\pi} \frac{1}{1 + x^2} \quad (x \in \mathbb{R}).$$

Voici quelques propriétés de la loi de Cauchy.

a) Le graphe de f ressemble à celui de la densité de la loi normale, mais approche l'axe des x si lentement que l'espérance mathématique de X *n'existe pas*.

b) *Interprétation de la notation* $C(0, 1)$. — Les nombres 0 et 1 ne sont plus, ici, l'espérance mathématique et l'écart-type (qui n'existent pas), mais admettent l'interprétation suivante :

$M = 0$ est la *médiane* de X ; en effet, $P\{X \leq 0\} = P\{X \geq 0\} = 1/2$;

$Q_1 = -1$, $Q_3 = +1$ sont respectivement le *premier* et le *troisième quartile* de X ; en effet, $P\{X \leq -1\} = P\{X \geq +1\} = 1/4$. On prend l'*écart interquartile* $Q_3 - Q_1 = 2$ comme mesure de la dispersion.

c) La variable X n'admet pas de fonction génératrice des moments ; en effet, l'expression

$$g(u) = \frac{1}{\pi} \int_{\mathbb{R}} \frac{e^{ux}}{1 + x^2} dx$$

² Laplace (Pierre-Simon, marquis de). — *Théorie analytique des Probabilités*. — Paris, 1814.

ne prend une valeur finie pour aucun nombre réel $u \neq 0$. Il convient ici d'introduire la fonction caractéristique qui est donnée par $\varphi(t) = e^{-|t|}$ ($t \in \mathbb{R}$). On constate qu'elle n'est *pas* dérivable pour $t = 0$, ce qui révèle le caractère « pathologique » de la loi $C(0, 1)$.

d) Considérons une variable aléatoire Y de loi $C(0, 1)$ et deux nombres réels α, β avec $\beta > 0$. On dit que la variable aléatoire $X = \alpha + \beta Y$ suit la *loi de Cauchy* $C(\alpha, \beta)$. Sa densité est donnée par :

$$f(x) = \frac{1}{\pi\beta} \frac{1}{1 + \left(\frac{x - \alpha}{\beta}\right)^2} \quad (x \in \mathbb{R}).$$

Les paramètres α, β admettent les interprétations suivantes : $M = \alpha$ est la médiane de X ; de plus, $Q_1 = \alpha - \beta$; $Q_3 = \alpha + \beta$ sont respectivement le *premier* et le *troisième quartile* de X ; $Q_3 - Q_1 = 2\beta$ est l'écart interquartile de X .

e) La fonction caractéristique de la loi $C(\alpha, \beta)$ est donnée par :

$$\varphi(t) = e^{it\alpha - \beta|t|} \quad (t \in \mathbb{R}).$$

Il résulte de la forme de cette fonction caractéristique que si (X_1, X_2) est un couple de variables aléatoires indépendantes de lois respectives $C(\alpha_1, \beta_1)$, $C(\alpha_2, \beta_2)$ (α_1, α_2 réels et $\beta_1 > 0, \beta_2 > 0$), alors la somme $X = X_1 + X_2$ a pour loi $C(\alpha, \beta)$, où $\alpha = \alpha_1 + \alpha_2, \beta = \beta_1 + \beta_2$.

Cette propriété a les conséquences surprenantes suivantes : soient X une variable aléatoire de loi $C(\alpha, \beta)$ et (X_1, \dots, X_n) un système de n variables aléatoires indépendantes admettant toutes même loi que X . Alors

la somme $X_1 + \dots + X_n$ a même loi que nX ;

la moyenne arithmétique $\bar{X} = (X_1 + \dots + X_n)/n$ a même loi que X .

Occurrences de loi de Cauchy

PROPOSITION 7.1. — *Soit V une variable aléatoire uniformément répartie sur $] -\pi/2, +\pi/2[$. Alors $X = \operatorname{tg} V$ suit la loi $C(0, 1)$.*

Démonstration. — Pour tout x réel on a :

$$P\{X \leq x\} = P\{V \leq \operatorname{Arctg} x\} = \frac{1}{\pi} \left(\frac{\pi}{2} + \operatorname{Arctg} x \right) = \frac{1}{2} + \frac{1}{\pi} \operatorname{Arctg} x ;$$

d'où, en dérivant, $f(x) = \frac{1}{\pi} \frac{1}{1 + x^2}$. \square

Remarque. — On vérifie sans peine que $1/\operatorname{tg} V$ a même fonction de répartition que $\operatorname{tg} V$. Il en résulte que *si X suit la loi $C(0, 1)$, il en est de même de $1/X$.*

Remarque. — La loi de $X = |Y|$, où Y suit la loi $C(0, 1)$, est appelée *loi de Lorentz* dans les ouvrages de physique.

8. La loi gamma. — Rappelons que la *fonction gamma* (ou seconde fonction d'Euler) est définie pour tout nombre réel $p > 0$ par

$$\Gamma(p) = \int_0^{\infty} e^{-x} x^{p-1} dx$$

et a les propriétés élémentaires : $\Gamma(p+1) = p\Gamma(p)$ ($p > 0$); $\Gamma(n) = (n-1)!$ (n entier ≥ 1); $\Gamma(1) = 1$; $\Gamma(1/2) = \sqrt{\pi}$.

La fonction gamma permet d'introduire une classe de lois de probabilité dépendant de deux paramètres strictement positifs dont le domaine d'application est extrêmement vaste.

Définition. — Une variable aléatoire X à valeurs dans $[0, +\infty[$ est dite suivre la loi $\Gamma(p, \lambda)$ (*gamma de paramètres* $p > 0, \lambda > 0$) si elle est absolument continue et admet pour densité :

$$f(x) = \begin{cases} \frac{\lambda}{\Gamma(p)} e^{-\lambda x} (\lambda x)^{p-1}, & \text{si } x \geq 0; \\ 0, & \text{sinon.} \end{cases}$$

On voit que pour $p = 1$ la loi $\Gamma(1, \lambda)$ coïncide avec la loi $\mathcal{E}(\lambda)$.

Les propriétés de la loi gamma sont les suivantes :

- a) $\mathbb{E}[X] = \frac{p}{\lambda}, \quad \text{Var } X = \frac{p}{\lambda^2}.$
 b) $g(u) = \mathbb{E}[e^{uX}] = \left(\frac{\lambda}{\lambda - u}\right)^p \quad (u \in]-\infty, \lambda[).$

Démonstration. — L'intégrale dans l'équation

$$g(u) = \frac{\lambda}{\Gamma(p)} \int_0^\infty e^{-x(\lambda-u)} (\lambda x)^{p-1} dx$$

est convergente si et seulement si $\lambda - u > 0$. Plaçons-nous dans ce cas et faisons le changement de variable $x(\lambda - u) = t$. On obtient :

$$\begin{aligned} g(u) &= \frac{\lambda}{\Gamma(p)} \int_0^\infty e^{-t} \left(\frac{\lambda t}{\lambda - u}\right)^{p-1} \frac{dt}{\lambda - u} = \left(\frac{\lambda}{\lambda - u}\right)^p \frac{1}{\Gamma(p)} \int_0^\infty e^{-t} t^{p-1} dt \\ &= \left(\frac{\lambda}{\lambda - u}\right)^p. \quad \square \end{aligned}$$

c) Il résulte de la forme de la fonction génératrice des moments que si (X_1, X_2) est un couple de variables aléatoires indépendantes de lois respectives $\Gamma(p_1, \lambda), \Gamma(p_2, \lambda)$ ($p_1, p_2, \lambda > 0$, avec le même λ), alors la somme $X_1 + X_2$ a pour loi $\Gamma(p_1 + p_2, \lambda)$.

Passons en revue certaines lois obtenues en faisant varier les paramètres p, λ dans la loi gamma.

- 1) La loi $\Gamma(1, \lambda)$ ($\lambda > 0$) coïncide avec la loi $\mathcal{E}(\lambda)$.
- 2) La loi $\Gamma(n, \lambda)$ (n entier $> 0, \lambda > 0$) coïncide avec la loi de la somme de n variables aléatoires indépendantes admettant toutes la loi $\mathcal{E}(\lambda)$.
- 3) Soit Y une variable aléatoire de loi $\mathcal{N}(0, 1)$. Alors $X = Y^2$ a pour loi $\Gamma(1/2, 1/2)$.

Démonstration. — La fonction génératrice des moments de X est égale à :

$$g(u) = \mathbb{E}[e^{uX}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{uy^2 - y^2/2} dy = \frac{2}{\sqrt{2\pi}} \int_0^\infty e^{-(y^2/2)(1-2u)} dy.$$

L'intégrale au dernier membre est convergente si et seulement si $1 - 2u > 0$. Plaçons-nous dans ce cas et faisons le changement de variable $y\sqrt{1 - 2u} = t$. Il vient :

$$g(u) = \frac{1}{\sqrt{1 - 2u}} \frac{2}{\sqrt{2\pi}} \int_0^\infty e^{-t^2/2} dt = \frac{1}{\sqrt{1 - 2u}} \quad (u \in] - \infty, 1/2[).$$

Or ceci est la fonction génératrice des moments de la loi $\Gamma(1/2, 1/2)$. \square

4) Soit (Y_1, \dots, Y_n) un système de n variables aléatoires indépendantes admettant toutes la loi $\mathcal{N}(0, 1)$. Alors $X = Y_1^2 + \dots + Y_n^2$ a pour loi $\Gamma(n/2, 1/2)$. Cette loi est appelée la loi du *chi-deux* à n degrés de liberté et est notée χ_n^2 . Pour $n = 2$ c'est la loi $\Gamma(1, 1/2)$, qui coïncide avec la loi $\mathcal{E}(1/2)$. On voit que si (X_1, X_2) est un couple de variables aléatoires, indépendantes, de lois respectives $\chi_{n_1}^2, \chi_{n_2}^2$ avec n_1, n_2 entiers strictement positifs, alors la somme $X_1 + X_2$ suit la loi $\chi_{n_1+n_2}^2$ (loi d'addition du chi-deux).

Occurrences de la loi gamma. — Nous avons vu que l'on a de bonnes raisons de supposer que la durée de vie d'un organisme (d'un appareil, d'un atome radioactif, ...) suit une loi exponentielle; l'inconvénient, en faisant cette hypothèse, est que l'on suppose, de façon implicite, l'absence de vieillissement de cet organisme (de cet appareil, de cet atome radioactif, ...). Une hypothèse plus réaliste consiste à supposer que la durée de vie suit une loi gamma $\Gamma(p, \lambda)$, avec p légèrement plus grand que 1 et $\lambda > 0$ (cf. Exercice 7).

9. La loi bêta. — La fonction bêta (première fonction d'Euler) $B(r, s)$ est définie pour tout couple (r, s) de réels strictement positifs par :

$$B(r, s) = \int_0^1 x^{r-1}(1-x)^{s-1} dx.$$

Elle est reliée à la fonction gamma par la relation :

$$B(r, s) = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}.$$

Il en résulte que $B(r, s) = B(s, r)$.

Définition. — Une variable aléatoire à valeurs dans $[0, 1]$ est dite suivre la loi $B(r, s)$ (bêta de paramètres $r > 0, s > 0$) si elle est absolument continue et admet pour densité :

$$f(x) = \frac{1}{B(r, s)} x^{r-1}(1-x)^{s-1} I_{[0,1]}(x).$$

On voit que pour $r = s = 1$ la loi $B(1, 1)$ coïncide avec la loi uniformément répartie sur $[0, 1]$.

L'espérance mathématique et la variance sont données par :

$$\mathbb{E}[X] = \frac{r}{r+s}, \quad \text{Var } X = \frac{rs}{(r+s)^2(r+s+1)}.$$

On constate que $\text{Var } X < \mathbb{E}[X][1 - \mathbb{E}[X]]$.

Occurrences de la loi bêta. — Soit (X_1, \dots, X_n) un système de n variables aléatoires indépendantes toutes uniformément distribuées sur $]0, 1[$. Alors les variables aléatoires

$$Y = \min(X_1, \dots, X_n), \quad Z = \max(X_1, \dots, X_n)$$

suivent respectivement les lois $B(1, n)$ et $B(n, 1)$ (cf. exercice 5). Les statistiques d'ordre associées à ce système suivent également des lois bêta.

10. — Soit $g(x) = \exp\left(-\left(\frac{1}{x} + \frac{1}{1-x}\right)\right) I_{]0,1[}(x)$,

$$f(x) = \frac{g(x)}{\int_{\mathbb{R}} g(t) dt}.$$

On vérifie que f est la densité de probabilité d'une variable aléatoire à valeurs dans $]0, 1[$, et qui est C^∞ dans tout \mathbb{R} .

11. Les lois de l'Arc sinus

a) La loi $B(1/2, 1/2)$, de densité $f(x) = \frac{1}{\pi} \frac{1}{\sqrt{x(1-x)}}$ ($0 < x < 1$) est appelée «loi de l'Arc sinus», en raison du fait que sa fonction de répartition est donnée par $F(x) = (2/\pi) \text{Arc sin}(\sqrt{x})$ ($0 < x < 1$). Elle occupe une place importante dans la théorie des fluctuations.

b) On rencontre d'autres lois, également liées à l'Arc sinus, qui font l'objet de la définition suivante :

Définition. — Une variable aléatoire X , à valeurs dans $] -1, +1[$, est dite suivre la loi (A_1) de l'Arc sinus si elle est absolument continue et admet pour densité :

$$f_X(x) = \frac{1}{\pi} \frac{1}{\sqrt{1-x^2}} \quad (-1 < x < +1). \quad (A_1)$$

La variable aléatoire $|X|$, à valeurs dans $]0, 1[$, suit la loi de densité :

$$f_{|X|}(x) = \frac{2}{\pi} \frac{1}{\sqrt{1-x^2}} \quad (0 < x < +1). \quad (A_2)$$

On l'appelle la loi (A_2) de l'Arc sinus.

Occurrences des lois (A_1) , (A_2) . — Les variables aléatoires suivantes suivent la loi (A_1) :

$\sin U$, où U est uniforme dans $] -\pi/2, +\pi/2[$;

$\cos U$, où U est uniforme dans $]0, \pi[$;

$\sin U$, $\cos U$ où U est uniforme dans $]0, 2\pi[$.

En outre, la loi (A_1) étant symétrique, chacune de ces variables aléatoires a même loi que son opposée.

Les variables aléatoires suivantes suivent la loi (A_2) :

$\sin U$, où U est uniforme dans $]0, \pi[$;

$\cos U$, où U est uniforme dans $] - \pi/2, +\pi/2[$.

Ces exemples montrent, s'il en était besoin, que des variables aléatoires différentes peuvent avoir même loi.

Soit U une variable aléatoire uniforme dans $]0, 2\pi[$; alors $X = \cos U$, $Y = \sin U$ suivent toutes deux la loi (A_1) . En outre, $X^2 - Y^2 = \cos^2 U - \sin^2 U = \cos(2U)$ et $2XY = 2 \sin U \cos U = \sin(2U)$ suivent également la loi (A_1) . Comme $X^2 + Y^2 = 1$, la variable aléatoire

$$Z = X^2 - Y^2 = \frac{X^2 - Y^2}{X^2 + Y^2} = \frac{1 - \left(\frac{Y}{X}\right)^2}{1 + \left(\frac{Y}{X}\right)^2} = \frac{1 - \operatorname{tg}^2 U}{1 + \operatorname{tg}^2 U}$$

suit la loi (A_1) . Comme U est uniforme sur $]0, 2\pi[$, la variable aléatoire $T = \operatorname{tg} U$ suit la loi de Cauchy $\mathcal{C}(0, 1)$. On a donc démontré la proposition suivante.

PROPOSITION 11.1. — *Soit T une variable aléatoire de loi $\mathcal{C}(0, 1)$; alors la variable aléatoire $Z = \frac{1 - T^2}{1 + T^2}$ suit la loi (A_1) de l'Arc sinus.*

COMPLÉMENTS ET EXERCICES

1. — Les seules solutions $r(\cdot)$, continues à droite, non identiquement nulles, de l'identité $[\forall x, y \geq 0, \quad r(x + y) = r(x)r(y)]$ sont de la forme $r(x) = e^{\alpha x}$ ($\alpha \in \mathbb{R}$).

2. *La loi de Pareto.* — Soit Y une variable aléatoire exponentielle de loi $\mathcal{E}(\lambda)$ ($\lambda > 0$). La variable aléatoire $X = e^Y$ est appelée *variable aléatoire de Pareto*, de loi de Pareto $\mathcal{P}(\lambda, 1)$.

a) La fonction de survie de X est donnée par $r(x) = \begin{cases} x^{-\lambda}, & \text{si } x \geq 1; \\ 1, & \text{si } x < 1. \end{cases}$
et sa densité par $f(x) = \begin{cases} \lambda/x^{\lambda+1}, & \text{si } x \geq 1; \\ 0, & \text{sinon.} \end{cases}$

b) On a : $\mathbb{E}[X] = \frac{\lambda}{\lambda - 1}$, si $\lambda > 1$ et $\mathbb{E}[X] = +\infty$, si $\lambda \leq 1$.

c) Pour tout entier $k \geq 1$ évaluer $\mathbb{E}[X^k]$ en déterminant la fonction de survie de X^k . En déduire qu'une variable aléatoire de Pareto n'admet pas de fonction génératrice des moments.

3. *La loi de Weibull.* — Une variable aléatoire X est dite *de Weibull de paramètres* (α, λ) ($\alpha > 0, \lambda > 0$), si la variable X^α suit une loi exponentielle

de paramètre $\lambda > 0$. Sa fonction de survie est donnée par $r(x) = e^{-\lambda x^\alpha}$ si $x \geq 0$ et 1 si $x < 0$; sa densité par $f(x) = \alpha \lambda e^{-\lambda x^\alpha} x^{\alpha-1}$ si $x \geq 0$ et 0 si $x < 0$; enfin, son espérance mathématique par $\mathbb{E}[X] = (1/\lambda^{1/\alpha})\Gamma(1 + (1/\alpha))$.

4. *La loi logistique standard.* — Une variable aléatoire X à valeurs dans \mathbb{R} est dite *logistique standard*, si elle est de la forme $X = -\text{Log}(e^Y - 1)$, où Y est une variable aléatoire exponentielle, de paramètre 1. Pour tout x réel, sa fonction de répartition est donnée par $F(x) = 1/(1 + e^{-x})$, sa fonction de survie par $r(x) = e^{-x}/(1 + e^{-x})$ et sa densité par $f(x) = F(x)r(x) = (1/2)/(1 + \text{ch } x)$. On voit que $f(x)$ est paire, on montre que $\text{Log } f(x)$ est concave.

5. — Soit (X_1, \dots, X_n) un système de n variables aléatoires indépendantes toutes uniformément distribuées sur $[0, 1]$. Alors les variables aléatoires $Y = \min(X_1, \dots, X_n)$ et $Z = \max(X_1, \dots, X_n)$ suivent respectivement les lois $B(1, n)$ et $B(n, 1)$.

6. *La moyenne géométrique.* — Soit X une variable aléatoire telle que $\mathbb{E}[|X|^r] < +\infty$ pour tout r appartenant à $[0, r_0[$ ($r_0 > 0$). Considérons, pour tout r appartenant à l'intervalle ouvert $]0, r_0[$, l'écart $e_r = (\mathbb{E}[|X|^r])^{1/r}$. D'après les inégalités sur les moyennes, la fonction $r \rightarrow e_r$ est croissante; il en résulte que la limite $\lim_{r \downarrow 0} e_r$ existe et est finie; nous la désignerons par e_0 et l'appellerons la *moyenne géométrique de X* .

a) Soit X une variable aléatoire de loi $\mathcal{N}(0, 1)$; calculer sa moyenne géométrique.

b) Soit X une variable aléatoire de loi $\mathcal{E}(\lambda)$; calculer sa moyenne géométrique.

c) Soit X une variable aléatoire de loi de Cauchy $\mathcal{C}(0, 1)$; montrer qu'elle admet une moyenne géométrique.

d) Construire la loi d'une variable aléatoire qui n'admet *pas* de moyenne géométrique.

7. — Soit X une variable aléatoire à valeurs dans $[0, +\infty[$, absolument continue de densité f ; supposons que sa fonction de survie $r(x) = \mathbb{P}\{X > x\}$ soit strictement positive pour tout $x > 0$. On appelle *taux de défaillance* de X la fonction $\rho(x) = f(x)/r(x)$. Prenons pour X une variable aléatoire de loi $\Gamma(p, \lambda)$ ($p, \lambda > 0$). Montrer que son taux de défaillance est strictement croissant, constant ou strictement décroissant, suivant que $p > 1$, $p = 1$ ou $0 < p < 1$. On notera que le cas $p = 1$ correspond à la loi $\mathcal{E}(\lambda)$.

8. — Soit X une variable aléatoire de loi $\Gamma(p, \lambda)$ ($p, \lambda > 0$). Montrer que pour tout r tel que $p + r > 0$ on a :

$$\mathbb{E}[X^r] = \frac{1}{\lambda^r} \frac{\Gamma(p+r)}{\Gamma(p)}.$$

9. a) Soit X une variable aléatoire suivant la première loi de Laplace, c'est-à-dire de densité $f(x) = (1/2) \exp(-|x|)$ ($x \in \mathbb{R}$). Calculer la fonction génératrice de $|X|$; en déduire sa loi.
- b) Soit (X_1, X_2) un couple de variables aléatoires indépendantes de loi $\mathcal{E}(1)$. Calculer les fonctions génératrices des variables aléatoires $X_1 + X_2$, $X_1 - X_2$, $|X_1 - X_2|$; en déduire leurs lois.
- c) La première loi de Laplace \mathcal{L} admet pour fonction génératrice $g(u) = 1/(1-u^2)$ ($|u| < 1$). Cette fonction peut se factoriser de deux manières différentes :

$$\alpha) \quad \frac{1}{1-u^2} = \frac{1}{1-u} \frac{1}{1+u}; \quad \beta) \quad \frac{1}{1-u^2} = \left(\frac{1}{\sqrt{1-u^2}} \right)^2.$$

α) Montrer que \mathcal{L} est la loi de la symétrisée d'une variable aléatoire de loi $\mathcal{E}(1)$.

β) Montrer que \mathcal{L} est la loi de $X_1 X_2 + X_3 X_4$ (ou de $\begin{vmatrix} X_1 & X_2 \\ X_3 & X_4 \end{vmatrix}$), où (X_1, X_2, X_3, X_4) est un système de quatre variables aléatoires indépendantes de la loi $\mathcal{N}(0, 1)$ (cf. Exercice 9 du chap. 13).

On est ici en présence d'un phénomène qui ne se produit pas pour une loi normale : une loi normale n'admet pas de « facteurs » non dégénérés autres que normaux. Ceci est un argument supplémentaire qui justifie le remplacement de la première loi de Laplace par la seconde loi, à savoir la loi normale, dans la théorie des erreurs.

10. a) Soit (X, Y) un couple de variables aléatoires, *indépendantes*, à *valeurs positives*, où Y suit la loi exponentielle $\mathcal{E}(\lambda)$ ($\lambda > 0$). Désignons par $L(u)$ la transformée de Laplace de X , c'est-à-dire $L(u) = \mathbb{E}[e^{-uX}]$ ($u \geq 0$). Montrer que $P\{Y > X\} = L(\lambda)$.
- b) Soit (X_1, \dots, X_n, Y) un système de $(n+1)$ variables aléatoires *indépendantes*, à *valeurs positives*, où Y suit la loi exponentielle $\mathcal{E}(\lambda)$ ($\lambda > 0$). Montrer que

$$P\{Y > X_1 + \dots + X_n\} = P\{Y > X_1\} \dots P\{Y > X_n\}.$$

Cas particulier. — Pour $n = 2$ on a

$$(1) \quad P\{Y > X_1 + X_2\} = P\{Y > X_1\} P\{Y > X_2\}.$$

Or $P\{Y > X\} = L(\lambda) > 0$; la relation (1) est donc équivalente à la suivante

$$(2) \quad P\{Y > X_1 + X_2 \mid Y > X_2\} = P\{Y > X_1\}.$$

Si (X_1, X_2, Y) est un système de variables aléatoires *indépendantes*, à *valeurs positives*, où Y suit une loi exponentielle, on a alors la propriété d'absence de mémoire sous la forme généralisée (2).

- c) Soit (X_1, \dots, X_n) un système de variables aléatoires *indépendantes*, admettant toutes la loi $\mathcal{E}(\lambda)$ ($\lambda > 0$). On pose $M = \max_{1 \leq k \leq n} X_k$. Calculer $P\{M > \sum_{k=1}^n X_k - M\}$.

11. — Soit X une variable aléatoire uniformément répartie sur $[0, 1]$. Calculer son écart e_r d'ordre r par rapport à l'origine; en déduire sa moyenne géométrique e_0 .

12. — Soit (Y_1, \dots, Y_n) un système de n variables aléatoires indépendantes, identiquement distribuées, de loi commune $\mathcal{N}(0, 1)$. La variable aléatoire $X^2 = Y_1^2 + \dots + Y_n^2$ suit la loi χ_n^2 , c'est-à-dire la loi $\Gamma(n/2, 1/2)$, dont nous désignerons la densité par g . La variable aléatoire $X = \sqrt{Y_1^2 + \dots + Y_n^2}$ admet alors la loi de densité

$$f(x) = 2xg(x^2) = \frac{1}{2^{(n/2)-1}} \frac{1}{\Gamma(n/2)} x^{n-1} e^{-x^2/2} \quad (x \geq 0).$$

a) Pour $n = 2$, la loi de $X = \sqrt{Y_1^2 + Y_2^2}$ a pour densité $f(x) = xe^{-x^2/2}$ ($x \geq 0$) (loi de Rayleigh). Elle coïncide avec la loi de Weibull de paramètres $\alpha = 2$, $\lambda = 1/2$.

b) Pour $n = 3$, la loi de $X = \sqrt{Y_1^2 + Y_2^2 + Y_3^2}$ a pour densité $f(x) = \sqrt{\frac{2}{\pi}} x^2 e^{-x^2/2}$ ($x \geq 0$) (loi de Maxwell).

c) La densité conjointe de (Y_1, \dots, Y_n) est donnée par

$$h(y_1, \dots, y_n) = \frac{1}{(\sqrt{2\pi})^n} \exp\left(-\frac{y_1^2 + \dots + y_n^2}{2}\right) = \frac{1}{(\sqrt{2\pi})^n} e^{-x^2/2},$$

où l'on a posé $x^2 = y_1^2 + \dots + y_n^2$. On voit que $h(y_1, \dots, y_n)$ ne dépend que de $x \geq 0$; on la notera $h(x)$. Désignons par $A_n(x)$ l'aire de la sphère $S_n(0, x)$ ($x > 0$). Alors $f(x) = A_n(x)h(x)$, c'est-à-dire

$$\frac{1}{2^{(n/2)-1}} \frac{1}{\Gamma(n/2)} x^{n-1} e^{-x^2/2} = A_n(x) \frac{1}{2^{n/2} \pi^{n/2}} e^{-x^2/2},$$

d'où

$$A_n(x) = 2 \frac{\pi^{n/2}}{\Gamma(n/2)} x^{n-1}.$$

Le volume $V_n(x)$ de la boule $B_n(0, x)$ ($x > 0$) s'obtient alors immédiatement :

$$V_n(x) = \int_0^x A_n(t) dt = \frac{\pi^{n/2}}{\Gamma(1 + n/2)} x^n.$$

13. a) Soit (X_1, X_2) un couple de variables aléatoires indépendantes uniformément réparties sur $[0, 1]$. Déterminer la densité et la fonction caractéristique de $X = X_1 - X_2$ et de $2X$.

b) Soit (Y_1, Y_2) un couple de variables aléatoires indépendantes uniformément réparties sur $[-1, +1]$. Posons $Y = Y_1 + Y_2$. Montrer que $\mathcal{L}(Y) = \mathcal{L}(2X)$.

**LOIS DE PROBABILITÉ DE FONCTIONS
DE VARIABLES ALÉATOIRES**

Nous nous proposons dans ce chapitre de déterminer les lois de certaines fonctions de variables aléatoires. Nous nous restreignons au cas où ces variables aléatoires sont absolument continues.

1. Cas à une dimension

THÉORÈME 1.1. — *Désignons par S, T deux intervalles ouverts, finis ou non, de \mathbb{R} . Soit X une variable aléatoire réelle à valeurs dans S , absolument continue, de densité f ; soient u une bijection continûment dérivable de S sur T et $h = u^{-1}$ la bijection inverse de T sur S . Alors $Y = u \circ X$ est une variable aléatoire réelle à valeurs dans T , absolument continue, dont la densité g est donnée par :*

$$g(y) = f(h(y)) |h'(y)| I_T(y).$$

Démonstration. — On a, pour tout borélien $A \subset T$,

$$P\{Y \in A\} = P\{X \in h(A)\} = \int_{\mathbb{R}} f(x) I_{h(A)}(x) dx,$$

d'où, en faisant le changement de variables $x = h(y)$,

$$P\{Y \in A\} = \int_{\mathbb{R}} f(h(y)) |h'(y)| I_A(y) dy. \quad \square$$

Nous donnons d'abord quelques exemples, qui sont de simples applications du Théorème 1.1.

Exemple 1. — Soit X une variable aléatoire réelle, de densité f . Alors $Y = e^X$ est une variable aléatoire réelle à valeurs strictement positives, dont la densité g est donnée par :

$$g(y) = \frac{1}{y} f(\text{Log } y) I_{]0, +\infty[}(y).$$

Cas particulier. — Prenons $\mathcal{L}(X) = \mathcal{N}(\mu, \sigma)$ ($\mu \in \mathbb{R}, \sigma > 0$); alors $Y = e^X$ admet pour densité

$$g(y) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{y} \exp\left(-\frac{1}{2}\left(\frac{\text{Log } y - \mu}{\sigma}\right)^2\right) I_{]0, +\infty[}(y).$$

La variable aléatoire Y est appelée *variable aléatoire log-normale*, de paramètres μ, σ .

Exemple 2. — Soit X une variable aléatoire réelle, de densité f . Alors $Y = \frac{1}{X}$ est une variable aléatoire réelle, de densité :

$$g(y) = \frac{1}{y^2} f\left(\frac{1}{y}\right) \quad (y \neq 0).$$

(En toute rigueur, il faudrait n'utiliser le Théorème 1.1 que pour $\mathbb{R} \setminus \{0\}$. Cependant, la variable aléatoire X étant absolument continue, on a $P\{X = 0\} = 0$, de sorte que l'on peut négliger l'origine.)

Cas particulier 1. — Prenons $\mathcal{L}(X) = \mathcal{N}(0, 1)$; alors $Y = \frac{1}{X}$ admet pour densité :

$$g(y) = \frac{1}{\sqrt{2\pi}} \frac{1}{y^2} \exp\left(-\frac{1}{2y^2}\right) \quad (y \neq 0).$$

On constate que $\mathbb{E}[|Y|] = \int_{\mathbb{R}} |y| g(y) dy = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{1}{|y|} \exp\left(-\frac{1}{2y^2}\right) dy = +\infty$. Ainsi l'inverse d'une variable aléatoire normale réduite n'admet pas d'espérance mathématique.

Cas particulier 2. — Prenons pour $\mathcal{L}(X)$ la loi de Cauchy $\mathcal{C}(0, 1)$. Alors $Y = \frac{1}{X}$ admet pour densité :

$$g(y) = \frac{1}{y^2} \frac{1}{\pi} \frac{1}{1 + \frac{1}{y^2}} = \frac{1}{\pi} \frac{1}{1 + y^2}.$$

On constate que Y suit encore la loi de Cauchy $\mathcal{C}(0, 1)$.

Remarque. — Dans le cas où l'application u n'est pas bijective, il n'y a pas de méthode générale pour déterminer la densité de $u \circ X$, mais, dans la plupart des cas particuliers qui se présentent, on peut imaginer un traitement adéquat.

Exemple 1. — Soit X une variable aléatoire réelle, absolument continue, de densité f . Soit à déterminer la densité g de $Y = |X|$. On voit que $u(x) = |x|$ n'est pas une bijection de \mathbb{R} sur \mathbb{R}^+ . On peut alors adopter la méthode suivante : commençons par calculer la fonction de répartition F_Y de Y :

$$F_Y(y) = \begin{cases} 0, & \text{pour } y \leq 0; \\ P\{Y \leq y\} = P\{-y \leq X \leq +y\}, & \text{pour } y > 0. \end{cases}$$

Comme la loi de probabilité de X est diffuse, on peut encore écrire

$$F_Y(y) = \begin{cases} 0, & \text{pour } y \leq 0; \\ F_X(y) - F_X(-y), & \text{pour } y > 0. \end{cases}$$

On en déduit la densité f_Y de Y par dérivation :

$$f_Y(y) = \begin{cases} 0, & \text{pour } y \leq 0; \\ f(y) + f(-y), & \text{pour } y > 0. \end{cases}$$

Si la variable aléatoire X est paire, on a :

$$f_Y(y) = \begin{cases} 0, & \text{pour } y \leq 0; \\ 2f(y), & \text{pour } y > 0. \end{cases}$$

Par exemple, si $\mathcal{L}(X) = \mathcal{N}(0, 1)$, la variable $Y = |X|$ a pour densité

$$f_Y(y) = \begin{cases} 0, & \text{pour } y \leq 0; \\ \frac{2}{\sqrt{2\pi}} e^{-y^2/2}, & \text{pour } y > 0. \end{cases}$$

Exemple 2. — Soit X une variable aléatoire réelle, absolument continue, de densité f . Soit à calculer la densité de $Y = X^2$. De nouveau, $u(x) = x^2$ n'est pas une bijection de \mathbb{R} sur \mathbb{R}^+ .

Commençons par calculer la fonction de répartition F_Y de Y :

$$F_Y(y) = \begin{cases} 0, & \text{pour } y \leq 0; \\ \mathbb{P}\{Y \leq y\} = \mathbb{P}\{-\sqrt{y} \leq X \leq +\sqrt{y}\}, & \text{pour } y > 0. \end{cases}$$

De même, puisque la loi de probabilité de X est diffuse,

$$F_Y(y) = \begin{cases} 0, & \text{pour } y \leq 0; \\ F_X(\sqrt{y}) - F_X(-\sqrt{y}), & \text{pour } y > 0. \end{cases}$$

Par dérivation,

$$f_Y(y) = \begin{cases} 0, & \text{pour } y \leq 0; \\ \frac{1}{2\sqrt{y}} (f(\sqrt{y}) + f(-\sqrt{y})), & \text{pour } y > 0. \end{cases}$$

Si la variable aléatoire X est paire, on a :

$$f_Y(y) = \begin{cases} 0, & \text{pour } y \leq 0; \\ \frac{1}{\sqrt{y}} f(\sqrt{y}), & \text{pour } y > 0. \end{cases}$$

Par exemple, si $\mathcal{L}(X) = \mathcal{N}(0, 1)$, la variable $Y = X^2$ a pour densité

$$f_Y(y) = \begin{cases} 0, & \text{pour } y \leq 0; \\ \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} e^{-y/2}, & \text{pour } y > 0. \end{cases}$$

C'est la loi du Chi-deux à un degré de liberté.

2. Cas à deux dimensions

THÉORÈME 2.1. — Désignons par S et T deux ouverts de \mathbb{R}^2 et soit (X, Y) un couple de variables aléatoires, à valeurs dans S , absolument continu, de densité conjointe f . Soit

$$G : (x, y) \mapsto (u, v) = (u(x, y), v(x, y))$$

une bijection continûment différentiable de S sur T ; soit

$$H = G^{-1} : (u, v) \mapsto (x, y) = (h_1(u, v), h_2(u, v))$$

la bijection inverse de T sur S . (Les dérivées partielles $D_u h_1$, $D_v h_1$, $D_u h_2$, $D_v h_2$ sont donc elles-mêmes continues.) On note

$$J = \frac{D(h_1, h_2)}{D(u, v)} = \frac{D(x, y)}{D(u, v)} = \begin{vmatrix} D_u h_1 & D_v h_1 \\ D_u h_2 & D_v h_2 \end{vmatrix}$$

le jacobien de la bijection inverse H .

Alors $(U, V) = G \circ (X, Y) = (u \circ (X, Y), v \circ (X, Y))$ est un couple de variables aléatoires réelles à valeurs dans T , absolument continu, et dont la densité conjointe g est donnée par

$$g(u, v) = f(h_1(u, v), h_2(u, v)) |J| I_T(u, v).$$

Démonstration. — On a, pour tout borélien $A \subset T$,

$$P\{(U, V) \in A\} = P\{(X, Y) \in H(A)\} = \iint_{\mathbb{R}^2} f(x, y) I_{H(A)}(x, y) dx dy,$$

d'où, en faisant le changement de variables $x = h_1(u, v)$, $y = h_2(u, v)$,

$$P\{(U, V) \in A\} = \iint_{\mathbb{R}^2} f(h_1(u, v), h_2(u, v)) |J| I_A(u, v) du dv. \quad \square$$

Remarque. — Dans la dernière intégrale écrite, on a utilisé la formule du changement de variables dans les intégrales doubles, qui exige que le jacobien J ne s'annule jamais sur A . Pour avoir un champ d'application plus général, on peut supposer que $J \neq 0$, sauf sur ensemble négligeable I tel que $H(I)$ est encore négligeable.

Exemple 1. — Soit (X, Y) un couple de variables aléatoires indépendantes, dont chacune suit la loi $\mathcal{N}(0, 1)$. La loi du couple a donc pour densité

$$f(x, y) = \frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right).$$

On lui associe le couple (R, Θ) défini par

$$R = (X^2 + Y^2)^{1/2}, \quad \Theta = \text{Arc tg } \frac{Y}{X}.$$

(1) *Loi conjointe de (R, Θ) .* — D'abord

$$(x, y) \mapsto (r, \theta) = ((x^2 + y^2)^{1/2}, \text{Arc tg } \frac{y}{x})$$

est une bijection continûment différentiable de $S = \mathbb{R}^2 \setminus \{0\}$ sur $T =]0, +\infty[\times]0, 2\pi[$. La bijection inverse est

$$(r, \theta) \mapsto (x, y) = (r \cos \theta, r \sin \theta),$$

d'où l'on déduit

$$J = \begin{vmatrix} x'_r & x'_\theta \\ y'_r & y'_\theta \end{vmatrix} = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} = r.$$

La densité conjointe de (R, Θ) est donc donnée par

$$\begin{aligned} g(r, \theta) &= f(r \cos \theta, r \sin \theta) |J| \\ &= \frac{1}{2\pi} e^{-r^2/2} r \quad (r, \theta) \in]0, +\infty[\times]0, 2\pi[. \end{aligned}$$

(2) *Lois marginales.* — Les densités marginales de R et de Θ s'en déduisent immédiatement :

$$g(r, \cdot) = \int_0^{2\pi} g(r, \theta) d\theta = e^{-r^2/2} r \quad (r \in]0, +\infty[)$$

$$g(\cdot, \theta) = \int_0^\infty g(r, \theta) dr = \frac{1}{2\pi} \quad (\theta \in [0, 2\pi[).$$

Remarque. — La loi de R , de densité $g(r, \cdot) = e^{-r^2/2} r$ ($r \in]0, +\infty[$) est appelée *loi de Rayleigh*. La fonction de survie de R est donnée par

$$P\{R > r\} = \int_r^\infty e^{-t^2/2} t dt = e^{-r^2/2} \quad (r > 0);$$

et son espérance mathématique par

$$\mathbb{E}[R] = \int_0^\infty P\{R > r\} dr = \int_0^\infty e^{-r^2/2} dr = \sqrt{\frac{\pi}{2}}.$$

(3) *Les variables aléatoires R et Θ sont indépendantes.* — Il résulte de (1) et (2) que $g(r, \theta) = g(r, \cdot)g(\cdot, \theta)$, $(r, \theta) \in]0, +\infty[\times [0, 2\pi[$, ce qui assure l'indépendance de R et Θ , donc aussi celle de R et Y/X .

Exemple 2. — Soit (X, Y) un couple de variables aléatoires indépendantes, dont chacune suit la loi $\mathcal{N}(0, 1)$. Posons $U = (X + Y)/\sqrt{2}$, $V = (X - Y)/\sqrt{2}$. Alors le couple (U, V) est formé de variables aléatoires *indépendantes*, dont chacune suit la loi $\mathcal{N}(0, 1)$.

Démonstration. — La densité conjointe de (X, Y) est

$$f(x, y) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + y^2)\right).$$

Faisons le changement de variables $u = (x + y)/\sqrt{2}$, $v = (x - y)/\sqrt{2}$, qui établit une bijection de \mathbb{R}^2 sur \mathbb{R}^2 . On calcule successivement $x = (u + v)/\sqrt{2}$, $y = (u - v)/\sqrt{2}$, $D(x, y)/D(u, v) = -1$, d'où la densité conjointe $g(u, v)$ de (U, V) :

$$g(u, v) = f\left(\frac{u+v}{\sqrt{2}}, \frac{u-v}{\sqrt{2}}\right) \cdot 1 = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left(\left(\frac{u+v}{\sqrt{2}}\right)^2 + \left(\frac{u-v}{\sqrt{2}}\right)^2\right)\right)$$

$$= \left(\frac{1}{\sqrt{2\pi}} e^{-u^2/2}\right) \cdot \left(\frac{1}{\sqrt{2\pi}} e^{-v^2/2}\right). \quad \square$$

3. Loi de probabilité d'une fonction de deux variables aléatoires.

Soient (X, Y) un couple de variables aléatoires réelles, *absolument continu*, de densité conjointe f et $u : \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction mesurable. On se propose de déterminer la loi de probabilité de la variable aléatoire $U = u \circ (X, Y)$, moyennant des conditions de régularité appropriées sur la fonction u . A cet effet, nous considérons U comme la première variable aléatoire marginale du couple (U, V) , avec $U = u \circ (X, Y)$ et $V = Y$.

(a) *Loi de probabilité du couple (U, V) .* — On suppose que $(x, y) \mapsto (u, v) = (u(x, y), y)$ est une bijection continûment différentiable de \mathbb{R}^2 sur \mathbb{R}^2 et l'on introduit la bijection inverse $(u, v) \mapsto (x, y) = (h(u, v), v)$. Il en résulte $J = \frac{D(x, y)}{D(u, v)} = \begin{vmatrix} h'_u & h'_v \\ 0 & 1 \end{vmatrix} = h'_u$. D'après le Théorème 2.1, le couple (U, V) est absolument continu et sa densité conjointe g est donnée par

$$g(u, v) = f(h(u, v), v) |h'_u(u, v)| \quad ((u, v) \in \mathbb{R}^2).$$

(b) *Loi de probabilité de U .* — Elle est donnée par sa densité

$$g(u) = g(u, \cdot) = \int_{\mathbb{R}} f(h(u, v), v) |h'_u(u, v)| dv \quad (u \in \mathbb{R}).$$

Exemple 1 (loi de la somme). — Soient $u = x + y$ et $v = y$, de sorte que $U = X + Y$ et $x = u - v$, $y = v$, d'où $J = \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1$ et

$$g(u) = \int_{\mathbb{R}} f(u - v, v) dv \quad (u \in \mathbb{R}).$$

Cas particulier (produit de convolution). — Dans le cas où le couple (X, Y) est *indépendant*, sa densité conjointe se factorise : $f(x, y) = f_1(x)f_2(y)$, où f_1, f_2 sont les densités de X, Y , respectivement, et

$$g(u) = \int_{\mathbb{R}} f_1(u - v) f_2(v) dv \quad (u \in \mathbb{R}).$$

On dit que $g = f_1 * f_2$ est le *produit de convolution* de f_1, f_2 .

Application. — Prenons $\mathcal{L}(X) = \mathcal{N}(\mu_1, \sigma_1)$, $\mathcal{L}(Y) = \mathcal{N}(\mu_2, \sigma_2)$ ($\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$). On peut vérifier que $\mathcal{L}(X + Y) = \mathcal{N}(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$. Le calcul direct est fastidieux, il est préférable d'utiliser la technique des fonctions génératrices ou des fonction caractéristiques.

Exemple 2 (loi du produit). — Prenons $u = xy$ et $v = y$, de sorte que $U = XY$ et $x = u/v$, $y = v$, d'où $J = \begin{vmatrix} 1/v & -u/v^2 \\ 0 & 1 \end{vmatrix} = 1/v$ et

$$g(u) = \int_{\mathbb{R}} f\left(\frac{u}{v}, v\right) \frac{1}{|v|} dv \quad (u \in \mathbb{R}).$$

Dans le cas où (X, Y) est indépendant, on a $f(x, y) = f_1(x)f_2(y)$ et

$$g(u) = \int_{\mathbb{R}} f_1(u/v) f_2(v) \frac{1}{|v|} dv \quad (u \in \mathbb{R}).$$

Application. — Prenons $\mathcal{L}(X) = \mathcal{N}(0, 1)$, $\mathcal{L}(Y) = \mathcal{N}(0, 1)$. La densité $g(u)$ de $U = XY$ est donnée par

$$\begin{aligned} g(u) &= \frac{1}{2\pi} \int_{\mathbb{R}} \exp\left(-\frac{1}{2}\left(\frac{u^2}{v^2} + v^2\right)\right) \frac{1}{|v|} dv \\ &= \frac{1}{\pi} \int_0^{\infty} \exp\left(-\frac{1}{2}\left(\frac{u^2}{v^2} + v^2\right)\right) \frac{1}{v} dv. \end{aligned}$$

Cette fonction admet une valeur finie pour tout $u \neq 0$. Pour un tel u on peut faire le changement de variable $v^2 = |u|t$ et l'on obtient

$$g(u) = \frac{1}{2\pi} \int_0^\infty \exp\left(-\frac{|u|}{2}\left(t + \frac{1}{t}\right)\right) \frac{dt}{t}.$$

On ne peut exprimer l'intégrale à l'aide des seules fonctions élémentaires; on montre qu'elle s'exprime à l'aide des fonctions de Bessel (*cf.* Exercice 8, chap. 13).

Exemple 3 (loi du rapport). — Considérons $u = \frac{x}{y}$, de sorte que $U = \frac{X}{Y}$.

Alors $u = \frac{x}{y}$, $v = y$, d'où $x = uv$, $y = v$, $J = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v$ et

$$g(u) = \int_{\mathbb{R}} f(uv, v) |v| dv \quad (u \in \mathbb{R}).$$

Dans le cas particulier où (X, Y) est indépendant, on a $f(x, y) = f_1(x)f_2(y)$ et

$$g(u) = \int_{\mathbb{R}} f_1(uv)f_2(v) |v| dv \quad (u \in \mathbb{R}).$$

Application. — Prenons $\mathcal{L}(X) = \mathcal{N}(0, 1)$, $\mathcal{L}(Y) = \mathcal{N}(0, 1)$. La densité $g(u)$ de $U = X/Y$ est donnée par

$$g(u) = \frac{1}{2\pi} \int_{\mathbb{R}} \exp\left(-\frac{v^2}{2}(1 + u^2)\right) |v| dv = \frac{1}{\pi} \int_0^\infty \exp\left(-\frac{v^2}{2}(1 + u^2)\right) v dv,$$

d'où, en faisant le changement de variables $v^2(1 + u^2)/2 = t$,

$$g(u) = \frac{1}{\pi} \frac{1}{1 + u^2} \int_0^\infty e^{-t} dt = \frac{1}{\pi} \frac{1}{1 + u^2} \quad (u \in \mathbb{R}).$$

On constate que le *rapport de deux variables aléatoires, indépendantes, normales, réduites, suit la loi de Cauchy* $\mathcal{C}(0, 1)$

Cette propriété a des conséquences surprenantes :

1) Les variables aléatoires X/Y et Y/X ont évidemment la même loi. Donc *l'inverse d'une variable aléatoire de Cauchy* $\mathcal{C}(0, 1)$ *est encore une variable aléatoire de Cauchy* $\mathcal{C}(0, 1)$.

2) Soit Z une variable aléatoire de Cauchy $\mathcal{C}(0, 1)$. Elle a même loi que X/Y , où X, Y sont des variables aléatoires indépendantes normales

réduites. De là $\frac{1+Z}{1-Z}$ a même loi que $\frac{1 + \frac{Y}{X}}{1 - \frac{Y}{X}} = \frac{X+Y}{X-Y} = \frac{\frac{X+Y}{\sqrt{2}}}{\frac{X-Y}{\sqrt{2}}}$.

Or $\left(\frac{X+Y}{\sqrt{2}}, \frac{X-Y}{\sqrt{2}}\right)$ est un couple de variables aléatoires indépendantes normales réduites. Donc $\frac{1+Z}{1-Z}$ suit encore la loi de Cauchy $\mathcal{C}(0, 1)$.

3) La variable aléatoire $1 + Z = 1 + \frac{Y}{X} = \frac{X + Y}{Y}$ est le rapport de deux variables aléatoires *symétriques*, mais n'est *pas* elle-même symétrique, puisqu'elle suit la loi de Cauchy $\mathcal{C}(1, 1)$.

COMPLÉMENTS ET EXERCICES

1. — Soit X une variable aléatoire suivant la loi de Cauchy $\mathcal{C}(0, 1)$. Alors la variable aléatoire $Y = \frac{1 + X}{1 - X}$ suit encore la loi de Cauchy $\mathcal{C}(0, 1)$.

2. — Soit X une variable aléatoire uniformément répartie sur l'intervalle $] -\pi/2, +\pi/2[$. Alors la variable aléatoire $Y = \operatorname{tg} X$ suit la loi de Cauchy $\mathcal{C}(0, 1)$.

3. — Soit (X, Y) un couple de variables aléatoires *indépendantes*, dont chacune suit la loi $\mathcal{E}(\lambda)$ ($\lambda > 0$). Alors la variable aléatoire $U = \frac{X}{Y}$, a pour densité $f(u) = \frac{1}{(1 + u)^2}$ ($u \geq 0$). Elle n'admet pas d'espérance mathématique finie. La variable aléatoire $V = U + 1 = \frac{X + Y}{Y}$ admet pour densité $g(v) = 1/v^2$ ($v \geq 1$); elle suit la loi de Pareto $\mathcal{P}(1, 1)$.

4. — Soit (X, Y) un couple de variables aléatoires *indépendantes*, de lois marginales $\mathcal{L}(X) = \Gamma(r, \lambda)$, $\mathcal{L}(Y) = \Gamma(s, \lambda)$ ($r, s, \lambda > 0$). Posons $U = X + Y$, $V = \frac{X}{X + Y}$.

a) Le couple (U, V) est indépendant.

b) $\mathcal{L}(U) = \Gamma(r + s, \lambda)$, $\mathcal{L}(V) = \mathcal{B}(r, s)$ (loi bêta). On constate que la loi marginale de V ne dépend *pas* de $\lambda > 0$.

5. — Soit (U, Y) un couple de variables aléatoires *indépendantes*, U étant uniforme sur $[0, 1]$ et Y absolument continue de densité g . On considère la variable aléatoire $X = UY$ et on désigne par f sa densité.

a) Calculer f en fonction de g .

b) On suppose que le support de Y soit $[0, +\infty[$. Montrer que f est dérivable et que f et g sont liées par la relation $xf'(x) + g(x) = 0$. En déduire que f admet un mode et un seul situé en $x = 0$.

c) On suppose que le support de Y soit \mathbb{R} . Montrer que f admet encore un mode et un seul situé en $x = 0$.

d) On prend $g(x) = \frac{x^2}{\sqrt{2\pi}} e^{-x^2/2}$, où $x \in \mathbb{R}$. Montrer que f est la densité de $\mathcal{N}(0, 1)$.

6. — Soit (X, Y) un couple de variables aléatoires *indépendantes* dont chacune suit la loi $\mathcal{N}(0, 1)$. Soit $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ une matrice 2×2 orthogonale. On pose $\begin{pmatrix} U \\ V \end{pmatrix} = A \begin{pmatrix} X \\ Y \end{pmatrix}$, c'est-à-dire $U = aX + bY$, $V = cX + dY$.

a) Le couple (U, V) est encore formé de variables aléatoires *indépendantes* dont chacune suit la loi $\mathcal{N}(0, 1)$.

b) Si T suit la loi de Cauchy $\mathcal{C}(0, 1)$, il en est de même de $Z = \frac{a + bT}{c + dT}$.

7. — Soit (X, Y) un couple de variables aléatoires *indépendantes*, dont chacune suit la loi $\mathcal{N}(0, 1)$. On pose $U = 2X$, $V = X - Y$.

a) Déterminer la densité conjointe du couple (U, V) , ainsi que les densités marginales de U et de V .

b) Déterminer la densité conditionnelle de U conditionnellement à l'évènement $\{V = 0\}$.

c) Déterminer la densité conditionnelle de $X + Y$ conditionnellement à $\{V = 0\}$.

d) On constate que les densités conditionnelles trouvées en b) et c) sont égales entre elles et que leur valeur commune est la densité (non conditionnelle) de $X + Y$; en d'autres termes, on constate que

$$\mathcal{L}(2X | X - Y = 0) = \mathcal{L}(X + Y | X - Y = 0) = \mathcal{L}(X + Y).$$

Aurait-on pu prévoir ce résultat ?

8. — Soit (U_1, \dots, U_n) un système de n variables aléatoires *indépendantes* admettant toutes la loi uniforme sur $[0, 1]$. La loi de $X = \prod_{i=1}^n U_i$ a pour densité

$$f(x) = \begin{cases} \frac{1}{(n-1)!} \left(\text{Log} \left(\frac{1}{x} \right) \right)^{n-1}, & \text{si } 0 < x \leq 1; \\ 0, & \text{sinon.} \end{cases}$$

9. — Soit X une variable aléatoire suivant la loi de Cauchy $\mathcal{C}(0, 1)$.

a) Montrer que X et $1/X$ ont même loi.

b) On pose : $Y = \begin{cases} X, & \text{avec probabilité } 1/2; \\ 1/X, & \text{avec probabilité } 1/2. \end{cases}$

Montrer que $\mathcal{L}(Y) = \mathcal{C}(0, 1)$.

10. — Soit (X, Y) un couple de variables aléatoires indépendantes dont chacune suit la loi $\mathcal{N}(0, 1)$. On pose $U = XY$ et $V = X/Y$.

1) Déterminer la densité conjointe de (U, V) .

2) En déduire les densités marginales de U et de V .

11. (A. Joffe). — Soit (X, Y) un couple de variables aléatoires indépendantes de loi commune $\mathcal{N}(0, 1)$.

1) Puisque X et Y sont indépendantes, on a $\mathcal{L}(|X| | Y = 0) = \mathcal{L}(|X|)$. Cette loi admet pour densité : $f(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2}$ ($x \geq 0$).

2) Introduisons les coordonnées polaires

$$R = \sqrt{X^2 + Y^2}, \quad \Theta = \text{Arctg}(Y/X).$$

On a vu dans ce présent chapitre, § 2, Exemple 1, que les variables aléatoires R et Θ sont indépendantes. Par conséquent $\mathcal{L}(R | \Theta = 0) = \mathcal{L}(R)$. La loi de R est la loi de Rayleigh, de densité $g(x) = xe^{-x^2/2}$ ($x \geq 0$). Cet exemple illustre le fait que la densité conditionnelle n'a pas de sens intrinsèque et doit être définie *au moyen d'une densité conjointe dans un système de coordonnées donné*.

12. — Soit (X, Y) un couple de variables aléatoires indépendantes dont chacune suit la loi $\mathcal{N}(0, 1)$. Déterminer la loi des variables aléatoires :

a) $U = \frac{X}{|Y|}$;

b) $Z = \frac{X + Y}{|X - Y|}$.

(La loi de Z est la loi de Student à un degré de liberté.)

13. — Soit (X, Y) un couple de variables aléatoires indépendantes, suivant chacune la loi de Cauchy $C(0, 1)$.

La moyenne harmonique $H = \left[\frac{1}{2} \left(\frac{1}{X} + \frac{1}{Y} \right) \right]^{-1}$ suit encore la loi de Cauchy $C(0, 1)$.

Cet énoncé se vérifie facilement, puisque l'inverse d'une variable aléatoire de Cauchy est encore de Cauchy et que la moyenne arithmétique de deux variables aléatoires de Cauchy, indépendantes, est encore de Cauchy.

CONVERGENCES STOCHASTIQUES

Dans ce chapitre, nous considérons des suites (X_n) ($n \geq 1$) de variables aléatoires définies sur un même espace probabilisé $(\Omega, \mathfrak{A}, \mathbb{P})$ et nous étudions le comportement asymptotique de telles suites lorsque n tend vers l'infini. Plusieurs types de convergence se sont imposés (convergence en loi, en probabilité, presque sûre, en moyenne d'ordre $r > 0$, et bien d'autres). Nous passons en revue les plus importants de ces types.

1. Convergence en loi, ou convergence étroite

Définition. — Donnons-nous :

a) une suite (X_n) ($n \geq 1$) de variables aléatoires définies sur un espace probabilisé $(\Omega, \mathfrak{A}, \mathbb{P})$ (on désigne par (F_n) ($n \geq 1$) la suite des fonctions de répartition correspondantes);

b) une variable aléatoire X définie sur le même espace probabilisé $(\Omega, \mathfrak{A}, \mathbb{P})$ (on désigne par F sa fonction de répartition; on notera que $F(+\infty) = 1$ et que $F(-\infty) = 0$).

On dit que la suite (X_n) ($n \geq 1$) *converge en loi (ou étroitement) vers X* , et l'on écrit $X_n \xrightarrow{\mathcal{L}} X$, ou $\mathcal{L}(X_n) \rightarrow \mathcal{L}(X)$, si en tout point x de continuité de F (en abrégé : $x \in \mathcal{C}(F)$), on a $F_n(x) \rightarrow F(x)$, lorsque n tend vers l'infini.

(En toute rigueur, il faudrait dire que la suite de lois $\mathcal{L}(X_n)$ converge en loi, ou étroitement, vers la loi $\mathcal{L}(X)$, mais la terminologie que nous adoptons est courante et justifiée.)

Remarque. — Supposons que $X_n \xrightarrow{\mathcal{L}} X$ et que les X_n , ainsi que X , soient du premier ordre (i.e., aient des espérances mathématiques finies). Il n'en résulte pas que la suite $(\mathbb{E}[X_n])$ ($n \geq 1$) converge, ni que, si elle converge, elle ait pour limite $\mathbb{E}[X]$; les situations les plus diverses peuvent se présenter.

Par exemple, soit (a_n) ($n \geq 1$) une suite de nombres réels strictement positifs; associons-lui une suite (X_n) ($n \geq 1$) de variables aléatoires de lois $(1/n)\varepsilon_{a_n} + (1 - (1/n))\varepsilon_0$. On vérifie que la suite (X_n) converge en loi vers $X = 0$ (donc $\mathbb{E}[X] = 0$) et que $\mathbb{E}[X_n] = a_n/n$ pour tout $n \geq 0$. On a les comportements suivants :

- si $a_n = \sqrt{n}$, alors $\mathbb{E}[X_n] = 1/\sqrt{n} \rightarrow 0 = \mathbb{E}[X]$;
- si $a_n = n$, alors $\mathbb{E}[X_n] = 1 \rightarrow 1 \neq \mathbb{E}[X]$;
- si $a_n = n^2$, alors $\mathbb{E}[X_n] = n \rightarrow +\infty \neq \mathbb{E}[X]$;
- si $a_n = n[2 + (-1)^n]$, la suite $\mathbb{E}[X_n] = 2 + (-1)^n$ oscille.

Exemple 1. — Prenons X_n uniformément répartie sur $[0, n]$ ($n \geq 1$). La suite (X_n) ($n \geq 1$) ne converge *pas* vers une limite. En effet, pour tout x réel et $n \geq 1$, on a :

$$F_n(x) = \begin{cases} 0, & \text{si } x < 0; \\ x/n, & \text{si } 0 \leq x < n; \\ 1, & \text{si } n \leq x. \end{cases}$$

Il en résulte que pour tout x on a : $F_n(x) \rightarrow F(x) = 0$, lorsque n tend vers l'infini. Or la limite $F(x) = 0$ n'est *pas* la fonction de répartition d'une loi de probabilité. On dit que la suite des lois de X_n converge *faiblement* vers la mesure nulle. Nous n'étudierons pas ici ce type de convergence.

Exemple 2. — Prenons X_n de loi $\frac{1}{2}(\varepsilon_{(-1/n)} + \varepsilon_{(1/n)})$ ($n \geq 1$). La suite (X_n) ($n \geq 1$) converge en loi vers la variable aléatoire $X = 0$. En effet,

$$F_n(x) = \begin{cases} 0, & \text{si } x < -1/n; \\ 1/2, & \text{si } -1/n \leq x < +1/n; \\ 1, & \text{si } 1/n \leq x. \end{cases}$$

Il en résulte que pour tout réel x , on a :

$$F_n(x) \rightarrow F^*(x) = \begin{cases} 0, & \text{si } x < 0; \\ 1/2, & \text{si } x = 0; \\ 1, & \text{si } x > 0. \end{cases}$$

lorsque n tend vers l'infini. La fonction F^* n'est *pas* la fonction de répartition d'une loi de probabilité (elle n'est *pas* continue à droite à l'origine). Toutefois, en désignant par F la fonction de répartition de la variable aléatoire $X = 0$, la fonction F^* coïncide avec F , sauf en $x = 0$, c'est-à-dire *sauf au point de discontinuité* de F . Pour tout x de l'ensemble $\mathcal{C}(F)$, on a donc

$$F_n(x) \rightarrow F(x), \quad \text{où } F(x) = \begin{cases} 0, & \text{si } x < 0; \\ 1, & \text{si } x \geq 0; \end{cases}$$

ce qui montre que $X_n \xrightarrow{\mathcal{L}} 0$.

Exemple 3. — Le présent exemple fait intervenir la « fonction de Dirac » introduite par les physiciens. Pour chaque $n \geq 1$, prenons X_n de loi $\mathcal{N}(0, \sigma_n)$ ($\sigma_n > 0$) et supposons que σ_n tende vers 0 lorsque n tend vers l'infini. Alors la suite (X_n) ($n \geq 1$) converge en loi vers la variable aléatoire $X = 0$. En effet, pour tout réel x et tout $n \geq 1$, on a :

$$F_n(x) = \frac{1}{\sigma_n \sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{u^2}{2\sigma_n^2}\right) du = \Phi\left(\frac{x}{\sigma_n}\right).$$

Il en résulte que pour tout réel x , on a :

$$F_n(x) \longrightarrow F^*(x) = \begin{cases} 0, & \text{si } x < 0; \\ 1/2, & \text{si } x = 0; \\ 1, & \text{si } x > 0. \end{cases}$$

La fonction F^* n'est *pas* la fonction de répartition d'une loi de probabilité (elle n'est *pas* continue à droite à l'origine). Toutefois, en désignant par F la

fonction de répartition de la variable aléatoire $X = 0$, la fonction F^* coïncide avec F , sauf en $x = 0$, c'est-à-dire sauf au point de discontinuité de F . On a donc pour tout x de l'ensemble $\mathcal{C}(F)$,

$$F_n(x) \rightarrow F(x), \quad \text{où } F(x) = \begin{cases} 0, & \text{si } x < 0; \\ 1, & \text{si } x \geq 0; \end{cases}$$

ce qui montre que $X_n \xrightarrow{\mathcal{L}} 0$.

Le théorème de Paul Lévy dont nous donnons l'énoncé ci-après précise les relations entre convergence en loi d'une suite de variables aléatoires et convergence de la suite des fonctions caractéristiques correspondantes.

THÉORÈME (Paul Lévy)

1) Soit (X_n) une suite de variables aléatoires, qui converge en loi vers une variable aléatoire X . Alors la suite (φ_n) des fonctions caractéristiques correspondantes converge vers la fonction caractéristique φ de X , et ceci uniformément dans tout intervalle fini.

2) Soient (X_n) une suite de variables aléatoires et (φ_n) la suite des fonctions caractéristiques correspondante. Supposons que la suite (φ_n) converge, au sens de la convergence simple, vers une fonction φ , dont la partie réelle $\Re\varphi$ est continue à l'origine. Alors

a) φ est une fonction caractéristique, c'est-à-dire il existe une loi de probabilité μ (et une seule) dont φ soit la fonction caractéristique.

b) La suite (X_n) ($n \geq 1$) converge en loi vers μ .

La partie 2) de ce théorème fournit un outil puissant pour établir la convergence en loi d'une suite de variables aléatoires. Elle est principalement utilisée pour démontrer des versions du théorème "central limit." On trouvera une démonstration d'une version de cette partie 2) au paragraphe 9 de ce chapitre.

2. Convergence en probabilité

Définition. — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires définies sur un même espace probabilisé $(\Omega, \mathfrak{A}, P)$.

a) On dit que la suite (X_n) ($n \geq 1$) converge en probabilité vers 0 lorsque n tend vers l'infini, et l'on écrit $X_n \xrightarrow{p} 0$, si pour tout $\varepsilon > 0$, on a :

$$\lim_{n \rightarrow \infty} P\{|X_n| > \varepsilon\} = 0.$$

b) Soit X une variable aléatoire définie sur le même espace probabilisé $(\Omega, \mathfrak{A}, P)$. On dit que la suite (X_n) ($n \geq 1$) converge en probabilité vers X , et l'on écrit $X_n \xrightarrow{p} X$, si $X_n - X \xrightarrow{p} 0$.

Remarque. — Supposons que $X_n \xrightarrow{p} X$ et que les X_n , ainsi que X , soient du premier ordre. Il n'en résulte pas que la suite $(\mathbb{E}[X_n])$ ($n \geq 1$) converge vers $\mathbb{E}[X]$, ni que, si elle converge, elle ait pour limite $\mathbb{E}[X]$; les situations les plus diverses peuvent se présenter.

Exemple. — Reprenons l'exemple qui illustre la Remarque du paragraphe 1 relative à la convergence en loi.

1) On a $X_n \xrightarrow{p} X = 0$, puisque pour tout $\varepsilon > 0$:

$$P\{X_n > \varepsilon\} \leq P\{X_n > 0\} = \frac{1}{n} \rightarrow 0.$$

2) La suite $(\mathbb{E}[X_n])$ a les différents comportements décrits ci-dessus.

Donnons tout d'abord deux propriétés de la convergence en probabilité.

THÉORÈME 2.1. — *Soit $(M_n = (X_n, Y_n))$ ($n \geq 1$) une suite de points aléatoires qui converge en probabilité vers le point aléatoire $M = (X, Y)$ (c'est-à-dire pour tout $\varepsilon > 0$ $\lim_n P\{|M_n - M| > \varepsilon\} = 0$, ce qui entraîne que l'on a simultanément $X_n \xrightarrow{p} X$ et $Y_n \xrightarrow{p} Y$). Soit d'autre part $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction continue en tout point (x, y) de \mathbb{R}^2 . Alors la suite des variables aléatoires $h(X_n, Y_n)$ ($n \geq 1$) converge en probabilité vers la variable aléatoire $h(X, Y)$.*

En particulier, si $X_n \xrightarrow{p} X$ et si f est une fonction réelle, continue en tout point de la droite réelle, alors $f \circ X_n \xrightarrow{p} f \circ X$.

Si $X = c$ (c réel), l'hypothèse $X_n \xrightarrow{p} c$ peut être remplacée par $X_n \xrightarrow{\mathcal{L}} c$.

THÉORÈME 2.2. — *Si la suite de variables aléatoires (X_n) ($n \geq 1$) converge en probabilité vers la variable aléatoire X et si $P\{X = 0\} = 0$, alors $1/X_n \xrightarrow{p} 1/X$.*

Les démonstrations de ces deux théorèmes peuvent être trouvées dans l'ouvrage de Fourgeaud-Fuchs.¹ Nous verrons que ce sont en fait des conséquences du Théorème 4.6 ci-dessous. Le corollaire suivant est une simple conséquence de ces deux théorèmes.

COROLLAIRE. — *Si la suite de points aléatoires $(M_n = (X_n, Y_n))$ ($n \geq 1$) converge en probabilité vers le point aléatoire $M = (X, Y)$, alors*

- 1) $X_n + Y_n \xrightarrow{p} X + Y$;
- 2) $\lambda X_n \xrightarrow{p} \lambda X$ ($\lambda \in \mathbb{R}$) ;
- 3) $X_n Y_n \xrightarrow{p} XY$;
- 4) $X_n / Y_n \xrightarrow{p} X / Y$, si $P\{Y = 0\} = 0$.

Remarque. — Le corollaire montre que la convergence en probabilité est compatible avec les opérations algébriques élémentaires. *Il n'en est pas ainsi de la convergence en loi.*

THÉORÈME 2.3 (Critère de convergence en probabilité). — *Soit (X_n) ($n \geq 1$) une suite de variables aléatoires ; s'il existe $r > 0$ tel que la suite de terme général $(\mathbb{E}[|X_n|^r])$ ($n \geq 1$) tende vers 0, alors la suite (X_n) ($n \geq 1$) converge en probabilité vers 0.*

¹ Fourgeaud (C.), Fuchs (A.). — *Statistique.* — Dunod, Paris, 1972, pp. 27–29.

Démonstration. — On a, d'après l'inégalité de Bienaymé-Tchebychev, pour tout $\varepsilon > 0$

$$P\{|X_n| \geq \varepsilon\} \leq \frac{\mathbb{E}[|X_n|^r]}{\varepsilon^r} \longrightarrow 0. \quad \square$$

3. Convergence en moyenne d'ordre $r > 0$

Définition. — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires définies sur un même espace probabilisé $(\Omega, \mathfrak{A}, P)$. On suppose qu'il existe $r > 0$ tel que pour tout $n \geq 1$ le moment $\mathbb{E}[|X_n|^r]$ soit fini.

a) On dit que la suite (X_n) ($n \geq 1$) *converge vers 0 en moyenne d'ordre r* , si $\mathbb{E}[|X_n|^r] \rightarrow 0$ lorsque n tend vers l'infini.

b) Soit X une autre variable aléatoire définie sur le même espace probabilisé $(\Omega, \mathfrak{A}, P)$. On dit que la suite (X_n) ($n \geq 1$) *converge vers X en moyenne d'ordre r* , si la suite $(X_n - X)$ ($n \geq 1$) converge vers 0 en moyenne d'ordre r .

Remarque 1. — Supposons que $X_n \rightarrow X$ en moyenne d'ordre r ; il n'en résulte pas que le moment $\mathbb{E}[|X|^r]$ soit fini, mais si ce moment est fini, alors $\mathbb{E}[|X_n|^r] \rightarrow \mathbb{E}[|X|^r]$.

Remarque 2. — Ce type de convergence est essentiellement utilisé pour $r = 2$, auquel cas on parle de convergence *en moyenne quadratique*.

4. Convergence presque sûre

Définition. — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires définies sur un même espace probabilisé $(\Omega, \mathfrak{A}, P)$.

a) On dit que la suite (X_n) ($n \geq 1$) *converge vers 0 presque sûrement*, et l'on écrit $X_n \xrightarrow{p.s.} 0$, s'il existe un ensemble P -négligeable $A \in \mathfrak{A}$ tel que pour tout $\omega \in \Omega \setminus A$, on ait $X_n(\omega) \rightarrow 0$, lorsque n tend vers l'infini.

b) Soit X une autre variable aléatoire définie sur le même espace probabilisé $(\Omega, \mathfrak{A}, P)$. On dit que la suite (X_n) ($n \geq 1$) *converge vers X presque sûrement*, et l'on écrit $X_n \xrightarrow{p.s.} X$, si la suite $(X_n - X)$ ($n \geq 1$) converge vers 0 presque sûrement.

Remarque. — Il résulte directement de la définition que les Théorèmes 2.1 et 2.2 ainsi que leurs corollaires sont valables pour la convergence presque sûre. Néanmoins, cette définition n'est *pas* opératoire et il convient, dans certains cas, de lui en substituer une autre, équivalente, qui a l'avantage de fournir instantanément des critères de convergence presque sûre. Nous introduisons cette nouvelle définition après le commentaire qui suit.

Commentaire de la définition. — Posons pour tout $\varepsilon > 0$

$$E_n(\varepsilon) = \{|X_n| > \varepsilon\}, \quad E(\varepsilon) = \limsup_{n \rightarrow \infty} E_n(\varepsilon) = \bigcap_{n \geq 1} \bigcup_{k \geq n} E_k(\varepsilon).$$

Introduisons l'ensemble de convergence de la suite (X_n) ($n \geq 1$) vers 0

$$C = \bigcap_{\varepsilon > 0} \bigcup_{n \geq 1} \bigcap_{k \geq n} \{|X_k| \leq \varepsilon\};$$

ainsi que son complémentaire, l'ensemble de divergence

$$D = C^c = \bigcup_{\varepsilon > 0} \bigcap_{n \geq 1} \bigcup_{k \geq n} \{|X_k| > \varepsilon\} = \bigcup_{\varepsilon > 0} E(\varepsilon).$$

On voit que $0 < \varepsilon < \varepsilon' \implies E(\varepsilon') \subset E(\varepsilon)$, de sorte que $(E(\varepsilon))$ ($\varepsilon > 0$) est une famille croissante lorsque $\varepsilon \downarrow 0$; il en résulte que :

a) l'ensemble D peut s'écrire $D = \bigcup_l E(1/l)$, avec $l \geq 1$ entier. Par conséquent, D (donc C) est *mesurable*. (Cette observation a été faite pour la première fois par Kolmogorov dans son ouvrage fondamental.²)

b) $D = \lim_{\varepsilon \downarrow 0} E(\varepsilon)$.

Il est clair que $X_n \xrightarrow{p.s.} 0$, si et seulement si $P(C) = 1$, c'est-à-dire si et seulement si $P(D) = 0$. Or cette dernière propriété est susceptible d'une interprétation intéressante faisant l'objet du théorème suivant.

THÉORÈME 4.1. — *Avec les notations ci-dessus, les deux propriétés suivantes sont équivalentes :*

a) $P(D) = 0$;

b) *pour tout $\varepsilon > 0$ on a $P(E(\varepsilon)) = 0$.*

Démonstration

a) \implies b) Il résulte de la relation $D = \bigcup_{\varepsilon > 0} E(\varepsilon)$ que pour tout $\varepsilon > 0$ on a $P(E(\varepsilon)) \leq P(D)$; d'où le résultat.

b) \implies a) Il résulte de la relation $D = \lim_{\varepsilon \downarrow 0} E(\varepsilon)$ que l'on a : $P(D) = \lim_{\varepsilon \downarrow 0} P(E(\varepsilon))$; d'où le résultat. \square

Ce théorème permet d'adopter dorénavant la définition suivante, beaucoup plus opératoire, de la convergence presque sûre.

Définition. — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires. Posons, pour tout $\varepsilon > 0$

$$E_n(\varepsilon) = \{|X_n| > \varepsilon\}, \quad E(\varepsilon) = \limsup_{n \rightarrow \infty} E_n(\varepsilon).$$

a) On dit que la suite (X_n) ($n \geq 1$) *converge vers 0 presque sûrement*, si pour tout $\varepsilon > 0$ on a $P(E(\varepsilon)) = 0$.

b) Soit X une autre variable aléatoire, définie sur le même espace probabilisé que les X_n . On dit que la suite *converge vers X presque sûrement*, si $X_n - X \xrightarrow{p.s.} 0$.

THÉORÈME 4.2. — *Soit (X_n) ($n \geq 1$) une suite de variables aléatoires. Les deux propriétés suivantes sont équivalentes.*

a) *La suite (X_n) ($n \geq 1$) converge vers 0 presque sûrement.*

² Kolmogorov (A. N.). — *Grundbegriffe der Wahrscheinlichkeitsrechnung.* — Berlin, Springer, 1933.

b) La suite de terme général $Y_n = \sup_{k \geq n} |X_k|$ converge vers 0 en probabilité.

Il en résulte immédiatement que la convergence presque sûre implique la convergence en probabilité.

Démonstration. — Avec les notations ci-dessus, on a :

$$\bigcup_{k \geq n} E_k(\varepsilon) = \bigcup_{k \geq n} \{|X_k| > \varepsilon\} = \left\{ \sup_{k \geq n} |X_k| > \varepsilon \right\}.$$

Cette suite d'ensembles est décroissante lorsque n croît, de sorte que pour tout $\varepsilon > 0$ on a :

$$E(\varepsilon) = \lim_{n \rightarrow \infty} \left\{ \sup_{k \geq n} |X_k| > \varepsilon \right\} \quad \text{et} \quad P(E(\varepsilon)) = \lim_{n \rightarrow \infty} P \left\{ \sup_{k \geq n} |X_k| > \varepsilon \right\}.$$

Cette égalité, valable pour tout $\varepsilon > 0$, montre que a) \Leftrightarrow b). \square

Nous donnons maintenant deux critères de convergence presque sûre.

PROPOSITION 4.3. — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires. Si, pour tout $\varepsilon > 0$, la série de terme général $P\{|X_n| > \varepsilon\}$ est convergente, alors la suite (X_n) ($n \geq 1$) converge vers 0 presque sûrement.

Démonstration. — Avec les notations ci-dessus on a, pour tout $\varepsilon > 0$ et tout $n \geq 1$

$$P(E(\varepsilon)) \leq \sum_{k \geq n} P(E_k(\varepsilon)).$$

Or le second membre, reste d'ordre n d'une série convergente, tend vers 0 lorsque n tend vers l'infini; le premier membre, étant indépendant de n , est donc nul. D'où pour tout $\varepsilon > 0$ l'identité $P(E(\varepsilon)) = 0$, c'est-à-dire $X_n \xrightarrow{p.s.} 0$. \square

PROPOSITION 4.4. — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires. S'il existe $r > 0$ tel que la série de terme général $\mathbb{E}[|X_n|^r]$ soit convergente, alors la suite (X_n) ($n \geq 1$) converge vers 0 presque sûrement.

Démonstration. — On a, d'après l'inégalité de Bienaymé-Tchebychev, pour tout $\varepsilon > 0$

$$P\{|X_n| \geq \varepsilon\} \leq \frac{\mathbb{E}[|X_n|^r]}{\varepsilon^r};$$

d'où le résultat en vertu de la Proposition 4.3. \square

Pour terminer cette section, nous donnons quelques liens entre convergence presque sûre et convergence en probabilité.

THÉORÈME 4.5. — Soit $X_n \xrightarrow{p} 0$, il existe une suite partielle (X_{n_k}) extraite de (X_n) telle que $X_{n_k} \xrightarrow{p.s.} 0$.

Démonstration. — Soient $\varepsilon > 0$ et (η_k) une suite de nombres strictement positifs tels que $\sum_{k \geq 1} \eta_k < +\infty$. Par hypothèse, pour tout $k \geq 1$ il existe

$n_k \geq 1$ tel que $P\{|X_{n_k}| > \varepsilon\} < \eta_k$. On peut toujours supposer que $n_k < n_{k+1}$ pour tout $k \geq 1$. On a alors pour tout $\varepsilon > 0$ les relations

$$P\left\{\sup_{k \geq n} |X_{n_k}| > \varepsilon\right\} = P\left\{\bigcup_{k \geq n} \{|X_{n_k}| > \varepsilon\}\right\} \leq \sum_{k \geq n} \eta_k.$$

Or le dernier membre tend vers 0 lorsque n tend vers l'infini. Le résultat en découle en vertu du Théorème 4.1. \square

THÉORÈME 4.6. — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires ; alors les deux propriétés suivantes sont équivalentes :

- a) $X_n \xrightarrow{p} 0$;
- b) De toute suite partielle extraite de (X_n) on peut extraire une suite partielle qui converge presque sûrement vers 0.

Démonstration

a) \Rightarrow b) Soit (X_{a_n}) une suite partielle extraite de (X_n) ; il est clair que $X_{a_n} \xrightarrow{p} 0$. Le Théorème 4.5 appliqué à la suite (X_{a_n}) montre qu'il existe une suite partielle extraite de (X_{a_n}) qui converge vers 0 presque sûrement.

b) \Rightarrow a) Supposons que a) ne soit pas satisfaite, c'est-à-dire qu'il existe $\varepsilon, \eta > 0$ tels que, quel que soit $N > 0$, il existe un entier $n \geq N$ pour lequel $P\{|X_n| > \varepsilon\} > \eta$. On en déduit qu'il existe une suite partielle (X_{a_n}) extraite de (X_n) telle que, pour tout $n \geq 1$, on ait $P\{|X_{a_n}| > \varepsilon\} > \eta$. Pour toute suite partielle (X_{b_k}) extraite de (X_{a_n}) on a alors, pour tout $k \geq 1$, l'inégalité $P\{|X_{b_k}| > \varepsilon\} > \eta$. Il en résulte que la suite (X_{b_k}) ne converge pas vers 0 en probabilité, donc non plus presque sûrement, ce qui contredit b). \square

Remarque. — Le Théorème 4.6 fournit des démonstrations immédiates des Théorèmes 2.1 et 2.2. Il suffit, en effet, d'observer que si $X_n \xrightarrow{p.s.} X$, alors, pour toute fonction continue f , on a : $f \circ X_n \xrightarrow{p.s.} f \circ X$.

5. Comparaison des divers types de convergence. — On a essentiellement le diagramme suivant :

$$\begin{array}{c} \text{Conv. en moyenne d'ordre } r \implies \text{Conv. en probabilité} \implies \text{Conv. en loi} \\ \uparrow \\ \text{Conv. presque sûre} \end{array}$$

Il est évident *a priori* que le mode de convergence en loi est le plus faible, puisque sa définition ne fait intervenir que les lois des X_n et ne fait pas référence au triplet fondamental.

Notation. — Dans ce qui suit (X_n) , (F_n) ($n \geq 1$) désignent une suite de variables aléatoires et la suite de leurs fonctions de répartition associées.

THÉORÈME 5.1. — Soit $r > 0$; la convergence en moyenne d'ordre r entraîne la convergence en probabilité.

Démonstration. — Elle résulte de l'inégalité de Bienaymé-Tchebychev, puisque pour tout $\varepsilon > 0$ on a :

$$P\{|X_n| \geq \varepsilon\} \leq \frac{\mathbb{E}[|X_n|^r]}{\varepsilon^r} \longrightarrow 0. \quad \square$$

Remarque. — La réciproque n'est pas exacte. Prenons $r = 1$ et considérons une suite de variables aléatoires (X_n) de lois $\frac{1}{n}\varepsilon_{n^2} + (1 - \frac{1}{n})\varepsilon_0$. On vérifie que cette suite converge vers 0 en probabilité, mais non en moyenne d'ordre 1. La réciproque est exacte dans le cas particulier où la suite (X_n) est presque sûrement bornée (cf. Exercice 4).

THÉORÈME 5.2. — *La convergence en probabilité entraîne la convergence en loi.*

La démonstration repose sur le lemme suivant.

LEMME 5.3. — *Étant donné un couple (X, Y) de variables aléatoires, on a, pour tout $\eta > 0$*

$$|F_X(x) - F_Y(x)| \leq F_X(x + \eta) - F_X(x - \eta) + P\{|X - Y| > \eta\}.$$

Démonstration

a) D'abord on a les relations :

$$\begin{aligned} \{Y \leq x\} &= \{Y \leq x, X \leq x + \eta\} + \{Y \leq x, X > x + \eta\} \\ &\subset \{X \leq x + \eta\} + \{|X - Y| > \eta\}; \end{aligned}$$

d'où

$$F_Y(x) \leq F_X(x + \eta) + P\{|X - Y| > \eta\}.$$

b) On montre de façon analogue que

$$F_X(x - \eta) \leq F_Y(x) + P\{|X - Y| > \eta\}.$$

c) Il résulte de a) et b) que :

$$F_X(x - \eta) - P\{|X - Y| > \eta\} \leq F_Y(x) \leq F_X(x + \eta) + P\{|X - Y| > \eta\}.$$

d) Or on a trivialement :

$$F_X(x - \eta) \leq F_X(x) \leq F_X(x + \eta).$$

e) Il résulte de c) et d) que :

$$|F_X(x) - F_Y(x)| \leq F_X(x + \eta) - F_X(x - \eta) + P\{|X - Y| > \eta\}. \quad \square$$

Pour démontrer le Théorème 5.2 on applique le lemme en prenant $Y = X_n$. On obtient donc, pour tout $n \geq 0$ et tout $\eta > 0$,

$$|F_X(x) - F_{X_n}(x)| \leq F_X(x + \eta) - F_X(x - \eta) + P\{|X - X_n| > \eta\}.$$

Supposons que x soit un point de continuité de F_X ; alors pour tout $\varepsilon > 0$, il existe $\eta(\varepsilon)$ tel que $F_X(x + \eta) - F_X(x - \eta) < \varepsilon$. Supposons que $X_n \xrightarrow{p} X$;

alors on peut associer au couple $(\varepsilon, \eta(\varepsilon))$ un nombre $N(\varepsilon) > 0$ tel que, pour tout $n \geq N$, on ait $\mathbb{P}\{|X - X_n| \geq \eta\} < \varepsilon$. Il en résulte que, si x est un point de continuité de F_X , on a, pour tout $n \geq N$, l'inégalité $|F_X(x) - F_{X_n}(x)| < 2\varepsilon$. \square

Remarque 1. — La réciproque n'est pas exacte : une suite de variables aléatoires peut converger en loi sans converger en probabilité. En voici un exemple dont la simplicité extrême montre ce qui sépare les deux modes de convergence. Soit X une variable aléatoire de loi $\frac{1}{2}(\varepsilon_0 + \varepsilon_1)$ et $Y = 1 - X$. Alors X et Y ont même loi et $|X - Y| = 1$. Considérons la suite (X_n) ($n \geq 1$) définie par $X_n = Y$ pour tout $n \geq 1$. Alors la suite (X_n) ($n \geq 1$) converge en loi vers X (trivialement), mais elle ne converge pas vers X en probabilité, puisque $|X_n - X| = |Y - X| = 1$.

Remarque 2. — La réciproque est exacte dans le cas particulier où la limite X se réduit à une variable aléatoire presque certaine. Nous allons montrer que si la suite (X_n) converge en loi vers 0, alors elle converge en probabilité vers 0. Supposons donc que l'on ait :

$$F_n(x) \rightarrow \begin{cases} 1, & \text{si } x > 0; \\ 0, & \text{si } x < 0. \end{cases}$$

On peut écrire, pour tout $\varepsilon > 0$ et tout $\eta > 0$ tel que $\varepsilon - \eta > 0$:

$$\begin{aligned} \mathbb{P}\{|X_n| > \varepsilon\} &= \mathbb{P}\{X_n > \varepsilon\} + \mathbb{P}\{X_n < -\varepsilon\} \\ &\leq \mathbb{P}\{X_n > \varepsilon - \eta\} + \mathbb{P}\{X_n \leq -\varepsilon\}. \end{aligned}$$

d'où

$$\mathbb{P}\{|X_n| > \varepsilon\} \leq 1 - F_n(\varepsilon - \eta) + F_n(-\varepsilon) \rightarrow 0. \quad \square$$

THÉORÈME 5.4. — *La convergence presque sûre entraîne la convergence en probabilité.*

Démonstration. — Ceci résulte immédiatement du Théorème 4.2. \square

Remarque 1. — La réciproque n'est pas exacte : une suite de variables aléatoires peut converger vers 0 en probabilité sans converger vers 0 presque sûrement ; il peut même arriver qu'aucune de ses réalisations ne converge vers 0, comme le montre l'exemple suivant (dit de la "bosse glissante").

Prenons pour Ω l'intervalle $[0, 1]$, pour \mathfrak{A} la tribu borélienne, pour \mathbb{P} la mesure de Lebesgue sur $([0, 1], \mathfrak{A})$ et considérons la suite à deux indices d'applications de Ω dans \mathbb{R} définies par :

$$\begin{aligned} X_{11} &= I_{[0,1]}; & X_{21} &= I_{[0,1/2[}; & X_{22} &= I_{[1/2,1]}; \\ X_{31} &= I_{[0,1/3[}; & X_{32} &= I_{[1/3,2/3[}; & X_{33} &= I_{[2/3,1]}; \dots \end{aligned}$$

Les graphes des X_{nk} ($n \geq 1, 1 \leq k \leq n$) sont des "bosses glissantes" qui s'amincissent lorsque n tend vers l'infini. On peut renuméroter la suite à

deux indices (X_{nk}) selon l'ordre lexicographique de façon à avoir une suite (Y_n) à un indice. On voit que :

- 1) la suite (Y_n) ne converge en aucun point $\omega \in [0, 1]$;
- 2) la suite (Y_n) converge en probabilité vers 0 ; en effet, pour tout ε de l'intervalle $]0, 1[$ on a pour tout $n \geq 1$ et pour tout k tel que $1 \leq k \leq n$ la relation $P\{|X_{nk}| > \varepsilon\} = 1/n$; d'où $P\{|Y_n| > \varepsilon\}$ tend vers 0, lorsque n tend vers l'infini.

Remarque 2. — L'exemple de la bosse glissante fournit également un exemple du fait suivant : *la convergence en moyenne quadratique n'entraîne pas la convergence presque sûre.* En effet,

- 1) la suite (Y_n) ne converge en aucun point ω de l'intervalle $[0, 1]$;
- 2) on a pour tout $n \geq 1$ et tout k tel $1 \leq k \leq n$ la relation $\mathbb{E}[|X_{nk}|^2] = 1/n$, d'où $\mathbb{E}[|Y_n|^2]$ tend vers 0, lorsque n tend vers l'infini ; ce qui montre que la suite (Y_n) ($n \geq 1$) converge vers 0 *en moyenne quadratique* (et aussi en moyenne d'ordre 1).

6. Convergence en loi de variables aléatoires à valeurs entières et absolument continues

THÉORÈME 6.1. — *Donnons-nous une suite (X_n) ($n \geq 1$) de variables aléatoires à valeurs dans \mathbb{Z} et une variable aléatoire X à valeurs dans \mathbb{Z} . Notons $(p_{n,k}, k \in \mathbb{Z})$ la loi de X_n ($n \geq 1$) et $(\alpha_k, k \in \mathbb{Z})$ la loi de X . Alors les deux propriétés suivantes sont équivalentes :*

- a) pour tout $k \in \mathbb{Z}$ on a : $\lim_{n \rightarrow \infty} p_{n,k} = \alpha_k$;
- b) $X_n \xrightarrow{\mathcal{L}} X$ ($n \rightarrow \infty$) (convergence en loi).

Démonstration

a) \Rightarrow b) D'abord $|p_{n,k} - \alpha_k| = p_{n,k} + \alpha_k - 2p_{n,k} \wedge \alpha_k$; d'où

$$\sum_{k \in \mathbb{Z}} |p_{n,k} - \alpha_k| = \sum_{k \in \mathbb{Z}} p_{n,k} + \sum_{k \in \mathbb{Z}} \alpha_k - 2 \sum_{k \in \mathbb{Z}} p_{n,k} \wedge \alpha_k$$

et, puisque $(p_{n,k})$ et (α_k) sont des lois de probabilité,

$$\sum_{k \in \mathbb{Z}} |p_{n,k} - \alpha_k| = 2 - 2 \sum_{k \in \mathbb{Z}} p_{n,k} \wedge \alpha_k.$$

Or $0 \leq p_{n,k} \wedge \alpha_k \leq \alpha_k$, avec $\sum_{k \in \mathbb{Z}} \alpha_k = 1$. De plus, pour tout $k \in \mathbb{Z}$ la suite $(p_{n,k} \wedge \alpha_k)$ tend vers α_k lorsque n tend vers l'infini. D'après le théorème de convergence dominée, il s'ensuit que $\sum_{k \in \mathbb{Z}} p_{n,k} \wedge \alpha_k \rightarrow \sum_{k \in \mathbb{Z}} \alpha_k = 1$, lorsque n tend vers l'infini., d'où $\sum_{k \in \mathbb{Z}} |p_{n,k} - \alpha_k| \rightarrow 0$ ($n \rightarrow \infty$).

Pour tout x réel, posons $F_n(x) = \sum_{k \leq x} p_{n,k}$, $F(x) = \sum_{k \in \mathbb{Z}} \alpha_k$. Alors pour tout x réel, on a :

$$|F_n(x) - F(x)| \leq \sum_{k \leq x} |p_{n,k} - \alpha_k| \leq \sum_{k \in \mathbb{Z}} |p_{n,k} - \alpha_k| \rightarrow 0$$

lorsque n tend vers l'infini, c'est-à-dire $X_n \xrightarrow{\mathcal{L}} X$.

b) \Rightarrow a) Désignons par F_n la fonction de répartition de X_n et par F celle de X . Alors pour tout $k \in \mathbb{Z}$ on a : $p_{n,k} = F_n(k) - F_n(k-1) \rightarrow F(k) - F(k-1) = \alpha_k$, lorsque n tend vers l'infini. On vérifie que $\sum_{k \in \mathbb{Z}} \alpha_k = F(+\infty) - F(-\infty) = 1$. \square

Remarque. — Ce théorème, joint au Théorème 4.2 du chapitre 9, fournit le critère suivant de convergence en loi d'une suite de variables aléatoires à valeurs entières positives.

CRITÈRE. — Soient (X_n) une suite de variables aléatoires à valeurs dans \mathbb{N} , de fonction génératrice G_n et X une variable aléatoire à valeurs dans \mathbb{N} , de fonction génératrice G . Si, pour tout $u \in]0, 1[$, on a $\lim_{n \rightarrow \infty} G_n(u) = G(u)$, alors $X_n \xrightarrow{\mathcal{L}} X$ ($n \rightarrow \infty$).

THÉORÈME 6.2 (théorème de Scheffé). — Donnons-nous une suite (X_n) ($n \geq 1$) de variables aléatoires absolument continues et X une variable aléatoire absolument continue. Notons f_n (resp. f) la densité de X_n (resp. X) et μ_n (resp. μ) sa loi. Supposons que pour presque tout x réel on ait : $f_n(x) \rightarrow f(x)$ lorsque n tend vers l'infini. Alors

a) $\|f_n - f\|_1 = \int_{\mathbb{R}} |f_n(x) - f(x)| dx \rightarrow 0$, lorsque n tend vers l'infini, c'est-à-dire $f_n \rightarrow f$ dans L^1 .

b) $\lim_{n \rightarrow \infty} \sup_{B \in \mathcal{B}^1} |\mu_n(B) - \mu(B)| = 0$, c'est-à-dire $\mu_n \rightarrow \mu$ en variation.

c) $X_n \xrightarrow{\mathcal{L}} X$ (en loi).

Démonstration

a) D'abord $|f_n - f| = f_n + f - 2f_n \wedge f$, d'où $\int_{\mathbb{R}} |f_n - f| dx = \int_{\mathbb{R}} f_n dx + \int_{\mathbb{R}} f dx - 2 \int_{\mathbb{R}} f_n \wedge f dx$. Puisque f_n et f sont des densités de probabilité, on en tire : $\|f_n - f\|_1 = 2 - 2 \int_{\mathbb{R}} f_n \wedge f dx$. Or pour tout $n \geq 1$ on a $0 \leq f_n \wedge f \leq f$ avec f intégrable et pour tout réel x on a la convergence : $(f_n \wedge f)(x) \rightarrow f(x)$ presque partout, lorsque n tend vers l'infini. En vertu du théorème de convergence dominée, on en conclut : $\int_{\mathbb{R}} f_n \wedge f dx \rightarrow \int_{\mathbb{R}} f dx = 1$, lorsque n tend vers l'infini. D'où $\|f_n - f\|_1 \rightarrow 0$ lorsque n tend vers l'infini.

b) Pour tout $B \in \mathcal{B}^1$ on a : $|\mu_n(B) - \mu(B)| = \left| \int_B (f_n - f) dx \right| \leq \int_B |f_n - f| dx \leq \int_{\mathbb{R}} |f_n - f| dx$, d'où $\sup_{B \in \mathcal{B}^1} |\mu_n(B) - \mu(B)| \leq \|f_n - f\|_1 \rightarrow 0$ lorsque n tend vers l'infini.

c) Pour tout réel x , posons $F_n(x) = \mu_n(]-\infty, x])$ et $F(x) = \mu(]-\infty, x])$. En appliquant b) à $B =]-\infty, x]$, on a, pour tout réel x , la convergence $|F_n(x) - F(x)| \rightarrow 0$ lorsque n tend vers l'infini, c'est-à-dire $X_n \xrightarrow{\mathcal{L}} X$. \square

Remarque. — Le fait que $X_n \xrightarrow{\mathcal{L}} X$ n'implique pas que pour tout réel x on ait $f_n(x) \rightarrow f(x)$ lorsque n tend vers l'infini.

Exemple. — Pour chaque $n \geq 1$, soit X_n une variable aléatoire de densité $f_n(x) = \begin{cases} 1 - \cos(2\pi nx), & \text{si } x \in [0, 1]; \\ 0, & \text{sinon.} \end{cases}$

a) La suite $(X_n)_{n \geq 1}$ converge en loi vers une variable aléatoire uniformément répartie sur $[0, 1]$, admettant donc une densité, à savoir $f(x) = I_{[0,1]}(x)$. En effet, pour tout $x \in [0, 1]$, on a $\int_0^x f_n(t) dt = x - \frac{\sin(2\pi nx)}{2\pi n}$, qui tend vers x lorsque n tend vers l'infini.

b) Or la suite $(f_n(x))$ ne converge pour aucune valeur de $x \in]0, 1[$.

7. Convergences étroite et presque sûre

7.1. *Inverse d'une fonction de répartition.* — Soit F la fonction de répartition d'une loi de probabilité μ sur \mathbb{R} . Pour tout $u \in]0, 1[$ l'ensemble $\{x : F(x) \geq u\}$ est un *intervalle* de \mathbb{R} , non borné, admettant un plus petit élément. Si l'on note $F^{-1}(u)$ cet élément minimum, on a :

$$\{x : F(x) \geq u\} = [F^{-1}(u), +\infty[.$$

On définit de la sorte une application F^{-1} , croissante de $]0, 1[$ dans \mathbb{R} . Cette application coïncide avec l'inverse de F , lorsque F applique bijectivement \mathbb{R} sur $]0, 1[$ (c'est-à-dire lorsque F est continue et strictement croissante). Dans le cas général, on l'appelle l'*inverse généralisée au sens de Paul Lévy*.

Il résulte de cette définition que, pour tout réel $u \in]0, 1[$ et tout couple (a, b) de réels tels que $a < b$, on a l'équivalence :

$$(7.1) \quad F(a) < u \leq F(b) \iff a < F^{-1}(u) \leq b.$$

7.2. *Construction d'une variable aléatoire de loi donnée.* — Conservons les mêmes notations que dans la sous-section précédente.

THÉORÈME 7.1. — *Considérons l'espace probabilisé $(]0, 1[, \mathcal{B}(]0, 1[), \mathbb{P})$, où \mathbb{P} désigne la restriction de la mesure de Lebesgue à la tribu $\mathcal{B}(]0, 1[)$. Alors en tant que variable aléatoire réelle sur cet espace, l'application F^{-1} admet F comme fonction de répartition, donc μ comme loi.*

Démonstration. — Pour tout réel x , la relation (7.1) implique :

$$\mathbb{P}\{F^{-1} \leq x\} = \mathbb{P}\{u : F^{-1}(u) \leq x\} = \mathbb{P}\{u : u \leq F(x)\} = F(x). \quad \square$$

7.3. Le théorème de Skorohod

THÉORÈME 7.2 (Skorohod). — *Soit (μ_n) une suite de lois de probabilité sur \mathbb{R} , qui converge étroitement vers une loi de probabilité μ . On peut alors définir, sur un même espace probabilisé, une suite de variables aléatoires (X_n) et une variable aléatoire X , telles que pour chaque n la variable X_n ait pour loi μ_n et X ait pour loi μ et telles que l'on ait : $X_n \xrightarrow{p.s.} X$.*

Démonstration. — Désignons par F_n la fonction de répartition de μ_n , par F celle de μ , et par \mathcal{C} l'ensemble des points de continuité de F . Désignons enfin par F_n^{-1} l'inverse généralisée de F_n et par F^{-1} celle de F . Associons à μ_n la variable aléatoire $X_n = F_n^{-1}$ définie sur $(]0, 1[, \mathcal{B}(]0, 1[), \lambda)$, où λ

désigne la mesure de Lebesgue, et à μ la variable aléatoire $X = F^{-1}$ définie sur le même espace.

Pour montrer que l'on a $X_n \xrightarrow{p.s.} X$, il suffit de montrer que la suite $(F_n^{-1}(u))$ converge vers $F^{-1}(u)$ en tout point $u \in]0, 1[$, où F^{-1} est continue (puisque le complémentaire de l'ensemble de ces points est de mesure de Lebesgue nulle). Soit donc $u \in]0, 1[$ un tel point. Si a, b sont deux éléments de \mathcal{C} tels que

$$(7.2) \quad a < F^{-1}(u) < b,$$

on peut trouver un point v tel que $u < v < 1$ et tel que l'on ait $a < F^{-1}(u) \leq F^{-1}(v) \leq b$, c'est-à-dire $F(a) < u < v \leq F(b)$. Puisque a et b sont dans \mathcal{C} , on a, pour n assez grand, les inégalités $F_n(a) < u \leq F_n(b)$, c'est-à-dire

$$(7.3) \quad a < F_n^{-1}(u) \leq b.$$

Les relations (7.2) et (7.3) suffisent pour conclure. \square

Remarque 1. — Convenons de dire qu'une variable aléatoire X' est une version d'une variable aléatoire X , si X' a même loi que X . (Il n'est naturellement pas nécessaire que X et X' soient définis sur un même espace probabilisé.) Avec cette terminologie, le Théorème 7.2 s'énonce ainsi :

Étant donné une suite de variables aléatoires (X_n) , qui converge en loi vers une variable aléatoire X , il existe des versions X'_n, X' de X_n, X , définies sur un même espace probabilisé, telles que $X'_n \xrightarrow{p.s.} X'$.

Remarque 2. — Voici un résultat, qui est une conséquence immédiate du théorème de Skorohod, dont la démonstration directe serait fastidieuse :

Supposons que $X_n \xrightarrow{\mathcal{L}} X$ et soit $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue. Alors $g \circ X_n \xrightarrow{\mathcal{L}} g \circ X$.

8. La convergence en loi d'un point de vue fonctionnel. — Nous nous proposons de donner une définition équivalente de la convergence en loi faisant intervenir une classe de "fonctions test." Une telle classe, notée \mathcal{H} , est formée de fonctions *continues, bornées* sur \mathbb{R} , possédant la propriété suivante :

(D) Pour tout couple (a, b) de réels, avec $a < b$, il existe un élément $f \in \mathcal{H}$ tel que l'on ait : $I_{]-\infty, a]} \leq f \leq I_{]-\infty, b]}$.

On pourra, par exemple, prendre pour \mathcal{H} l'une quelconque des trois classes suivantes :

- a) la classe de *toutes* les fonctions continues bornées sur \mathbb{R} ;
- b) la classe, plus restreinte, des fonctions lipschitziennes et bornées sur \mathbb{R} ;
- c) la classe, encore plus restreinte, des fonctions de la forme :

$$x \mapsto 1 \wedge \left[\frac{(b-x)^+}{b-a} \right], \quad (a < b).$$

THÉORÈME 8.1. — *Donnons-nous une suite (X_n) de variables aléatoires réelles et une variable aléatoire X , non nécessairement définies sur un même espace probabilisé, et choisissons une classe \mathcal{H} de fonctions continues, bornées sur \mathbb{R} , possédant la propriété (D). Alors les deux propriétés suivantes sont équivalentes :*

- 1) $X_n \xrightarrow{\mathcal{L}} X$;
- 2) Pour tout $f \in \mathcal{H}$, on a : $\mathbb{E}[f \circ X_n] \rightarrow \mathbb{E}[f \circ X]$ ($n \rightarrow \infty$).

Démonstration

1) \Rightarrow 2) Grâce au Théorème 7.2, il existe des versions X'_n, X' de X_n, X , définies sur un même espace probabilisé, telles que X'_n converge vers X' presque sûrement. On voit alors que pour tout $f \in \mathcal{H}$ la suite $(f \circ X'_n)$ converge presque sûrement vers $f \circ X'$. Puisque f est bornée, il résulte du théorème de convergence dominée que $\mathbb{E}[f \circ X'_n] \rightarrow \mathbb{E}[f \circ X']$; la même relation a naturellement lieu pour X_n, X , de sorte qu'en définitive on a : $\mathbb{E}[f \circ X_n] \rightarrow \mathbb{E}[f \circ X]$.

2) \Rightarrow 1) Désignons par F la fonction de répartition de X et par F_n celle de X_n . Soient x un point de continuité pour F et δ un réel strictement positif. En vertu de la propriété (D), il existe deux éléments f, g de \mathcal{H} tels que $I_{]-\infty, x-\delta]} \leq f \leq I_{]-\infty, x]} \leq g \leq I_{]-\infty, x+\delta]}$. On a alors, pour tout n , $f \circ X_n \leq I_{\{X_n \leq x\}} \leq g \circ X_n$ et par conséquent $\mathbb{E}[f \circ X_n] \leq F_n(x) \leq \mathbb{E}[g \circ X_n]$. On en déduit, en faisant tendre n vers l'infini,

$$\mathbb{E}[f \circ X] \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq \mathbb{E}[g \circ X]$$

d'où, a fortiori,

$$F(x - \delta) \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq F(x + \delta).$$

Il ne reste plus qu'à faire tendre δ vers 0 pour obtenir la convergence de $(F_n(x))$ vers $F(x)$. \square

Remarque. — On voit sans peine que le Théorème 8.1 est encore valable si, au lieu de supposer les fonctions de la classe \mathcal{H} continues et bornées, on suppose que chacune d'entre elles est borélienne, bornée et telle que l'ensemble de ses points de discontinuité est négligeable pour la loi de X .

9. Le théorème de Paul Lévy. — Nous avons rencontré au paragraphe 6 de ce chapitre un critère de convergence en loi d'une suite de variables aléatoires à valeurs dans \mathbb{N} , critère qui utilise les *fonctions génératrices*. Dans le cas général, on dispose d'un critère analogue, où les fonctions génératrices sont remplacées par les *fonctions caractéristiques*. Ce critère, dont la démonstration est plus élaborée, porte le nom de théorème de Paul Lévy. Nous en donnons ci-après une version dont la démonstration est due essentiellement à Giorgio Letta.

THÉORÈME 9.1. — *Donnons-nous une suite de variables aléatoires (X_n) et une variable aléatoire X . Pour chaque entier n notons μ_n la loi de X_n et*

$\hat{\mu}_n$ la transformée de Fourier de μ_n . De même, désignons par μ et $\hat{\mu}$ la loi de X et la transformée de Fourier de μ . Si $\hat{\mu}_n \rightarrow \hat{\mu}$, au sens de la convergence simple, alors $X_n \xrightarrow{\mathcal{L}} X$.

Démonstration. — Elle repose sur deux lemmes. Le premier lemme est le Lemme 7.2 du chapitre 13, utilisé pour démontrer que la fonction caractéristique déterminait la mesure. Par commodité ici, appelons-le “Lemme 1.” Le second lemme s’énonce comme suit.

LEMME 2. — *Donnons-nous une suite (μ_n) de mesures de probabilité sur \mathbb{R} , une mesure de probabilité μ sur \mathbb{R} et une densité de probabilité g sur \mathbb{R} vérifiant la propriété ancipitale du Lemme 1, à savoir qu’elle est aussi, à un facteur près, la fonction caractéristique d’une densité de probabilité sur \mathbb{R} . Si $\hat{\mu}_n \rightarrow \hat{\mu}$ au sens de la convergence simple, alors $\mu_n * g \rightarrow \mu * g$ étroitement.*

Démonstration. — Désignons par h_n, h les densités de $\mu_n * g, \mu * g$, respectivement. On a, d’après le Lemme 1

$$h_n(u) = c \int_{\mathbb{R}} e^{iux} f(x) \hat{\mu}_n(-x) dx.$$

Or, pour tout n , la fonction $|\hat{\mu}_n|$ est majorée par 1 (qui est intégrable pour la mesure $f\lambda$) et $\hat{\mu}_n \rightarrow \hat{\mu}$ au sens de la convergence simple. D’après le théorème de convergence dominée (Théorème 9.3 du chap. 10), on a donc

$$h_n(u) \rightarrow c \int_{\mathbb{R}} e^{iux} f(x) \hat{\mu}(-x) dx = h(u),$$

lorsque n tend vers l’infini. Il résulte alors du théorème de Scheffé 6.2 que $\mu_n * g \rightarrow \mu * g$ étroitement. \square

Revenons à présent à la démonstration du Théorème 9.1.

1) Pour tout $\varepsilon > 0$, on peut construire une variable aléatoire Z , indépendante de la suite (X_n) et de X , admettant une densité g ayant la propriété ancipitale et telle que $\mathbb{E}[|Z|] < \varepsilon$. (Par exemple, si Y est une variable aléatoire de loi $\mathcal{N}(0, 1)$, la variable aléatoire $Z = \varepsilon Y$ convient, puisqu’une densité normale a la propriété ancipitale). Le Lemme 2 peut alors s’écrire :

$$(9.1) \quad \forall \varepsilon > 0 \quad X_n + Z \xrightarrow{\mathcal{L}} X + Z.$$

2) Soit \mathcal{H} la classe des fonctions lipschitziennes bornées sur \mathbb{R} . Nous allons montrer que pour tout f dans \mathcal{H} on a :

$$(9.2) \quad \mathbb{E}[f \circ X_n] \rightarrow \mathbb{E}[f \circ X].$$

En effet, soit f bornée lipschitzienne de constante l ; alors

$$\begin{aligned} |\mathbb{E}[f \circ X_n] - \mathbb{E}[f \circ X]| &\leq |\mathbb{E}[f \circ X_n] - \mathbb{E}[f \circ (X_n + Z)]| \\ &\quad + |\mathbb{E}[f \circ (X_n + Z)] - \mathbb{E}[f \circ (X + Z)]| + |\mathbb{E}[f \circ (X + Z)] - \mathbb{E}[f \circ X]| \end{aligned}$$

Le fait que f est lipschitzienne de constante l implique que le premier et le troisième terme du second membre sont majorés tous deux par $l \mathbb{E}[|Z|] \leq l\varepsilon$. Le deuxième terme du second membre tend vers 0 lorsque n tend vers l'infini, en vertu du Théorème 8.1 et de (9.1). Puisque $\varepsilon > 0$ est arbitraire, cela suffit pour démontrer (9.2).

3) D'après le Théorème 8.1, la propriété que pour tout $f \in \mathcal{H}$ on ait $\mathbb{E}[f \circ X_n] \rightarrow \mathbb{E}[f \circ X]$ est équivalente à $X_n \xrightarrow{\mathcal{L}} X$. Ceci achève la démonstration du Théorème 9.1.

COROLLAIRE. — Soient μ et ν deux lois de probabilité sur \mathbb{R} . Si $\hat{\mu} = \hat{\nu}$, alors $\mu = \nu$.

Démonstration. — Prenons la suite de terme général $\mu_n = \mu$ ($n \geq 1$); alors la suite $(\hat{\mu}_n)$ est constante et $\hat{\mu}_n \rightarrow \hat{\mu} = \hat{\nu}$. D'après le Théorème 9.1, la suite (μ_n) converge étroitement à la fois vers μ et vers ν , d'où $\mu = \nu$, d'après l'unicité de la convergence étroite. \square

COMPLÉMENTS ET EXERCICES

1. — Soit une suite de points aléatoires $(M_n = (X_n, Y_n))$ qui converge en probabilité vers un point aléatoire $M = (X, Y)$ (ceci entraîne que $X_n \xrightarrow{p} X$ et $Y_n \xrightarrow{p} Y$). Montrer directement que :

a) $X_n + Y_n \xrightarrow{p} X + Y$; b) $X_n Y_n \xrightarrow{p} XY$.

2. — Il n'est pas vrai en général que $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} Y$ impliquent $X_n + Y_n \xrightarrow{\mathcal{L}} X + Y$, mais on a le résultat suivant :

Soit une suite de points aléatoires $(M_n = (X_n, Y_n))$ telle que $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{p} 0$ (la variable X étant définie sur le même espace probabilisé que les X_n); alors

a) $X_n + Y_n \xrightarrow{\mathcal{L}} X$; b) $X_n Y_n \xrightarrow{p} 0$, donc aussi $X_n Y_n \xrightarrow{\mathcal{L}} 0$.

Il en résulte que si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} c$ (c réel), alors

a') $X_n + Y_n \xrightarrow{\mathcal{L}} X + c$; b') $X_n Y_n \xrightarrow{\mathcal{L}} cX$.

3. (*La convergence presque sûre n'entraîne pas la convergence en moyenne quadratique*). — Considérons une suite de variables aléatoires (X_n) de loi $P_{X_n} = (1 - 1/n^2)\varepsilon_0 + (1/2n^2)(\varepsilon_{-n} + \varepsilon_{+n})$. La suite (X_n) tend vers 0 presque sûrement, mais non en moyenne quadratique.

4. — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires presque sûrement bornées. Montrer que si $X_n \xrightarrow{p} X$, alors pour tout $r > 0$ on a $\mathbb{E}[|X_n - X|^r] \rightarrow 0$.

5. — Pour tout entier $n \geq 0$ et tout nombre p tel que $0 \leq p \leq 1$, on pose $B(n, p; k) = \binom{n}{k} p^k q^{n-k}$. Montrer que si n tend vers $+\infty$ et p tend vers 0 de sorte que $np = \lambda$ reste constant, alors, pour tout $k \geq 0$, on a $B(n, p; k) \rightarrow \pi(k; \lambda) = e^{-\lambda} \lambda^k / k!$. Si donc pour chaque entier n la variable X_n est une variable binomiale de paramètres $p = \lambda/n$, n , alors la suite (X_n) converge en loi vers une variable aléatoire de Poisson de paramètre λ .

6. — Soient X une variable aléatoire *centrée* et ε un nombre strictement positif.

a) On pose : $g(\varepsilon) = \mathbb{E}[e^{\varepsilon X}]$. Établir l'inégalité :

$$\mathbb{P} \left\{ X \geq \frac{t + \text{Log } g(\varepsilon)}{\varepsilon} \right\} \leq e^{-t}, \quad \text{pour } t > 0.$$

b) On pose $g^*(\varepsilon) = \mathbb{E}[e^{-\varepsilon X}]$. Établir l'inégalité :

$$\mathbb{P} \left\{ X \leq -\frac{t + \text{Log } g^*(\varepsilon)}{\varepsilon} \right\} \leq e^{-t}, \quad \text{pour } t > 0.$$

7. — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires du second ordre vérifiant $\sum_{n \geq 1} \mathbb{E}[X_n^2] < +\infty$. Montrer que

a) $X_n \rightarrow 0$ presque sûrement ;

b) $X_n \rightarrow 0$ en moyenne quadratique.

Ceci montre que si une suite de variables aléatoires vérifie la Proposition 4.2 (deuxième critère de convergence presque sûre) pour $r = 2$, alors elle converge aussi en moyenne quadratique.

8. — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires du second ordre. Posons $\mathbb{E}[X_n] = \mu_n$, $\text{Var } X_n = \sigma_n^2$ et supposons que $|\mu_n| \rightarrow +\infty$ et que $\sigma_n^2 / |\mu_n| = O(1)$. Montrer qu'alors $X_n / \mu_n \rightarrow 1$ en moyenne quadratique, donc aussi en probabilité.

9. — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires *décroissante*. Montrer que si $X_n \xrightarrow{p} 0$, alors $X_n \xrightarrow{p.s.} 0$.

10. — Prenons pour espace probabilisé $([0, 1], \mathcal{B}^1, \lambda)$, où λ désigne la mesure de Lebesgue sur $[0, 1]$. On considère la suite (X_n) ($n \geq 1$) des variables aléatoires définies sur cet espace par :

$$X_n(x) = \begin{cases} 1/\sqrt{x}, & \text{si } 0 < x < 1/n; \\ 0, & \text{si } 1/n \leq x \leq 1. \end{cases}$$

Montrer que $X_n \xrightarrow{p} 0$, mais que X_n ne tend pas vers 0 en moyenne quadratique. (Voir aussi exercice 17.)

11. — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires; posons $Y_n = \frac{1}{n} \sum_{k=1}^n X_k$ ($n \geq 1$). De $X_n \xrightarrow{p} 0$, on ne peut déduire $Y_n \xrightarrow{p} 0$. Autrement dit, le théorème de Césaro n'est pas vrai pour la convergence en probabilité; en revanche, il l'est pour la convergence presque sûre.
 [Prendre pour loi de X_n la loi $(1/n)\varepsilon_n + (1 - 1/n)\varepsilon_0$ ($n \geq 1$) et supposer les X_n indépendantes.]

12. — Soient U une variable aléatoire uniformément répartie sur $[0, 1]$ et (U_n) ($n \geq 1$) une suite de variables aléatoires indépendantes ayant chacune la même loi que U . Soit d'autre part Y une variable aléatoire exponentielle de paramètre 1. Pour tout $n \geq 1$, on pose $Z_n = n \min(U_1, \dots, U_n)$. Montrer que $Z_n \xrightarrow{\mathcal{L}} Y$.

13. — Soit X une variable aléatoire exponentielle de paramètre $\lambda > 0$. Déterminer la loi de la variable aléatoire $e^{-\lambda X}$.

14. — On désigne par (X_n) ($n \geq 1$) une suite de variables aléatoires indépendantes, toutes exponentielles de même paramètre $\lambda > 0$. Déterminer les limites en loi des suites de terme général :

- a) $A_n = n \min(e^{-\lambda X_1}, \dots, e^{-\lambda X_n})$;
- b) $B_n = n^{1/\lambda} \min(e^{-X_1}, \dots, e^{-X_n})$;
- c) $C_n = n^{-1/\lambda} \max(e^{X_1}, \dots, e^{X_n})$;
- d) $D_n = \max(X_1, \dots, X_n) - \text{Log } n$, lorsque le paramètre λ vaut 1.

[Utiliser les exercices 12 et 13.]

15. — Soient X une variable aléatoire à valeurs dans $[0, +\infty[$ et (X_n) ($n \geq 1$) une suite de variables aléatoires indépendantes, admettant toutes même loi que X . Montrer que

- a) si $P\{X > x\} = o(1/x)$ lorsque x tend vers l'infini, alors $Z_n = \frac{1}{n} \max(X_1, \dots, X_n) \xrightarrow{\mathcal{L}} 0$;
- b) si $P\{X > x\} \sim \alpha/x^\lambda$ lorsque x tend vers l'infini, avec $\alpha, \lambda > 0$, alors

$$Z_n = \frac{1}{n^{1/\lambda}} \max(X_1, \dots, X_n) \xrightarrow{\mathcal{L}} Y,$$

où Y est une variable aléatoire de Fréchet, dont la fonction de répartition, pour $x > 0$, est donnée par $P\{Y \leq x\} = e^{-\alpha x^{-\lambda}}$.

16. — Reprenons les notations de l'Exercice 15, mais supposons X à valeurs dans \mathbb{R} et de loi *symétrique*. Alors les affirmations a) et b) de l'Exercice 15 sont encore valables. Nous le montrons pour b) :

Pour $x < 0$, on a $P\{Z_n \leq x\} = (P\{X \leq n^{1/\lambda}x\})^n$, qui est égal, puisque $\mathcal{L}(X)$ est symétrique, à $(P\{X > n^{1/\lambda}|x|\})^n \sim \left(\frac{\alpha}{n(|x|)^\lambda}\right)^n$, qui tend vers 0 lorsque n tend vers l'infini.

Pour $x > 0$, on a $P\{Z_n \leq x\} = (P\{X \leq n^{1/\lambda}x\})^n = (1 - P\{X > n^{1/\lambda}x\})^n = \left(1 - \frac{\alpha}{nx^\lambda} + o(1/n)\right)^n$, une expression qui tend vers $e^{-\alpha x^{-\lambda}}$ lorsque n tend vers l'infini.

Pour a) on peut prendre comme exemple pour $\mathcal{L}(X)$ la première loi de Laplace, ou bien $\mathcal{N}(0, 1)$ et pour b) la loi de Cauchy $\mathcal{C}(0, 1)$ avec $\alpha = 1/\pi$ et $\lambda = 1$.

17. (E. Khalili). — On conserve les mêmes hypothèses que dans l'exercice 10 au sujet de la suite des variables aléatoires (X_n) .

a) Calculer explicitement la fonction de répartition F_n de X_n et en déduire $X_n \xrightarrow{\mathcal{L}} 0$.

b) Montrer que $X_n \xrightarrow{p.s.} 0$.

Les indications données ci-après tiennent lieu de solution. Pour évaluer $F_n(y)$, distinguer les quatre cas : $y < 0$, $y = 0$, $0 < y \leq \sqrt{n}$, $\sqrt{n} < y$. On obtient : $F_n(y) = 0$, $1 - \frac{1}{n}$, $1 - \frac{1}{n}$ ou $1 - \frac{1}{y^2}$, selon que $y < 0$, $y = 0$, $0 < y \leq \sqrt{n}$ ou $\sqrt{n} < y$. On a donc $\lim_n F_n(y) = 0$ pour $y < 0$ et $\lim_n F_n(y) = 1$ pour $y \geq 0$.

Pour b) il suffit de remarquer que pour $0 < x \leq 1$, on a $X_n(x) \rightarrow 0$, lorsque n tend vers l'infini et que $\lambda\{]0, 1]\} = 1$.

18. — Soit (X_n) ($n \geq 0$) une suite de variables aléatoires absolument continues, de support \mathbb{R} , la densité de X_n étant donnée par

$$f_n(x) = \begin{cases} n/2\pi, & \text{si } x = 0; \\ \frac{1 - \cos(nx)}{n\pi x^2}, & \text{si } x \neq 0. \end{cases}$$

1) Vérifier que pour tout $n \geq 1$ la fonction f_n est une densité de probabilité.

2) On pose $F(x) = \int_{-\infty}^x f_n(t) dt$. Montrer que

$$\lim_{n \rightarrow \infty} F_n(x) = \begin{cases} 0, & \text{si } x < 0; \\ 1/2, & \text{si } x = 0; \\ 1, & \text{si } x > 0; \end{cases}$$

c'est-à-dire que $X_n \xrightarrow{\mathcal{L}} 0$. (On rappelle que $\frac{1}{\pi} \int_{-\infty}^{+\infty} \left(\frac{\sin t}{t}\right)^2 dt = 1$.)

Remarque 1. — La suite (f_n) vérifie $\lim_{n \rightarrow \infty} f_n(x) = \begin{cases} +\infty, & \text{si } x = 0; \\ 0, & \text{si } x \neq 0; \end{cases}$ et la limite en loi de la suite (X_n) n'est pas absolument continue.

Remarque 2. — La loi de X_n admet pour fonction caractéristique $\varphi_n(t) = \left(1 - \frac{|t|}{n}\right) I_{[-n, +n]}(t)$. C'est une *loi triangulaire* de Khintchine.

LOIS DES GRANDS NOMBRES

A l'origine du calcul des probabilités se trouve le désir de fournir des modèles à certains faits expérimentaux que l'on appelait confusément « lois empiriques du hasard » et qui se manifestent par une étonnante stabilité des fréquences d'un évènement quand on donne à celui-ci un grand nombre d'occasions de se produire. Ainsi, par exemple, on s'est aperçu il y a fort longtemps qu'en lançant un grand nombre de fois une pièce de monnaie parfaite, la fréquence d'apparition de « pile » se stabilise autour de la valeur $\frac{1}{2}$, valeur qu'on était tenté de définir comme la « probabilité » d'apparition de « pile ».

C'est J. Bernoulli (*Ars Conjectandi*, 1713) qui, le premier, a fourni un modèle rendant compte de ce phénomène; il a introduit un mode de convergence proche de la convergence en probabilité et il a montré que la suite des fréquences d'apparition de « pile » converge, selon ce mode, vers la valeur $\frac{1}{2}$, la probabilité théorique d'apparition de « pile ». Les arguments de Bernoulli sont de nature combinatoire et fort compliqués. Ils ont été considérablement simplifiés par Tchebychev, grâce à l'inégalité qui porte son nom et que lui-même a introduite à cette occasion. Le problème étudié par J. Bernoulli a été considérablement généralisé et a donné naissance aux différentes versions de la *loi des grands nombres*.

Soit (X_n) ($n \geq 1$) une suite de variables aléatoires réelles centrées. On cherche des conditions suffisantes pour que la suite de variables aléatoires

$$\left(\frac{1}{n} \sum_{k=1}^n X_k \right) \quad (n \geq 1)$$

converge vers 0 au sens de l'un des types de convergence introduits au chapitre 16. Seules les convergences en probabilité et presque sûre ont fait l'objet d'une étude systématique. On parle respectivement de loi *faible* et de loi *forte* des grands nombres.

Définition. — La suite (X_n) ($n \geq 1$) satisfait la *loi faible des grands nombres* si la suite de terme général $\frac{1}{n} \sum_{k=1}^n X_k$ converge vers 0 *en probabilité*. La suite (X_n) ($n \geq 1$) satisfait la *loi forte des grands nombres* si la suite de terme général $\frac{1}{n} \sum_{k=1}^n X_k$ converge vers 0 *presque sûrement*.

1. La loi faible des grands nombres. — Il existe plusieurs conditions suffisantes pour qu'une suite de variables aléatoires (X_n) ($n \geq 1$) satisfasse

la loi faible des grands nombres. Voici quelques énoncés, où l'on convient de poser pour tout $n \geq 1$:

$$(1.1) \quad S_n = \sum_{k=1}^n X_k, \quad Y_n = \frac{S_n}{n}.$$

THÉORÈME 1.1 (Loi faible des grands nombres dans L^2 pour variables aléatoires deux à deux non corrélées). — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires de L^2 , deux à deux non corrélées, centrées. Pour tout $n \geq 1$ on pose $\text{Var } X_n = \sigma_n^2 < +\infty$. Si $(1/n^2) \sum_{k=1}^n \sigma_k^2$ tend vers 0 lorsque n tend vers l'infini, alors Y_n tend vers 0 dans L^2 , d'où aussi $Y_n \rightarrow 0$ en probabilité.

Démonstration. — Puisque les X_n sont deux à deux non corrélées, on a, pour tout $n \geq 1$

$$\mathbb{E}[Y_n^2] = \text{Var } Y_n = \frac{1}{n^2} \text{Var } S_n = \frac{1}{n^2} \sum_{k=1}^n \sigma_k^2,$$

d'où $\mathbb{E}[Y_n^2] \rightarrow 0$ lorsque n tend vers l'infini, c'est-à-dire $Y_n \rightarrow 0$ dans L^2 . La convergence en probabilité de Y_n vers 0 s'en déduit immédiatement en vertu de l'inégalité de Bienaymé-Tchebychev. \square

Remarques. — Le Théorème 1.1 est naturellement valable si les variables aléatoires X_n sont *indépendantes* ou simplement *indépendantes deux à deux*.

APPLICATION 1.2. — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires de L^2 deux à deux non corrélées. Pour tout $n \geq 1$ posons $\mathbb{E}[X_n] = \mu_n$ et supposons que la suite de terme général $\frac{1}{n} \sum_{k=1}^n \mu_k$ tend vers μ et que $(1/n^2) \sum_{k=1}^n \sigma_k^2$ tend vers 0, lorsque n tend vers l'infini. Alors la suite $(\frac{1}{n} \sum_{k=1}^n X_k)$ tend vers μ dans L^2 , donc aussi en probabilité.

Démonstration. — Appliquons le Théorème 1.1 à la suite $(X_n - \mu_n)$ ($n \geq 1$) de variables aléatoires centrées; on a :

$$\frac{1}{n} \sum_{k=1}^n (X_k - \mu_k) = \frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{n} \sum_{k=1}^n \mu_k \rightarrow 0$$

dans L^2 , donc en probabilité. Le résultat en découle. \square

Le théorème suivant concernant des suites de variables *identiquement distribuées* est un corollaire du Théorème 1.1.

THÉORÈME 1.3 (Loi faible des grands nombres dans L^2 pour les variables aléatoires deux à deux non corrélées, identiquement distribuées). — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires de L^2 , centrées, deux à deux non corrélées et identiquement distribuées. Alors $Y_n \rightarrow 0$ dans L^2 , d'où aussi $Y_n \rightarrow 0$ en probabilité.

Démonstration. — Pour tout $n \geq 1$ posons $\text{Var } X_n = \sigma_n^2 = \sigma^2 < +\infty$. Alors

$$\frac{1}{n^2} \sum_{k=1}^n \sigma_k^2 = \frac{\sigma^2}{n} \rightarrow 0;$$

d'où le résultat en vertu du Théorème 1.1. \square

Remarque 1. — Le Théorème 1.3 est naturellement valable si les variables aléatoires X_n sont *indépendantes* ou seulement *indépendantes deux à deux*.

Remarque 2. — La suite de terme général $\mathbb{E}[Y_n^2]$ tend vers 0 *en décroissant*; en effet, $\mathbb{E}[Y_n^2] = \sigma^2/n \downarrow 0$.

APPLICATION 1.4. — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires de L^2 , deux à deux non corrélées, identiquement distribuées. Désignons par μ l'espérance mathématique commune des X_n ; alors $\frac{1}{n} \sum_{k=1}^n X_k$ tend vers μ dans L^2 , donc aussi en probabilité.

Démonstration. — Appliquons le Théorème 1.3 à la suite $(X_n - \mu)$ ($n \geq 1$) de variables aléatoires *centrées*; il vient

$$\frac{1}{n} \sum_{k=1}^n (X_k - \mu) = \frac{1}{n} \sum_{k=1}^n X_k - \mu \rightarrow 0$$

dans L^2 , donc aussi en probabilité. \square

APPLICATION 1.5. — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires indépendantes, identiquement distribuées de loi commune $p\varepsilon_1 + q\varepsilon_0$ ($0 \leq p \leq 1$, $p+q=1$). Alors $\frac{1}{n} \sum_{k=1}^n X_k$ tend vers p dans L^2 , donc aussi en probabilité.

C'est l'exemple classique du jeu de « pile » ou « face », dû à Bernoulli.

Nous venons de voir que si les variables aléatoires sont de L^2 , la démonstration de la loi faible des grands nombres (Théorèmes 1.1 et 1.3) est particulièrement simple. En fait, on peut s'affranchir de cette hypothèse et supposer que les variables aléatoires sont *seulement de L^1* , pourvu qu'elles soient, en outre, indépendantes deux à deux et *identiquement distribuées*. La démonstration de la loi faible des grands nombres dans ce cas est plus élaborée et fait appel à des techniques de troncature et de centrage, comme nous allons le voir.

THÉORÈME 1.6 (Loi faible des grands nombres dans L^1 pour variables aléatoires indépendantes deux à deux, identiquement distribuées). — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires de L^1 , centrées, indépendantes deux à deux et identiquement distribuées. En utilisant les notations (1.1), on a alors : $Y_n \rightarrow 0$ dans L^1 , donc aussi $Y_n \rightarrow 0$ en probabilité.

Démonstration. — Si les X_n étaient dans L^2 , le théorème serait démontré, car, d'après le Théorème 1.3, on aurait $Y_n \rightarrow 0$ dans L^2 , donc dans L^1 . L'idée de la démonstration est de se ramener au cas de L^2 par des techniques de *troncature* et de *centrage*. Nous avons besoin du lemme suivant.

LEMME 1.7. — Pour tout $\varepsilon > 0$ il existe une fonction f borélienne et bornée sur \mathbb{R} , telle que $f \circ X_1$ soit centrée et vérifie la relation

$$\|X_1 - f \circ X_1\|_1 < \varepsilon.$$

En outre, f ne dépend que de la loi de X_1 .

Démonstration du lemme

a) Donnons-nous $\varepsilon > 0$; puisque X_1 est dans L^1 , on peut choisir $c > 0$ assez grand pour que la fonction

$$g(x) = x I_{[-c,+c]} = \begin{cases} x, & \text{si } |x| \leq c; \\ 0, & \text{sinon;} \end{cases}$$

vérifie

$$\|X_1 - g \circ X_1\|_1 = \int_{\{|x|>c\}} |x| d\mu(x) < \varepsilon.$$

b) La fonction g ne convient pas nécessairement, puisque $g \circ X_1$ n'est pas nécessairement centrée; pour réaliser la condition de centrage, on introduit la fonction

$$f(x) = g(x) - m, \quad \text{où } m = \mathbb{E}[g \circ X_1],$$

c'est-à-dire

$$f(x) = x I_{[-c,+c]}(x) - \int_{[-c,+c]} x d\mu(x).$$

c) La fonction f convient pour c assez grand; en effet, $f \circ X_1$ est centrée par construction; pour voir qu'elle vérifie $\|X_1 - f \circ X_1\|_1 < \varepsilon$, on opère comme suit. On choisit c tel que $\|X_1 - g \circ X_1\|_1 < \varepsilon$; ceci est possible d'après a). On a alors, puisque X_1 est centrée, les relations

$$|m| = |\mathbb{E}[X_1] - m| = |\mathbb{E}[X_1] - \mathbb{E}[g \circ X_1]| \leq \|X_1 - g \circ X_1\|_1 < \varepsilon.$$

d'où, en définitive,

$$\|X_1 - f \circ X_1\|_1 \leq \|X_1 - g \circ X_1\|_1 + |m| < 2\varepsilon. \quad \square$$

Revenons à la démonstration du Théorème 1.6. Posons $X'_n = f \circ X_n$, $S'_n = X'_1 + \dots + X'_n$ et $Y'_n = S'_n/n$. Les variables aléatoires X'_n sont centrées, indépendantes deux à deux et identiquement distribuées. Elles sont, de plus, bornées, donc dans L^2 . Il résulte alors du Théorème 1.3 que $Y'_n \rightarrow 0$ dans L^2 , donc dans L^1 . D'autre part,

$$\|Y_n - Y'_n\|_1 \leq \frac{1}{n} \sum_{k=1}^n \|X_k - X'_k\|_1.$$

Or, pour tout $k = 1, \dots, n$, l'expression $\|X_k - X'_k\|_1$ ne dépend que de la loi commune des X_n ; toutes ces quantités sont donc égales, d'où

$$\|Y_n - Y'_n\|_1 \leq \|X_1 - X'_1\|_1 < \varepsilon.$$

Enfin, on a

$$\|Y_n\|_1 \leq \|Y_n - Y'_n\|_1 + \|Y'_n\|_1,$$

d'où, pour n assez grand, $\|Y_n\|_1 < 2\varepsilon$; ainsi la suite de terme général $\|Y_n\|_1 = \mathbb{E}[|Y_n|]$ converge bien vers 0 lorsque n tend vers l'infini. \square

Remarque 1. — Le Théorème 1.6 est naturellement valable si les variables aléatoires X_n sont indépendantes.

Remarque 2. — Dans le cas où les variables aléatoires X_n sont *indépendantes*, la suite de terme général $\mathbb{E}[|Y_n|] = \|Y_n\|_1$ tend vers 0 *en décroissant*.

Cette remarque se démontre comme suit. On a :

$$Y_{n-1} = \frac{n}{n-1}Y_n - \frac{X_n}{n-1},$$

d'où

$$\mathbb{E}[Y_{n-1} | Y_n] = \frac{n}{n-1}Y_n - \frac{1}{n-1}\mathbb{E}[X_n | Y_n].$$

Or $\mathbb{E}[X_1 | Y_n] = \dots = \mathbb{E}[X_n | Y_n]$, puisque les variables aléatoires X_1, \dots, X_n sont indépendantes et identiquement distribuées; d'où

$$Y_n = \mathbb{E}[Y_n | Y_n] = \frac{1}{n}(\mathbb{E}[X_1 | Y_n] + \dots + \mathbb{E}[X_n | Y_n]) = \mathbb{E}[X_n | Y_n].$$

Il en résulte que

$$\mathbb{E}[Y_{n-1} | Y_n] = \frac{n}{n-1}Y_n - \frac{1}{n-1}Y_n = Y_n;$$

d'où

$$|Y_n| \leq \mathbb{E}[|Y_{n-1}| | Y_n].$$

En prenant l'espérance mathématique des deux côtés, il vient

$$\mathbb{E}[|Y_n|] \leq \mathbb{E}[|Y_{n-1}|]. \quad \square$$

2. La loi forte des grands nombres. — Nous commençons par donner un énoncé de la loi forte des grands nombres pour les variables aléatoires de L^2 . (On peut trouver sa démonstration dans l'ouvrage de Fourgeaud-Fuchs (*op. cit.*)).

THÉORÈME 2.1 (Loi forte des grands nombres pour les variables aléatoires de L^2). — *Soit (X_n) ($n \geq 1$) une suite de variables aléatoires de L^2 , centrées et indépendantes. Pour tout $n \geq 1$, on pose $\text{Var } X_n = \sigma_n^2 < +\infty$ et, comme précédemment*

$$(2.1) \quad S_n = \sum_{k=1}^n X_k, \quad Y_n = \frac{S_n}{n} \quad (n \geq 1).$$

Si la série $\sum_{n \geq 1} \sigma_n^2/n^2$ est convergente, alors la suite (Y_n) tend vers 0 presque sûrement.

THÉORÈME 2.2 (Rajchman). — *Soit (X_n) ($n \geq 1$) une suite de variables aléatoires de L^2 , centrées et indépendantes. Pour tout $n \geq 1$, on pose $\text{Var } X_n = \sigma_n^2$ et on reprend les notations (2.1) ci-dessus. Si $\sup_n \sigma_n^2 < +\infty$, alors*

- a) $Y_n \rightarrow 0$ presque sûrement.
- b) $Y_n \rightarrow 0$ dans L^2 .

Démonstration

a) Posons $\sigma^2 = \sup_n \sigma_n^2 < +\infty$; alors $\sum_{n \geq 1} \frac{\sigma_n^2}{n^2} \leq \sigma^2 \sum_{n \geq 1} \frac{1}{n^2} < \infty$; d'où $Y_n \rightarrow 0$ presque sûrement, d'après le Théorème 2.1.

b) On a : $\mathbb{E}[Y_n^2] = \text{Var } Y_n = \frac{1}{n^2} \sum_{k=1}^n \sigma_k^2 \leq \frac{\sigma^2}{n} \rightarrow 0$; d'où $Y_n \rightarrow 0$ dans L^2 , d'après le Théorème 1.1. \square

Remarque 1. — Rajchman a établi les mêmes conclusions en remplaçant « indépendantes » par « deux à deux non corrélées ».

Remarque 2. — Il en résulte que, dans l'énoncé du théorème de Bernoulli, on peut remplacer la convergence en probabilité par la convergence presque sûre (E. Borel).

THÉORÈME 2.3 (Loi forte des grands nombres pour variables aléatoires de L^1 (Kolmogorov)). — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires de L^1 , centrées, indépendantes et identiquement distribuées. Utilisant les notations (2.1) ci-dessus, on a : $Y_n \rightarrow 0$ presque sûrement.

Démonstration (L. Pratelli, non publié).

a) On sait (Théorème 4.2 du chap. 16) que la propriété $Y_n \xrightarrow{p.s.} 0$ équivaut à l'énoncé :

$$\text{pour tout } \varepsilon > 0, \text{ on a } \mathbb{P}\left\{\sup_{k \geq m} |Y_k| > \varepsilon\right\} \longrightarrow 0 \quad (m \rightarrow \infty).$$

b) Démontrons le lemme suivant.

LEMME 2.4. — Pour tout $m \geq 1$, tout $\varepsilon > 0$ on a :

$$\varepsilon \mathbb{P}\left\{\sup_{k \geq m} |Y_k| > \varepsilon\right\} \leq \|Y_m\|_1,$$

c'est-à-dire que si $Y_m \rightarrow 0$ dans L^1 , alors $Y_m \rightarrow 0$ presque sûrement.

c) Le résultat découle de a) et b) et du Théorème 1.6 (loi faible des grands nombres dans L^1).

Démonstration du lemme. — On va démontrer l'énoncé suivant (équivalent au lemme) : Pour tout couple d'entiers (m, n) tel que $1 \leq m \leq n$ et tout $\varepsilon > 0$ on a :

$$\varepsilon \mathbb{P}\left\{\sup_{m \leq k \leq n} |Y_k| > \varepsilon\right\} \leq \|Y_m\|_1.$$

Introduisons l'ensemble $T_n = \sup\{k : 1 \leq k \leq n, |Y_k| > \varepsilon\}$ (avec la convention $\sup \emptyset = -\infty$) et posons $A = \{\sup_{m \leq k \leq n} |Y_k| > \varepsilon\}$. Alors

$$A = \{T_n \geq m\} = \sum_{m \leq k \leq n} \{T_n = k\} \quad \text{et} \quad \varepsilon \mathbb{P}(A) = \varepsilon \sum_{m \leq k \leq n} \mathbb{P}\{T_n = k\}.$$

Or, par définition de T_n , on a, pour tout k tel que $m \leq k \leq n$ les relations :

$$\begin{aligned} \varepsilon P\{T_n = k\} &\leq \int_{\{T_n=k\}} |Y_k| dP = \int_{\{T_n=k, Y_k>0\}} Y_k dP + \int_{\{T_n=k, Y_k<0\}} (-Y_k) dP \\ &= B + C. \end{aligned}$$

Calculons séparément B et C . D'abord

$$B = \frac{1}{k} \sum_{j=1}^k \int_{\{T_n=k, Y_k>0\}} X_j dP.$$

Or, puisque les X_n sont indépendantes et identiquement distribuées, toutes les intégrales du second membre sont égales. Ce second membre est donc aussi égal à la moyenne arithmétique de k nombres tous égaux à l'intégrale $\int_{\{T_n=k, Y_k>0\}} X_1 dP$. C'est donc aussi la moyenne arithmétique de m ($m \leq k$) nombres tous égaux à cette quantité. Par conséquent,

$$B = \frac{1}{m} \sum_{j=1}^m \int_{\{T_n=k, Y_k>0\}} X_1 dP = \int_{\{T_n=k, Y_k>0\}} Y_m dP.$$

On opère de façon analogue pour C et l'on obtient :

$$C = \int_{\{T_n=k, Y_k<0\}} (-Y_m) dP.$$

Finalement

$$\varepsilon P\{T_n = k\} \leq B + C = \int_{\{T_n=k\}} |Y_m| dP;$$

d'où, en sommant sur k ,

$$\varepsilon P(A) \leq \sum_{m \leq k \leq n} \int_{\{T_n=k\}} |Y_m| dP \leq \mathbb{E}[|Y_m|] = \|Y_m\|_1. \quad \square$$

COROLLAIRE 2.5. — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires de L^1 , indépendantes et identiquement distribuées. Alors

$$Y_n = \frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{p.s.} \mathbb{E}[X_1].$$

Ce corollaire admet une réciproque; cf. Exercice 3.

3. Les lemmes de Borel-Cantelli

LEMME 3.1 (Borel-Cantelli). — Soit (A_n) ($n \geq 1$) une suite d'évènements; posons $A^* = \limsup_n A_n$.

a) Si $\sum_{n \geq 1} P(A_n) < +\infty$, alors $P(A^*) = 0$. Autrement dit, avec une probabilité égale à 1, au plus un nombre fini d'évènements A_n se réalisent.

b) Supposons les évènements A_n indépendants deux à deux.

Si $\sum_{n \geq 1} P(A_n) = +\infty$, alors $P(A^*) = 1$. Autrement dit, avec une probabilité égale à 1, une infinité d'évènements A_n se réalisent.

Démonstration.

a) On a $A^* = \bigcap_{n \geq 1} \bigcup_{k \geq n} A_k$, d'où, pour tout $n \geq 1$, $P(A^*) \leq P(\bigcup_{k \geq n} A_k) \leq \sum_{k \geq n} P(A_k)$. Le dernier membre étant le reste d'ordre n d'une série *convergente*, tend vers 0 lorsque n tend vers l'infini; d'où $P(A^*) = 0$.

b) Posons $S_n = I_{A_1} + \dots + I_{A_n}$. Par hypothèse, on a :

$$\mathbb{E}[S_n] = \sum_{k=1}^n \mathbb{E}[I_{A_k}] = \sum_{k=1}^n P(A_k) \uparrow +\infty.$$

Puisque les A_n sont indépendants deux à deux, on a également :

$$\text{Var } S_n = \sum_{k=1}^n \text{Var } I_{A_k} \leq \sum_{k=1}^n \mathbb{E}[I_{A_k}^2] = \sum_{k=1}^n \mathbb{E}[I_{A_k}] = \mathbb{E}[S_n].$$

En posant $T_n = S_n/\mathbb{E}[S_n]$, il vient

$$\mathbb{E}[(T_n - 1)^2] = \text{Var } T_n = \frac{\text{Var } S_n}{(\mathbb{E}[S_n])^2} \leq \frac{1}{\mathbb{E}[S_n]},$$

expression qui tend vers 0 lorsque n tend vers l'infini. Ainsi $T_n - 1 \rightarrow 0$ dans L^2 , donc en probabilité; d'où $T_n \rightarrow 1$ en probabilité.

On peut donc extraire de la suite (T_n) une suite partielle (T_{n_k}) telle que $T_{n_k} \rightarrow 1$ *presque sûrement*, lorsque k tend vers l'infini. Comme l'hypothèse $\sum_{n \geq 1} P(A_n) = +\infty$ est équivalente à $\mathbb{E}[S_{n_k}] \uparrow +\infty$ ($k \rightarrow +\infty$), il en résulte que $S_{n_k} \uparrow +\infty$ ($k \rightarrow +\infty$) *presque sûrement*, propriété qui équivaut à $P(A^*) = 1$. \square

Remarque. — La réciproque de a) n'est *pas* vraie. Pour le voir, prenons pour espace probabilisé $(\Omega, \mathfrak{A}, P)$, où $\Omega = [0, 1]$, où \mathfrak{A} est l'ensemble des boréliens de $[0, 1]$ et où P est la mesure de Lebesgue sur $[0, 1]$. Considérons la suite d'évènements $(A_n = [0, 1/n])$ ($n \geq 1$). On vérifie que cette suite est décroissante, donc $A^* = \bigcap_{n \geq 1} A_n = \{0\}$ et $P(A^*) = 0$, alors que

$$\sum_{n \geq 1} P(A_n) = \sum_{n \geq 1} \frac{1}{n} = +\infty.$$

L'hypothèse d'indépendance dans b) est donc essentielle.

Application. — Considérons une suite de parties indépendantes de « pile » ou « face », la probabilité d'apparition de « pile » en une partie étant p ($0 < p < 1$). Soit A un *mot* de longueur $l \geq 1$, c'est-à-dire une suite de l termes dont chacun est soit « pile », soit « face ». Désignons par A_1 l'évènement consistant en le fait que le mot A se réalise dans les l premières parties, par A_2 l'évènement consistant en le fait que le mot A se réalise dans les l parties suivantes, etc. Les évènements A_1, A_2, \dots sont indépendants et pour tout $n \geq 1$ on a $P(A_n) = P(A_1) > 0$, d'où $\sum_{n \geq 1} P(A_n) = +\infty$. Il résulte du lemme (partie b)) qu'avec une probabilité égale à 1, le mot A se réalise *une infinité de fois* au cours du jeu. Un argument analogue montre que si un singe

dactylographe tape « au hasard » sur une machine à écrire, alors, avec une probabilité égale à 1, tout texte, de quelque longueur que ce soit, se réalise *une infinité de fois* au cours de la frappe.¹

Le lemme de Borel-Cantelli a pour conséquence le théorème suivant.

THÉORÈME 3.2 (Loi (0, 1) de E. Borel). — *Soient (A_n) ($n \geq 1$) une suite d'évènements indépendants deux à deux et A^* l'évènement $A^* = \limsup_n A_n$. Alors $P(A^*)$ ne peut prendre que la valeur 0 ou 1, et ceci selon que la série de terme général $P(A_n)$ est convergente ou divergente.*

Ce théorème est un premier exemple de la célèbre loi (0, 1) de Kolmogorov, qui affirme que certains évènements (« terminaux ») ne peuvent admettre pour probabilité que 0 ou 1.

Comme illustration de ce théorème, montrons que si (X_n) ($n \geq 1$) est une suite de variables aléatoires indépendantes telle que la suite (Y_n) ($n \geq 1$), avec $Y_n = \frac{1}{n} \sum_{k=1}^n X_k$, converge presque sûrement vers une limite Y , alors Y est presque sûrement constante. Pour le voir, remarquons tout d'abord que, pour tout $k \geq 1$, le système (X_1, \dots, X_k) est indépendant de $Y = \lim_n (X_1 + \dots + X_n)/n = \lim_n (X_{k+1} + \dots + X_{k+n})/n$, donc Y_k est indépendante de Y . Pour tout x réel l'évènement $\{Y_k \leq x\}$ est donc indépendant de $\{Y \leq x\}$. (L'évènement $\{Y \leq x\}$ est bien un évènement « terminal ».) On a alors $P(\{Y_k \leq x\} \cap \{Y \leq x\}) = P\{Y_k \leq x\}P\{Y \leq x\}$ pour tout x réel. En faisant tendre k vers l'infini, on en déduit $P\{Y \leq x\} = (P\{Y \leq x\})^2$ et donc $P\{Y \leq x\} = 0$ ou 1 pour tout x . Comme l'application $x \mapsto P\{Y \leq x\}$ est une fonction de répartition, elle est nécessairement égale à l'échelon-unité. Ainsi $Y = \text{constante}$. \square

COMPLÉMENTS ET EXERCICES

1. — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires indépendantes, de même loi, de L^2 . On pose $m = \mathbb{E}[X_1]$, $\sigma^2 = \text{Var } X_1$. Pour tout entier $n \geq 2$, on définit les variables aléatoires :

$$Y_n = \frac{1}{n} \sum_{k=1}^n X_k, \quad Z_n = \frac{1}{n-1} \sum_{k=1}^n (X_k - Y_n)^2.$$

- a) Calculer $\mathbb{E}[Z_n]$.
- b) Montrer que $Z_n \xrightarrow{p.s.} \sigma^2$, lorsque n tend vers l'infini.

¹ Borel (Émile). — *Le hasard*. — Paris, Librairie Félix Alcan, 1938.

2. — On se place dans les hypothèses du Théorème 1.6, mais on suppose, en plus, que les variables aléatoires X_n sont *indépendantes* et non seulement indépendantes deux à deux. Montrer directement, en utilisant les fonctions caractéristiques, que $Y_n \xrightarrow{p} 0$.

3. — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires indépendantes, identiquement distribuées. On suppose que $Y_n = (1/n) \sum_{k=1}^n X_k \xrightarrow{p.s.} Y$. Démontrer les propriétés suivantes :

- a) $\sum_{n \geq 1} P\{|X_n| \geq n\} < +\infty$;
- b) les X_n sont intégrables;
- c) $Y = \text{constante}$ presque sûrement.

4. — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires. On pose $S_n = X_1 + \dots + X_n$. Montrer que si $S_n/\sqrt{n} \xrightarrow{\mathcal{L}} Y$, alors $S_n/n \xrightarrow{p} 0$, c'est-à-dire la suite (X_n) ($n \geq 1$) vérifie la loi faible des grands nombres.

5. — Le modèle de « pile » ou « face » de Bernoulli fournit une démonstration remarquable, due à Bernstein, du théorème d'approximation de Weierstrass, qui affirme qu'une fonction *continue* sur un intervalle *borné* peut être approchée par des polynômes, et ceci uniformément sur cet intervalle.

Soit (X_n) ($n \geq 1$) une suite de variables aléatoires indépendantes, identiquement distribuées, de loi commune $p\varepsilon_1 + q\varepsilon_0$ ($0 \leq p \leq 1$, $p + q = 1$). On pose $Y_n = (1/n) \sum_{k=1}^n X_k$; le théorème de Bernoulli affirme que $Y_n \xrightarrow{p} p$. Soit $h : [0, 1] \rightarrow \mathbb{R}$ une fonction *continue*, donc *bornée*. Montrons que $\mathbb{E}[h \circ Y_n] \rightarrow h(p)$ ($n \rightarrow \infty$) et ceci uniformément en $p \in [0, 1]$.

Démonstration. — Pour tout $\delta > 0$, en désignant par μ la loi de Y_n , on a l'évaluation : $|\mathbb{E}[h \circ Y_n - h(p)]| \leq \mathbb{E}[|h \circ Y_n - h(p)|] = A + B$, où $A = \int_{\{|x-p| \leq \delta\}} |h(x) - h(p)| d\mu(x)$ et $B = \int_{\{|x-p| > \delta\}} |h(x) - h(p)| d\mu(x)$.

La fonction h , continue sur $[0, 1]$, est uniformément continue : quel que soit $\varepsilon > 0$, il existe $\delta(\varepsilon) > 0$ tel que $|x - p| \leq \delta$ entraîne $|h(x) - h(p)| < \varepsilon$. De là, $A < \varepsilon$.

Fixons ε , donc δ . Soit M une borne supérieure de $|h|$ sur $[0, 1]$. On a : $B \leq 2M \int_{\{|x-p| > \delta\}} d\mu(x) = 2MP\{|Y_n - p| > \delta\}$, une quantité qui d'après l'inégalité de Bienaymé-Tchebychev est majorée par $2M \text{Var } Y_n / \delta^2 \leq 2M pq / (n\delta^2) \leq 2M / (n\delta^2)$. La dernière quantité est indépendante de p et tend vers 0 lorsque n tend vers l'infini. Il en est de même de B et ceci uniformément en p .

Il en résulte que $\mathbb{E}[h \circ Y_n]$ tend, uniformément en p , vers $h(p)$, lorsque n tend vers l'infini. Or $Y_n = S_n/n$ et $\mathcal{L}(S_n) = B(n, p)$. De là, $\mathbb{E}[h \circ Y_n] = \sum_{k=0}^n h(k/n) \binom{n}{k} p^k (1-p)^{n-k}$ tend, uniformément en $p \in [0, 1]$, vers $h(p)$. C'est

l'énoncé du théorème de Weierstrass. De plus, les polynômes approchants sont explicités. On les appelle les *polynômes de Bernstein*. \square

6. — Considérons la boule $B_n(0, R)$ de \mathbb{R}^n ($n \geq 1$), de centre 0 et de rayon $R \geq 0$. Son volume est donné par $V_n(R) = \pi^{n/2} R^n / \Gamma(1 + n/2)$ (cf. Exercice 12, chap. 14). Projetons ce volume sur l'un des axes de coordonnées (que nous prendrons comme axe des x); on obtient une distribution de masses sur \mathbb{R} , qui admet une densité $g_n(x, R)$, que l'on norme pour avoir une densité de *probabilité* : $f_n(x, R) = g_n(x, R) / V_n(R)$. Prenons $R = \sqrt{n}$; il est remarquable de constater que la suite des densités de probabilité $f_n(x, \sqrt{n})$ converge ponctuellement, lorsque n tend vers l'infini, vers la densité de la loi normale $\mathcal{N}(0, 1)$; autrement dit, pour tout x réel, on a :

$$f_n(x, \sqrt{n}) \rightarrow \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad (n \rightarrow \infty).$$

7. — Soit (u_n) ($n \geq 1$) une suite de nombres réels tels que pour tout $n \geq 1$ on ait $0 < u_n \leq 1$; soit (X_n) ($n \geq 1$) une suite de variables aléatoires indépendantes telles que pour tout $n \geq 1$ la loi de X_n soit la loi : $u_n \varepsilon_{1/u_n} + (1 - u_n) \varepsilon_0$. Alors

- 1) Pour tout $n \geq 1$ on a $\mathbb{E}[X_n] = 1$.
- 2) $X_n \xrightarrow{p} 0$ si et seulement si $u_n \rightarrow 0$.
- 3) $X_n \xrightarrow{p.s.} 0$ si et seulement si $\sum_{n \geq 1} u_n < +\infty$.

On note que si (u_n) ($n \geq 1$) est une suite telle que la *série* de terme général u_n est convergente, il résulte de 3) et du théorème de Césaro que l'on a $\frac{X_1 + \cdots + X_n}{n} \xrightarrow{p.s.} 0$ et pourtant pour tout $n \geq 1$ on a $\mathbb{E}[X_n] = 1$.

**LE RÔLE CENTRAL DE LA LOI NORMALE;
LE THÉORÈME « CENTRAL LIMIT »**

Le théorème « central limit » donne des conditions suffisantes dans lesquelles une somme finie de variables aléatoires réelles convenablement normalisée suit approximativement une loi normale. Comme décrit dans l'aperçu historique reproduit ci-après, la mise en place de ce théorème apparaît dès le dix-neuvième siècle avec Gauss et Laplace, mais il faut attendre le vingtième siècle pour que les critères d'application soient explicités.

1. Aperçu historique. — La loi normale est associée aux noms prestigieux de Gauss et de Laplace; ces mathématiciens ont tous deux introduit cette loi, mais leurs approches étaient fondamentalement différentes.

L'approche de Gauss. — La loi normale a été introduite par Gauss à propos d'un problème d'estimation de paramètre.¹ Soit θ une quantité (inconnue) dont n observations indépendantes ont fourni les n approximations x_1, \dots, x_n . Gauss se propose d'estimer θ à partir des seules valeurs observées x_1, \dots, x_n .

a) Une première estimation de θ est fournie par la *méthode des moindres carrés*; elle consiste à prendre pour estimation de θ la valeur $\tilde{\theta}$ qui réalise le minimum de la fonction $\theta \mapsto \sum_{k=1}^n (x_k - \theta)^2$; on a immédiatement $\tilde{\theta} = (x_1 + \dots + x_n)/n = \bar{x}$.

b) Gauss propose une autre méthode, qui, dans le langage actuel, est connue sous le nom de *méthode du maximum de vraisemblance*. Il introduit une fonction $f(x)$ qui représente la (densité de) probabilité de l'erreur x commise lors d'une observation; il fait sur f les hypothèses (H) suivantes :

$$f > 0, \int_{-\infty}^{+\infty} f(x) dx = 1;$$

f est paire (i.e. on commet aussi souvent une erreur positive qu'une erreur négative);

$f(x)$ est décroissante lorsque $|x| \rightarrow +\infty$ (i.e. on commet moins souvent de grandes erreurs que des petites).

Les erreurs afférentes aux valeurs observées x_1, \dots, x_n de θ sont $x_1 - \theta, \dots, x_n - \theta$, de sorte que la (densité de) probabilité pour que ces erreurs

¹ Gauss (C.F.). — *Theoria motus corporum coelestium*, Liber II, Section III, 1809; surtout les paragraphes 175, 176, 177, 178.

aient été commises simultanément est, compte tenu de l'indépendance des observations :

$$(1.1) \quad L(x_1, \dots, x_n; \theta) = f(x_1 - \theta) \dots f(x_n - \theta).$$

C'est la *fonction de vraisemblance* dans le langage moderne.

Il est à présent naturel de chercher les densités f vérifiant les hypothèses (H) telles que la fonction $L(\theta)$ définie par (1.1) prenne sa valeur maximale au point $\theta = \bar{\theta} = \bar{x}$. En langage moderne, *on cherche les densités f vérifiant (H) telles que l'estimation de θ par la méthode des moindres carrés coïncide avec son estimation par le maximum de vraisemblance*. Gauss montre que les seules densités qui satisfont ces conditions sont les densités *normales* (centrées).

c) Écrivons (1.1) sous la forme :

$$\text{Log } L(\theta) = \text{Log } f(x_1 - \theta) + \dots + \text{Log } f(x_n - \theta).$$

On cherche f vérifiant (H) telle que pour tout n et toute suite (x_1, \dots, x_n) , on ait

$$\left[\frac{d}{d\theta} \text{Log } L(\theta) \right]_{\theta=\bar{x}} = 0,$$

c'est-à-dire en posant $g = \text{Log } f$

$$g'(x_1 - \bar{x}) + \dots + g'(x_n - \bar{x}) = 0.$$

Pour résoudre cette équation fonctionnelle, nous prenons en particulier : $x_2 = \dots = x_n = x_1 - ny$ avec y réel. Alors $\bar{x} = x_1 - (n-1)y$; d'où $x_1 - \bar{x} = (n-1)y$ et $x_2 - \bar{x} = \dots = x_n - \bar{x} = -y$. On a alors, pour tout n et tout y

$$g'[(n-1)y] + (n-1)g'(-y) = 0.$$

Comme f est paire, on a $g'(-y) = -g'(y)$, d'où $g'[(n-1)y] = (n-1)g'(y)$, ou encore

$$\frac{g'[(n-1)y]}{(n-1)y} = \frac{g'(y)}{y} = k;$$

soit encore $g'(x) = kx$ et $\text{Log } f(x) = g(x) = k(x^2/2) + C$. En adaptant les constantes de telle façon que f vérifie les hypothèses (H), on voit que f est la *densité d'une loi normale* (centrée). \square

Remarque. — Voyons sur un exemple que si f n'est pas une densité normale, l'estimation \bar{x} par la méthode des moindres carrés est en général différente de l'estimation par le maximum de vraisemblance. Prenons

$$f(x) = \frac{1}{2} e^{-|x|} \quad (\text{densité de la première loi de Laplace}).$$

La fonction de vraisemblance est :

$$L(x_1, \dots, x_n; \theta) = f(x_1 - \theta) \dots f(x_n - \theta) = \left(\frac{1}{2}\right)^n \exp\left(-\sum_{k=1}^n |x_k - \theta|\right).$$

Toute quantité qui maximise $L(\theta)$ minimise $\sum_{k=1}^n |x_k - \theta|$ et réciproquement. Or toute quantité $\tilde{\theta}$ qui minimise $\sum_{k=1}^n |x_k - \theta|$ est une médiane $M(x_1, \dots, x_n)$ de (x_1, \dots, x_n) . Ainsi l'estimation de θ par la méthode des moindres carrés est la moyenne arithmétique \bar{x} de (x_1, \dots, x_n) , alors que l'estimation de θ par le maximum de vraisemblance est n'importe quelle médiane $M(x_1, \dots, x_n)$ de (x_1, \dots, x_n) . Il serait intéressant de chercher si la densité de la première loi de Laplace est la seule densité (paire) pour laquelle l'estimation par le maximum de vraisemblance est une médiane de (x_1, \dots, x_n) . Le problème est, à notre connaissance, toujours ouvert.

L'approche de Laplace. — Elle est centrée sur l'approximation de la loi binomiale par la loi normale. En langage moderne, il s'agit d'établir le théorème suivant :

Soit (X_n) ($n \geq 1$) une suite de variables aléatoires dont le terme général X_n suit la loi binomiale $B(n, p)$ ($n \geq 1, 0 < p < 1, q = 1 - p$). On pose $Y_n = (X_n - np) / \sqrt{npq}$; alors la suite (Y_n) ($n \geq 1$) converge en loi vers une variable aléatoire de loi $\mathcal{N}(0, 1)$.

Dès 1733, de Moivre a démontré ce théorème dans le cas particulier $p = 1/2$ par des calculs laborieux d'estimation des coefficients binomiaux; sa démonstration fut incluse dans son ouvrage fondamental.² Le cas général $0 < p < 1$ fut traité par Laplace³ par des méthodes où pointe déjà la théorie des fonctions caractéristiques, qui permet à l'heure actuelle de démontrer ce théorème sans aucune difficulté. La contribution de Laplace ne s'arrête pas là. Il a, en effet, mis la loi normale en rapport avec la théorie des erreurs d'observations et ceci par une analyse très fine dont nous donnons ci-dessous un aperçu (avec des notations modernes).

La loi des erreurs. — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires indépendantes, centrées, du second ordre. On pose $\sigma_k^2 = \text{Var } X_k$ et l'on suppose $0 < \sigma_k < +\infty$. Introduisons pour tout entier $n \geq 1$ les variables aléatoires

$$S_n = X_1 + \dots + X_n, \quad Y_n = \frac{S_n}{\sigma(S_n)}, \quad \text{où } \sigma^2(S_n) = \sigma_1^2 + \dots + \sigma_n^2.$$

On cherche à quelles conditions la suite (Y_n) ($n \geq 1$) converge en loi vers une variable aléatoire de loi $\mathcal{N}(0, 1)$. Notons $\psi = \text{Log } \varphi$ la seconde fonction caractéristique. Il résulte de la Proposition 5.3, chap. 13 que pour tout $k \geq 1$ la fonction $\psi_k = \psi_{X_k}$ est de la forme

$$\psi_k = -\frac{\sigma_k^2 t^2}{2} [1 + \varepsilon_k(\sigma_k t)],$$

² de Moivre (A.). — *The doctrine of chances or a method of calculating the probabilities of events in play.* — London, Millar, 1756, p. 235–283. Réimpression par Chelsea, New York, 1967.

³ Laplace (Pierre-Simon, marquis de). — *Théorie analytique des probabilités.* — Paris, Courcier, 1820, p. 83–84. Voir aussi *Œuvres complètes*, vol. VII. — Paris, Gauthier-Villars, 1896.

où $\varepsilon_k(t) \rightarrow 0$ pour $t \rightarrow 0$. On a ensuite (cf. Théorème 5.1, chap. 13)

$$\begin{aligned}\psi_{Y_n}(t) &= \sum_{k=1}^n \psi_k\left(\frac{t}{\sigma(S_n)}\right) = -\frac{t^2}{2} - \frac{t^2}{2} \sum_{k=1}^n \left(\frac{\sigma_k}{\sigma(S_n)}\right)^2 \varepsilon_k\left(\frac{\sigma_k}{\sigma(S_n)}t\right) \\ &= -\frac{t^2}{2} - \frac{t^2}{2} R_n.\end{aligned}$$

Pour que (Y_n) converge en loi vers une variable aléatoire de loi $\mathcal{N}(0,1)$, il faut et il suffit que $\psi_{Y_n}(t) \rightarrow -t^2/2$, pour tout réel t , c'est-à-dire que R_n tende vers 0, pour tout t , lorsque n tend vers l'infini. Il en est bien ainsi si les deux propriétés suivantes sont satisfaites :

$$1) \quad \sup_{k=1, \dots, n} \frac{\sigma_k}{\sigma(S_n)} \rightarrow 0 \quad (n \rightarrow \infty),$$

c'est-à-dire *le plus grand des écarts-types individuels est négligeable devant l'écart-type de la somme S_n* .

2) Lorsque t tend vers 0, les $\varepsilon_k(t)$ convergent vers 0 uniformément en k , c'est-à-dire pour tout $\varepsilon > 0$, il existe $\delta(\varepsilon) > 0$ tel que pour tout t satisfaisant $|t| < \delta(\varepsilon)$, on a pour tout $k \geq 1$ l'inégalité $|\varepsilon_k(t)| < \varepsilon$.

En effet, soit t un réel non nul. Prenons N_0 assez grand pour que pour tout $n \geq N_0$ on ait $\sup_{k=1, \dots, n} \frac{\sigma_k}{\sigma(S_n)} \leq \frac{\delta(\varepsilon)}{|t|}$. On aura alors, pour tout $n \geq N_0$ et tout $k = 1, \dots, n$, les inégalités

$$\frac{\sigma_k}{\sigma(S_n)} |t| \leq \delta(\varepsilon), \quad \varepsilon_k\left(\frac{\sigma_k}{\sigma(S_n)}t\right) < \varepsilon,$$

d'où

$$R_n = \sum_{k=1}^n \left(\frac{\sigma_k}{\sigma(S_n)}\right)^2 \varepsilon_k\left(\frac{\sigma_k}{\sigma(S_n)}t\right) \leq \varepsilon \sum_{k=1}^n \left(\frac{\sigma_k}{\sigma(S_n)}\right)^2 = \varepsilon.$$

Laplace avait bien vu l'importance de la condition 1) pour assurer la convergence de la loi des Y_n vers $\mathcal{N}(0,1)$ et il s'était rendu compte que s'il y avait une erreur disproportionnée par rapport aux autres, cette erreur ferait prévaloir sa loi. Cependant, G. Darmonis,⁴ à qui nous devons cette analyse, a fait remarquer que Laplace n'avait pas vu que la condition 1) n'était *pas suffisante*, et qu'il fallait lui en adjoindre une autre, par exemple, la condition 2).

2. Le théorème « central limit ». — Nous allons étudier dans ce paragraphe la tendance d'une suite de variables aléatoires vers une variable aléatoire normale et nous allons établir quelques théorèmes fournissant des conditions suffisantes pour que cette tendance ait lieu. Tout théorème de ce type porte le nom de théorème « central limit ». C'est Polyà qui, en 1920,

⁴ Darmonis (G.). — *Cours de calcul des probabilités*. — Paris, Cours polycopié, 1955.

a désigné ce théorème par le nom de *zentraler Grenzwertsatz*,⁵ traduit en anglais par «central limit theorem» et en français par «théorème central limite». Nous n'utiliserons pas cette dernière traduction.

THÉORÈME 2.1 (Lindeberg-Lévy, 1920). — *Soit (X_n) ($n \geq 1$) une suite de variables aléatoires indépendantes, identiquement distribuées, de L^2 . Posons :*

$$\begin{aligned} \mu &= \mathbb{E}[X_1], & \sigma^2 &= \text{Var } X_1 \quad (0 < \sigma < +\infty); \\ S_n &= X_1 + \cdots + X_n, & \bar{X}_n &= \frac{S_n}{n}, & Y_n &= \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}. \end{aligned}$$

Alors la suite (Y_n) ($n \geq 1$) converge en loi vers une variable aléatoire de loi $\mathcal{N}(0, 1)$, c'est-à-dire que pour tout réel x on a :

$$\mathbb{P}\{Y_n \leq x\} \rightarrow \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \quad (n \rightarrow +\infty).$$

Démonstration. — Désignons par φ la fonction caractéristique de $X_1 - \mu$; cette variable aléatoire étant centrée, de L^2 , la formule de Taylor jusqu'à l'ordre deux de $\varphi(t)$ au voisinage de $t = 0$ fournit (cf. Proposition 5.3, chap. 13)

$$\varphi(t) = 1 - \frac{t^2}{2}\sigma^2 + o(t^2) \quad (t \rightarrow 0).$$

Or

$$Y_n = \sum_{k=1}^n \frac{X_k - \mu}{\sigma\sqrt{n}},$$

d'où

$$\varphi_{Y_n}(t) = \left(\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \rightarrow e^{-t^2/2};$$

ce qui établit le résultat en vertu du théorème de Paul Lévy. \square

Remarque. — L'importance du Théorème 2.1 réside dans le fait que pour n «grand» les lois de probabilité (en général compliquées) de S_n , \bar{X}_n peuvent être approchées par les lois normales $\mathcal{N}(n\mu, \sigma\sqrt{n})$, $\mathcal{N}(\mu, \sigma/\sqrt{n})$. Le Théorème 2.1 comporte un certain nombre de cas particuliers que nous allons examiner.

Cas particulier 1 (de Moivre-Laplace). — Prenons pour loi commune des X_n la loi de Bernoulli $p\varepsilon_1 + q\varepsilon_0$ ($p, q > 0, p + q = 1$). Alors S_n suit la loi binomiale $B(n, p)$. On a $\mathbb{E}[S_n] = np$, $\text{Var } S_n = npq$, et le Théorème 2.1 affirme :

PROPOSITION 2.2. — *La suite des variables aléatoires de terme général $Y_n = (S_n - np)/\sqrt{npq}$ converge en loi vers une variable aléatoire de loi $\mathcal{N}(0, 1)$.*

⁵ Polyà (G.). — Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem, *Math. Z.*, t. 8 (1920), p. 171–181.

Il en résulte que, pour n « grand » la loi binomiale $B(n, p)$ peut être approchée par la loi normale $\mathcal{N}(np, \sqrt{npq})$.

Cas particulier 2. — Prenons pour loi commune des X_n la loi de Poisson π_1 , de paramètre 1. Alors S_n suit la loi de Poisson π_n de paramètre n ; on a $\mathbb{E}[S_n] = n$, $\text{Var } S_n = n$, et le Théorème 2.1 affirme :

PROPOSITION 2.3. — *La suite des variables aléatoires de terme général $Y_n = (S_n - n)/\sqrt{n}$ converge en loi vers une variable aléatoire de loi $\mathcal{N}(0, 1)$.*

Il en résulte que, pour n « grand », la loi de Poisson π_n peut être approchée par la loi normale $\mathcal{N}(n, \sqrt{n})$.

Remarque 1 (Bernstein). — Il résulte de la Proposition 2.3 que lorsque n tend vers l'infini, on a :

$$\mathbb{P}\left\{\frac{S_n - n}{\sqrt{n}} \leq 0\right\} \rightarrow \int_{-\infty}^0 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \frac{1}{2},$$

c'est-à-dire que

$$\mathbb{P}\{S_n \leq n\} = e^{-n} \sum_{k=0}^n \frac{n^k}{k!} \rightarrow \frac{1}{2} \quad (n \rightarrow \infty).$$

(On remarquera que ce dernier passage à la limite n'est pas aisé à établir sans faire appel au théorème « central limit ».)

Remarque 2. — La Proposition 2.3 peut être étendue au cas d'une famille (X_λ) ($\lambda > 0$) de variables aléatoires, où le terme général X_λ suit la loi de Poisson π_λ de paramètre λ . On peut montrer directement (cf. Exercice 1) que, lorsque λ tend vers l'infini, $(X_\lambda - \lambda)/\sqrt{\lambda}$ converge en loi vers une variable aléatoire de loi $\mathcal{N}(0, 1)$, de sorte que, pour λ « grand », la loi de Poisson π_λ peut être approchée par la loi normale $\mathcal{N}(\lambda, \sqrt{\lambda})$.

Cas particulier 3. — Prenons pour loi commune des X_n la loi exponentielle $\mathcal{E}(\lambda)$, de paramètre $\lambda > 0$. Alors S_n suit la loi gamma $\Gamma(n, \lambda)$; on a $\mathbb{E}[S_n] = n/\lambda$, $\text{Var } S_n = n/\lambda^2$, et le Théorème 2.1 affirme :

PROPOSITION 2.4. — *La suite des variables aléatoires de terme général $Y_n = \frac{S_n - n/\lambda}{\sqrt{n}/\lambda}$ converge en loi vers une variable aléatoire de loi $\mathcal{N}(0, 1)$.*

Il en résulte que, pour n « grand », la loi gamma $\Gamma(n, \lambda)$ peut être approchée par la loi normale $\mathcal{N}(n/\lambda, \sqrt{n}/\lambda)$.

Remarque. — La Proposition 2.4 peut être étendue au cas d'une famille (X_p) ($p > 0$) de variables aléatoires, où le terme général X_p suit la loi gamma $\Gamma(p, \lambda)$, avec $\lambda > 0$ fixé. On peut montrer directement (cf. Exercice 2) que, lorsque p tend vers l'infini, λ restant fixé, la variable aléatoire $\frac{X_p - p/\lambda}{\sqrt{p}/\lambda}$ converge en loi vers une variable aléatoire de loi $\mathcal{N}(0, 1)$, de sorte que, pour p « grand », la loi $\Gamma(p, \lambda)$ peut être approchée par la loi normale $\mathcal{N}(p/\lambda, \sqrt{p}/\lambda)$.

3. Le théorème « central limit » et la formule de Stirling. — Soit (X_p) ($p > 0$) une famille de variables aléatoires dont le terme général X_p suit la loi $\Gamma(p+1, 1)$. Lorsque p tend vers l'infini, $(X_p - (p+1))/\sqrt{p+1}$ converge en loi vers une variable aléatoire de loi $\mathcal{N}(0, 1)$; il est clair que $(X_p - p)/\sqrt{p}$ a le même comportement asymptotique lorsque p tend vers l'infini. Nous allons montrer que cette dernière propriété est équivalente à la formule de Stirling.

THÉORÈME 3.1. — Soit (X_p) ($p > 0$) une famille de variables aléatoires dont le terme général X_p suit la loi $\Gamma(p+1, 1)$. Alors chacune des deux propriétés suivantes peut être déduite de l'autre :

- a) $(X_p - p)/\sqrt{p}$ converge en loi vers une variable aléatoire de loi $\mathcal{N}(0, 1)$, lorsque p tend vers l'infini (théorème « central limit ») ;
- b) $\Gamma(p+1) \sim p^p e^{-p} \sqrt{2\pi p}$, lorsque p tend vers l'infini (formule de Stirling).

Démonstration. — Désignons les densités de X_p et de $(X_p - p)/\sqrt{p}$ respectivement par f_p et g_p ; on a :

$$f_p(x) = \frac{1}{\Gamma(p+1)} e^{-x} x^p I_{[0, +\infty[}(x); \quad g_p(x) = \sqrt{p} f_p(p + x\sqrt{p}).$$

Pour tout réel x , on peut choisir $p > 0$ assez grand pour que $p + x\sqrt{p} > 0$; on a alors

$$g_p(x) = \sqrt{p} \frac{1}{\Gamma(p+1)} e^{-(p+x\sqrt{p})} (p+x\sqrt{p})^p = \frac{p^p e^{-p} \sqrt{2\pi p}}{\Gamma(p+1)} \frac{1}{\sqrt{2\pi}} r_p(x),$$

où

$$r_p(x) = e^{-x\sqrt{p}} \left(1 + \frac{x}{\sqrt{p}}\right)^p;$$

de là :

$$(3.1) \quad \int_{-\infty}^x g_p(u) du = \frac{p^p e^{-p} \sqrt{2\pi p}}{\Gamma(p+1)} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x r_p(u) du.$$

Un calcul élémentaire montre que $r_p(u) \rightarrow e^{-u^2/2}$ lorsque p tend vers l'infini; d'où, en appliquant le théorème de convergence dominée, on a pour tout réel x

$$(3.2) \quad \int_{-\infty}^x r_p(u) du \rightarrow \int_{-\infty}^x e^{-u^2/2} du,$$

lorsque p tend vers l'infini. Le théorème résulte alors de (3.1) et (3.2). \square

4. Le théorème de Lindeberg. — Dans le théorème suivant, on suppose donnée une suite (X_k) ($k \geq 1$) de variables aléatoires, de L^2 , centrées et indépendantes. On pose $\text{Var } X_k = \sigma_k^2$ et on désigne par μ_k la loi de probabilité de X_k . Pour tout $n \geq 1$ on note S_n la somme partielle $S_n = \sum_{k=1}^n X_k$ et on pose $\text{Var } S_n = \sum_{k=1}^n \sigma_k^2 = C_n^2$.

THÉORÈME 4.1. — Si la « condition de Lindeberg » suivante est satisfaite, à savoir que pour tout $\varepsilon > 0$, on a

$$\frac{1}{C_n^2} \sum_{k=1}^n \int_{|x| \geq \varepsilon C_n} x^2 d\mu_k(x) \rightarrow 0,$$

lorsque n tend vers l'infini, alors la suite des variables aléatoires $Z_n = S_n/C_n$ converge en loi vers une variable aléatoire de loi $\mathcal{N}(0, 1)$.

Démonstration

a) La condition de Lindeberg implique que $\frac{1}{C_n^2} \sup_{1 \leq k \leq n} \sigma_k^2 \rightarrow 0$ lorsque n tend vers l'infini. En effet,

$$\begin{aligned} \sup_{1 \leq k \leq n} \sigma_k^2 &= \sup_{1 \leq k \leq n} \left(\int_{|x| < \varepsilon C_n} x^2 d\mu_k + \int_{|x| \geq \varepsilon C_n} x^2 d\mu_k \right) \\ &\leq \varepsilon^2 C_n^2 + \sum_{k=1}^n \int_{|x| \geq \varepsilon C_n} x^2 d\mu_k; \end{aligned}$$

d'où :

$$\frac{1}{C_n^2} \sup_{1 \leq k \leq n} \sigma_k^2 \leq \varepsilon^2 + \frac{1}{C_n^2} \sum_{k=1}^n \int_{|x| \geq \varepsilon C_n} x^2 d\mu_k.$$

Le résultat découle de la condition de Lindeberg et du fait que $\varepsilon > 0$ est arbitraire.

b) Considérons la variable aléatoire $Z_n = \frac{S_n}{C_n} = \sum_{k=1}^n \frac{X_k}{C_n}$; sa fonction caractéristique est donnée par $\varphi_{Z_n}(t) = \prod_{k=1}^n \varphi_{X_k}(t/C_n)$. D'après la partie a), pour tout nombre réel t on peut trouver n assez grand pour que, pour tout $k = 1, \dots, n$ on ait $3(t^2/C_n^2)\sigma_k^2 \leq 1$. Il résulte alors du Théorème 4.7, chap. 13, appliqué à chaque X_k , que $\text{Log } \varphi_{Z_n}(t)$ existe et admet la représentation

$$\text{Log } \varphi_{Z_n}(t) = \sum_{k=1}^n \text{Log } \varphi_{X_k}\left(\frac{t}{C_n}\right),$$

avec

$$\begin{aligned} \text{Log } \varphi_{Z_n}(t) &= -\frac{t^2}{2} \frac{\sum_{k=1}^n \sigma_k^2}{C_n^2} - t^2 \left(\frac{1}{C_n^2} \sum_{k=1}^n \int_0^1 (1-u) \mathbb{E} \left[X_k^2 \left(e^{ituX_k/C_n} - 1 \right) \right] du \right) \\ &\quad + 3t^4 \frac{\sum_{k=1}^n \sigma_k^4}{C_n^4} \theta \quad [\text{avec } |\theta| \leq 1] \\ &= -\frac{t^2}{2} - t^2 A_n + 3t^4 B_n. \end{aligned}$$

Pour démontrer le théorème, il suffit de montrer que A_n et B_n tendent vers 0 lorsque n tend vers l'infini.

c) Montrons que $A_n \rightarrow 0$. Posons $g(x, t, u) = x^2(e^{itux/C_n} - 1)$; on a :

$$\mathbb{E}[X_k^2(e^{itux_k/C_n} - 1)] = \int_{|x| < \varepsilon C_n} g d\mu_k + \int_{|x| \geq \varepsilon C_n} g d\mu_k.$$

Si $|x| < \varepsilon C_n$, on a $|g| \leq x^2 |tux|/C_n$, puisque $|e^{i\alpha} - 1| \leq |\alpha|$. De là $|g| \leq x^2 |tu| \varepsilon$. Si $|x| \geq \varepsilon C_n$, on a $|g| \leq 2x^2$, puisque $|e^{i\alpha} - 1| \leq 2$. On en déduit :

$$\begin{aligned} \left| \mathbb{E}[X_k^2(e^{itux_k/C_n} - 1)] \right| &\leq |tu| \varepsilon \sigma_k^2 + 2 \int_{|x| \geq \varepsilon C_n} x^2 d\mu_k; \\ |A_n| &\leq \varepsilon |t| + \frac{1}{C_n^2} \sum_{k=1}^n \int_{|x| \geq \varepsilon C_n} x^2 d\mu_k; \end{aligned}$$

d'où $A_n \rightarrow 0$, en vertu de la condition de Lindeberg et du fait que $\varepsilon > 0$ est arbitraire.

d) Montrons que $B_n \rightarrow 0$. On a

$$|B_n| \leq \frac{\sum_{k=1}^n \sigma_k^4}{C_n^4} \leq \frac{\sup_{1 \leq k \leq n} \sigma_k^2}{C_n^2} \frac{\sum_{k=1}^n \sigma_k^2}{C_n^2} = \frac{\sup_{1 \leq k \leq n} \sigma_k^2}{C_n^2};$$

d'où $B_n \rightarrow 0$, en vertu de la partie a). \square

Remarques. — Nous avons vu dans la partie a) de la démonstration que la condition de Lindeberg implique que $(1/C_n^2) \sup_{1 \leq k \leq n} \sigma_k^2 \rightarrow 0$, lorsque n tend vers l'infini; c'est-à-dire que la plus grande variance de X_1, \dots, X_n est asymptotiquement négligeable devant la variance de la somme $S_n = X_1 + \dots + X_n$. Le rôle de cette dernière condition est mis en relief par le théorème suivant dû à Feller.⁶

THÉORÈME 4.1'. — *Adoptons les hypothèses et les notations du théorème 4.1. Alors les deux propriétés suivantes sont équivalentes :*

- 1) *la condition de Lindeberg;*
- 2) a) *la suite de terme général $Y_n = S_n/C_n$ converge en loi vers une variable aléatoire de loi $\mathcal{N}(0, 1)$;*
 b) $\frac{1}{C_n^2} \sup_{1 \leq k \leq n} \sigma_k^2 \rightarrow 0 \quad (n \rightarrow \infty)$.

⁶ Feller (William). — *An Introduction to Probability Theory and its Applications*, vol. 2. J. Wiley, 1962, p. 491.

5. Un complément au théorème de Lindeberg-Lévy. — Dans ce paragraphe, nous donnons une condition nécessaire et suffisante pour que le théorème « central limit » soit vrai, lorsque les variables aléatoires sont indépendantes et *identiquement distribuées*. L'énoncé en question est le suivant.

THÉORÈME 5.1. — *Soit (X_n) ($n \geq 1$) une suite de variables aléatoires réelles, indépendantes et identiquement distribuées, telles que $X_1 \in L^1$ et $\mathbb{E}[X_1] = 0$. Alors les deux propriétés suivantes sont équivalentes :*

- a) *la suite de terme général $Y_n = \frac{X_1 + \dots + X_n}{\sqrt{n}}$ converge en loi vers une variable aléatoire de loi $\mathcal{N}(0, 1)$;*
 b) *$X_1 \in L^2$ et $\mathbb{E}[X_1^2] = 1$.*

Démonstration (L. Pratelli, communication privée). — Puisque l'implication b) \Rightarrow a) n'est autre que le théorème de Lindeberg-Lévy, il suffit de démontrer a) \Rightarrow b). A cet effet, nous utiliserons les deux lemmes suivants.

LEMME 5.2. — *Soit X une variable aléatoire de loi μ et de fonction caractéristique φ . Alors on a, dans $[0, +\infty]$,*

$$\lim_{u \rightarrow 0} 2 \frac{1 - \Re \varphi(u)}{u^2} = \int_{\mathbb{R}} x^2 d\mu(x).$$

Démonstration. — Pour $u \neq 0$ on a :

$$I(u) = 2 \frac{1 - \Re \varphi(u)}{u^2} = 2 \int_{\mathbb{R}} \frac{1 - \cos ux}{u^2} d\mu(x).$$

Le terme $(1 - \cos ux)/u^2$ est compris entre 0 et $x^2/2$ et converge vers $x^2/2$, lorsque $u \rightarrow 0$. Dans $[0, +\infty]$, on a donc

$$\int_{\mathbb{R}} x^2 d\mu(x) \leq \liminf_{u \rightarrow 0} I(u) \leq \limsup_{u \rightarrow 0} I(u) \leq \int_{\mathbb{R}} x^2 d\mu(x),$$

où la première inégalité résulte du Lemme de Fatou, la seconde étant triviale. Le lemme en découle. \square

LEMME 5.3. — *Soit (X, Y) un couple de variables aléatoires réelles indépendantes, tel que $X + Y$ appartienne à L^2 . Alors X et Y appartiennent à L^2 .*

Démonstration. — Montrons, par exemple, que Y appartient à L^2 . Désignons par μ la loi de X . Alors pour tout x réel on a :

$$\int_{\mathbb{R}} \mathbb{E}[(x + Y)^2] d\mu(x) = \mathbb{E}[(X + Y)^2] < +\infty ;$$

de sorte que, pour μ -presque tout x (donc pour un x au moins!), la somme $x + Y$ appartient à L^2 . \square

L'implication a) \Rightarrow b) du Théorème 5.1 se démontre alors comme suit. Désignons par φ la fonction caractéristique de X_1 et plaçons-nous d'abord dans le cas où la loi μ de X_1 est *symétrique*, ce qui implique que φ est *réelle*.

La condition a) entraîne que l'on a : $\lim_{n \rightarrow \infty} (\varphi(1/\sqrt{n}))^n = e^{-1/2}$. Or pour n assez grand, on a $\varphi(1/\sqrt{n}) > 0$ et on peut passer aux logarithmes. On obtient $n \operatorname{Log} \varphi(1/\sqrt{n}) \sim -1/2$, d'où $\varphi(1/\sqrt{n}) - 1 \sim -1/(2n)$ et donc $2 \frac{1 - \varphi(1/\sqrt{n})}{1/n} \sim 1$. Par le Lemme 5.2 on en déduit

$$\int_{\mathbb{R}} x^2 d\mu(x) = \lim_{n \rightarrow +\infty} 2 \frac{1 - \varphi(1/\sqrt{n})}{1/n} = 1.$$

Passons maintenant au cas général. La condition a) entraîne que la suite (Z_n) ($n \geq 1$) définie par

$$\begin{aligned} Z_n &= \frac{(X_1 - X_2) + (X_3 - X_4) + \cdots + (X_{2n-1} - X_{2n})}{\sqrt{2n}} \\ &= \frac{1}{\sqrt{2}} \left(\frac{X_1 + X_3 + \cdots + X_{2n-1}}{\sqrt{n}} - \frac{X_2 + X_4 + \cdots + X_{2n}}{\sqrt{n}} \right) \end{aligned}$$

converge en loi vers une variable normale de loi $\mathcal{N}(0, 1)$. D'autre part, si l'on pose $Y_n = (X_{2n-1} - X_{2n})/\sqrt{2}$, on peut écrire Z_n sous la forme $Z_n = (Y_1 + \cdots + Y_n)/\sqrt{n}$, où les Y_n sont des variables aléatoires indépendantes, de même loi *symétrique, centrée*. On peut donc leur appliquer le résultat de a). On trouve ainsi $\mathbb{E}[Y_1^2] = 1$ et le Lemme 5.3 permet d'en déduire $\mathbb{E}[X_1^2] = 1$. \square

6. Le théorème de Liapounov. — Le théorème de Lindeberg admet le corollaire suivant, connu sous le nom de théorème de Liapounov.

THÉORÈME 6.1. — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires indépendantes, de $L_{2+\delta}$ pour un certain $\delta > 0$, et centrées. Posons $\operatorname{Var} X_k = \sigma_k^2$. Introduisons la suite des sommes partielles $S_n = \sum_{k=1}^n X_k$ et posons $\operatorname{Var} S_n = \sum_{k=1}^n \sigma_k^2 = C_n^2$. Enfin, supposons vérifiée la « condition de Liapounov » :

$$\frac{1}{(C_n)^{2+\delta}} \sum_{k=1}^n \mathbb{E}[|X_k|^{2+\delta}] \rightarrow 0 \quad (n \rightarrow \infty).$$

Alors la suite de variables aléatoires $(Z_n = S_n/C_n)$ converge en loi vers une variable aléatoire de loi $\mathcal{N}(0, 1)$.

Démonstration. — Il suffira de montrer que la condition de Liapounov implique celle de Lindeberg. Notons μ_k la loi de probabilité de X_k . On a pour tout $\delta > 0$

$$\begin{aligned} \int_{|x| \geq \varepsilon C_n} x^2 d\mu_k &\leq \int_{|x| \geq \varepsilon C_n} x^2 \left(\frac{|x|}{\varepsilon C_n} \right)^\delta d\mu_k \\ &= \frac{1}{(\varepsilon C_n)^\delta} \int_{|x| \geq \varepsilon C_n} |x|^{2+\delta} d\mu_k \leq \frac{1}{(\varepsilon C_n)^\delta} \mathbb{E}[|X_k|^{2+\delta}]; \end{aligned}$$

d'où

$$\frac{1}{C_n^2} \sum_{k=1}^n \int_{|x| \geq \varepsilon C_n} x^2 d\mu_k \leq \frac{1}{\varepsilon^\delta} \frac{1}{(C_n)^{2+\delta}} \sum_{k=1}^n \mathbb{E}[|X_k|^{2+\delta}]. \quad \square$$

Remarque. — Si les X_n sont en outre *identiquement distribuées*, la condition de Liapounov est automatiquement vérifiée et la suite (Z_n) converge en loi vers une variable aléatoire normale de loi $\mathcal{N}(0, 1)$.

APPLICATION 6.2. — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires indépendantes, *identiquement distribuées*, de loi de probabilité commune $\frac{1}{2}(\varepsilon_{-1} + \varepsilon_{+1})$. On lui associe la suite de variables aléatoires indépendantes, de terme général $Y_n = n^\alpha X_n$, où α est une constante positive. Alors la suite des variables aléatoires de terme général $\frac{1}{n^{\alpha+1/2}} \sum_{k=1}^n Y_k$ converge en loi vers une variable aléatoire de loi $\mathcal{N}(0, 1/\sqrt{2\alpha+1})$.

Démonstration

a) Rappelons que la somme $s_\alpha = \sum_{k=1}^n k^\alpha$ est équivalente à $n^{\alpha+1}/(\alpha+1)$, lorsque n tend vers l'infini.

b) Posons $S_n = \sum_{k=1}^n Y_k$; comme Y_n , la variable S_n est *centrée* et l'on a : $\text{Var } S_n = \sum_{k=1}^n \text{Var } Y_k = \sum_{k=1}^n k^{2\alpha}$; d'où, d'après a)

$$\text{Var } S_n \sim \frac{n^{2\alpha+1}}{2\alpha+1} \quad \text{et} \quad \sigma(S_n) \sim \frac{n^{\alpha+1/2}}{\sqrt{2\alpha+1}},$$

lorsque n tend vers l'infini.

c) Pour tout $\delta > 0$ on a $\mathbb{E}[|Y_k|^{2+\delta}] = k^{\alpha(2+\delta)}$; d'où

$$\sum_{k=1}^n \mathbb{E}[|Y_k|^{2+\delta}] = \sum_{k=1}^n k^{\alpha(2+\delta)} \sim \frac{n^{\alpha(2+\delta)+1}}{\alpha(2+\delta)+1}.$$

Posons toujours $C_n = \sigma(S_n)$. Le rapport de Liapounov devient alors, pour tout $\delta > 0$:

$$\begin{aligned} \frac{1}{(C_n)^{2+\delta}} \sum_{k=1}^n \mathbb{E}[|X_k|^{2+\delta}] \\ \sim \frac{(2\alpha+1)^{(2+\delta)/2}}{n^{(\alpha+1/2)(2+\delta)}} \frac{n^{\alpha(2+\delta)+1}}{\alpha(2+\delta)+1} = \frac{(2\alpha+1)^{(2+\delta)/2}}{\alpha(2+\delta)+1} \frac{1}{n^{\delta/2}}, \end{aligned}$$

expression qui tend vers 0 lorsque n tend vers l'infini.

d) Le théorème de Liapounov permet donc d'affirmer que

$$\frac{1}{C_n} \sum_{k=1}^n Y_k \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Or $C_n = \sigma(S_n) \sim n^{\alpha+1/2}/\sqrt{2\alpha+1}$, lorsque n tend vers l'infini. On a donc :

$$\sqrt{2\alpha+1} \frac{1}{n^{\alpha+1/2}} \sum_{k=1}^n Y_k \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1),$$

ou encore :

$$\frac{1}{n^{\alpha+1/2}} \sum_{k=1}^n Y_k \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1/\sqrt{2\alpha+1}). \quad \square$$

Remarque. — Pour $\alpha = 1$, on a $Y_n = nX_n$ et

$$\frac{1}{n^{3/2}} \sum_{k=1}^n Y_k \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1/\sqrt{3}).$$

C'est un résultat que l'on utilise dans la théorie du test de Wilcoxon.⁷

APPLICATION 6.3. — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires indépendantes, la loi de X_n étant $p_n\varepsilon_1 + q_n\varepsilon_0$ ($0 < p_n < 1$, $p_n + q_n = 1$). Supposons que $\sum_{n \geq 1} p_n q_n = +\infty$. Alors la suite des variables aléatoires de terme général $Z_n = \sum_{k=1}^n (X_k - p_k) / \sqrt{\sum_{k=1}^n p_k q_k}$ converge en loi vers une variable aléatoire normale de loi $\mathcal{N}(0, 1)$.

Démonstration. — Comme $\mathbb{E}[X_k] = p_k$ et $\text{Var } X_k = p_k q_k$, en posant $S_n = \sum_{k=1}^n (X_k - p_k)$, on a : $(C_n)^2 = \text{Var } S_n = \sum_{k=1}^n p_k q_k$. D'autre part, $\mathbb{E}[|X_k - p_k|^3] = p_k q_k^3 + q_k p_k^3 = p_k q_k (p_k^2 + q_k^2) \leq p_k q_k$. Enfin, la suite $(X_n - p_n)$ vérifie la condition de Liapounov pour $\delta = 1$, puisque

$$\frac{1}{(C_n)^3} \sum_{k=1}^n \mathbb{E}[|X_k - p_k|^3] \leq \frac{\sum_{k=1}^n p_k q_k}{(\sum_{k=1}^n p_k q_k)^{3/2}} = \frac{1}{(\sum_{k=1}^n p_k q_k)^{1/2}},$$

expression qui tend vers 0 lorsque n tend vers l'infini. Il en résulte que $Z_n = S_n/C_n$ converge en loi vers une variable aléatoire normale de loi $\mathcal{N}(0, 1)$. \square

COMPLÉMENTS ET EXERCICES

1. — Soit (X_λ) ($\lambda > 0$) une famille de variables aléatoires, où le terme général X_λ suit la loi de Poisson π_λ . Montrer que $(X_\lambda - \lambda)/\sqrt{\lambda}$ converge en loi vers une variable aléatoire de loi $\mathcal{N}(0, 1)$, lorsque λ tend vers l'infini.

⁷ Siegel (Sidney). — *Non-Parametric Statistics for the Behavioral Sciences*. — McGraw-Hill, 1956, p. 79.

2. — Soit (X_p) ($p > 0$) une famille de variables aléatoires, où le terme général X_p suit la loi gamma $\Gamma(p, \lambda)$ ($\lambda > 0$). Montrer que $\frac{X_p - (p/\lambda)}{\sqrt{p/\lambda}}$ converge en loi vers une variable aléatoire de loi $\mathcal{N}(0, 1)$, lorsque p tend vers l'infini.

3. — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires indépendantes, la loi de X_n étant donnée par $(1/n)\varepsilon_1 + (1 - 1/n)\varepsilon_0$. On pose $S_n = \sum_{k=1}^n X_k$.

a) Montrer que $\mathbb{E}[S_n] \sim \text{Log } n$, $\text{Var } S_n \sim \text{Log } n$.

b) Montrer que $Y_n = \frac{S_n - \text{Log } n}{\sqrt{\text{Log } n}}$ converge en loi vers une variable aléatoire de loi $\mathcal{N}(0, 1)$.

4. — Soit $0 < p < 1$, $p + q = 1$. Montrer que si $p > q$, alors

$$\sum_{k=[n/2]+1}^n \binom{n}{k} p^k q^{n-k}$$

tend vers 1, lorsque n tend vers l'infini.

5. a) Soient (Y_n) ($n \geq 1$) une suite de variables aléatoires et (N_n) ($n \geq 1$) une suite de variables aléatoires, indépendante de la suite (Y_n) ($n \geq 1$), à valeurs dans \mathbb{N}^* et telles que $N_n \xrightarrow{p} +\infty$ lorsque n tend vers l'infini (c'est-à-dire telle que pour tout $k \geq 1$ on a $\lim_{n \rightarrow +\infty} \mathbb{P}\{N_n \geq k\} = 1$.) Montrer que, si la suite (Y_n) ($n \geq 1$) converge en loi vers une limite Y , alors la suite (Y_{N_n}) converge en loi vers la même limite Y .

b) Soit (X_n) ($n \geq 1$) une suite de variables aléatoires de L^2 , indépendantes, identiquement distribuées, centrées et réduites. On pose $Y_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k$. Soit d'autre part (N_n) ($n \geq 1$) une suite ayant les propriétés de a). Montrer que l'on a la version suivante du Théorème « central limit » $Y_{N_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$ lorsque n tend vers l'infini.

LA LOI DU LOGARITHME ITÉRÉ

Une des préoccupations majeures des probabilistes a été de tout temps l'étude des fluctuations de sommes $S_n = X_1 + \dots + X_n$, convenablement normées, où les termes sont issus d'une suite (X_n) de variables aléatoires *indépendantes, identiquement distribuées*. Les premières études portèrent sur le cas particulier où les X_n sont des variables aléatoires *de Bernoulli centrées*. Dès 1909, elles donnèrent lieu à un premier résultat : la *loi forte des grands nombres* de Borel, qui affirme que $S_n/n \xrightarrow{p.s.} 0$. Ce résultat était pourtant bien maigre comparé au but que les mathématiciens de ce début de siècle s'étaient fixé : il affirme, en effet, que $S_n = o(n)$ p.s., une propriété peu apte à décrire le comportement de la suite (S_n) de façon satisfaisante.

Cependant l'élan était donné; les mathématiciens les plus éminents s'intéressèrent au problème et obtinrent des résultats de plus en plus précis. Citons, en particulier, Hausdorff (1913) qui montra que pour tout $\varepsilon > 0$ on a $S_n = o(n^{(1/2)+\varepsilon})$ p.s., puis Hardy et Littlewood (1914) qui établirent que l'on a en fait $S_n = O(\sqrt{n \log n})$ p.s. Un point culminant fut atteint en 1924 lorsque Khintchine énonça sa désormais célèbre *loi du logarithme itéré*, qui fait l'objet du Théorème 3.3 du présent chapitre. La démonstration qui en est donnée ici reprend les principales étapes du *cheminement historique* qui a conduit à ce théorème.

1. Notations et lemmes préliminaires. — Soit (Y_n) ($n \geq 1$) une suite de variables aléatoires de *Bernoulli, indépendantes, identiquement distribuées*, de loi commune $\frac{1}{2}(\varepsilon_1 + \varepsilon_0)$. Pour tout $n \geq 1$, on désigne par X_n la variable aléatoire *centrée, réduite* $X_n = 2Y_n - 1$. On note $g(u)$ la fonction génératrice des moments de X_1 , à savoir

$$g(u) = g_{X_1}(u) = \mathbb{E}[e^{uX_1}] = \frac{1}{2}(e^u + e^{-u}) = \operatorname{ch} u \quad (u \in \mathbb{R})$$

et pour tout $n \geq 1$ on pose : $S_n = \sum_{k=1}^n X_k$.

LEMME 1.1. — *Pour tout $u \in \mathbb{R}$, on a : $g(u) \leq e^{u^2/2}$.*

Démonstration. — Il suffit de comparer terme à terme les développements en série de puissances de $g(u) = \operatorname{ch} u$ et de $e^{u^2/2}$, qui s'écrivent :

$$g(u) = \operatorname{ch} u = \sum_{k \geq 0} \frac{u^{2k}}{(2k)!} \quad \text{et} \quad e^{u^2/2} = \sum_{k \geq 0} \frac{u^{2k}}{2^k k!}.$$

Le résultat en découle, puisque pour tout $k \geq 0$ on a l'inégalité : $\frac{1}{(2k)!} \leq \frac{1}{2^k k!}$. \square

Remarques. — Posons $S_n^* = S_n/\sqrt{n}$, $g_n(u) = g_{S_n^*}(u) = (g(u/\sqrt{n}))^n$. Il résulte du Lemme 1.1 que l'on a l'inégalité

$$g_n(u) \leq e^{u^2/2},$$

pour tout $u \in \mathbb{R}$. Or, d'après le théorème « central limit »,

$$S_n^* \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad (n \rightarrow \infty).$$

D'où, pour tout $u \in \mathbb{R}$,

$$g_n(u) \rightarrow e^{u^2/2} \quad (n \rightarrow \infty).$$

On voit ainsi que $g_n(u)$ tend vers $e^{u^2/2}$ par valeurs inférieures.

LEMME 1.2. — Pour tout $a > 0$ et pour tout $n \geq 1$, on a :

$$(1.1) \quad \mathbb{P}\{S_n > a\} \leq e^{-a^2/(2n)};$$

$$(1.2) \quad \mathbb{P}\{|S_n| > a\} \leq 2e^{-a^2/(2n)}.$$

Démonstration. — Pour tout $a > 0$ et tout $u > 0$, les deux événements $\{S_n > a\}$ et $\{e^{uS_n} > e^{ua}\}$ sont équivalents. D'après l'inégalité de Markov, on peut écrire :

$$\mathbb{P}\{S_n > a\} \leq \frac{\mathbb{E}[e^{uS_n}]}{e^{ua}} = \frac{g(u)^n}{e^{ua}},$$

d'où, d'après le Lemme 1.1,

$$\mathbb{P}\{S_n > a\} \leq e^{(nu^2/2) - ua}.$$

Cette inégalité est valable pour tout $u > 0$. Choisissons $u > 0$ de telle sorte que le second membre soit minimum, c'est-à-dire considérons la valeur u_0 qui annule la dérivée de l'exposant. On trouve $u_0 = a/n$ et la valeur de l'exposant devient $-a^2/(2n)$. Ceci entraîne l'inégalité (1.1). La variable aléatoire S_n étant symétrique, pour tout $n \geq 1$, on a $\mathbb{P}\{S_n < -a\} = \mathbb{P}\{S_n > a\}$, ce qui entraîne l'inégalité (1.2). \square

LEMME 1.3. — Pour tout $a > 0$, tout $n \geq 1$ et tout $u \geq 0$, on a :

$$(1.3) \quad \mathbb{P}\left\{\sup_{1 \leq k \leq n} S_k \geq a\right\} \leq \frac{\mathbb{E}[e^{uS_n}]}{e^{ua}}.$$

Démonstration. — Introduisons les ensembles suivants :

$$\begin{aligned} A_0 &= \{S_1 < a, \dots, S_n < a\}, & A_1 &= \{S_1 \geq a\}, \\ A_k &= \{S_1 < a, \dots, S_{k-1} < a, S_k \geq a\} \quad (k = 2, \dots, n). \end{aligned}$$

Les ensembles A_0, \dots, A_n sont disjoints, leur réunion est Ω , et l'on a :

$$\sum_{1 \leq k \leq n} A_k = \left\{ \sup_{1 \leq k \leq n} S_k \geq a \right\}.$$

Pour tout $u \geq 0$ on peut donc écrire :

$$(1.4) \quad \mathbb{E}[e^{uS_n}] \geq \sum_{k=1}^n \int_{A_k} e^{uS_n} d\mathbf{P} = \sum_{k=1}^n \int_{\Omega} e^{uS_n} I_{A_k} d\mathbf{P}.$$

Pour tout $k = 1, \dots, n$, faisons la décomposition $S_n = S_k + R_k$, où $R_k = X_{k+1} + \dots + X_n$ (pour $k = n$, on pose $R_n = 0$). Il vient alors, pour tout $k = 1, \dots, n$,

$$\int_{\Omega} e^{uS_n} I_{A_k} d\mathbf{P} = \int_{\Omega} \left(e^{uS_k} I_{A_k} \right) e^{uR_k} d\mathbf{P}.$$

Or les deux variables aléatoires $e^{uS_k} I_{A_k}$ et e^{uR_k} sont *indépendantes* (la première est une fonction de X_1, \dots, X_k , la seconde une fonction de X_{k+1}, \dots, X_n). On a donc

$$\begin{aligned} \int_{\Omega} e^{uS_n} I_{A_k} d\mathbf{P} &= \int_{\Omega} e^{uS_k} I_{A_k} d\mathbf{P} \int_{\Omega} e^{uR_k} d\mathbf{P} \\ &\geq e^{ua} \mathbf{P}(A_k) (g(u))^{n-k}, \end{aligned}$$

d'où, puisque $g(u) \geq 1$,

$$\int_{\Omega} e^{uS_n} I_{A_k} d\mathbf{P} \geq e^{ua} \mathbf{P}(A_k).$$

On a alors, d'après (1.4),

$$\begin{aligned} \mathbb{E}[e^{uS_n}] &\geq e^{ua} \sum_{k=1}^n \mathbf{P}(A_k) = e^{ua} \mathbf{P}\left(\sum_{1 \leq k \leq n} A_k \right) \\ &= e^{ua} \mathbf{P}\left\{ \sup_{1 \leq k \leq n} S_k \geq a \right\}. \quad \square \end{aligned}$$

LEMME 1.4. — *Pour tout $a > 0$ et pour tout $n \geq 1$, on a :*

$$(1.5) \quad \mathbf{P}\left\{ \sup_{1 \leq k \leq n} S_k > a \right\} \leq e^{-a^2/(2n)};$$

$$(1.6) \quad \mathbf{P}\left\{ \sup_{1 \leq k \leq n} |S_k| > a \right\} \leq 2e^{-a^2/(2n)}.$$

Démonstration. — En majorant le second membre de l'inégalité (1.3) du Lemme 1.3 comme dans la démonstration du Lemme 1.2, on obtient l'inégalité (1.5). La variable aléatoire S_n étant symétrique, on a, en outre,

$$\mathbf{P}\left\{ \inf_{1 \leq k \leq n} S_k \leq -a \right\} \leq \mathbf{P}\left\{ \sup_{1 \leq k \leq n} S_k \geq a \right\};$$

d'où l'inégalité (1.6). \square

2. Résultats intermédiaires

THÉORÈME 2.1 (loi forte des grands nombres, E. Borel, 1909). — Lorsque n tend vers l'infini, on a : $S_n/n \xrightarrow{p.s.} 0$, c'est-à-dire $S_n = o(n)$ p.s.

Démonstration. — Dans le Lemme 1.2 (2), faisons $a = n\varepsilon$. Alors pour tout $\varepsilon > 0$, l'inégalité $P\{|S_n/n| > \varepsilon\} \leq 2e^{-(\varepsilon^2/2)n}$ est vérifiée. Comme pour tout $\varepsilon > 0$, le second membre est le terme général d'une série convergente, on a, toujours pour tout $\varepsilon > 0$: $\sum_{n \geq 1} P\{|S_n/n| > \varepsilon\} < \infty$; ceci implique le Théorème 2.1. \square

En 1914, Hardy et Littlewood¹ ont montré que $S_n = O(\sqrt{n \log n})$ p.s., un résultat qui a été affiné en 1922 par Steinhaus.

THÉORÈME 2.2 (Steinhaus, 1922). — Presque sûrement, on a :

$$\limsup_{n \rightarrow \infty} \left| \frac{S_n}{\sqrt{2n \log n}} \right| \leq 1.$$

Ceci équivaut à dire que pour tout $c > 1$, presque sûrement seul un nombre fini parmi les événements $E_n = \{|S_n| > c\sqrt{2n \log n}\}$ peut se réaliser.

Démonstration. — Appliquons le Lemme 1.2 (2) avec $a = c\sqrt{2n \log n}$. On obtient : $P(E_n) \leq 2e^{-c^2 \log n} = 2n^{-c^2}$. Or, pour tout $c > 1$, ce dernier membre est le terme général d'une série convergente. Il en est de même de $P(E_n)$ et le lemme de Borel-Cantelli permet de conclure. \square

3. La loi du logarithme itéré²

THÉORÈME 3.1. — Presque sûrement, on a :

$$\limsup_{n \rightarrow \infty} \left| \frac{S_n}{\sqrt{2n \log \log n}} \right| \leq 1.$$

Ceci équivaut à dire que pour tout $c > 1$, presque sûrement seul un nombre fini parmi les événements $A_n = \{|S_n| > c\sqrt{2n \log \log n}\}$ peut se réaliser.

Démonstration. — Prenons $c > 1$ et un nombre γ tel que $1 < \gamma < c$. Pour $r \geq 1$, désignons par n_r l'entier le plus proche de γ^r et introduisons l'évènement

$$B_r = \left\{ \sup_{n_r < n \leq n_{r+1}} |S_n| > \sqrt{2n_r \log \log n_r} \right\}.$$

Il est clair que si l'évènement B_r ne se réalise que pour un nombre fini d'indices r , alors l'évènement A_n ne se réalise que pour un nombre fini

¹ Hardy (G.H.) and Littlewood (J.E.). — Some problems of Diophantine approximation, *Acta Math.*, vol. **37** (1914), p. 155–239.

² Le présent paragraphe s'inspire de la présentation donnée par Feller *An Introduction to Probability and its Applications*, vol. 1. — Wiley, New York, 1966, p. 192–195.

d'indices n et le théorème est démontré. Il suffit donc, en vertu du lemme de Borel-Cantelli, de montrer que la série $\sum P(B_r)$ est convergente. Or ceci est réalisé. Pour le voir, on applique le Lemme 1.4 (2) pour $a = c\sqrt{2n_r \log \log n_r}$:

$$P(B_r) \leq 2e^{-c^2(n_r/n_{r+1}) \log \log n_r} = 2\left(\frac{1}{\log n_r}\right)^{c^2(n_r/n_{r+1})}.$$

Or $(n_r/n_{r+1}) \sim (1/\gamma) > (1/c)$, d'où, pour r assez grand, $n_r/n_{r+1} > 1/c$. On aura donc, pour r assez grand,

$$P(B_r) \leq 2\left(\frac{1}{\log n_r}\right)^c \sim 2\left(\frac{1}{r \log \gamma}\right)^c.$$

Or, pour tout $c > 1$, la dernière expression est le terme général d'une série convergente. Il en est de même de $P(B_r)$. Le théorème en résulte en vertu du lemme de Borel-Cantelli. \square

THÉORÈME 3.2. — *Pour tout c tel que $0 < c < 1$, l'évènement $A_n = \{S_n > c\sqrt{2n \log \log n}\}$ a lieu pour une infinité d'indices.*

Démonstration. — Prenons $0 < c < 1$; notons, de plus, γ un entier et η un réel tels que $\gamma \geq 2$ et $0 < c < \eta < (\gamma - 1)/\gamma < 1$. Posons, enfin, $n_r = \gamma^r$ ($r \geq 1$).

a) Si l'évènement A_{n_r} a lieu pour une infinité d'indices r , alors l'évènement A_n a lieu pour une infinité d'indices n .

b) Posons $D_r = S_{n_r} - S_{n_{r-1}} = \sum_{k=n_{r-1}+1}^{n_r} X_k$. Alors pour tout $r \geq 1$,

les variables D_r et $S_{n_{r-1}}$ sont *indépendantes*; de même, les D_r ($r \geq 1$) sont *indépendantes* entre elles. On en déduit, en posant

$$B_r = \{D_r > \eta\sqrt{2n_r \log \log n_r}\}, \quad C_r = \{S_{n_{r-1}} > -(\eta - c)\sqrt{2n_r \log \log n_r}\},$$

l'inclusion $B_r \cap C_r \subset A_{n_r}$.

c) Nous allons voir que, pour un choix judicieux de η , l'évènement C_r a lieu presque sûrement pour tout indice r , sauf peut-être pour un nombre *fini* d'entre eux. En effet, d'après le Théorème 3.1, l'évènement $E_r = \{|S_{n_{r-1}}| < 2\sqrt{2n_{r-1} \log \log n_{r-1}}\}$ a lieu presque sûrement pour tout indice r , sauf peut-être pour un nombre *fini* d'entre eux.

Choisissons à présent η suffisamment proche de 1 pour que $1 - \eta < ((\eta - c)/r)^2$. Alors

$$4n_{r-1} = 4\frac{n_r}{\gamma} < 4n_r(1 - \eta) < n_r(\eta - c)^2,$$

et l'on a :

$$\begin{aligned} E_r &= \{|S_{n_{r-1}}| < 2\sqrt{2n_{r-1} \log \log n_{r-1}}\} \\ &\subset \{|S_{n_{r-1}}| < (\eta - c)\sqrt{2n_r \log \log n_r}\} \\ &\subset \{S_{n_{r-1}} > -(\eta - c)\sqrt{2n_r \log \log n_r}\} = C_r. \end{aligned}$$

L'inclusion $E_r \subset C_r$ fournit le résultat

d) Nous allons montrer que $\sum P(B_r) = +\infty$. Puisque les B_r ($r \geq 1$) sont indépendants, le lemme de Borel-Cantelli entraîne que l'évènement B_r a lieu, presque sûrement, pour une infinité d'indices r . En effet, D_r est une variable aléatoire de variance $n_r - n_{r-1}$. La variable réduite étant $D_r^* = D_r / \sqrt{n_r - n_{r-1}}$, on peut récrire B_r sous la forme

$$B_r = \left\{ D_r^* > \eta \sqrt{2 \frac{n_r}{n_r - n_{r-1}} \log \log n_r} \right\}.$$

Or $n_r / (n_r - n_{r-1}) = \gamma / (\gamma - 1) < 1/\eta$. D'où

$$B_r \supset \{ D_r^* > \sqrt{\eta} \sqrt{2 \log \log n_r} \} = \{ D_r^* > \sqrt{\eta} \sqrt{2 \log(r \log \gamma)} \},$$

pour $0 < \eta < 1$. Or $D_r^* \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$, lorsque r tend vers l'infini. On en déduit que la série de terme général $P\{D_r^* > \sqrt{\eta} \sqrt{2 \log(r \log \gamma)}\}$ est divergente. Un argument plus complet est donné dans les compléments à la page 258.

e) Il résulte de c), d) et de l'inclusion $B_r \cup C_r \subset A_{n_r}$ que l'évènement A_{n_r} a lieu presque sûrement pour une infinité d'indices r . \square

THÉORÈME 3.2'. — *Pour tout c tel que $0 < c < 1$, l'évènement $A'_n = \{S_n < -c\sqrt{2n \log \log n}\}$ a lieu pour une infinité d'indices.*

Démonstration. — Elle découle du Théorème 3.2, puisque les variables aléatoires S_n sont symétriques. \square

Les Théorèmes 3.1, 3.2, 3.2' se résument finalement en l'énoncé suivant.

THÉORÈME 3.3 (Loi du logarithme itéré, Khintchine, 1924). — *Presque sûrement, on a :*

$$\limsup_{n \rightarrow \infty} \frac{S_n}{\sqrt{2n \log \log n}} = 1 \quad \text{et} \quad \liminf_{n \rightarrow \infty} \frac{S_n}{\sqrt{2n \log \log n}} = -1.$$

En d'autres termes, pour tout $\varepsilon > 0$, presque sûrement, la suite de terme général S_n dépasse la valeur $(1 + \varepsilon)\sqrt{2n \log \log n}$ au plus *un nombre fini* de fois; en revanche, presque sûrement, elle dépasse *une infinité de fois* la valeur $(1 - \varepsilon)\sqrt{2n \log \log n}$. De même, presque sûrement, elle est inférieure à la valeur $-(1 + \varepsilon)\sqrt{2n \log \log n}$ au plus *un nombre fini* de fois; en revanche, presque sûrement, elle est inférieure *une infinité de fois* à la valeur $-(1 - \varepsilon)\sqrt{2n \log \log n}$.

COROLLAIRE. — *Avec une probabilité égale à 1, la suite (S_n) prend une infinité de fois toute valeur entière.*

Les recherches ultérieures essayèrent de s'affranchir de l'hypothèse « classique » que les variables aléatoires Y_n suivaient la loi de Bernoulli de paramètre $\frac{1}{2}$. Parmi les nombreux résultats obtenus dans cette direction, citons le résultat suivant.

THÉORÈME 3.4 (Hartman-Wintner,³ 1941). — Soit (X_n) ($n \geq 1$) une suite de variables aléatoires indépendantes, identiquement distribuées, centrées, de L^2 , d'écart-type commun $\sigma > 0$. Posons $S_n = \sum_{k=1}^n X_k$ ($n \geq 1$). Alors, presque sûrement, on a :

$$\limsup_{n \rightarrow \infty} \frac{S_n}{\sqrt{2n \log \log n}} = \sigma \quad \text{et} \quad \liminf_{n \rightarrow \infty} \frac{S_n}{\sqrt{2n \log \log n}} = -\sigma.$$

COMPLÉMENTS ET EXERCICES

1. — Soit (Y_n) ($n \geq 1$) une suite de variables aléatoires indépendantes, identiquement distribuées, de loi commune $p\varepsilon_1 + q\varepsilon_0$ ($0 < p < 1$, $p + q = 1$).

On pose : $X_n = (Y_n - p)/\sqrt{pq}$ et $S_n = \sum_{k=1}^n X_k$ ($n \geq 1$).

a) Évaluer la fonction génératrice des moments $g(u)$ de X_1 .

b) Pour tout $u \in \mathbb{R}$, on a : $g(u) \geq 1$. (Le résultat est banal lorsque $p = q = \frac{1}{2}$.)

Solution. — On a pour tout u réel $g(u) = g_{X_1}(u) = \mathbb{E}[e^{uX_1}] = p \exp\left(u \frac{q}{\sqrt{pq}}\right) + q \exp\left(-u \frac{p}{\sqrt{pq}}\right)$. Posons $a = \exp\left(u \frac{q}{\sqrt{pq}}\right)$, $b = \exp\left(-u \frac{p}{\sqrt{pq}}\right)$.

Alors $g(u) = pa + qb$ n'est autre que la moyenne arithmétique de a et b , tandis que $a^p b^q$ (qui vaut 1) est la moyenne géométrique de a et b . D'après l'inégalité classique sur les moyennes, on en déduit : $g(u) \geq 1$.

2. — On conserve les notations de l'exercice précédent. L'inégalité $g(u) \leq e^{u^2/2}$ est satisfaite pour tout $u \geq 0$, lorsque $p \geq q$ et pour tout $u \leq 0$, lorsque $p \leq q$. (Lorsque $p = q = \frac{1}{2}$, cette inégalité fait l'objet du Lemme 1.1.).

Solution. — Faisons la démonstration pour $p \geq q$, $u \geq 0$. Introduisons la fonction $f(u) = (u^2/2) - \text{Log } g(u)$, de sorte que

$$f'(u) = u - \frac{g'(u)}{g(u)}, \quad f''(u) = 1 + \frac{g'^2(u) - g(u)g''(u)}{g^2(u)}.$$

Comme

$$g'(u) = \sqrt{pq} \left(\exp\left(u \frac{q}{\sqrt{pq}}\right) - \exp\left(-u \frac{p}{\sqrt{pq}}\right) \right)$$

et

$$g''(u) = q \exp\left(u \frac{q}{\sqrt{pq}}\right) + p \exp\left(-u \frac{p}{\sqrt{pq}}\right),$$

on en tire

³ Hartmann (Ph.) and Wintner (A.). — On the law of the iterated logarithm, *Amer. J. Math.*, vol. **63**, p. 169-176.

$$g'^2(u) - g(u)g''(u) = -\exp\left(-\frac{u}{\sqrt{pq}}(p-q)\right).$$

Donc $f''(u) = 1 - \frac{1}{g^2(u)} \exp\left(-\frac{u}{\sqrt{pq}}(p-q)\right)$. D'après l'exercice 1 et les hypothèses $p - q \geq 0$, $u \geq 0$, on en déduit :

$$f''(u) \geq 1 - \exp\left(-\frac{u}{\sqrt{pq}}(p-q)\right) \geq 0.$$

La fonction $u \mapsto f(u)$ ($u \geq 0$) est donc *convexe*; de plus $f(0) = 0$ et $f'(0) = 0$. On en déduit $f(u) \geq 0$ pour tout $u \geq 0$.

3. *Divergence de la série de terme général* $\mathbb{P}\{D_r^* > \sqrt{\eta}\sqrt{2\log(r\log\gamma)}\}$ (cf. Théorème 3.2). — Désignons par F_r^* la fonction de répartition de D_r^* et par Φ celle de $\mathcal{N}(0, 1)$. Comme D_r a même loi que $S_{n_r - n_{r-1}}$, la majoration de Berry-Esséen⁴ fournit $\sup_{-\infty < x < +\infty} |F_r^*(x) - \Phi(x)| \leq A(n_r - n_{r-1})^{-1/2}$, où A est une constante positive. Posons ensuite

$$(1) \quad x_r = \sqrt{\eta}\sqrt{2\log(r\log\gamma)} = \sqrt{\eta}\sqrt{2\log\log n_r}.$$

Il vient :

$$(2) \quad \begin{aligned} \mathbb{P}\{D_r^* > x_r\} &= 1 - F_r^*(x_r) \geq 1 - \Phi(x_r) - |F_r^*(x_r) - \Phi(x_r)| \\ &\geq 1 - \Phi(x_r) - A(n_r - n_{r-1})^{-1/2}. \end{aligned}$$

Or, lorsque $r \rightarrow \infty$, on a : $1 - \Phi(x_r) \sim (2\pi)^{-1/2}x_r^{-1}\exp(-x_r^2/2)$; d'où, en remplaçant x_r par sa valeur tirée de (1)

$$(3) \quad 1 - \Phi(x_r) \sim k \frac{1}{\sqrt{\log\log n_r}} \frac{1}{(\log n_r)^\eta},$$

où k est une constante. Or,

$$(4) \quad A(n_r - n_{r-1})^{-1/2} = A\sqrt{\gamma/(\gamma-1)}(1/\sqrt{n_r}).$$

Il résulte alors de (3) et (4) que le terme $A(n_r - n_{r-1})^{-1/2}$ est négligeable devant $1 - \Phi(x_r)$. Le dernier membre de (2) est donc équivalent à $1 - \Phi(x_r)$.

Or, d'après (3), l'expression $1 - \Phi(x_r)$ est le terme général d'une série *divergente*. Il en est donc de même du premier membre de (2), à savoir $\mathbb{P}\{D_r^* > x_r\}$.

⁴ Feller (William). — *An Introduction to Probability Theory and its Applications*, vol. 2. J. Wiley, 1962, p. 515.

APPLICATIONS DES PROBABILITÉS : PROBLÈMES RÉSOLUS

Dans ce dernier chapitre on trouve les énoncés et les solutions complètes de plusieurs problèmes du calcul des probabilités dont la solution fait appel aux différentes techniques et méthodes présentées dans le livre. Ces problèmes fournissent une *ouverture* vers d'autres branches des mathématiques. Ils sont de nature très différente. On y trouve : le problème des rencontres revisité, un problème de temps d'atteinte, un modèle pour l'acheminement du courrier par voie hiérarchique, une ouverture vers les fractions continues, une application de la formule de Bernstein, un regard vers le modèle de diffusion d'Ehrenfest et un problème de probabilité géométrique.

1. Le problème des rencontres revisité. — Dans l'exercice 2 du chapitre 4 nous avons étudié le problème classique des rencontres. Nous nous proposons ici de prolonger cette étude en étudiant les propriétés de la variable aléatoire « nombre de rencontres », en particulier son comportement asymptotique. Comme la formule de Poincaré (*cf.* chap. 3, § 3) et ses extensions (*cf.* chap. 3, exercice 10) jouent un rôle initial essentiel dans cette étude, nous nous proposons de redémontrer ces extensions par une nouvelle méthode reposant sur l'algèbre des indicateurs et un argument de comptage.

1.1. *Les extensions de la formule de Poincaré.* — Soient $n \geq 1$ et n événements E_1, \dots, E_n . On désigne par A_r ($0 \leq r \leq n$) l'évènement « sur ces n événements exactement r se produisent ». Dans l'exercice 10 du chapitre 3 on a donné une formule pour la probabilité $P(A_r)$. Retrouvons cette formule en établissant l'identité suivante sur les indicateurs

$$(1.1.1) \quad I_{A_r} = \sum_{k=r}^n (-1)^{k-r} \binom{k}{r} \sum_{1 \leq i_1 < \dots < i_k \leq n} I_{E_{i_1} \dots E_{i_k}},$$

d'où l'on déduit la formule cherchée

$$(1.1.2) \quad P(A_r) = \sum_{k=r}^n (-1)^{k-r} \binom{k}{r} \sum_{1 \leq i_1 < \dots < i_k \leq n} P(E_{i_1} \dots E_{i_k}),$$

en prenant l'espérance mathématique des deux membres de (1.1.1).

Démonstration. — Soit ω un élément de l'ensemble fondamental Ω sous-jacent. Il existe un sous-ensemble unique L de $[n]$, de cardinal l ($0 \leq l \leq n$),

tel que $\omega \in \prod_{i \in L} E_i \times \prod_{i \in L^c} E_i^c$. Si $l < r$ on a évidemment $I_{A_r}(\omega) = 0$ et également $I_{E_{i_1} \dots E_{i_k}}(\omega) = 0$ pour toute suite $1 \leq i_1 < \dots < i_k \leq n$ telle que $k \geq r$. Les deux membres de l'identité sont donc nuls.

Si $l \geq r$, on a $I_{A_r}(\omega) = 1$ ou 0 suivant que $l = r$ ou $l > r$. Par ailleurs, $I_{E_{i_1} \dots E_{i_k}}(\omega) = 1$ si et seulement si $k \leq l$ et $\{i_1, \dots, i_k\} \subset L$. Pour k fixé tel que $r \leq k \leq l$ il y a donc $\binom{l}{k}$ suites $(i_1 < \dots < i_k)$ telles que $I_{E_{i_1} \dots E_{i_k}}(\omega) = 1$. Le second membre de (1.1.1) appliqué à cet élément ω s'écrit :

$$\sum_{k=r}^l (-1)^{k-r} \binom{k}{r} \binom{l}{k} = \binom{l}{r} \sum_{k=r}^l (-1)^{k-r} \binom{l-r}{k-r} = \binom{l}{r} \sum_{j=0}^{l-r} (-1)^j \binom{l-r}{j},$$

une expression qui vaut 1 ou 0 suivant que $l = r$ ou $l > r$. \square

Utilisant la même technique retrouvons la formule pour la probabilité de l'évènement B_n «il se produit un nombre impair d'évènements parmi E_1, \dots, E_n », en établissant la formule

$$(1.1.3) \quad I_{B_n} = \sum_{k=1}^n (-1)^{k-1} 2^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} I_{E_{i_1} \dots E_{i_k}},$$

d'où l'on déduit :

$$(1.1.4) \quad P(B_n) = \sum_{k=1}^n (-1)^{k-1} 2^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P(E_{i_1} \dots E_{i_k}).$$

Reprenons les mêmes notations que ci-dessus pour le sous-ensemble L de cardinal l correspondant à l'épreuve ω . On a $I_{B_n}(\omega) = 1$ si et seulement si l est impair. Par ailleurs, le membre de gauche de (1.1.3) appliqué à l'épreuve ω vaut

$$\sum_{k=1}^l (-1)^{k-1} 2^{k-1} \binom{l}{k} = -\frac{1}{2} \sum_{k=1}^l (-2)^k \binom{l}{k} = -\frac{1}{2} [(-1)^l - 1],$$

qui est bien égal à 1 ou à 0 suivant que l est impair ou pair.

1.2. Le nombre de rencontres. — Rappelons les énoncés de base sur le problème des rencontres (cf. chap. 4, exercice 2). Une urne contient n boules numérotées de 1 à n ($n \geq 1$). On les extrait successivement *sans remise* et après chaque tirage on note le numéro de la boule tirée. On dit qu'il y a *rencontre au rang i* ($1 \leq i \leq n$) si la boule tirée au $i^{\text{ième}}$ tirage porte le numéro i .

Désignons par E_i l'évènement «il y a rencontre au rang i ». On a : $P(E_i) = (n-1)!/n!$ ($1 \leq i \leq n$), puis $P(E_i E_j) = (n-2)!/n!$ ($1 \leq i < j \leq n$), \dots , enfin $P(E_1 \dots E_n) = 1/n!$

Nous nous proposons d'étudier la variable aléatoire $X_n = \sum_{k=1}^n I_{E_k}$ qui représente le *nombre total de rencontres* dans ces n tirages. Elle prend ses valeurs dans $\{0, 1, \dots, n\}$. Commençons par énoncer quelques propriétés qui peuvent s'établir sans connaître la loi de probabilité de X_n .

PROPOSITION 1.2.1

- 1) Pour tout $n \geq 1$, $\mathbb{E}[X_n] = 1$.
- 2) Pour tout $n \geq 2$, $\text{Var } X_n = 1$.
- 3) Pour tout $x > 1$ et tout $n \geq 2$ on a la majoration :

$$\mathbb{P}\{X_n > x\} \leq \frac{1}{(x-1)^2}.$$

Démonstration. — Pour tout $k = 1, \dots, n$, on a $\mathbb{E}[I_{E_k}^2] = \mathbb{E}[I_{E_k}] = \mathbb{P}(E_k) = 1/n$ et aussi $\mathbb{E}[I_{E_i} I_{E_j}] = \mathbb{P}(E_i E_j) = 1/(n(n-1))$ pour $i \neq j$. D'où

$$\mathbb{E}[X_n] = \sum_{k=1}^n \mathbb{E}[I_{E_k}] = n \frac{1}{n} = 1;$$

$$\mathbb{E}[X_n^2] = \sum_{k=1}^n \mathbb{E}[I_{E_k}^2] + 2 \sum_{1 \leq i < j \leq n} \mathbb{E}[I_{E_i} I_{E_j}] = n \frac{1}{n} + 2 \binom{n}{2} \frac{1}{n(n-1)} = 2;$$

$$\text{Var } X_n = \mathbb{E}[X_n^2] - (\mathbb{E}[X_n])^2 = 1.$$

Enfin, l'inégalité de Bienaymé-Tchebychev fournit la majoration :

$$\mathbb{P}\{X_n > x\} = \mathbb{P}\{X_n - \mathbb{E}[X_n] > x - 1\} \leq \frac{\text{Var } X_n}{(x-1)^2} = \frac{1}{(x-1)^2}. \quad \square$$

Pour déterminer la loi de probabilité de X_n on utilise l'extension (1.1.2) de la formule de Poincaré.

THÉORÈME 1.2.2. — La loi de probabilité de X_n est donnée par :

$$(1.2.1) \quad \mathbb{P}\{X_n = r\} = \begin{cases} \frac{1}{r!} \sum_{i=0}^{n-r} \frac{(-1)^i}{i!}, & \text{si } r \in \{0, 1, \dots, n\}; \\ 0, & \text{sinon.} \end{cases}$$

Démonstration. — Utilisons la formule (1.1.2) en prenant pour E_i l'évènement «rencontre au rang i » ($i = 1, \dots, n$). Alors A_r est l'évènement $\{X_n = r\}$. D'autre part, pour $1 \leq i_1 < \dots < i_k \leq n$ on a $\mathbb{P}(E_{i_1} \dots E_{i_k}) = (n-k)!/n!$. D'où, pour $r = 0, 1, \dots, n$

$$\begin{aligned} \mathbb{P}\{X_n = r\} &= \sum_{k=r}^n (-1)^{k-r} \binom{k}{r} \binom{n}{k} \frac{(n-k)!}{n!} \\ &= \sum_{k=r}^n (-1)^{k-r} \frac{k!}{r!(k-r)!} \frac{n!}{k!(n-k)!} \frac{(n-k)!}{n!} \\ &= \frac{1}{r!} \sum_{k=r}^n \frac{(-1)^{k-r}}{(k-r)!} = \frac{1}{r!} \sum_{i=0}^{n-r} \frac{(-1)^i}{i!}. \quad \square \end{aligned}$$

Remarque 1. — Une autre manière d'établir la formule (1.2.1) est de faire seulement appel à la formule de Poincaré. D'après cette formule, la probabilité qu'il n'y ait aucune rencontre vaut $p_n = \sum_{i=0}^n (-1)^i / i!$ (cf. chap. 4, exercice 2, b)). Une suite de n tirages peut être identifiée à une permutation de $(1, 2, \dots, n)$. Le nombre $n! \times p_n$ est le nombre de permutations présentant 0 rencontre; ce sont les permutations qu'on appelle aussi les *dérangements* de l'ensemble $[n]$. Une permutation présentant r rencontres est alors caractérisée par la donnée d'un couple (J, d) , où J est un sous-ensemble de $[n]$ de cardinal r et où d est un dérangement de $[n] \setminus J$. Leur nombre est donc $\binom{n}{r} \times (n-r)! p_{n-r}$. La probabilité d'avoir exactement r rencontres est ainsi égale à $\binom{n}{r} \times (n-r)! p_{n-r} / n! = (1/r!) \times (\sum_{i=0}^{n-r} (-1)^i / i!)$, qui est la formule (1.2.1).

Remarque 2. — Le théorème 1.2.2 fournit l'identité

$$\sum_{r=0}^n \frac{1}{r!} \sum_{i=0}^{n-r} \frac{(-1)^i}{i!} = 1.$$

Cas particuliers. — Il est intéressant de noter les formules :

$$P\{X_n = 0\} = \sum_{i=0}^n \frac{(-1)^i}{i!} \quad \text{et} \quad P\{X_n = 1\} = \sum_{i=0}^{n-1} \frac{(-1)^i}{i!},$$

expressions qui tendent toutes deux vers e^{-1} lorsque n tend vers l'infini. On a encore $P\{X_n = n-1\} = 0$ (s'il y a au moins $(n-1)$ rencontres, il y en a nécessairement n) et évidemment $P\{X_n = n\} = 1/n!$

THÉORÈME 1.2.3. — *La fonction génératrice de X_n est donnée par :*

$$G(s) = \mathbb{E}[s^{X_n}] = \sum_{k=0}^n \frac{(s-1)^k}{k!}.$$

Démonstration. — Par définition

$$G(s) = \sum_{r=0}^n s^r P\{X_n = r\} = \sum_{r=0}^n \sum_{i=0}^{n-r} \frac{s^r}{r!} \frac{(-1)^i}{i!},$$

d'où, en faisant le changement d'indice $i = k - r$,

$$G(s) = \sum_{r=0}^n \sum_{k=r}^n \frac{s^r}{r!} \frac{(-1)^{k-r}}{(k-r)!} = \sum_{r=0}^n \sum_{k=r}^n \frac{1}{k!} \binom{k}{r} s^r (-1)^{k-r},$$

puis en échangeant l'ordre des sommations,

$$G(s) = \sum_{k=0}^n \frac{1}{k!} \sum_{r=0}^k \binom{k}{r} s^r (-1)^{k-r} = \sum_{k=0}^n \frac{(s-1)^k}{k!}. \quad \square$$

COROLLAIRE. — *Tous les moments factoriels de X_n sont égaux à 1 jusqu'à l'ordre n (inclus), à 0 après l'ordre n .*

Démonstration. — Faisons le changement de variable $s = 1 + u$ dans la fonction génératrice. On obtient :

$$G(1 + u) = \sum_{k=0}^n \frac{u^k}{k!} = 1 + \sum_{k=1}^n 1 \cdot \frac{u^k}{k!},$$

d'où le résultat en vertu de la Proposition 2.5 du chapitre 9. \square

Cas particuliers. — Pour $n \geq 2$ on retrouve les résultats de la Proposition 1.2.1 : $\mathbb{E}[X_n] = 1$ et $\text{Var } X_n = \mathbb{E}[X_n^2] - (\mathbb{E}[X_n])^2 = \mathbb{E}[X_n^2] - \mathbb{E}[X_n] = \mathbb{E}[X_n(X_n - 1)] = 1$.

THÉORÈME 1.2.4. — *La suite (X_n) ($n \geq 1$) tend en loi vers une variable aléatoire de Poisson de paramètre 1.*

Démonstration. — Fixons $r \geq 0$ et prenons $n > r$. Alors $P\{X_n = r\} = (1/r!) \sum_{i=0}^{n-r} (-1)^i / i!$, une quantité qui tend vers $e^{-1}/r!$ lorsque n tend vers l'infini.

Une autre démonstration consiste à faire appel au Théorème 4.1 du chapitre 9 en considérant la fonction génératrice $G(s) = \sum_{k=0}^n (s-1)^k / k!$ et en constatant que, lorsque n tend vers l'infini, $G(s)$ tend vers e^{s-1} , qui est la fonction génératrice de la loi de Poisson $\mathcal{P}(1)$. \square

Il est bon de noter, en parallèle avec le Théorème 1.2.4, que tous les moments factoriels de $\mathcal{P}(1)$ sont égaux à 1.

On peut estimer la vitesse de convergence de la loi $\mathcal{L}(X_n)$ de X_n vers $\mathcal{P}(1)$.

THÉORÈME 1.2.5. — *Soit Y une variable aléatoire de loi $\mathcal{P}(1)$. Introduisons la distance en variation d définie par*

$$d(X_n, Y) = \frac{1}{2} \sum_{k \geq 0} |P\{X_n = k\} - P\{Y = k\}|.$$

Alors

$$d(X_n, Y) \leq \frac{2^n}{(n+1)!} + \frac{1}{2} \frac{1}{(n+1)!} \sim \frac{2^n}{(n+1)!}.$$

Démonstration. — Ecrivons :

$$2d(X_n, Y) = \sum_{k=0}^n \left| \frac{1}{k!} \left(\sum_{i=0}^{n-k} \frac{(-1)^i}{i!} - e^{-1} \right) \right| + e^{-1} \sum_{k \geq n+1} \frac{1}{k!} = A + B.$$

D'abord la série $\sum_{i \geq 0} (-1)^i / i!$ (de somme e^{-1}) est une série alternée. La valeur absolue de son terme général tend vers 0 en décroissant. En remplaçant sa

somme par une somme partielle, on commet une erreur dont la valeur absolue est majorée par la valeur absolue du premier terme négligé. On a donc :

$$\begin{aligned} A &\leq \sum_{k=0}^n \frac{1}{k!} \frac{1}{(n+1-k)!} = \frac{1}{(n+1)!} \sum_{k=0}^n \binom{n+1}{k} \\ &\leq \frac{1}{(n+1)!} \sum_{k=0}^{n+1} \binom{n+1}{k} = \frac{2^{n+1}}{(n+1)!}. \end{aligned}$$

On a d'autre part

$$B = e^{-1} \sum_{k \geq n+1} \frac{1}{k!} = \frac{\Gamma(n+1, 1)}{\Gamma(n+1)},$$

où

$$\Gamma(n+1, 1) = \int_0^1 t^n e^{-t} dt \leq \int_0^1 t^n dt = \frac{1}{n+1}.$$

Il en résulte

$$B \leq \frac{1}{(n+1)\Gamma(n+1)} = \frac{1}{(n+1)!}. \quad \square$$

Terminons ce paragraphe par une étude asymptotique de la probabilité qu'il y ait un nombre *impair* de rencontres. Comme on va le voir, la valeur limite de cette probabilité est strictement inférieure à $\frac{1}{2}$.

En particulierisant la formule (1.1.4) au problème des rencontres, B_n est l'évènement «il y a un nombre impair de rencontres»; d'autre part $P(E_{i_1} \cdots E_{i_k}) = (n-k)!/k!$. On en tire :

$$\begin{aligned} P(B_n) &= \sum_{k=1}^n (-2)^{k-1} \binom{n}{k} \frac{(n-k)!}{n!} = \sum_{k=1}^n (-2)^{k-1} \frac{n!}{k!(n-k)!} \frac{(n-k)!}{n!} \\ &= -\frac{1}{2} \sum_{k=1}^n \frac{(-2)^k}{k!} = -\frac{1}{2} \left(\sum_{k=0}^n \frac{(-2)^k}{k!} - 1 \right) = \frac{1}{2} - \frac{1}{2} \sum_{k=0}^n \frac{(-2)^k}{k!}, \end{aligned}$$

une quantité qui tend vers $\frac{1}{2} - \frac{1}{2}e^{-2} \approx 0,43233$ lorsque n tend vers l'infini.

2. Un problème de temps d'atteinte. — Nous nous proposons d'étudier, dans un cas très particulier, le problème du temps d'atteinte d'un processus stochastique, c'est-à-dire d'une famille (S_t) ($t \in T$) de variables aléatoires réelles, dans un sous-ensemble fixé D de \mathbb{R} . Il s'agit de donner des informations sur le *premier* instant t pour lequel S_t appartient à D .

Nous prenons ici pour ensemble T l'ensemble \mathbb{N} des entiers positifs et pour processus stochastique la famille (S_n) ($n \in \mathbb{N}$) des variables aléatoires définie par $S_0 = 0$ et $S_n = X_1 + \cdots + X_n$ ($n \geq 1$), où (X_n) ($n \geq 1$) est une suite de variables aléatoires, *indépendantes, uniformément distribuées sur l'intervalle* $[0, 1]$. Enfin, on prend pour D la demi-droite ouverte $D =]1, +\infty[$.

Nous nous proposons donc d'étudier la loi de probabilité du premier instant N où la somme $S_n = X_1 + \dots + X_n$ atteindra D , autrement dit d'étudier la variable aléatoire N définie par

$$(2.1) \quad N = \inf\{n \geq 1 : S_n > 1\}.$$

Il est clair que N prend ses valeurs dans $\{2, 3, \dots\}$. D'autre part, par définition même, on a les propriétés

$$(2.2) \quad S_{N-1} \leq 1, \quad S_N > 1,$$

ainsi que l'identité ensembliste

$$(2.3) \quad \{N > n\} = \{S_n \leq 1\} \quad (n \geq 0).$$

Notons que la dernière identité est bien valable pour $n = 0$ puisque l'on a posé : $S_0 = 0$.

LEMME 2.1. — Posons $F_n(t) = P\{S_n \leq t\}$ ($t \geq 0$). Alors, pour tout $t \in [0, 1]$, on a : $F_n(t) = t^n/n!$

Démonstration. — Pour $n = 1$ la propriété est vraie. Elle découle de l'expression-même de la fonction de répartition de X_1 . Procédons par récurrence sur n en supposant que l'on a $F_n(t) = t^n/n!$ pour $t \in [0, 1]$. Pour tout $t \in [0, 1]$ il vient alors

$$\begin{aligned} F_{n+1}(t) &= P\{S_{n+1} \leq t\} = P\{S_n + X_{n+1} \leq t\} \\ &= \int_0^t P\{S_n + X_{n+1} \leq t \mid X_{n+1} = x\} dx \\ &= \int_0^t P\{S_n \leq t - x \mid X_{n+1} = x\} dx, \end{aligned}$$

d'où, puisque S_n et X_{n+1} sont indépendantes

$$F_{n+1}(t) = \int_0^t P\{S_n \leq t - x\} dx,$$

soit, en vertu de l'hypothèse de récurrence

$$F_{n+1}(t) = \int_0^t \frac{(t-x)^n}{n!} dx = \frac{t^{n+1}}{(n+1)!}. \quad \square$$

Remarque. — L'expression de la fonction de répartition de S_n pour le seul intervalle $[0, 1]$ suffit pour le calcul de la loi de probabilité de N . Notons que la densité de S_n sur tout l'intervalle $[0, n]$ a été donnée dans l'exercice 7 du chapitre 11.

PROPOSITION 2.2. — La variable aléatoire N admet

$$a) \text{ la fonction de survie } P\{N > n\} = \frac{1}{n!} \quad (n \geq 0),$$

$$\text{b) la loi de probabilité } \begin{cases} P\{N = 0\} = P\{N = 1\} = 0, \\ P\{N = n\} = \frac{n-1}{n!} \quad (n \geq 2). \end{cases}$$

Démonstration. — Pour $n = 0$ on a $P\{N > 0\} = 1 = 1/0!$. Pour $n \geq 1$ il résulte de (2.3) que $P\{N > n\} = P\{S_n \leq 1\} = F_n(1)$. Or, d'après le Lemme 2.1, $F(1) = 1/n!$

Pour obtenir la loi de probabilité de N on note d'abord que $P\{N = n\} = 0$ pour $n = 0, 1$. Par ailleurs, pour $n \geq 2$, il vient :

$$P\{N = n\} = P\{N > n-1\} - P\{N > n\} = \frac{1}{(n-1)!} - \frac{1}{n!} = \frac{n-1}{n!}. \quad \square$$

PROPOSITION 2.3. — *L'espérance mathématique de N est donnée par : $\mathbb{E}[N] = e$.*

Démonstration. — En effet, $\mathbb{E}[N] = \sum_{n \geq 0} P\{N > n\}$, d'où, d'après la Proposition 2.2 a), $\mathbb{E}[N] = \sum_{n \geq 0} (1/n!) = e$. \square

PROPOSITION 2.4. — *Désignons par $G(s) = \mathbb{E}[s^N]$ la fonction génératrice de N . Alors*

$$G(1+u) = 1 + \sum_{r \geq 1} r e \frac{u^r}{r!}.$$

Il en résulte que, pour tout $r \geq 1$, le moment factoriel d'ordre r de N est donné par : $\mathbb{E}[N(N-1)\cdots(N-r+1)] = r e$.

Démonstration. — En effet,

$$\begin{aligned} G(s) &= \sum_{n \geq 2} \frac{n-1}{n!} s^n = s \sum_{n \geq 2} \frac{s^{n-1}}{(n-1)!} - \sum_{n \geq 2} \frac{s^n}{n!} \\ &= s \sum_{n \geq 1} \frac{s^n}{n!} - \sum_{n \geq 2} \frac{s^n}{n!} = s(e^s - 1) - (e^s - 1 - s) \\ &= (s-1)e^s + 1, \end{aligned}$$

d'où, en faisant $s = 1 + u$,

$$\begin{aligned} G(1+u) &= u e^{1+u} + 1 = 1 + u e \sum_{n \geq 0} \frac{u^n}{n!} \\ &= 1 + e \sum_{n \geq 0} \frac{u^{n+1}}{n!}, \end{aligned}$$

enfin en faisant le changement d'indice $n+1 = r$

$$G(1+u) = 1 + e \sum_{r \geq 1} \frac{u^r}{(r-1)!} = 1 + \sum_{r \geq 1} r e \frac{u^r}{r!}.$$

L'expression du moment factoriel d'ordre $r \geq 1$ résulte alors de la Proposition 2.5 du chap. 9. \square

3. Acheminement du courrier par voie hiérarchique. — Un fonctionnaire reçoit une lettre qui doit être acheminée au ministre par voie hiérarchique. Il est impératif que la lettre parvienne au ministre avant 1 heure, après avoir transité par les n échelons réglementaires E_1, \dots, E_n . Le dernier échelon E_n symbolise le bureau du ministre. On fait les hypothèses suivantes :

a) La lettre parvient au fonctionnaire de base, à l'échelon E_0 , à une date uniformément répartie entre 0 heure et 1 heure.

b) La lettre est transmise immédiatement à l'échelon E_1 (on convient que l'apposition de la signature du fonctionnaire et l'envoi physique de la lettre sont instantanés) et y parvient à une date uniformément répartie entre l'instant où elle a été envoyée de l'échelon E_0 et 1 heure.

c) Pour tout $k = 1, \dots, (n - 1)$ le fonctionnaire à l'échelon E_k transmet la lettre à l'échelon E_{k+1} et la lettre y parvient à une date uniformément répartie entre l'instant où elle quitte l'échelon E_k et 1 heure.

On désigne par X_0 la date à laquelle la lettre parvient à l'échelon E_0 et par X_1, \dots, X_n les dates auxquelles elle parvient aux échelons E_1, \dots, E_n . On se propose d'étudier la loi de probabilité de X_n .

Dans ce modèle, on impose que la lettre parvienne au ministre avant 1 heure, de sorte que $X_n < 1$. On pose : $Y_n = 1 - X_n > 0$. Les fonctionnaires situés aux échelons élevés vont disposer de moins en moins de temps pour faire transiter la lettre à l'échelon supérieur. C'est la rançon du pouvoir !

PROPOSITION 3.1. — *La variable aléatoire $Y_n = 1 - X_n$ admet la représentation suivante :*

$$Y_n = U_0 U_1 \cdots U_n,$$

où (U_0, U_1, \dots, U_n) est un système de $(n + 1)$ variables aléatoires indépendantes, chacune étant uniformément répartie sur l'intervalle $]0, 1]$.

La loi de Y_n est explicitement donnée dans l'exercice 8 du chapitre 15.

Démonstration. — Posons $Y_k = 1 - X_k$ ($k = 0, 1, \dots, n$). On a par définition : $0 < Y_n \leq Y_{n-1} \leq \dots \leq Y_1 \leq Y_0 \leq 1$. Introduisons les variables aléatoires $U_0 = Y_0, U_1 = Y_1/Y_0, \dots, U_n = Y_n/Y_{n-1}$, qui prennent leurs valeurs dans $]0, 1]$. On a naturellement $Y_n = U_0 U_1 \cdots U_n$ et tout revient à montrer que le système de variables (U_0, U_1, \dots, U_n) ainsi défini est un système de variables aléatoires indépendantes, chacune uniformément répartie sur $]0, 1]$.

Pour ne pas alourdir les notations et sans néanmoins nuire à la généralité faisons la démonstration pour $n = 2$. Soient y_0, y_1, y_2 trois nombres vérifiant $0 < y_2 \leq y_1 \leq y_0 \leq 1$. On a, par définition, en utilisant les notations usuelles sur les densités :

$$f_{Y_0}(y_0) = I_{]0,1]}(y_0), \quad f_{Y_1|Y_0}(y_1|y_0) = \frac{1}{y_0} I_{]0,1]} \left(\frac{y_1}{y_0} \right),$$

$$f_{Y_2|Y_1, Y_0}(y_2|y_1, y_0) = f_{Y_2|Y_1}(y_2|y_1) = \frac{1}{y_1} I_{]0,1]} \left(\frac{y_2}{y_1} \right),$$

d'où la densité conjointe de (Y_0, Y_1, Y_2) :

$$\begin{aligned} f_{(Y_0, Y_1, Y_2)}(y_0, y_1, y_2) &= f_{Y_0}(y_0) f_{Y_1 | Y_0}(y_1 | y_0) f_{Y_2 | Y_1, Y_0}(y_2 | y_1, y_0) \\ &= \frac{1}{y_0 y_1} I_{]0,1]}(y_0) I_{]0,1]} \left(\frac{y_1}{y_0} \right) I_{]0,1]} \left(\frac{y_2}{y_1} \right). \end{aligned}$$

Pour obtenir la densité conjointe de (U_0, U_1, U_2) on fait le changement de variables $u_0 = y_0$, $u_1 = y_1/y_0$, $u_2 = y_2/y_1$. Les variables u_0, u_1, u_2 sont toutes comprises entre 0 et 1 et l'on a : $y_0 = u_0$, $y_1 = u_0 u_1$, $y_2 = u_0 u_1 u_2$. Le jacobien de cette dernière transformation est donnée par

$$J = \frac{D(y_0, y_1, y_2)}{D(u_0, u_1, u_2)} = u_0^2 u_1.$$

Par suite la densité conjointe de (U_0, U_1, U_2) est donnée par :

$$\begin{aligned} g_{(U_0, U_1, U_2)}(u_0, u_1, u_2) &= f_{(Y_0, Y_1, Y_2)}(y_0, y_1, y_2) |J| \\ &= \frac{1}{u_0^2 u_1} I_{]0,1]}(u_0) I_{]0,1]}(u_1) I_{]0,1]}(u_2) u_0^2 u_1 \\ &= I_{]0,1]}(u_0) I_{]0,1]}(u_1) I_{]0,1]}(u_2). \quad \square \end{aligned}$$

Il résulte de la précédente proposition que pour tout $r > 0$ la variable aléatoire Y_n admet un moment d'ordre r donné par

$$\mathbb{E}[Y_n^r] = \prod_{k=0}^n \mathbb{E}[U_k^r] = (\mathbb{E}[U_0^r])^{n+1} = \frac{1}{(r+1)^{n+1}}.$$

PROPOSITION 3.2

- $Y_n \rightarrow 0$ en moyenne d'ordre $r > 0$, donc en probabilité.
- $Y_n \xrightarrow{p.s.} 0$.

Démonstration. — La première assertion résulte du fait que $1/(r+1)^{n+1}$ tend vers 0 quand n tend vers l'infini. Quant à la seconde, on note que la série de terme général $1/(r+1)^{n+1}$ est convergente. Le résultat en découle en vertu de la Proposition 4.4 du chapitre 16. \square

En fait, on peut démontrer un résultat encore plus fort à savoir que $Y_n = o(1/2^{n+1})$ presque sûrement, comme il résulte de la proposition suivante.

PROPOSITION 3.3. — *La suite de terme général $Z_n = 2^{n+1} Y_n$ converge presque sûrement vers 0.*

Démonstration. — Posons $V_k = 2U_k$ ($k = 0, \dots, n$). Alors (V_0, V_1, \dots, V_n) est un système de $(n+1)$ variables aléatoires indépendantes, dont chacune est uniformément distribuée sur $[0, 2]$. On a donc $Z_n = 2^{n+1} Y_n = V_0 V_1 \cdots V_n$, d'où $\mathbb{E}[Z_n^{1/2}] = \prod_{k=0}^n \mathbb{E}[V_k^{1/2}] = (\mathbb{E}[V_0^{1/2}])^{n+1}$. Or $\mathbb{E}[V_0^{1/2}] = \frac{1}{2} \int_0^2 \sqrt{x} dx = 2\sqrt{2}/3 = a < 1$. On en déduit que la série de terme général $\mathbb{E}[Z_n^{1/2}] = a^{n+1}$ est convergente. La Proposition 4.4 du chapitre 16, appliquée à $r = \frac{1}{2}$ permet de conclure. \square

Remarque. — Le fait que la suite de terme général Z_n converge presque sûrement peut être déduit de la théorie des *martingales*. En effet (V_0, V_1, \dots, V_n) est un système de $(n+1)$ variables aléatoires indépendantes, identiquement distribuées, à *valeurs positives* et d'*espérance mathématique égale à 1*. La suite de terme général $Z_n = V_0 V_1 \cdots V_n$ vérifie alors trivialement les propriétés : $\mathbb{E}[Z_n] = 1$, $\mathbb{E}[Z_n | Z_0, Z_1, \dots, Z_{n-1}] = Z_{n-1}$ ($n \geq 1$). C'est la définition-même d'une *martingale positive*. Or un théorème classique affirme que cette martingale converge presque sûrement vers 0, sauf dans le cas banal où les termes de la suite dont on prend les produits partiels sont presque sûrement égaux à la constante 1.¹

4. Fractions continues. — Soit (q_n) ($0 \leq n \leq N$) (resp. (q_n) ($n \geq 0$)) une suite finie (resp. infinie) d'entiers tels que $q_0 \geq 0$ et tels que $q_n \geq 1$ pour tout $n = 1, 2, \dots, N$ (resp. pour tout $n \geq 1$). Pour tout n tel que $0 \leq n \leq N$ (resp. pour tout $n \geq 1$) le nombre rationnel $[q_0; q_1, \dots, q_n]$ défini par

$$(4.1) \quad [q_0; q_1, \dots, q_n] = q_0 + \frac{1}{q_1 + \frac{1}{\ddots + \frac{1}{q_n}}},$$

est appelé *convergent* d'ordre n de la suite (q_n) ($0 \leq n \leq N$) (resp. de la suite (q_n) ($n \geq 0$)). On dit que la suite formée par ces convergents est une *fraction continue finie* ou une *fraction continue (infinie)*, suivant que la suite initiale des q_n est finie ou infinie. Les nombres entiers q_0, q_1, q_2, \dots s'appellent les *quotients partiels* de la fraction continue.

La *valeur* d'une fraction continue finie est définie comme son convergent de plus grand ordre, soit $[q_0; q_1, \dots, q_N]$ avec les notations précédentes. On démontre (voir, par exemple, Hardy et Wright,² chap. 10) que la suite des convergents d'une fraction continue *infinie* converge vers une limite, disons x . On dit alors que la fraction continue a pour *valeur* x . Il est coutumier de représenter la valeur x comme $[q_0; q_1, q_2, \dots]$, ou comme

$$(4.2) \quad x = q_0 + \frac{1}{q_1 + \frac{1}{\ddots + \frac{1}{q_n + \frac{1}{q_{n+1} + \frac{1}{\ddots}}}}}$$

¹ Voir, par exemple, Neveu (Jacques). — *Martingales à temps discret*. — Paris, Masson, 1972.

² Hardy (G.H.) and Wright (E.M.). — *An introduction to the theory of numbers*. Oxford Univ. Press, new edition 1979. L'ouvrage publié pour la première fois en 1938 a eu de multiples rééditions et reste un grand classique.

On démontre (*op. cit.*) que pour tout nombre rationnel r il existe exactement deux fractions continues finies de valeur r . On démontre encore (*op. cit.*) que pour tout nombre irrationnel x il existe une fraction continue *infinie* unique ayant pour valeur x .

Autrement dit à tout nombre irrationnel x correspond une suite unique (q_n) ($n \geq 0$) de nombres entiers strictement positifs telle que l'identité (4.2) soit satisfaite.

Du fait de l'unicité du développement de x en fraction continue, on a les propriétés suivantes :

$$(4.3) \quad 0 < x < 1 \Rightarrow q_0 = 0 \text{ et } q_1 = \text{partie entière de } \frac{1}{x}.$$

Dans ce paragraphe, partant d'un nombre x supposé être la réalisation d'une variable aléatoire X (dans une épreuve ω), nous donnons explicitement une loi de probabilité de X telle que les différents quotients partiels q_1, q_2, \dots soient les réalisations de variables aléatoires Q_1, Q_2, \dots ayant toutes une loi commune. Nous distinguons deux cas, suivant que la variable aléatoire X est à valeurs dans $]0, 1[$ ou dans $]1, +\infty[$.

4.1. *Variable aléatoire à valeurs dans $]0, 1[$.* — Considérons une variable aléatoire X à valeurs dans $]0, 1[$, de loi *diffuse* et de fonction de répartition F . La probabilité pour qu'elle prenne une valeur dans \mathbb{Q} est nulle. On peut donc, avec une probabilité égale à 1, la développer en fraction continue *infinie* :

$$X = \frac{1}{Q_1 + \frac{1}{Q_2 + \frac{1}{\ddots}}}} = [0, Q_1, Q_2, \dots],$$

où les quotients partiels Q_1, Q_2, \dots sont des variables aléatoires à valeurs dans $\mathbb{N}^* = \{1, 2, \dots\}$.

D'après (4.3), Q_1 est égal à la partie entière de $1/X$. Par conséquent, la différence $1/X - Q_1$ est une variable aléatoire à valeurs dans $]0, 1[$, notée Y . Exprimons d'abord la loi de probabilité conjointe de (Q_1, Y) en fonction de F . Pour tout entier $k \geq 1$ et tout $y \in]0, 1[$ on a :

$$\begin{aligned} \{Q_1 = k, Y \leq y\} &= \left\{k \leq \frac{1}{X} < k+1, \frac{1}{X} - k \leq y\right\} \\ &= \left\{k \leq \frac{1}{X} \leq k+y\right\} = \left\{\frac{1}{k+y} \leq X \leq \frac{1}{k}\right\}. \end{aligned}$$

D'où la loi de probabilité conjointe de (Q_1, Y) :

$$h(k, y) = P\{Q_1 = k, Y \leq y\} = F\left(\frac{1}{k}\right) - F\left(\frac{1}{k+y}\right).$$

On en déduit les lois marginales :

$$(4.1.1) \quad \pi(k) = P\{Q_1 = k\} = h(k, 1) = F\left(\frac{1}{k}\right) - F\left(\frac{1}{k+1}\right);$$

$$(4.1.2) \quad r(k) = \sum_{n \geq k} \pi(n) = P\{Q_1 \geq k\} = F\left(\frac{1}{k}\right);$$

$$(4.1.3) \quad G(y) = P\{Y \leq y\} = \sum_{k \geq 1} h(k, y) = \sum_{k \geq 1} \left(F\left(\frac{1}{k}\right) - F\left(\frac{1}{k+y}\right) \right).$$

THÉORÈME 4.1.1 (Gauss). — Soit F_1 la fonction de répartition définie par

$$(4.1.4) \quad F_1(x) = \begin{cases} 0, & \text{si } x \leq 0; \\ \frac{1}{\text{Log } 2} \text{Log}(1+x), & \text{si } 0 < x < 1; \\ 1, & \text{si } x \geq 1. \end{cases}$$

Si X est une variable aléatoire admettant F_1 pour fonction de répartition, alors les quotients partiels Q_1, Q_2, \dots de son développement en fraction continue sont identiquement distribués. Leur loi commune est donnée par :

$$P\{Q_1 \geq k\} = \frac{1}{\text{Log } 2} \text{Log}\left(1 + \frac{1}{k}\right) \quad (k = 1, 2, \dots).$$

Démonstration. — Substituons l'expression de F_1 donnée par (4.1.4) à F dans les formules (4.1.1)—(4.1.3). On obtient pour $k \in \mathbb{N}^*$ et $y \in]0, 1[$ les expressions :

$$(4.1.5) \quad \begin{aligned} \pi(k) &= P\{Q_1 = k\} = \frac{1}{\text{Log } 2} \left(\text{Log}\left(1 + \frac{1}{k}\right) - \text{Log}\left(1 + \frac{1}{k+1}\right) \right); \\ r(k) &= P\{Q_1 \geq k\} = \frac{1}{\text{Log } 2} \text{Log}\left(1 + \frac{1}{k}\right); \\ G(y) &= \frac{1}{\text{Log } 2} \sum_{k \geq 1} \left(\text{Log}\left(1 + \frac{1}{k}\right) - \text{Log}\left(1 + \frac{1}{k+y}\right) \right) \\ &= \frac{1}{\text{Log } 2} \sum_{k \geq 1} \left(\text{Log}\left(1 + \frac{y}{k}\right) - \text{Log}\left(1 + \frac{y}{k+1}\right) \right) \\ &= \frac{1}{\text{Log } 2} \text{Log}(1+y). \end{aligned}$$

On constate que $G(y) = F_1(y)$: la variable aléatoire Y a même loi que X . On effectue alors sur Y la même opération que sur X en introduisant le deuxième quotient partiel Q_2 . Il admet donc la même loi que Q_1 . On peut ainsi continuer pour tous les autres quotients partiels. \square

4.2. *Variable aléatoire à valeurs dans $]1, +\infty[$.* — Supposons maintenant que X soit à valeurs dans $]1, +\infty[$, de loi diffuse, de fonction de répartition F . Avec une probabilité égale à 1 on peut développer X en fraction continue infinie :

$$X = [Q_0; Q_1, Q_2, \dots].$$

Ici Q_0 est égal à la partie entière de X , donc $Q_0 \in \mathbb{N}^*$ et la différence $X - Q_0$, que nous posons égale à Y , est une variable aléatoire à valeurs dans $]0, 1[$. Pour tout $k \in \mathbb{N}^*$ et tout $y \in]0, 1[$ on a :

$$\{Q_0 = k, Y \leq y\} = \{k \leq X < k + 1, X - k \leq y\} = \{k \leq X \leq k + y\}.$$

On en déduit la loi de probabilité conjointe de (Q_0, Y)

$$h(k, y) = P\{Q_0 = k, Y \leq y\} = F(k + y) - F(k),$$

puis les lois marginales :

$$(4.2.1) \quad \pi(k) = P\{Q_0 = k\} = h(k, 1) = F(k + 1) - F(k);$$

$$(4.2.2) \quad r(k) = \sum_{n \geq k} \pi(n) = P\{Q_0 \geq k\} = 1 - F(k);$$

$$(4.2.3) \quad G(y) = P\{Y \leq y\} = \sum_{k \geq 1} h(k, y) = \sum_{k \geq 1} (F(k + y) - F(k)).$$

Ces calculs permettent d'obtenir un résultat analogue au résultat énoncé dans le Théorème 4.1.1.

THÉORÈME 4.2.1 (Gauss). — *Soit F_2 la fonction de répartition définie par*

$$(4.2.4) \quad 1 - F_2(x) = \begin{cases} 1, & \text{si } x \leq 1; \\ \frac{1}{\text{Log } 2} \text{Log}\left(1 + \frac{1}{x}\right), & \text{si } x > 1. \end{cases}$$

Si X est une variable aléatoire admettant F_2 pour fonction de répartition, alors les quotients partiels Q_0, Q_1, Q_2, \dots de son développement en fraction continue sont identiquement distribués. Leur loi commune est donnée par :

$$P\{Q_0 \geq k\} = \frac{1}{\text{Log } 2} \text{Log}\left(1 + \frac{1}{k}\right) \quad (k = 1, 2, \dots).$$

Démonstration. — Comme pour le précédent théorème, on remplace la fonction F qui apparaît dans les formules (4.2.1)—(4.2.3) par la fonction F_2 donnée en (4.2.4). On obtient pour $k \in \mathbb{N}^*$ et $y > 1$ les formules :

$$(4.2.5) \quad \begin{aligned} \pi(k) &= \frac{1}{\text{Log } 2} \left(\text{Log}\left(1 + \frac{1}{k}\right) - \text{Log}\left(1 + \frac{1}{k+1}\right) \right); \\ r(k) &= P\{Q_0 \geq k\} = \frac{1}{\text{Log } 2} \text{Log}\left(1 + \frac{1}{k}\right); \\ G(y) &= P\{Y \leq y\} = \frac{1}{\text{Log } 2} \sum_{k \geq 1} \left(\text{Log}\left(1 + \frac{1}{k}\right) - \text{Log}\left(1 + \frac{1}{k+y}\right) \right). \end{aligned}$$

En utilisant le même calcul que dans la démonstration du Théorème 4.1.1 on en déduit :

$$\begin{aligned} G(y) &= \frac{1}{\text{Log } 2} \sum_{k \geq 1} \left(\text{Log} \left(1 + \frac{y}{k} \right) - \text{Log} \left(1 + \frac{y}{k+1} \right) \right) \\ &= \frac{1}{\text{Log } 2} \text{Log}(1+y). \end{aligned}$$

On constate que $G(y) = F_1(y)$ donnée dans le Théorème 4.1.1. On peut donc appliquer ce dernier théorème à Y et ainsi développer Y en fraction continue $[0; Q_1, Q_2, \dots]$. On constate, en comparant (4.1.5) et (4.2.5) que les quotients partiels Q_0, Q_1, Q_2, \dots ont la même loi. \square

Remarque. — Il est tout à fait remarquable de trouver cette même distribution pour tous les quotients partiels d'une fraction continue. Les démonstrations des deux théorèmes ne sont que de simples vérifications. Le plus difficile était d'imaginer des lois F_1 et F_2 qui convenaient. Nous sommes ici redevables du grand talent de Gauss.

5. Une application de la formule de Bernstein. — Une urne contient n boules numérotées de 1 à n ($n \geq 1$). On procède à une suite de tirages *avec remise* et l'on se propose d'étudier le nombre X de tirages nécessaires pour amener, *pour la première fois*, une boule déjà tirée. Comme toute suite de longueur $(n+1)$ dont les termes sont pris dans $\{1, 2, \dots, n\}$ contient au moins deux termes qui sont égaux (principe des tiroirs), la variable aléatoire X est à valeurs dans $\{2, \dots, n+1\}$.

On peut prendre comme ensemble fondamental Ω l'ensemble de toutes les suites de longueur $(n+1)$ dont les termes sont pris dans $\{1, 2, \dots, n\}$ et l'équipartition sur Ω . L'évènement $\{X > k\}$ ($k = 1, \dots, n$) est identifié au sous-ensemble de Ω de toutes les suites dont les k premiers termes sont distincts. Son cardinal est évidemment $(n!/(n-k)!)n^{n+1-k}$. On a donc

$$P\{X > k\} = \frac{n!}{(n-k)!} n^{n+1-k} \frac{1}{n^{n+1}} = \frac{1}{n^k} \frac{n!}{(n-k)!}, \quad k \in \{1, \dots, n\}.$$

Puisque $P\{X > 0\} = 1$, cette formule est encore vraie pour $k = 0$. C'est la *fonction de fiabilité* de X . La *loi de probabilité* de X vaut

$$P\{X = k\} = P\{X > k-1\} - P\{X > k\} = (k-1) \frac{n!}{n^k (n-k+1)!},$$

$$k \in \{2, \dots, n+1\}.$$

Son *espérance mathématique* est égale à

$$\mathbb{E}[X] = \sum_{k \geq 0} P\{X > k\} = n! \sum_{k=0}^n \frac{1}{n^k (n-k)!},$$

une expression qu'on peut récrire, en posant $n-k = j$,

$$\mathbb{E}[X] = \frac{n!}{n^n} \sum_{j=0}^n \frac{n^j}{j!}.$$

L'étude du comportement asymptotique de $\mathbb{E}[X]$ lorsque n tend vers l'infini ne paraît pas aisée; il est remarquable que la formule de Bernstein, établie au chap. 18, Remarque 1 de la Proposition 2.3 conduise au but. Rappelons cette formule :

$$e^{-n} \sum_{j=0}^n \frac{n^j}{j!} \rightarrow \frac{1}{2} \quad (n \rightarrow +\infty).$$

Il vient alors $\mathbb{E}[X] \sim \frac{1}{2} \frac{n!}{n^n} e^n$, d'où, en définitive, en utilisant la formule de Stirling $n! \sim (n/e)^n \sqrt{2\pi n}$, l'estimation : $\mathbb{E}[X] \sim \sqrt{\frac{\pi n}{2}} \quad (n \rightarrow +\infty)$.

6. Le modèle de la diffusion d'Ehrenfest. — Imaginons un certain nombre $a \geq 2$ de boules numérotées de 1 à a et réparties dans deux boîtes A et B . On considère l'opération suivante : à chaque instant entier (par exemple à chaque seconde à partir de l'instant 0) on choisit au hasard (c'est-à-dire avec l'équirépartition) un nombre entier dans $\{1, \dots, a\}$ et l'on déplace la boule portant le numéro prélevé de sa boîte dans l'autre. On procède à une suite illimitée d'opérations de ce type et l'on suppose que les tirages effectués à des instants distincts sont *indépendants*. On définit l'état du système (A, B) comme étant le nombre de boules situées dans la boîte A ; on voit qu'il y a $(a+1)$ états : $0, 1, \dots, a$. On désigne par X_n ($n \geq 0$) l'état du système à la date n .

Supposons qu'à un certain instant le système soit dans l'état $i \in \{0, 1, \dots, a\}$, c'est-à-dire qu'il y ait i boules dans la boîte A . A l'instant suivant, il sera nécessairement dans l'un des deux états $i-1, i+1$ selon que l'on aura tiré une boule de A ou de B . Il y a deux exceptions à cette règle :

$i = 0$, auquel cas seule la transition $0 \rightarrow 1$ est possible;

$i = a$, auquel cas seule la transition $a \rightarrow a-1$ est possible.

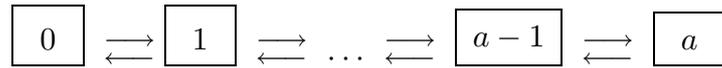
La probabilité p_{ij} pour que le système, se trouvant dans l'état i à la date n , passe dans l'état j à la date suivante $(n+1)$ est parfaitement définie; elle ne dépend que de i et de j , mais non de n , un fait dont nous avons tenu compte dans la notation. On l'appelle la *probabilité de passage* de l'état i à l'état j . La matrice $\mathcal{P} = (p_{ij})$ ($0 \leq i, j \leq a$) est appelée *matrice de passage*. Cette matrice est l'élément de base de la définition des chaînes de Markov homogènes, (cf. chap. 10, exercice 9), dont l'étude ne sera pas abordée ici, d'autant que la propriété principale du modèle d'Ehrenfest peut être démontrée sans faire appel à cette étude.

Dans le cas présent, on a :

$$\begin{cases} p_{i,i-1} = \frac{i}{a}, & i = 1, \dots, a; \\ p_{i,i+1} = 1 - \frac{i}{a}, & i = 0, \dots, a-1; \\ p_{i,j} = 0, & (i, j \in \{0, 1, \dots, a\}, |i-j| \neq 1). \end{cases}$$

(La première relation est encore valable pour $i = 0$, de même la deuxième relation est encore valable pour $i = a$, mais elles donnent toutes deux des valeurs nulles aux probabilités.)

En notant $i \rightarrow j$ le fait que $p_{ij} > 0$, on voit que l'on peut codifier les transitions en une opération dans le diagramme suivant :



Voici le théorème qui donne toute son importance au modèle d'Ehrenfest.

THÉORÈME 6.1. — On a : $\mathbb{E}[X_n] - \frac{a}{2} = \left(1 - \frac{2}{a}\right)^n \left(\mathbb{E}[X_0] - \frac{a}{2}\right)$.

Démonstration. — Par définition de l'espérance mathématique conditionnelle on a :

$$\begin{aligned} \mathbb{E}[X_n] &= \mathbb{E}[\mathbb{E}[X_n | X_{n-1}]] \\ &= \sum_{i=0}^a \mathbb{P}\{X_{n-1} = i\} \mathbb{E}[X_n | X_{n-1} = i]. \end{aligned}$$

Or

$$\begin{aligned} \mathbb{E}[X_n | X_{n-1} = i] &= p_{i,i-1} \times (i-1) + p_{i,i+1} \times (i+1) \\ &= \frac{i}{a}(i-1) + \left(1 - \frac{i}{a}\right)(i+1) = \left(1 - \frac{2}{a}\right)i + 1; \end{aligned}$$

d'où

$$\begin{aligned} \mathbb{E}[X_n] &= \left(1 - \frac{2}{a}\right) \sum_{i=0}^a i \mathbb{P}\{X_{n-1} = i\} + 1 \\ &= \left(1 - \frac{2}{a}\right) \mathbb{E}[X_{n-1}] + 1, \end{aligned}$$

ou encore

$$\mathbb{E}[X_n] - \frac{a}{2} = \left(1 - \frac{2}{a}\right) \left(\mathbb{E}[X_{n-1}] - \frac{a}{2}\right).$$

D'où le résultat par itération sur n . \square

Remarque 1. — Il résulte du théorème précédent que, quel que soit $\mathbb{E}[X_0]$, la suite de terme général $\mathbb{E}[X_n]$ converge *exponentiellement*, lorsque n tend vers l'infini, vers $\frac{a}{2}$, la moitié du nombre total de boules. Le cas où $X_0 = a$, c'est-à-dire le cas où à la date 0 toutes les boules sont dans la boîte A, est particulièrement intéressant. Dans ce cas $\mathbb{E}[X_n] - \frac{a}{2} = \left(1 - \frac{2}{a}\right)^n \frac{a}{2}$, d'où il résulte que $\mathbb{E}[X_n]$ converge *en décroissant* et *exponentiellement* vers $\frac{a}{2}$. Si l'on assimile les boules aux particules d'un gaz enfermé, à la date 0, dans la boîte A, ce gaz diffuse dans la boîte B et à la longue il y aura en moyenne autant de particules dans la boîte A que dans la boîte B. C'est exactement ce que le modèle d'Ehrenfest montre.

Remarque 2. — Il est clair que la suite (X_n) ($n \geq 0$) elle-même ne tend vers une limite, ni presque sûrement, ni même en probabilité, puisque pour tout $n \geq 1$ on a $|X_n - X_{n-1}| = 1$.

7. Vecteurs aléatoires uniformément répartis sur la sphère-unité de \mathbb{R}^n . — Les résultats de ce paragraphe seront utilisés dans le problème de probabilité géométrique étudié à la fin de ce chapitre.

Soient $n \geq 2$ et $X = (X_1, \dots, X_n)$ un vecteur aléatoire uniformément réparti sur la sphère-unité de \mathbb{R}^n . Comme les composantes X_1, \dots, X_n ont toutes même loi de probabilité, nous étudions la loi de probabilité de la première composante X_1 .

PROPOSITION 7.1. — *La loi de probabilité de X_1 admet une densité $f(x)$, nulle pour $|x| \geq 1$ et qui, pour $|x| < 1$, est donnée par :*

$$f(x) = c_n (1 - x^2)^{(n-3)/2}, \quad \text{où } c_n = \frac{1}{\sqrt{\pi}} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)}.$$

Pour $n = 2$ cette loi de probabilité est la loi A_1 de l'Arcsinus et pour $n = 3$ n'est autre que la distribution uniforme sur $] -1, +1[$.

Démonstration. — La loi de X_1 coïncide avec la projection orthogonale, sur l'axe $0x_1$, de la masse-unité uniformément répartie sur la sphère-unité. Déterminons cette projection. Le volume de la boule $B_n(0, R)$ est donné par

$$V_n(R) = \frac{\pi^{n/2}}{\Gamma(1 + n/2)} R^n$$

et l'aire de la sphère $S_{n-1}(0, R)$ par

$$\sigma_{n-1}(R) = \frac{d}{dR} V_n(R) = 2 \frac{\pi^{n/2}}{\Gamma(n/2)} R^{n-1}.$$

Prenons $R = 1$ et projetons la mesure aréolaire portée par $S_{n-1}(0, 1)$ sur l'axe $0x_1$. On obtient de la sorte une distribution de masses sur R , de densité $g(x)$, nulle pour $|x| \geq 1$ et qui, pour $|x| < 1$, est donnée par : $g(x) dx = \sigma_{n-2}(\sqrt{1-x^2}) ds$ avec $\sqrt{1-x^2} ds = dx$. D'où

$$g(x) = \sigma_{n-2}(\sqrt{1-x^2}) \frac{1}{\sqrt{1-x^2}} = 2 \frac{\pi^{(n-1)/2}}{\Gamma((n-1)/2)} (1-x^2)^{(n-3)/2}.$$

On en déduit la densité normée $f(x) = g(x)/\sigma_{n-1}(1)$, expression qui, après simplification, est celle de l'énoncé. \square

Dans l'énoncé suivant, E est un sous-espace vectoriel de \mathbb{R}^n , de dimension $k \geq 2$. On désigne par E^\perp son complément orthogonal et par p, q les projections orthogonales de \mathbb{R}^n sur E, E^\perp , respectivement. On pose ensuite

$$U = p \circ X, \quad V = q \circ X$$

et on désigne par U^* le vecteur aléatoire normé $|U|^{-1}U$ (défini hors de l'ensemble négligeable $\{|U| = 0\}$).

PROPOSITION 7.2. — *Le vecteur aléatoire U^* est uniformément réparti sur la sphère-unité de E .*

Démonstration. — Étant donnée une rotation ρ de E , considérons la rotation $\bar{\rho}$ de \mathbb{R}^n caractérisée par les conditions $p \circ \bar{\rho} = \rho \circ p$, $q \circ \bar{\rho} = q$. Posons $X' = \bar{\rho} \circ X$. On a alors

$$\rho \circ U^* = |p \circ X'|^{-1} p \circ X', \quad V = q \circ X'.$$

En outre, X' a même loi que X . Il en résulte que le couple $(\rho \circ U^*, V)$ a même loi conjointe que le couple (U^*, V) . Cela suffit pour conclure. \square

PROPOSITION 7.3. — *Pour tout $n \geq 2$ désignons par F_n la fonction de répartition de X_1 . Alors la suite (F_n) est croissante.*

Démonstration. — Plaçons-nous dans les hypothèses de la Proposition 7.2 et prenons pour E l'espace vectoriel engendré par les $(n-1)$ premiers éléments de la base canonique de \mathbb{R}^n . La première composante X_1 de X coïncide alors avec la première composante U_1 de U . D'autre part, $U_1^* = |U|^{-1} U_1$ est la première composante de $U^* = |U|^{-1} U$, qui, d'après la Proposition 7.2, est uniformément répartie sur la sphère-unité de \mathbb{R}^{n-1} . L'inégalité $F_{n-1} \leq F_n$ résulte alors du fait que l'on a : $X_1 = U_1 \leq |U|^{-1} U_1 = U_1^*$. \square

8. Un problème de probabilité géométrique. — Les « probabilités géométriques » étaient à la mode à l'époque où le calcul des probabilités était dans son enfance. Les problèmes de l'aiguille de Buffon, de la baguette brisée, ... ont exercé la sagacité de nombreux mathématiciens. Dans ce paragraphe, nous nous proposons de résoudre un problème qui s'apparente à cette discipline et qui a été posé par Williams.³

PROBLÈME. — *Trois navettes spatiales échouent sur une planète aux points P, Q, R ; ces points sont indépendants et uniformément distribués sur la surface de la planète (assimilée à une sphère de centre O et de rayon 1). Deux navettes « communiquent directement par radio » si l'angle qu'elles forment avec le centre O de la sphère est inférieur à $\frac{\pi}{2}$. Par exemple, P et Q communiquent directement si $\widehat{POQ} < \frac{\pi}{2}$. Alors la probabilité pour que les trois navettes communiquent entre elles est égale à $(\pi + 2)/(4\pi)$. Il est précisé que deux navettes peuvent communiquer soit directement, soit par l'intermédiaire de la troisième.*

La méthode de résolution que nous utilisons ci-après nous a été fournie par G. Letta et L. Pratelli.

Notons $u \cdot v$ le produit scalaire des deux vecteurs u, v de \mathbb{R}^3 et $U \cdot V$ celui des deux vecteurs aléatoires U, V dans \mathbb{R}^3 , puis désignons par S la sphère-unité de \mathbb{R}^3 . Sur un espace probabilisé $(\Omega, \mathfrak{A}, P)$ donnons-nous trois vecteurs aléatoires indépendants U, V, W , à valeurs dans S , dont chacun admet comme loi la répartition uniforme μ sur S et posons :

$$A = \{U \cdot V > 0\}, \quad B = \{V \cdot W > 0\}, \quad C = \{W \cdot U > 0\}.$$

³ Williams (David). — *Probability with martingales.* — Cambridge, Cambridge Math. Textbooks, 1994, exercice EG2, p. 224.

L'évènement dont on cherche la probabilité est donné par :

$$D = (A \cap B) \cup (B \cap C) \cup (C \cap A).$$

Il peut s'exprimer comme réunion de trois évènements *deux à deux disjoints* :

$$(8.1) \quad D = (A \cap B) \cup (B \cap C \cap A^c) \cup (C \cap A \cap B^c).$$

Pour calculer la probabilité de D nous utilisons les deux Lemmes suivants, dont la démonstration sera donnée dans l'Appendice.

LEMME 8.1. — *Si un vecteur aléatoire V , à valeurs dans S , admet comme loi la répartition uniforme μ , alors, pour tout élément $u \in S$, la variable aléatoire réelle $u \cdot V$ admet comme loi la répartition uniforme sur l'intervalle $[-1, +1]$.*

LEMME 8.2. — *Pour tout élément $v \in S$, désignons par H_v l'hémisphère ouvert, intersection de la sphère S et du demi-espace $\{u : u \cdot v > 0\}$. Etant donné un couple (v, w) d'éléments de S , tel que $v \cdot w > 0$, on désigne par α l'angle formé par ces deux vecteurs, soit $\alpha = \text{Arccos}(v \cdot w)$. On a alors $\mu(H_w \cap H_{-v}) = \alpha/(2\pi)$.*

Nous sommes à présent en mesure de calculer la probabilité de D en utilisant la formule (8.1)

a) D'après le Lemme 8.1 on a :

$$P(A \cap B) = \int_S \mu(dv) P\{U \cdot v > 0, v \cdot W > 0\} = \frac{1}{4}.$$

b) En utilisant le Lemme 8.2, puis le Lemme 8.1, il vient :

$$\begin{aligned} P(B \cap C \cap A^c) &= P\{V \cdot W > 0, W \cdot U > 0, U \cdot V < 0\} \\ &= \int_S \mu(dv) \int_{H_v} \mu(dw) P\{w \cdot U > 0, U \cdot v < 0\} \\ &= \int_S \mu(dv) \int_{H_v} \mu(dw) \mu(H_w \cap H_{-v}) \\ &= \int_S \mu(dv) \int_{H_v} \mu(dw) \frac{1}{2\pi} \text{Arccos}(v \cdot w) \\ &= \frac{1}{2\pi} \int_S \mu(dv) \int_{\{v \cdot W > 0\}} \text{Arccos}(v \cdot W) dP \\ &= \frac{1}{2\pi} \frac{1}{2} \int_0^1 \text{Arccos } t \, dt \\ &= \frac{1}{4\pi} \int_0^{\pi/2} x \sin x \, dx = \frac{1}{4\pi}. \end{aligned}$$

On obtient de même $P(C \cap A \cap B^c) = \frac{1}{4\pi}$, d'où

$$P(D) = \frac{1}{4} + \frac{1}{4\pi} + \frac{1}{4\pi} = \frac{\pi + 2}{4\pi} \approx 0,409155.$$

Généralisation. — Le problème précédent peut être généralisé en remplaçant \mathbb{R}^3 par \mathbb{R}^n ($n \geq 2$). Le Lemme 8.2 demeure valable tandis que le Lemme 8.1 doit être modifié de la manière suivante.

LEMME 8.1'. — *Si un vecteur aléatoire V , à valeurs dans la sphère-unité S de \mathbb{R}^n admet comme loi la répartition uniforme μ , alors, pour tout élément $u \in S$, la variable aléatoire réelle $u \cdot V$ admet comme densité la fonction f*

$$f(x) = \begin{cases} c_n(1 - x^2)^{(n-3)/2}, & \text{si } |x| < 1; \\ 0, & \text{sinon;} \end{cases}$$

où c_n désigne la constante de normalisation $c_n = \frac{1}{\sqrt{\pi}} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)}$.

Pour résoudre le problème généralisé, remarquons que l'on a exactement comme dans le cas $n = 3$ les évaluations :

$$(8.2) \quad \begin{aligned} P(A \cap B) &= \frac{1}{4}; \\ P(B \cap C \cap A^c) &= \frac{1}{2\pi} \int_S \mu(dv) \int_{\{v \cdot W > 0\}} \text{Arccos}(v \cdot W) dP. \end{aligned}$$

Pour calculer cette dernière probabilité on utilise le Lemme 8.1' :

$$\begin{aligned} P(B \cap C \cap A^c) &= \frac{1}{2\pi} \int_0^1 f(t) \text{Arccos } t dt \\ &= \frac{1}{2\pi} c_n \int_0^1 (1 - t^2)^{(n-3)/2} \text{Arccos } t dt = \frac{1}{2\pi} c_n \int_0^{\pi/2} x \sin^{n-2} x dx. \end{aligned}$$

On obtient la même valeur pour $P(C \cap A \cap B^c)$ et l'on trouve en définitive :

$$P(D) = \frac{1}{4} + \frac{1}{\pi} c_n \int_0^{\pi/2} x \sin^{n-2} x dx, \quad \text{avec} \quad c_n = \frac{1}{\sqrt{\pi}} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)}.$$

Cas particuliers :

$$\begin{aligned} n = 2, c_2 = \frac{1}{\pi}, \quad P(D) &= \frac{1}{4} + \frac{1}{\pi^2} \int_0^{\pi/2} x dx = \frac{1}{4} + \frac{1}{8} = \frac{3}{8} = 0,375. \\ n = 3, c_3 = \frac{1}{2}, \quad P(D) &= \frac{1}{4} + \frac{1}{2\pi} \int_0^{\pi/2} x \sin x dx = \frac{1}{4} + \frac{1}{2\pi} = \frac{\pi + 2}{4\pi} \approx 0,409155. \end{aligned}$$

Comportement asymptotique de la solution. — Désignons par p_n la probabilité que trois navettes atterrissant sur la sphère unité dans l'espace à n dimensions (!) communiquent par radio. Nous nous proposons de montrer que la suite (p_n) est croissante et de calculer sa limite. A cet effet, considérons, sur un espace probabilisé $(\Omega, \mathfrak{A}, P)$, une suite (X_n) ($n \geq 2$) de variables aléatoires réelles, où X_n est la première composante d'un vecteur aléatoire uniformément réparti sur la sphère-unité de \mathbb{R}^n . On a alors en vertu de (8.2)

$$p_n = \frac{1}{4} + \frac{1}{\pi} \int_{\{X_n > 0\}} \text{Arccos } X_n dP.$$

Posons $\delta_n = \frac{\pi}{4} - \frac{1}{\pi} \int_{\{X_n > 0\}} \text{Arccos } X_n dP$. La relation $P\{X_n > 0\} = \frac{1}{2}$ permet d'écrire :

$$\delta_n = \int_{\{X_n > 0\}} \left[\frac{\pi}{2} - \text{Arccos } X_n \right] dP = \int \left[\frac{\pi}{2} - \text{Arccos } X_n \right]^+ dP,$$

c'est-à-dire en posant $Y_n = \frac{\pi}{2} - \text{Arccos } X_n$,

$$(8.3) \quad \delta_n = \int Y_n^+ dP = \int_0^{\pi/2} P\{Y_n > y\} dy = \int_0^{\pi/2} P\{X_n > \sin y\} dy.$$

La relation évidente $\mathbb{E}[X_n^2] = 1/n$ montre que la suite (X_n) converge vers 0 dans L^2 , donc en probabilité. Il résulte alors de (8.3), en vertu du théorème de convergence dominée, que $\delta_n \rightarrow 0$, c'est-à-dire que

$$\lim_{n \rightarrow \infty} \int_{\{X_n > 0\}} \text{Arccos } X_n dP = \frac{\pi}{4} \quad \text{et donc} \quad \lim_{n \rightarrow \infty} p_n = \frac{1}{2}.$$

Reste à montrer que la suite (p_n) est croissante ou encore que la suite (δ_n) est décroissante. Or ceci résulte de (8.3) et de la Proposition 7.3 qui affirme que la suite des fonctions de répartition des variables X_n est croissante. \square

Remarque. — Le problème a encore un sens pour $n = 1$. La « sphère » est alors réduite aux deux points d'abscisses -1 et $+1$ et la condition « angle inférieur à $\pi/2$ » équivaut à « angle nul ». Les trois navettes communiquent donc si et seulement si elles atterrissent au même point. Par suite, $p_1 = 2(1/2^3) = 1/4 = 0,25$ et la suite (p_n) est bien croissante à partir de p_1 : $p_1 = 0,25$, $p_2 = 0,375$, $p_3 = 0,409$, ...

Appendice. — Nous donnons, pour terminer, une démonstration des deux lemmes 8.1' et 8.2 et traitons directement le cas $n \geq 2$. Le Lemme 8.1' n'est autre que la Proposition 7.1. Il suffit donc de démontrer le Lemme 8.2.

Démonstration du Lemme 8.2. — On peut supposer $u \neq w$. Dans le cas $n = 2$, une simple figure montre que le lemme est vrai. Dans le cas $n \geq 3$, désignons par E le plan (sous-espace vectoriel de \mathbb{R}^n de dimension 2) engendré par les deux vecteurs v, w . En outre, désignons par p la projection orthogonale de \mathbb{R}^n sur E . Considérons alors, sur un espace probabilisé $(\Omega, \mathfrak{A}, P)$, un vecteur aléatoire M uniformément réparti sur la sphère-unité de \mathbb{R}^n , posons $J = p \circ M$ et désignons par J^* le vecteur aléatoire normé $|J|^{-1}J$ (défini hors de l'ensemble négligeable $\{|J| = 0\}$). La Proposition 7.2 nous apprend que J^* est uniformément réparti sur le cercle-unité du plan E (à deux dimensions). Or on a :

$$\begin{aligned} \mu(H_w \cap H_{-v}) &= P\{M \in H_w \cap H_{-v}\} = P\{M \cdot w > 0, M \cdot v < 0\} \\ &= P\{J \cdot w > 0, J \cdot v < 0\} = \frac{\alpha}{2\pi}, \end{aligned}$$

où l'égalité finale résulte du cas particulier $n = 2$ déjà traité. \square

SOLUTIONS DES EXERCICES

Chapitre Premier

1. a) AB^cC^c ; b) AB^cC ; c) ABC ; d) $A \cup B \cup C$; e) $A^cBC \cup AB^cC \cup ABC^c \cup ABC$; f) $AB^cC^c \cup A^cBC^c \cup A^cB^cC \cup A^cB^cC^c$; g) $A^cB^cC^c$; h) $A^cBC \cup AB^cC \cup ABC^c$; i) $(ABC)^c$.
2. Considérons les « atomes » suivants dont la réunion est Ω : ABC , ABC^c , AB^cC , $[AB^cC^c]$, $[A^cBC]$, A^cBC^c , A^cB^cC , $A^cB^cC^c$. D'après l'énoncé, les atomes entre crochets sont vides. On a donc :
 - a) $AC^c = ABC^c \cup AB^cC^c = ABC^c$.
 - c) $B = ABC \cup ABC^c \cup A^cBC \cup A^cB^cC$. On voit que $AC^c \subset B$.
6. Considérons les « atomes » suivants dont la réunion est Ω : EFG , EFG^c , EF^cG , EF^cG^c , E^cFG , E^cFG^c , E^cF^cG , $E^cF^cG^c$. Chacun des événements A , B peut s'écrire comme réunion disjointe d'atomes : $A = EFG \cup EF^cG \cup E^cFG$, $B = EFG \cup EFG^c \cup EF^cG \cup EF^cG^c \cup E^cFG$.
 - a) On voit que $A \subset B$.
 - b) On a $A = B$, si et seulement si $EFG^c \cup EF^cG^c = \emptyset$, i.e. $EG^c = \emptyset$, i.e. $E \subset G$.
Autre démonstration utilisant les fonctions indicatrices :
 - a) $I_A = I_{E \cup F} I_G = (I_E + I_F - I_{EF}) I_G = I_{EG} + I_{FG} - I_{EFG}$ et $I_B = I_E + I_{FG} - I_{EFG}$. On voit que $I_A - I_B = I_{EG} - I_E \leq 0$, car $EG \subset E$; donc $A \subset B$.
 - b) On a $A = B$, si et seulement si $I_A - I_B = 0$, i.e. $I_{EG} - I_E = 0$, i.e. $EG = E$, i.e. $E \subset G$.
7. a) $A \triangle B = A^cB \cup AB^c$ et $A \triangle B^c = A^cB^c \cup AB$. D'où le résultat.
- b) Considérons la partition de Ω selon les atomes : $A^{\varepsilon_1} B^{\varepsilon_2} C^{\varepsilon_3}$, $\varepsilon_1, \varepsilon_2, \varepsilon_3 = 0, 1$, où l'on a posé A^ε égal à A ou à A^c , suivant que $\varepsilon = 1$ ou 0 . Alors chacun des événements peut s'écrire comme réunion disjointe d'atomes : $A \triangle B = A^cBC \cup A^cBC^c \cup AB^cC \cup AB^cC^c$; $A \triangle C = A^cBC \cup A^cB^cC \cup ABC^c \cup AB^cC^c$; d'où $(A \triangle B) \cap (A \triangle C) = A^cBC \cup AB^cC^c$. On voit de façon analogue que : $A \triangle (B \cup C) = A^cBC \cup A^cBC^c \cup A^cB^cC \cup AB^cC^c$. On a $(A \triangle B) \cap (A \triangle C) = A \triangle (B \cup C)$ si et seulement si $A^cBC^c \cup A^cB^cC = \emptyset$, i.e. $A^c(B \triangle C) = \emptyset$, i.e. $B \triangle C \subset A$.

Chapitre 2

2. $\{\emptyset, \{ab\}, \{c\}, \{a, b, c\}\}$.
3. b) $\mathfrak{F}_3 = \{\emptyset, \{a\}, \{c, d\}, \{b, e\}, \{a, b, e\}, \{a, c, d\}, \{b, c, d, e\}, \Omega\}$.
- d) $\mathfrak{F}_4 = \mathfrak{F}_2 \cup \mathfrak{F}_3 \cup \{\{e\}, \{a, e\}, \{b, c, d\}, \{a, b, c, d\}\}$.

4. a) Introduisons la relation d'équivalence R sur Ω définie par :

$$x R y \iff \forall A \in \mathfrak{A} \quad [(x \in A) \iff (y \in A)].$$

Les classes d'équivalence de cette relation appartiennent à \mathfrak{A} et forment la partition désirée.

b) Non.

5. L'algèbre \mathfrak{A} est engendrée par la famille (d'ensembles disjoints)

$\Pi = \{A_1^{\varepsilon_1} \cap \dots \cap A_n^{\varepsilon_n}\}$, où $\varepsilon_1, \dots, \varepsilon_n = 0, 1$ et où l'on a posé : $A^\varepsilon = A$ ou A^c , suivant que $\varepsilon = 1$ ou 0 . Les éléments de cette famille peuvent être appelés les « atomes » ; tout élément de \mathfrak{A} est réunion (finie) d'atomes. La famille Π a au plus 2^n éléments, donc \mathfrak{A} a au plus 2^{2^n} éléments.

6. Les dix propriétés se démontrent de la même façon. Montrons-le pour \mathcal{C}_3 :

a) $]a, b[= [a, b] \setminus [a, a]$, d'où $\mathfrak{T}(\mathcal{C}_3) \subset \mathcal{B}^1$.

b) $]a, b[=]a, b[\cup \left(\bigcap_{k \geq 1}]a - 1/k, a[\right)$, d'où $\mathcal{B}^1 \subset \mathfrak{T}(\mathcal{C}_3)$.

7. Dans cet exercice, « dénombrable » signifie « au plus dénombrable ». La condition est évidemment *suffisante*, puisque si Ω est dénombrable, alors $\mathfrak{T}(\mathcal{C}) = \mathfrak{P}(\Omega)$. Pour montrer qu'elle est *nécessaire*, on introduit la classe \mathfrak{A} des éléments de $\mathfrak{P}(\Omega)$ constituée des parties dénombrables de Ω , ainsi que de leurs complémentaires. Une simple vérification montre que \mathfrak{A} est une *tribu*. Supposons à présent que $\mathfrak{T}(\mathcal{C}) = \mathfrak{P}(\Omega)$. Il résulte alors de la suite d'inclusions $\mathcal{C} \subset \mathfrak{A} \subset \mathfrak{P}(\Omega)$ que $\mathfrak{T}(\mathcal{C}) = \mathfrak{T}(\mathfrak{A}) = \mathfrak{P}(\Omega)$. Or puisque \mathfrak{A} est une tribu, on a $\mathfrak{T}(\mathfrak{A}) = \mathfrak{A}$, d'où, en définitive, $\mathfrak{A} = \mathfrak{P}(\Omega)$. Ainsi toute partie de Ω est soit dénombrable, soit de complémentaire dénombrable. Or ceci ne peut avoir lieu que si Ω est lui-même dénombrable. Si, en effet, Ω est non-dénombrable, il existe un sous-ensemble $A \in \mathfrak{P}(\Omega)$ tel que A et A^c sont non-dénombrables.

8. Prenons pour ensemble fondamental l'ensemble à trois éléments $\Omega = \{a, b, c\}$ et désignons par \mathcal{A}_a (resp. \mathcal{A}_b , resp. \mathcal{A}_c) la tribu engendrée par $\{a\}$ (resp. $\{b\}$, resp. $\{c\}$). Par exemple, $\mathcal{A}_a = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$. Or l'ensemble $\mathcal{A}_b \cup \mathcal{A}_c$ n'est *pas* une tribu, puisqu'il contient bien les éléments $\{b\}$ et $\{c\}$, mais non $\{b, c\}$. De même, les ensembles $\mathcal{A}_c \cup \mathcal{A}_a$ et $\mathcal{A}_a \cup \mathcal{A}_b$ ne sont pas des tribus, mais la réunion $\mathcal{A}_a \cup \mathcal{A}_b \cup \mathcal{A}_c$ en est une.

Chapitre 3

4. Appliquer la formule de Poincaré (cf. Proposition 3.1) aux événements

A_1^c, \dots, A_n^c :

$$P(A_1^c \cup \dots \cup A_n^c) = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P(A_{i_1}^c \cap \dots \cap A_{i_k}^c).$$

D'où

$$\begin{aligned} 1 - P(A_1 \cap \dots \cap A_n) &= \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} (1 - P(A_{i_1} \cup \dots \cup A_{i_k})). \\ &= \sum_{k=1}^n (-1)^{k-1} \binom{n}{k} - \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P(A_{i_1} \cup \dots \cup A_{i_k}). \end{aligned}$$

Or $\sum_{k=1}^n (-1)^{k-1} \binom{n}{k} = 1 - (1-1)^n$; d'où le résultat.

5. a) Il s'agit de montrer que $d(A, B) \geq 0$, $d(A, A) = 0$ et $d(A, C) \leq d(A, B) + d(B, C)$. Il suffit de montrer cette dernière propriété; elle devient évidente, si l'on représente chacun des évènements $A \triangle C$, $A \triangle B$, $B \triangle C$ par les « atomes » : $A \triangle C = ABC^c \cup AB^cC^c \cup A^cBC \cup A^cB^cC$; $A \triangle B = AB^cC \cup AB^cC^c \cup A^cBC \cup A^cB^cC^c$; $B \triangle C = ABC^c \cup A^cBC^c \cup AB^cC \cup A^cB^cC$. On voit que $A \triangle C \subset (A \triangle B) \cup (B \triangle C)$. D'où le résultat. On peut établir cette dernière inclusion de façon plus rapide en observant que l'opération \triangle est associative et que $A \triangle A = \emptyset$ pour tout $A \in \mathfrak{A}$. On a alors : $A \triangle C = A \triangle \emptyset \triangle C = A \triangle (B \triangle B) \triangle C = (A \triangle B) \triangle (B \triangle C) \subset (A \triangle B) \cup (B \triangle C)$.
- b) On a $P(A \triangle B) = P(A) + P(B) - 2P(A \cap B)$, d'où le résultat en remarquant que $P(A \cap B) \leq P(A), P(B)$.
6. Démontrons la première des inégalités. On pose : $A_* = \liminf_n A_n = \bigcup_{n \geq 1} B_n$ et $B_n = \bigcap_{k \geq n} A_k$ ($n \geq 1$). Alors (B_n) est une suite *croissante* et $A_* = \lim_n B_n$. On a donc : $P(A_*) = P(\lim_n B_n) = \lim_n P(B_n)$. Or pour tout $n \geq 1$ on a : $B_n \subset A_n$, d'où $P(B_n) \leq P(A_n)$. Prenons la limite inférieure des deux côtés et remarquons que, la suite (B_n) étant croissante, on a : $\liminf_n P(B_n) = \lim_n P(B_n) = P(A_*)$. Il en résulte : $P(A_*) \leq \liminf_n P(A_n)$. Prenons $\Omega = \{-1, +1\}$, $P(\{-1\}) = 1/4$, $P(\{+1\}) = 3/4$, $A_n = \{(-1)^n\}$. Alors $\liminf_n A_n = \emptyset$, $\limsup_n A_n = \Omega$, $\liminf_n P(A_n) = 1/4$, $\limsup_n P(A_n) = 3/4$, d'où le résultat.
7. a) On montre que \mathfrak{A} est une algèbre de la façon suivante : $\Omega \in \mathfrak{A}$, car Ω est cofini; ensuite, \mathfrak{A} est stable par passage au complémentaire. Il reste à montrer que \mathfrak{A} est stable par réunion finie. Soient $A_1, \dots, A_n \in \mathfrak{A}$. Si tous les A_i sont finis, il en est de même de $\bigcup_{i=1}^n A_i$, qui est donc un élément de \mathfrak{A} . Supposons qu'il existe un A_i , soit A_1 , qui ne soit pas fini. Il est donc cofini. On a alors $\bigcup_{i=1}^n A_i \supset A_1$, $(\bigcup_{i=1}^n A_i)^c \subset A_1^c$ et A_1^c est fini; donc $\bigcup_{i=1}^n A_i$ est cofini, et par suite appartient à \mathfrak{A} . Montrons que \mathfrak{A} n'est pas une σ -algèbre. Prenons $\Omega = \mathbb{N}$; il suffit de remarquer que l'ensemble $\bigcup_{k \geq 0} \{2k\}$ des nombres pairs est une réunion dénombrable d'éléments de \mathfrak{A} (il n'est ni fini, ni cofini).
- b) Il suffit de remarquer que l'algèbre $\alpha(\mathcal{C})$ engendrée par \mathcal{C} contient l'ensemble \mathfrak{A} des parties finies et cofinies de Ω ; comme cet ensemble est une algèbre, on a : $\alpha(\mathcal{C}) = \mathfrak{A}$.
- c) On a tout d'abord $P \geq 0$ et $P(\emptyset) = 0$, puisque \emptyset est fini. Montrons que P est simplement additive, c'est-à-dire que pour toute suite finie (A_1, \dots, A_n) d'éléments deux à deux disjoints de \mathfrak{A} on a :

$$(1) \quad P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i).$$

Si tous les A_i sont finis, il en est de même de $\bigcup_{i=1}^n A_i$ et on a la relation (1) sous la forme $0 = 0 + \dots + 0$. Supposons qu'il existe un A_i , soit A_1 , qui ne soit pas fini; il est donc cofini. Puisque les A_i sont disjoints, cet A_1 est le seul élément de la suite (A_1, \dots, A_n) qui soit cofini; comme $\bigcup_{i=1}^n A_i$ est lui-même cofini, on a la relation (1) sous la forme $1 = 1 + 0 + \dots + 0$. Montrons enfin que P n'est pas σ -additive sur \mathfrak{A} . Prenons $\Omega = \mathbb{N}$; $A_n = \{n\}$ ($n \in \mathbb{N}$). La suite (A_n) est une suite d'éléments deux à deux disjoints de \mathfrak{A} dont la réunion \mathbb{N} appartient à \mathfrak{A} . Si P était σ -additive, on devrait avoir $P(\mathbb{N}) = P(\bigcup_{n \geq 0} \{n\}) = \sum_{n \geq 0} P(\{n\})$, c'est-à-dire $1 = 0$, ce qui est une contradiction.

8. Les démonstrations sont analogues à celles de l'exercice 7. Remarquons que si « dénombrable » signifie « infini dénombrable », alors \mathfrak{A} n'est pas une algèbre. En effet, puisque Ω a une puissance strictement supérieure à celle du dénombrable, l'ensemble Ω n'est ni dénombrable, ni codénombrable ($\Omega^c = \emptyset$ est fini, et non « infini dénombrable »); donc $\Omega \notin \mathfrak{A}$.
9. Considérons la partition $\Pi = \{A_1^{\varepsilon_1} \cap \dots \cap A_n^{\varepsilon_n} : \varepsilon_1, \dots, \varepsilon_n = 0, 1\}$ de Ω , où l'on a posé $A^\varepsilon = A$ ou A^c , suivant que $\varepsilon = 1$ ou 0 . Appelons « atomes » les éléments de cette partition (il y en a au plus 2^n). Tout élément de \mathfrak{A} est réunion (disjointe) d'atomes, de sorte que le premier membre de (1) s'écrit sous la forme $\sum_{i=1}^{m'} c'_i P(\alpha_i)$, où $m' > 0$, (c'_1, \dots, c'_m) est une suite de réels et $(\alpha_1, \dots, \alpha_m)$ est une suite d'atomes. La formule (1) s'écrit donc :

$$(1') \quad \sum_{i=1}^{m'} c'_i P(\alpha_i) \geq 0.$$

Revenons à présent à la démonstration. Il est clair que a) \Rightarrow b) et il suffit donc de démontrer b) \Rightarrow a). Supposons donc que (1') soit vraie pour tout P telle que $P(A_i) = 0$ ou 1 pour tout $i = 1, \dots, n$; alors, par définition des atomes, on a $P(\alpha) = 0$ ou 1 , pour toute atome α . En choisissant P de telle sorte que $P(\alpha_i) = 1$ ($i = 1, \dots, m'$), il résulte de (1') que $c'_i \geq 0$ ($i = 1, \dots, m'$). Mais alors on a (1') pour toute mesure P , ce qui est la propriété a).

10. D'après l'exercice 9, les quatre formules à prouver sont établies dans toute leur généralité, si l'on montre que pour tout l tel que $0 \leq l \leq n$ chacune d'elles est vraie pour la mesure de probabilité P telle que $P(A_1) = \dots = P(A_l) = 1$ et $P(A_{l+1}) = \dots = P(A_n) = 0$. Considérons donc une telle mesure de probabilité P définie à partir de l'indice l . D'abord les quatre formules sont banalement vraies si $l = 0$. Supposons $1 \leq l \leq n$ et considérons une sous-suite $1 \leq i_1 < \dots < i_k \leq n$ où $k \geq 1$. Si $i_k \leq l$, on a $P(A_{i_1}) = \dots = P(A_{i_k}) = 1$, d'où, pour $k \geq 2$, $P(A_{i_1} A_{i_2}) = P(A_{i_1}) + P(A_{i_2}) - P(A_{i_1} \cup A_{i_2}) = 2 - 1 = 1$ et donc $P(A_{i_1} \dots A_{i_k}) = 1$ par récurrence sur k . Au contraire, si $i_k \geq l + 1$, on a : $0 \leq P(A_{i_1} \dots A_{i_k}) \leq P(A_{i_k}) = 0$. Ainsi $P(A_{i_1} \dots A_{i_k}) = 1$ ou 0 suivant

que $\{i_1, \dots, i_k\}$ est ou n'est pas un sous-ensemble de $[l]$. La somme S_k^n est donc égale au nombre de sous ensembles $\{i_1, \dots, i_k\}$, de cardinal k , de l'ensemble $[l]$, c'est-à-dire égale à $\binom{l}{k}$ (voir Proposition 4.4.1 du chap. 4). Pour prouver la formule de Poincaré il suffit donc de vérifier l'identité $1 = \sum_{k=1}^n (-1)^{k-1} \binom{l}{k}$, qui est une banalité.

Pour démontrer l'identité sur les V_n^r observons que lorsque $l < r$, on a évidemment $V_n^r = 0$. Or pour $0 \leq k \leq n - r$ la quantité $S_{r+k}^n = \binom{l}{r+k}$ est nulle; le second membre de l'identité à prouver est nul lui aussi. Lorsque $l = r$, on a $P(A_1 \cdots A_l) = 1$, d'où $V_n^r = 1$. Par ailleurs, pour $1 \leq k \leq n - r$, on a $S_{r+k}^n = \binom{l}{r+k} = \binom{l}{l+k} = 0$ et $S_r^n = \binom{l}{r} = \binom{l}{l} = 1$. Le second membre vaut aussi 1. Enfin, lorsque $l > r$, on a $V_n^r = 0$. Quant au second membre, il vaut : $\sum_{k=0}^{n-r} (-1)^k \binom{r+k}{k} \binom{l}{r+k} = \sum_{k=0}^{l-r} (-1)^k \binom{r+k}{k} \binom{l}{r+k} = \binom{l}{r} \sum_{k=0}^{l-r} (-1)^k \binom{l-r}{k} = 0$.

Pour démontrer l'identité sur les W_n^r notons également que si $l < r$ les deux membres sont nuls et ils sont tous deux égaux à 1 si $l = r$. Lorsque $l > r$, on a $1 \geq W_n^r \geq P(A_1 \cdots A_l A_{l+1}^c \cdots A_n^c) = 1$. Quant au second membre, il est aussi égal à 1 d'après l'identité démontrée dans § 5.4.

Pour la quatrième identité observons que le membre de gauche vaut $\frac{1}{2}[1 - (-1)^l] = 1$ ou 0 suivant que l est impair ou pair. Le membre de droite vaut $\sum_{j=1}^l (-1)^{j-1} 2^{j-1} \binom{l}{j} = -\frac{1}{2} \sum_{j=1}^l (-2)^j \binom{l}{j} = -\frac{1}{2}[(-1)^l - 1]$.

11. Pour tout entier $n \geq 1$ on peut écrire : $A = A_1 + \cdots + A_n + R_n$, où $R_n = \bigcup_{k \geq n+1} A_k$. D'après c) on a donc $P(A) = P(A_1) + \cdots + P(A_n) + P(R_n)$, d'où, puisque $P(R_n) \geq 0$, l'inégalité $P(A) \geq P(A_1) + \cdots + P(A_n)$, d'où, en faisant tendre n vers l'infini, $P(A) \geq \sum_{n \geq 1} P(A_n)$.

Remarque. — La σ -additivité est équivalente à l'additivité simple, si pour la suite décroissante (R_n) ($n \geq 1$), on a $P(R_n) \rightarrow 0$.

12. En posant $f_a(x) = \sum_{n \geq 0} (a)_n (x^n/n!)$, on obtient $f'_a(x) = a f_{a+1}(x)$ ($|x| < 1$). D'autre part, $f_{a+1}(x) - f_a(x) = x f_{a+1}(x)$, d'où l'on déduit l'équation différentielle $f'_a(x) = a f_a(x)/(1 - x)$. Comme $f_a(0) = 1$, on en tire $f_a(x) = (1 - x)^{-a}$ ($|x| < 1$).

13. On a les développements :

$$\begin{aligned} \sin x &= x {}_0F_1\left(\begin{matrix} - \\ 3/2 \end{matrix}; -\frac{x^2}{4}\right) = \sum_{n \geq 0} (-1)^n \frac{x^{2n+1}}{(2n+1)!}; \\ \cos x &= {}_0F_1\left(\begin{matrix} - \\ 1/2 \end{matrix}; -\frac{x^2}{4}\right) = \sum_{n \geq 0} (-1)^n \frac{x^{2n}}{(2n)!}; \\ \ln(1+x) &= x {}_2F_1\left(\begin{matrix} 1, 1 \\ 2 \end{matrix}; -x\right) = \sum_{n \geq 1} (-1)^{n-1} \frac{x^n}{n} \quad (|x| < 1); \end{aligned}$$

$$\operatorname{arctg} x = x {}_2F_1\left(\begin{matrix} 1/2, 1 \\ 3/2 \end{matrix}; -x^2\right) = \sum_{n \geq 0} (-1)^n \frac{x^{2n+1}}{2n+1} \quad (|x| < 1);$$

$$\begin{aligned} \arcsin x &= x {}_2F_1\left(\begin{matrix} 1/2, 1/2 \\ 3/2 \end{matrix}; x^2\right) \\ &= x + \sum_{n \geq 1} \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2 \cdot 4 \cdot 6 \cdots (2n)} \frac{x^{2n+1}}{2n+1} \quad (|x| < 1). \end{aligned}$$

Chapitre 4

4. a) On prend pour Ω la classe des parties de quatre éléments de X , puis $\mathfrak{A} = \mathfrak{P}(\Omega)$ et pour P l'équirépartition sur Ω . Notons que $\operatorname{card} \Omega = \binom{4}{4}$.
- b) Soit $X_i = X \setminus \{(i, d), (i, g)\}$. Alors A_i est la classe des parties de Ω constituées de la $i^{\text{ième}}$ paire et d'une partie à deux éléments de X_i .
- c) $P(A_i) = \frac{\binom{18}{2}}{\binom{20}{4}} = \frac{3}{95}$.
- d) Pour $k = 2$ on a : $P(A_{i_1} \cap A_{i_2}) = \frac{\binom{16}{0}}{\binom{20}{4}} = \frac{1}{\binom{20}{4}}$; pour $k > 2$, la quantité est nulle.
- e) Soit $E = A_1 \cup \cdots \cup A_{10}$. D'après la formule de Poincaré

$$P(E) = \sum_{i=1}^{10} P(A_i) - \sum_{1 \leq i < j \leq 10} P(A_i \cap A_j) = 10 \frac{3}{95} - \frac{\binom{10}{2}}{\binom{20}{4}} = \frac{99}{323}.$$

5. On prend pour Ω la classe des parties à quatre éléments de l'ensemble de cinquante-deux cartes; le cardinal de Ω est $\binom{52}{4}$ et P est l'équirépartition sur Ω . La probabilité cherchée est $\frac{\binom{4}{2} \binom{48}{2}}{\binom{52}{4}} \approx 0,025$.
6. *Jeu de «Passe-Dix»*. Plus généralement, supposons que l'on prenne un nombre *impair* $(2k+1)$ de dés ($k \geq 0$). Notons Ω l'ensemble des suites $\omega = (\alpha_1, \dots, \alpha_{2k+1})$, où α_i est le numéro amené par le $i^{\text{ième}}$ dé ($i = 1, \dots, 2k+1$) et prenons l'équirépartition P sur Ω . Si A désigne l'évènement «la somme des points amenés est *strictement* supérieure à $3+7k$ », il s'agit de montrer que $P(A) = P(A^c)$. Or l'application $(\alpha_1, \dots, \alpha_{2k+1}) \mapsto (7-\alpha_1, \dots, 7-\alpha_{2k+1})$ est une *bijection* de A sur A^c , puisque $\alpha_1 + \cdots + \alpha_{2k+1} > 3+7k \Leftrightarrow (7-\alpha_1) + \cdots + (7-\alpha_{2k+1}) \leq 3+7k$. Autrement dit, le jeu de «Passe- $(3+7k)$ » est équitable. Le jeu de «Passe-Dix» en est le cas particulier pour $k = 1$.

Chapitre 5

3. L'ensemble \mathcal{C} des éléments ω de Ω pour lesquels la suite $(X_n(\omega))$ est convergente peut être représenté par :

$$\mathcal{C} = \bigcap_{l \geq 1} \bigcup_{n \geq 1} \bigcap_{k \geq n} \left\{ \omega : |X_k(\omega) - X_n(\omega)| < \frac{1}{l} \right\}.$$

5. $a = 0$; $F_Y(y) = P\{Y \leq y\} = H_b(y) = \begin{cases} 0, & \text{si } y < b; \\ 1, & \text{si } y \geq b. \end{cases}$
- $a > 0$; $F_Y(y) = P\{aX + b \leq y\} = P\left\{X \leq \frac{y-b}{a}\right\} = F\left(\frac{y-b}{a}\right).$

$$\begin{aligned}
 a < 0; F_Y(y) &= P\{aX + b \leq y\} = P\left\{X \geq \frac{y-b}{a}\right\} \\
 &= 1 - P\left\{X \leq \frac{y-b}{a}\right\} = 1 - F\left(\frac{y-b}{a} - 0\right).
 \end{aligned}$$

6. La fonction $\Phi_h(x)$ est croissante (elle est dérivable et sa dérivée est positive). On a d'autre part l'encadrement $F(x) \leq \Phi_h(x) \leq F(x+h)$, qui montre que $\lim_{x \rightarrow -\infty} \Phi_h(x) = 0$ et $\lim_{x \rightarrow +\infty} \Phi_h(x) = 1$.
8. a) On a, pour $0 \leq u \leq 1$, les relations : $G(u) = P\{U \leq u\} = P\{F \circ X \leq u\} = P\{X \leq F^{-1}(u)\} = F(F^{-1}(u)) = u$.
 b) En posant $h = F^{-1}$, on a : $P\{X \leq x\} = P\{X \in]-\infty, x]\} = P\{h \circ U \in]-\infty, x]\} = P\{U \in h^{-1}(]-\infty, x])\}$. On vérifie que $h^{-1}(]-\infty, x]) =]0, F(x)]$, d'où $P\{X \leq x\} = P\{U \in]0, F(x)]\} = F(x)$.
9. On a, par hypothèse, pour tout x , les relations : $F(2x) = P\{X \leq 2x\} = P\{2X \leq 2x\} = P\{X \leq x\} = F(x)$; d'où, en itérant : $F(x) = F(2^n x)$. En passant à la limite pour $n \rightarrow \infty$, il vient :

$$F(x) = \begin{cases} F(-\infty) = 0, & \text{pour } x < 0; \\ F(+\infty) = 1, & \text{pour } x > 0. \end{cases}$$

Comme F est en outre une fonction de répartition, elle est continue à droite; d'où $F(0) = 1$. On voit que $F = H_0$ est la fonction de répartition de ε_0 . De là $X = 0$, presque sûrement.

Chapitre 6

2. a) Soit \mathcal{E}_1 la classe des évènements qui sont indépendants de chaque évènement de \mathcal{C}_2 . On vérifie facilement que \mathcal{E}_1 est une *classe monotone*. Comme elle contient \mathcal{C}_1 , elle contient encore la classe monotone engendrée $\mathfrak{M}(\mathcal{C}_1)$. De là $\mathfrak{M}(\mathcal{C}_1)$ et \mathcal{C}_2 sont deux classes indépendantes. De la même façon, on démontre que la classe \mathcal{E}_2 des évènements qui sont indépendants de chaque évènement de $\mathfrak{M}(\mathcal{C}_1)$ est une classe monotone; elle contient \mathcal{C}_2 , donc aussi $\mathfrak{M}(\mathcal{C}_2)$. Par suite, $\mathfrak{M}(\mathcal{C}_1)$ et $\mathfrak{M}(\mathcal{C}_2)$ sont des classes indépendantes.
 b) Il suffit d'observer que, si \mathfrak{A} est une tribu, on a $\mathfrak{M}(\mathfrak{A}) = \mathfrak{T}(\mathfrak{A})$ (Proposition 7, chap. 2) et d'appliquer la partie a).
3. $P(A) = \frac{1}{3}$, $P(B) = \frac{1}{2}$, $P(AB) = \frac{1}{6}$. D'où $P(AB) = P(A)P(B)$.
5. Écrivons « A indépendant de BC », « B indépendant de CA », « C indépendant de AB » : $P(ABC) = P(A)P(BC) = P(B)P(CA) = P(C)P(AB)$, c'est-à-dire

$$\frac{P(ABC)}{P(A)P(B)P(C)} = \frac{P(BC)}{P(B)P(C)} = \frac{P(CA)}{P(C)P(A)} = \frac{P(AB)}{P(A)P(B)} = k. \quad (a)$$

Écrivons « A indépendant de $B \cup C$ » :

$$P(AB) + P(AC) = P(A)P(B) + P(A)P(C). \quad (b)$$

En substituant (a) dans (b), il vient $k = 1$.

6. Prendre A, B non indépendants et $C = B^c$.
7. Par hypothèse, $P(AB | C) = P(A | C)P(B | C)$, $P(AB | C^c) = P(A | C^c)P(B | C^c)$, $P(A | C) = P(A)$; d'où $P(AB) = P(AB | C)P(C) + P(AB | C^c)P(C^c) = P(A | C)P(B | C)P(C) + P(A | C^c)P(B | C^c)P(C^c) = P(A)P(BC) + P(A)P(BC^c) = P(A)P(B)$.
8. a) On divise l'ensemble des familles de deux enfants en quatre classes notées $(g, g), (g, f), (f, g), (f, f)$; les lettres g et f sont pour « garçon » et « fille », la première lettre d'un doublet se rapportant à l'aîné de la famille. On prend $\Omega = \{(g, g), (g, f), (f, g), (f, f)\}$ et sur Ω l'équirépartition P . Dans ces conditions :
- $\alpha)$ $P\{(g, g) | \{(g, g), (g, f), (f, g)\}\} = \frac{1}{3}$.
 $\beta)$ $P\{(g, g) | \{(g, g), (g, f)\}\} = \frac{1}{2}$.
- b) Un enfant issu d'une famille de deux enfants peut être classé suivant l'alternative : « il est garçon ou fille », « il a un frère ou une sœur ». Cela fait un ensemble Ω de $2^2 = 4$ profils. Si l'on prend l'équirépartition P sur Ω , la probabilité cherchée est $\frac{1}{2}$.
9. La variable aléatoire X doit être constante (la tribu engendrée $\mathfrak{T}(X)$ est indépendante d'elle-même, donc formée d'éléments de probabilité 0 ou 1).
10. Les trois variables sont indépendantes deux à deux (noter que le produit $X_1X_2X_3 = (X_1X_2)^2$ est identiquement égal à 1).
11. $F_Y(x) = F_1(x) \dots F_n(x)$ et $F_Z(x) = 1 - (1 - F_1(x)) \dots (1 - F_n(x))$.
12. a) $P_2(s) = \sum_{k=0}^s P_1(k)P_1(s-k)$; b) $P_r(k) = e^{-ra} \frac{(ra)^k}{k!}$.
13. On a : $\text{card } A_k = \binom{r+s-1}{r-1}$, d'où $P\{X_k = 1\} = \frac{\text{card } A_k}{\text{card } \Omega} = \frac{r}{r+s}$.
14. On trouve $P\{X_k = 1 | S_n = i\} = i/n$, quantité ne dépendant ni de r , ni de s , et ceci quelle que soit la nature des tirages (avec remise ou sans remise). C'est là un résultat surprenant.
16. On prend pour Ω l'ensemble des $3! = 6$ permutations de A, B, C : $ABC, ACB, BAC, BCA, CAB, CBA$. On prend pour P l'équirépartition sur Ω ; alors $P(E) = P(F) = \frac{1}{2}$, $P(E \cap F) = \frac{1}{3} \neq P(E)P(F)$; de là E et F ne sont pas indépendants.
17. a) Non. b) Oui (prendre, par exemple, l'équirépartition sur les « atomes » AB, AB^c, A^cB, A^cB^c).

Chapitre 7

2. $P\{T_r = r + k\} = \binom{r+k-1}{r-1} p^r q^k \quad (k \geq 0)$.
3. Considérons une suite de répétitions indépendantes d'une alternative dont les deux résultats sont : A « choix de la poche g », B « choix de la poche d », avec les probabilités $\frac{1}{2}$ et $\frac{1}{2}$. Désignons par T_s le nombre *minimum* de répétitions qu'il faut effectuer pour obtenir s fois le résultat A . La variable aléatoire T_s suit la loi binomiale négative : $P\{T_s = s + k\} = \binom{s+k-1}{s-1} \left(\frac{1}{2}\right)^{s+k}$ ($k \geq 0$).
- a) L'évènement « la poche g est reconnue vide et la poche d contient r allumettes » est équivalent à $\{T_{N+1} = (N+1) + (N-r)\}$, dont la probabilité est $u'_r = \binom{2N-r}{N} 2^{-(N+1)} 2^{-(N-r)}$. Le même raisonnement s'applique en échangeant poche g et d ; d'où $u_r = 2u'_r = \binom{2N-r}{N} 2^{-2N+r}$.
- b) $v_r = 2P\{T_N = N + (N-r)\} = \binom{2N-r-1}{N-1} 2^{-2N+r+1}$.
- c) $v = \sum_{r=1}^N v_r 2^{-(r+1)} = 2^{-2N} \sum_{r=1}^N \binom{2N-r-1}{N-1}$. (Supposons qu'au moment où la première boîte est vidée, l'autre contienne encore r allumettes ($r \geq 1$). Cette première boîte n'est pas la première à être reconnue vide, si et seulement si, lors des $(r+1)$ choix ultérieurs, le fumeur choisit l'autre boîte : probabilité égale à $2^{-(r+1)}$.)
- d) Utiliser l'identité $\binom{a-k}{N-1} + \binom{a-k}{N} = \binom{a-k+1}{N}$ et sommer de $k=0$ à $k=n$.
- e) Prenons dans d) $n=N$, $a=2N-1$: $\sum_{k=1}^N \binom{2N-1-k}{N-1} = \binom{2N}{N} - \binom{N-1}{N} - \binom{2N-1}{N-1} = \frac{1}{2} \binom{2N}{N}$, d'où le résultat.

Remarque 1. — Il résulte de a) que l'on a $\sum_{r=0}^N \binom{2N-r}{N} (1/2)^{2N-r} = 1$, qu'il n'est pas aisé de démontrer directement (voir Exercice 11).

Remarque 2. — La solution donnée ci-dessus repose sur la notion de variable aléatoire *binomiale négative*, dont la définition présuppose un espace probabilisé *infini*. Or on peut résoudre le problème dans le cadre d'un espace probabilisé *fini*, comme nous l'a indiqué Anatole Joffe.

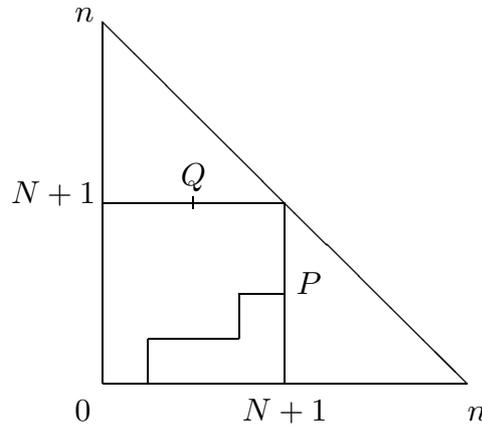
Le nombre d'allumettes qu'il est nécessaire de choisir pour que le fumeur s'aperçoive qu'une boîte est vide peut varier entre $N+1$ et $2N+1$. Prenons $n \geq 2N+1$ et assimilons le choix d'une poche à un jeu de « pile » ou « face » ; on peut modéliser le problème en prenant pour espace fondamental l'ensemble Ω_n des suites de longueur n dont les termes sont des 0 ou des 1, chaque suite étant affectée de la probabilité $1/2^n$, ce qui définit une mesure de probabilité P sur Ω_n .

Il y a bijection entre ces suites et les chemins latticiels de longueur n reliant l'origine à la droite $x+y=n$. L'espace qui nous intéresse est l'ensemble Ω' des chemins ω' partant de l'origine et aboutissant à l'un des côtés $x=N+1$, $y=N+1$, du carré $(0,0)$, $(0,N+1)$, $(N+1,0)$, $(N+1,N+1)$. Soit ω' un chemin de l'espace Ω' de longueur

$(N + 1) + (N - r) = 2N + 1 - r$, aboutissant donc, soit au point P $(N + 1, N - r)$, soit au point Q $(N - r, N + 1)$. La probabilité est donnée par $P(\{\omega'\}) = P\{\omega \in \Omega_n : \omega \text{ admet } \omega' \text{ comme section commençante}\}$, qui vaut $(1/2^n)2^{n-(2N+1-r)} = 2^{-(2N+1)+r}$. (On constate que cette expression ne dépend plus de n .)

Pour calculer u_r il faut compter le nombre de chemins ω' allant de $(0, 0)$ à P et de $(0, 0)$ à Q et ceci, *en touchant le bord du carré pour la première fois*. Cela signifie que le chemin doit se terminer par un segment horizontal s'il aboutit à P et par un segment vertical s'il aboutit à Q . Le nombre de ces chemins aboutissant à P est donc $\binom{2N-r}{N}$; de même pour ceux aboutissant à Q . D'où

$$u_r = 2 \binom{2N-r}{N} \frac{1}{2^{2N+1-r}} = \binom{2N-r}{N} \frac{1}{2^{2N-r}} \quad (r = 0, 1, \dots, N).$$



On voit de même, en raisonnant sur le carré $(0, 0)$, $(0, N)$, $(N, 0)$, (N, N) , que $v_r = \binom{2N-r-1}{N-1} (1/2^{2N-r-1})$ ($r = 1, \dots, N$). Les rubriques b) ... e) se traitent comme précédemment. On doit à un étudiant de licence de 1995 d'avoir calculé directement v en montrant que $v = u_0/2$. Sa méthode a consisté à considérer les deux événements :

A (de probabilité u_0) «le fumeur s'aperçoit, *pour la première fois*, que l'une des boîtes est vide, et l'autre est vide aussi, sans que le fumeur le sache»; B (de probabilité v) «la première boîte vidée n'est pas la première *reconnue* vide».

Lors de la réalisation de chacun de ces événements, on passe par le point (N, N) , mais dans le cas A , peu importe si l'on continue horizontalement ou verticalement; en revanche, dans le cas B , il faut continuer dans le sens opposé à l'arrivée. Il en résulte que $v = u_0/2$.

4. $\frac{b(k; n, p)}{b(k-1; n, p)} = 1 + \frac{(n+1)p - k}{kq}$. Le terme $b(k; n, p)$ est supérieur au précédent pour $k < (n+1)p$, inférieur pour $k > (n+1)p$. Si $(n+1)p = m$ est un entier, alors $b(m; n, p) = b(m-1; n, p)$. D'autre part, il existe un entier m et un seul vérifiant $(n+1)p - 1 < m \leq (n+1)p$.

5. Valeur maximum atteinte pour l'entier k défini par $\lambda - 1 < k \leq \lambda$.
6. $P = \binom{500}{3} \left(\frac{1}{365}\right)^3 \left(\frac{364}{365}\right)^{497}$. On a $n = 500$, $p = 1/365$, d'où $np = \lambda = 500/365$. En utilisant l'approximation de la loi de Poisson, on a : $P \approx e^{-\lambda} \lambda^3 / 3! \approx 0,11$.
7. Si a) est vraie, on a $p_n = e^{-\lambda} \lambda^n / n!$ pour tout $n \geq 0$, d'où b). Si b) est vraie, pour tout $n \geq 1$ on a : $\frac{p_n}{p_0} = \frac{p_1}{p_0} \frac{p_2}{p_1} \dots \frac{p_n}{p_{n-1}} = \frac{\lambda^n}{n!}$ et donc $p_n = p_0 \frac{\lambda^n}{n!}$, d'où $p_0 = e^{-\lambda}$, puisqu'il faut que $\sum_{n \geq 0} p_n = 1$.
8. On a : $\mathbb{E}[1/X] = \mathbb{E}[1/(1+(X-1))] = \mathbb{E}[\int_0^1 u^{X-1} du] = \int_0^1 \mathbb{E}[u^{X-1}] du = \int_0^1 (1/u) \mathbb{E}[u^X] du$. Or $\mathbb{E}[u^X] = G_X(u) = pu/(1-qu)$, d'où $\mathbb{E}[1/X] = \int_0^1 (p/(1-qu)) du = -(p/q) \text{Log } p$. Pour $p = \frac{1}{2}$, on trouve $\mathbb{E}[1/X] = \text{Log } 2$. [Noter que si X suit la loi géométrique $\sum_{k \geq 0} pq^k \varepsilon_k$ ($0 < p < 1$), alors $\mathbb{E}[1/X] = +\infty$; X partage cette propriété avec la variable exponentielle, dont elle est la version discrète.]
9. 2 a) La fonction génératrice de (N_1, N_2) est donnée par
- $$g(u, v) = \mathbb{E}[u^{N_1} v^{N_2}] = \sum_{n \geq 0} \mathbb{E}[u^{N_1} v^{N_2} | N = n] P\{N = n\}.$$
- Or
- $$\mathbb{E}[u^{N_1} v^{N_2} | N = n] = \sum_{k=0}^n u^k v^{n-k} \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^n \binom{n}{k} (pu)^k (qv)^{n-k} = (pu + qv)^n,$$
- d'où $g(u, v) = \sum_{n \geq 0} (pu + qv)^n P\{N = n\}$, c'est-à-dire en désignant par G la fonction génératrice de N , la relation $g(u, v) = G(pu + qv)$. Or si $\mathcal{L}(N) = \mathcal{P}(\lambda)$, on a $G(u) = e^{\lambda(u-1)}$, d'où $g(u, v) = e^{\lambda(pu+qv-1)} = e^{\lambda p(u-1)} e^{\lambda q(v-1)} = g(u, 1)g(1, v)$.
- 2 b) Supposons N_1, N_2 indépendantes; alors $g(u, v) = g(u, 1)g(1, v)$, c'est-à-dire d'après (1), en désignant par G la fonction génératrice de N , $G(pu + qv) = G(pu + q)G(p + qv)$. En posant $h(x) = G(1 - x)$, on a $h(pu + qv) = h(pu)h(qv)$, d'où $h(x + y) = h(x)h(y)$. C'est l'équation fonctionnelle de l'exponentielle; les seules solutions continues sont de la forme $h(x) = e^{-\lambda x}$ (λ réel); d'où $G(x) = h(1 - x) = e^{\lambda(x-1)}$. Comme G doit être une fonction génératrice, on doit avoir $\lambda > 0$ et N suit une loi de Poisson de paramètre λ .
10. 1) $P\{T_1 = n\} = P\{\varepsilon_1 = \dots = \varepsilon_{n-1} = 0; \varepsilon_n = 1\} = q^{n-1}p$ ($n \geq 1$).
- 2) Il s'agit de montrer que $P\{\tau_1 = i_1, \dots, \tau_n = i_n\} = \prod_{k=1}^n q^{i_k-1}p$, où $i_k \geq 1$ ($k = 1, \dots, n$). Pour le voir, on exprimera l'évènement $\{\tau_1 = i_1, \dots, \tau_n = i_n\}$ en fonction des ε_k : $\{\tau_1 = i_1, \dots, \tau_n = i_n\} = \{\varepsilon_1 = \dots = \varepsilon_{i_1-1} = 0, \varepsilon_{i_1} = 1; \varepsilon_{i_1+1} = \dots = \varepsilon_{i_1+i_2-1} = 0, \varepsilon_{i_1+i_2} = 1; \dots, \varepsilon_{i_1+\dots+i_{n-1}+1} = \dots = \varepsilon_{i_1+\dots+i_n-1} = 0, \varepsilon_{i_1+\dots+i_n} = 1\}$, d'où $P\{\tau_1 = i_1, \dots, \tau_n = i_n\} = q^{i_1-1}p \cdot q^{i_2-1}p \cdot \dots \cdot q^{i_n-1}p$.

- 3) Exprimons l'évènement $\{T_1 = t_1, \dots, T_n = t_n\}$ en fonction des ε_k :
 $\{T_1 = t_1, \dots, T_n = t_n\} = \{\tau_1 = t_1, \tau_2 = t_2 - t_1, \dots, \tau_n = t_n - t_{n-1}\}$,
d'où, en posant $t_0 = 0$, $P\{T_1 = t_1, \dots, T_n = t_n\} = \prod_{k=1}^n q^{t_k - t_{k-1} - 1} p =$
 $p^n q^{(t_1 - t_0 - 1) + \dots + (t_n - t_{n-1} - 1)} = p^n q^{t_n - n}$, si $0 < t_1 < \dots < t_n$, et 0 sinon.
- 4) D'abord, $P(A) = P\{N_m = n\} = P\{\sum_{k=1}^m \varepsilon_k = n\} = \binom{m}{n} p^n q^{m-n}$. Or
 $B \cap A = \{T_1 = t_1, \dots, T_n = t_n, N_m = n\} = \{T_1 = t_1, \dots, T_n = t_n,$
 $T_{n+1} > m\}$, d'où $P(B \cap A) = \sum_{k \geq m+1} P\{T_1 = t_1, \dots, T_n = t_n, T_{n+1} = k\}$,
une expression qui est nulle lorsque l'une au moins des inégalités $0 <$
 $t_1 < \dots < t_n \leq m$ n'est pas vérifiée, et qui lorsque toutes ces inégalités
sont vérifiées, vaut $\sum_{k \geq m+1} p^{n+1} q^{k-n-1} = \frac{p^{n+1} q^{m+1}}{q^{n+1} (1-q)} = p^n q^{m-n}$. Ainsi

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \begin{cases} \frac{1}{\binom{m}{n}}, & \text{si } 0 < t_1 < \dots < t_n \leq m; \\ 0, & \text{sinon.} \end{cases}$$

- 5) a) On commence par déterminer la loi conjointe de (U_n, V_n) ; à cet effet, on exprime l'évènement $\{U_n = i, V_n = j\}$ en fonction des ε_k . Pour $0 \leq i \leq n-1$ et $j \geq 1$ on a $\{U_n = i, V_n = j\} = \{\varepsilon_{n-i} = 1, \varepsilon_{n-i+1} = \dots = \varepsilon_n = 0; \varepsilon_{n+1} = \dots = \varepsilon_{n+j-1} = 0, \varepsilon_{n+j} = 1\}$ et pour $i = n$ et $j \geq 1$ on trouve $\{U_n = n, V_n = j\} = \{\varepsilon_1 = \dots = \varepsilon_n = 0; \varepsilon_{n+1} = \dots = \varepsilon_{n+j-1} = 0, \varepsilon_{n+j} = 1\}$. La loi conjointe de (U_n, V_n) est donnée pour $0 \leq i \leq n-1, j \geq 1$ par $P\{U_n = i, V_n = j\} = p^2 q^{i+j-1} = (pq^i)(pq^{j-1})$ et $P\{U_n = n, V_n = j\} = q^n pq^{j-1}$. On en déduit les lois marginales

$$P\{U_n = i\} = \sum_{j \geq 1} P\{U_n = i, V_n = j\} = \begin{cases} pq^i, & \text{si } 0 \leq i \leq n-1; \\ q^n, & \text{si } i = n; \end{cases}$$

$$P\{V_n = j\} = \sum_{i=0}^n P\{U_n = i, V_n = j\} = pq^{j-1} \quad (j \geq 1).$$

La loi conjointe est bien le produit des lois marginales. Les variables aléatoires U_n, V_n sont donc indépendantes. On remarquera, en outre, que U_n a même loi que $\inf(\tau_1 - 1, n)$ et que V_n a même loi que τ_1 (qui est indépendante de n).

- b) Il résulte de la remarque de 5a) que U_n converge en loi vers $\tau_1 - 1$, c'est-à-dire comme on peut le voir directement, $\lim_{n \rightarrow \infty} P\{U_n = i\} = pq^i$ ($i = 0, 1, \dots$).

11. Posons $G = \sum_{r=0}^N \binom{2N-r}{N} \left(\frac{1}{2}\right)^{2N-r}$. Alors, avec $N - r = i$, on a $G =$
 $\left(\frac{1}{2}\right)^N \sum_{i=0}^N \frac{\binom{N+1}{i}}{i!} \left(\frac{1}{2}\right)^i$. Or par l'identité binomiale

$(\frac{1}{2})^N \sum_{i=0}^{\infty} \frac{(N+1)_i}{i!} (\frac{1}{2})^i = (\frac{1}{2})^N (1 - \frac{1}{2})^{-(N+1)} = 2$. Il suffit de démontrer $(\frac{1}{2})^N \sum_{i=N+1}^{\infty} \frac{(N+1)_i}{i!} (\frac{1}{2})^i = 1$. Le membre de gauche se réécrit $(\frac{1}{2})^{2N+1} \frac{(N+1)_{N+1}}{(N+1)!} {}_2F_1(\frac{2N+2, 1}{N+2}; \frac{1}{2})$, qui vaut, d'après l'identité de Gauss, $(\frac{1}{2})^{2N+1} \frac{(N+1)_{N+1}}{(N+1)!} \frac{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2} + N + 1 + \frac{1}{2})}{\Gamma(\frac{1}{2} + N + 1)\Gamma(\frac{1}{2} + \frac{1}{2})} = 1$.

Chapitre 8

1. Loi Espérance Variance

$B(n, p)$	np	npq
π_λ	λ	λ

On constate que pour la loi de Poisson, l'espérance mathématique et la variance sont égales toutes les deux au paramètre.

2. Le nombre d'essais suit la loi uniforme, d'où son espérance $(n + 1)/2$.
3. b) $P\{L = n\} = p^n q + q^n p$ ($n \geq 1$) d'où $\mathbb{E}[L] = 2 + (p - q)^2/(pq) \geq 2$ et $\text{Var } L = 2 + (1 + pq)(p - q)^2/(p^2 q^2) \geq 2$. Pour déterminer la loi de M , on la considère comme loi marginale du couple (L, M) :

$$P\{L = l, M = n\} = p^l q^n p + q^l p^n q \quad (l, n \geq 1);$$

$$P\{M = n\} = \sum_{l \geq 1} P\{L = l, M = n\} = q^{n-1} p^2 + p^{n-1} q^2 \quad (n \geq 1).$$

d'où $\mathbb{E}[M] = 2$ (indépendant de p) et $\text{Var } M = 2 + 2(p - q)^2/(pq) \geq 2$.

- c) d) calculs sans difficultés. e) $\mathbb{E}[T] = q/p$.
- f) utiliser les relations : $k \binom{r}{k} = r \binom{r-1}{k-1}$ et $\binom{r+k-1}{k} = (-1)^k \binom{-r}{k}$.
4. Le nombre d'essais suit une loi géométrique de paramètre $p = 1/n$, d'où son espérance mathématique $1/p = n$.
9. a) On prend pour Ω l'ensemble des permutations des n chapeaux et pour P l'équipartition sur Ω .
- b) Soit A_k l'évènement « la $k^{\text{ième}}$ personne récupère son chapeau ». On a : $P(A_k) = (n - 1)!/n! = 1/n$ pour tout $k = 1, \dots, n$, puis $P(A_k \cap A_l) = (n - 2)!/n! = 1/(n(n - 1))$ pour $1 \leq k < l \leq n$, enfin $\mathbb{E}[X_k] = P(A_k) = 1/n$ et $\mathbb{E}[X_k X_l] = P(A_k A_l) = 1/(n(n - 1))$, d'où

$$\mathbb{E}[S_n] = \mathbb{E}[X_1] + \dots + \mathbb{E}[X_n] = 1;$$

$$\mathbb{E}[S_n^2] = \sum_{k=1}^n \mathbb{E}[X_k^2] + 2 \sum_{k < l} \mathbb{E}[X_k X_l] = n \frac{1}{n} + n(n - 1) \frac{1}{n(n - 1)} = 2.$$

$$\text{Var } S_n = \mathbb{E}[S_n^2] - (\mathbb{E}[S_n])^2 = 1,$$

toutes ces quantités étant indépendantes de n .

$$\begin{aligned} \text{c) } P\{S_n \geq 11\} &= P\{S_n - 1 \geq 10\} = P\{S_n - \mathbb{E}[S_n] \geq 10\} \\ &\leq P\{|S_n - \mathbb{E}[S_n]| \geq 10\} \leq \frac{\text{Var } S_n}{10^2} = \frac{1}{10^2}. \end{aligned}$$

10. a) $X+Z = 1-Y$, d'où $\text{Var}(X+Z) = \text{Var } X + \text{Var } Z + 2 \text{Cov}(X, Z) = \text{Var } Y$,
 $2 \text{Cov}(Z, X) = \text{Var } X - (\text{Var } Z - \text{Var } Y) \leq 0$. Même calcul pour $\text{Cov}(Y, Z)$.

b) $X+Y = 1-Z$, d'où $\text{Var}(X+Y) = \text{Var } X + \text{Var } Y + 2 \text{Cov}(X, Y) = \text{Var } Z$
 et $2 \text{Cov}(X, Y) = \text{Var } Z - (\text{Var } X + \text{Var } Y)$.

$$\begin{aligned} \text{c) } 2 \text{Cov}(X, Z) &= -[(\text{Var } Z - \text{Var } Y) + \text{Var } X] \leq 0; \\ 2 \text{Cov}(Y, Z) &= -[(\text{Var } Z - \text{Var } X) + \text{Var } Y] \leq 0; \\ 2 |\text{Cov}(X, Z)| &= \text{Var } Z - (\text{Var } Y - \text{Var } X) \leq \text{Var } Z; \\ 2 |\text{Cov}(Y, Z)| &= \text{Var } Z + (\text{Var } Y - \text{Var } X) \geq \text{Var } Z; \\ \text{d'où } |\text{Cov}(X, Z)| &\leq |\text{Cov}(Y, Z)|. \end{aligned}$$

11. $\{10 - n < X < 10 + n\} = \{|X - 10|/5 < n/5\}$. Soit n tel que
 $-0,99 \leq P\{|X - 10|/5 < n/5\}$, c'est-à-dire tel que

$$P\{|X - 10|/5 \geq n/5\} \leq 0,01.$$

Or d'après l'inégalité de Tchebychev :

$$P\{|X - 10|/5 \geq 5\} \leq 1/(n/5)^2 = 25/n^2.$$

L'avant-dernière inégalité est donc vérifiée dès que $25/n^2 \leq 0,01$, soit
 $n \geq 50$.

12. Pour tout $t > 0$, on a, d'après l'inégalité de Markov

$$P\{|X| \geq t\} \leq \frac{\mathbb{E}[|X|]}{t} = 0.$$

Or $\{|X| > 0\} = \bigcup_{n \geq 1} \{|X| \geq 1/n\}$, d'où

$$P\{|X| > 0\} \leq \sum_{n \geq 1} P\left\{|X| \geq \frac{1}{n}\right\} = 0.$$

14. D'après les inégalités sur les moyennes, on a : $(\mathbb{E}[1/X])^{-1} \leq \mathbb{E}[X]$.

15. Soit $P_X = \sum_k \alpha_k \varepsilon_{x_k}$ la loi de X . On a, pour tout $n \geq 1$

$$\sum_{|x_k| \geq n} |x_k|^r \alpha_k \geq n^r P\{|X| \geq n\}.$$

Or, si $\mathbb{E}[|X|^r] \leq +\infty$, alors $\lim_n \sum_{|x_k| \geq n} |x_k|^r \alpha_k \rightarrow 0$.

16. La covariance étant invariante par changement d'origine, il suffit de
 supposer que X_1, X_2, Y_1, Y_2 sont centrées; on a alors :

$$\begin{aligned} \text{Cov}(X_1 + Y_1, X_2 + Y_2) &= \mathbb{E}[(X_1 + Y_1)(X_2 + Y_2)] = \mathbb{E}[X_1 X_2] + \mathbb{E}[X_1 Y_2] + \\ &\mathbb{E}[Y_1 X_2] + \mathbb{E}[Y_1 Y_2] = \mathbb{E}[X_1 X_2] + \mathbb{E}[X_1] \mathbb{E}[Y_2] + \mathbb{E}[Y_1] \mathbb{E}[X_2] + \mathbb{E}[Y_1 Y_2] = \\ &\text{Cov}(X_1, X_2) + \text{Cov}(Y_1, Y_2). \end{aligned}$$

17. D'abord I_A et I_B sont indépendantes ssi pour tout $\varepsilon = 0$ ou 1 et tout $\varepsilon' = 0$ ou 1 on a $P\{I_A = \varepsilon, I_B = \varepsilon'\} = P\{I_A = \varepsilon\}P\{I_B = \varepsilon'\}$. Cette condition est équivalente à $P(AB) = P(A)P(B)$ (car les trois autres relations écrites avec A^c et B^c en découlent). Comme enfin $\mathbb{E}[I_A] = P(A)$, $\mathbb{E}[I_B] = P(B)$, $\mathbb{E}[I_A I_B] = P(AB)$, cette condition est équivalente à $\mathbb{E}[I_A I_B] = \mathbb{E}[I_A]\mathbb{E}[I_B]$.
18. On peut, sans nuire à la généralité, supposer X et Y centrées; alors $X + Y$ et $X - Y$ le sont aussi et l'on a : $\text{Cov}(X + Y, X - Y) = \mathbb{E}[(X + Y)(X - Y)] = \mathbb{E}[X^2 - Y^2] = \mathbb{E}[X^2] - \mathbb{E}[Y^2] = \text{Var } X - \text{Var } Y = 0$. [Si (X, Y) est un couple de variables aléatoires indépendantes, normales, centrées, réduites, le couple $(X + Y, X - Y)$ n'est pas seulement non corrélé, mais encore indépendant.]

Chapitre 9

1. $(ps + q)^n; \exp(\lambda(s - 1)); ps(1 - qs)^{-1}$.
2. La fonction génératrice de X vaut $G_X(u) = \frac{pu}{1 - qu}$ ($u \neq p/q$). Puisque q est dans l'intervalle $]0, 1[$, l'intervalle ouvert sur lequel G_X est définie contient le point $u = 1$. D'après la Proposition 2.5, tous les moments factoriels de X existent; le moment factoriel d'ordre r est égal au coefficient de $v^r/r!$ dans le développement de $G_X(1 + v)$ au voisinage de $v = 0$. Or $G_X(u) = -\frac{p}{q} + \frac{p}{q} \frac{1}{1 - qu}$ et $G_X(1 + v) = -\frac{1}{q} + \frac{1}{q} \frac{1}{1 - \frac{q}{p}v}$. Pour $-p/q < v < p/q$ on a le développement
- $$G_X(1 + v) = 1 + \sum_{r \geq 1} \frac{1}{q} \left(\frac{q}{p}\right)^r r! \cdot \frac{v^r}{r!}.$$
- D'où, pour tout $r \geq 1$, $\mathbb{E}[X(X - 1) \cdots (X - r + 1)] = \frac{1}{q} \left(\frac{q}{p}\right)^r r!$ Lorsque $p = q = \frac{1}{2}$, ce dernier moment factoriel vaut $2r!$
- On peut également trouver l'expression de la $r^{\text{ième}}$ dérivée de G_X et appliquer la Proposition 2.4.
5. b) Exprimant chaque coefficient binomial en fonction des factorielles montantes, on obtient :

$$G_H(s) = \frac{1}{(-N)_n} \sum_k (-M)_k (-n)_k (-1)^k (-N + M)_{n-k} \frac{s^k}{k!}.$$

Pour $n \leq N - M$ on a $(-N + M)_n = (-N + M)_{n-k} (-1)^k (N - M - n + 1)_k$ et on obtient l'expression demandée. Pour $n \geq N - M + 1$, on prend l'indice de sommation $l = k - (n - (N - M))$ qui varie dans l'intervalle $[0, \min\{(N - M), (N - n)\}]$. Remarquer que l'on a :

$$\frac{(n - N + M + 1)_{N-M}}{(M + 1)_{N-M}} = \frac{(n - N + M + 1)_{N-n}}{(n + 1)_{N-n}}.$$

- d) Utiliser l'identité de Chu-Vandermonde.
6. $s^b G(s); G(s^a)$.

7. a) $G(s)/(1-s)$; b) $sG(s)/(1-s)$; c) $(1-sG(s))/(1-s)$;
 d) $P\{X=0\}/s + (1-G(s)/s)/(1-s)$; e) $(G(s^{1/2}) + G(-s^{1/2}))/2$.
8. a) $P\{T_1 = m\} = \left(\frac{1}{3}\right)^{m-1} \left(\frac{2}{3}\right)$ ($m \geq 1$); $G_{T_1}(s) = 2s/((3-s))$;
 $\mathbb{E}[T_1] = \frac{3}{2}$; $\text{Var } T_1 = \frac{3}{4}$.
 b) $\left(\frac{1}{2}\right)^{n-m-1} \left(\frac{1}{3}\right)^m$.
 c) $P\{T_2 = n\} = 2\left[\left(\frac{1}{2}\right)^{n-1} - \left(\frac{1}{3}\right)^{n-1}\right]$ ($n \geq 2$);
 $G_{T_2}(s) = 2s^2/(6-5s+s^2)$; $\mathbb{E}[T_2] = 7/2$; $\text{Var } T_2 = 11/4$.
 d) $P\{X_n = 0\} = P\{T_1 = n\} + P\{T_2 = n\} = 4\left[\left(\frac{1}{2}\right)^n - \left(\frac{1}{3}\right)^n\right]$ ($n \geq 1$), ce qui détermine la loi de X_n , puisque cette loi est de Bernoulli.
9. Utiliser la Proposition 3.2.
10. a) Résulte de la Proposition 3.2.
 c) On a $x_{n+1} = G(x_n)$ pour $n \geq 1$ et G est continue.
 d) Noter que $G'(1) = \mu$. e) Considérer les cas $\mu \leq 1$ et $\mu \geq 1$.
 f) $\mathbb{E}[X_n] = \mu^n$ ($n \geq 2$); $\text{Var } X_n = \sigma^2 \mu^{n-1} (1 - \mu^n)/(1 - \mu)$ ($n \geq 1$).
 g) $G_n(s) = 1 - p^n + p^n s$.
11. $t_n = 1 - \left(1 - \frac{1}{2^n}\right)^a$; $H(s) = \sum_{j=1}^a \binom{a}{j} \left(1 - \frac{s}{2^j}\right)^{-1}$; $\mathbb{E}[T] = H(1) = \sum_{j=1}^a \binom{a}{j} (-1)^{j-1} \left(1 - \frac{1}{2^j}\right)^{-1}$.
12. a) On a $U(s) = \frac{P(s)}{(s-s_1)^{r_1} \dots (s-s_m)^{r_m}} = \sum_{1 \leq i \leq m} \sum_{1 \leq j \leq r_i} \frac{a_{ij}}{(s-s_i)^j}$. (1)
 En multipliant par $(s-s_1)^{r_1}$, puis en faisant $s = s_1$, il vient $a_{1,r_1} = \frac{P(s_1)}{(s_1-s_2)^{r_2} \dots (s_1-s_m)^{r_m}}$. Or $Q(s) = (s-s_1)^{r_1} \dots (s-s_m)^{r_m}$ et $Q^{(r_1)}(s_1) = r_1! (s_1-s_2)^{r_2} \dots (s_1-s_m)^{r_m}$, d'où $a_{1,r_1} = r_1! \frac{P(s_1)}{Q^{(r_1)}(s_1)}$.
 On obtient une formule analogue pour chaque a_{i,r_i} .
 b) Pour $|s| < |s_i|$ la formule du binôme donne :

$$\frac{1}{(s-s_i)^j} = (-1)^j s_i^{-j} \left(1 - \frac{s}{s_i}\right)^{-j} = (-1)^j s_i^{-j} \sum_{n \geq 0} \binom{j}{s_i} \frac{(j)_n}{n!}. \quad (2)$$
 En substituant dans (1) on obtient la formule demandée.
 c) Soit $|s_1| < |s_i|$ pour $i = 2, \dots, m$, la racine s_1 , étant simple, est nécessairement réelle. On voit sans peine que le terme prépondérant dans le membre de droite de (2) est celui correspondant à $i = 1, j = r_1$. (On remarquera que si $1 \leq j < r$, alors $(j)_n = o((r)_n)$ lorsque n tend vers l'infini.)
 d) Supposons que le degré de P soit $m+r$ ($r \geq 0$). Une division permet de représenter $U(s)$ comme somme d'un polynôme de degré r et d'une fraction rationnelle $P_1(s)/Q(s)$, où le degré de P_1 est strictement inférieur au degré de Q . Le polynôme n'affecte que les $r+1$ premiers termes de la suite (u_n) et la fraction rationnelle $P_1(s)/Q(s)$ peut être décomposée en éléments simples comme précédemment. Ainsi le résultat de c) subsiste.

13. a) Pour qu'au cours de n jets, avec $n \geq 3$, on n'amène aucun triplet FFF , il faut en particulier que les trois premiers jets n'amènent pas ce triplet. Il faut donc que la suite des n jets commence par P , ou FP , ou FFP ; ce sont des évènements de probabilité $1/2$, $1/4$, $1/8$, respectivement. Conditionnellement à l'un de ces évènements, la probabilité pour que les jets suivants n'amènent aucun triplet FFF est u_{n-1} , u_{n-2} , u_{n-3} , respectivement; d'où le résultat.

b)
$$U(s) - 1 - s - s^2 = \sum_{n \geq 3} u_n s^n = \frac{1}{2} \sum_{n \geq 3} u_{n-1} s^n + \frac{1}{4} \sum_{n \geq 3} u_{n-2} s^n + \frac{1}{8} \sum_{n \geq 3} u_{n-3} s^n = \frac{s}{2} (U(s) - 1 - s) + \frac{s^2}{4} (U(s) - 1) + \frac{s^3}{8} U(s).$$
 Le résultat en découle en résolvant par rapport à $U(s)$.

d) D'après la partie c) du problème 12, on a $u_n \sim -as_1^{-(n+1)}$, où $a = \frac{P(s_1)}{Q'(s_1)} = \frac{2s_1^2 + 4s_1 + 8}{4 + 4s_1 + 3s_1^2} = 1,236\dots$, d'où $u_n \sim \frac{1,236\dots}{(1,087\dots)^{n+1}}$ lorsque n tend vers l'infini.

15. Non. Supposons, en effet, le premier dé pipé selon (p_1, \dots, p_6) et le deuxième selon (q_1, \dots, q_6) ($p_i \geq 0$; $p_1 + \dots + p_6 = 1$) et ($q_i \geq 0$; $q_1 + \dots + q_6 = 1$). Désignons par X_1 (resp. X_2) le point amené par le premier (resp. le second) dé. Leurs fonctions génératrices sont données par : $G_{X_1}(s) = p_1 s + \dots + p_6 s^6 = s(p_1 + p_2 s + \dots + p_6 s^5) = s P_1(s)$ et $G_{X_2}(s) = q_1 s + \dots + q_6 s^6 = s P_2(s)$, où P_1, P_2 sont des polynômes de degré 5 en s , à coefficients réels. Par hypothèse, on a $G_{X_1+X_2}(s) = \frac{1}{11}(s^2 + \dots + s^{12}) = \frac{s^2}{11}(1 + s + \dots + s^{10})$. Si le problème était possible, on aurait : $G_{X_1+X_2} = G_{X_1} G_{X_2}$, c'est-à-dire

$$P_1(s) P_2(s) = \frac{1}{11}(1 + s + \dots + s^{10}) = \frac{1}{11} \frac{1 - s^{11}}{1 - s} = Q(s).$$

Or Q est un polynôme à coefficients réels admettant dix zéros complexes (non réels) conjugués deux à deux, alors que P_1 et P_2 , polynômes de degré 5, à coefficients réels, admettent chacun un zéro réel; d'où une contradiction.

16. $\alpha_k = \sum \frac{n!}{n_1! \dots n_6!}$, où la sommation est étendue à l'ensemble des suites (n_1, \dots, n_6) vérifiant $n_1 \geq 0, \dots, n_6 \geq 0$ et $n_1 + \dots + n_6 = k$.

17. λ^r .

18. a) $G_{S_r}(s) = \left(\frac{ps}{1 - qs} \right)^r$ ($|s| < 1, q = 1 - p$).

b) $G_{S_r}(s) = (ps)^r (1 - qs)^{-r} = (ps)^r \sum_{k \geq 0} \binom{-r}{k} (-qs)^k = \sum_{k \geq 0} \binom{-r}{k} p^k (-q)^r s^{r+k}$

d'où la loi de probabilité de S_r donnée par : $P\{S_r = r + k\} = \binom{-r}{k} p^k (-q)^r$ ($k \geq 0$). On voit que la loi $\Pi(r, p)$ de S_r est une loi binomiale négative de paramètres r, p .

c) Évident à partir de a).

19. a) On doit avoir $1 = k \sum_{n \geq 1} \frac{\theta^n}{n} = k \operatorname{Log} \frac{1}{1-\theta}$, d'où $k = -\frac{1}{\operatorname{Log}(1-\theta)}$.
- b) Pour $-1/\theta < u < 1/\theta$ on a : $G_X(u) = k \sum_{n \geq 1} \frac{(\theta u)^n}{n} = \frac{\operatorname{Log}(1-\theta u)}{\operatorname{Log}(1-\theta)}$.
- c) Comme $\theta \in]0, 1[$, la fonction G_X est définie dans un voisinage autour du point 1. Comme $G'_X(u) = k\theta/(1-\theta u)$ et $G''_X(u) = k\theta^2/(1-\theta u)^2$, on a par la Proposition 2.4 : $\mathbb{E}[X] = G'_X(1) = k\theta/(1-\theta)$ et $\operatorname{Var} X = G''_X(1) + G'_X(1) - (G'_X(1))^2 = k\theta(1-k\theta)/(1-\theta)^2$.
20. Désignons par C_1, \dots, C_N les champignons cueillis et notons $\{C_k = c\}$ l'évènement consistant en le fait que le $k^{\text{ième}}$ champignon cueilli est comestible. La probabilité pour que *tous* les champignons soient comestibles est :

$$\mathbb{P}\{C_1 = c, \dots, C_N = c\} = \sum_{n \geq 1} \mathbb{P}\{N = n\} \mathbb{P}\{C_1 = c, \dots, C_n = c \mid N = n\}.$$

En supposant les évènements $\{C_1 = c\}, \{C_2 = c\}, \dots$ indépendants entre eux et indépendants de N , on obtient : $\mathbb{P}\{C_1 = c, \dots, C_N = c\} = \sum_{n \geq 1} \mathbb{P}\{N = n\} (\mathbb{P}\{C_1 = c\})^n = \sum_{n \geq 1} \mathbb{P}\{N = n\} p^n = G(p)$.

Chapitre 10

1. a) L'ensemble \mathfrak{A}_n n'est autre que la tribu engendrée par π_n , c'est-à-dire l'image réciproque $\pi_n^{-1}(\mathfrak{B}(S^n))$. C'est donc une tribu.
- b) Pour $A \subset S^n$, on a $\pi_n^{-1}(A) = \pi_{n+1}^{-1}(A \times S)$.
- c) L'ensemble fondamental Ω appartient à \mathfrak{A}_1 , donc à \mathfrak{A} . Si C appartient à \mathfrak{A} , il appartient à \mathfrak{A}_n pour un certain $n \geq 1$; le complémentaire C^c appartient aussi à \mathfrak{A}_n , donc aussi à \mathfrak{A} . Enfin, si C et D sont des éléments de \mathfrak{A} , on peut supposer que C est un n -cylindre et D un m -cylindre et que $n \leq m$. Comme la suite (\mathfrak{A}_n) est croissante, les deux cylindres C et D appartiennent à \mathfrak{A}_m , ainsi que leur réunion, puisque \mathfrak{A}_m est une tribu. Par suite, la réunion appartient à \mathfrak{A} . Les axiomes sur les propriétés d'une algèbre sont donc vérifiés. En revanche, \mathfrak{A} n'est *pas* une tribu. Prenons, en effet, un élément fixé (x_1, x_2, \dots) de Ω . Pour tout $n \geq 1$, l'ensemble $C_n = \{\pi_n = (x_1, x_2, \dots, x_n)\}$ est un n -cylindre. Or l'intersection $\bigcap_{n \geq 1} C_n$ est simplement le singleton $\{\omega\}$ qui, lui, n'est pas un n -cylindre, quel que soit n .
2. D'abord \mathfrak{A}_n est la tribu engendrée par π_n . C'est donc la plus petite tribu rendant π_n mesurable. Toute tribu rendant *tous* les π_n mesurables doit donc contenir la réunion des \mathfrak{A}_n , c'est-à-dire \mathfrak{A} , et par conséquent aussi $\mathfrak{T} = \mathfrak{T}(\mathfrak{A})$. Chaque X_n est mesurable, car pour tout $T \subset S$, l'ensemble $X_n^{-1}(T)$, qu'on peut écrire $\pi_n^{-1}(S^{n-1} \times T)$, appartient à \mathfrak{A}_n et donc aussi à \mathfrak{T} . Enfin, tout cylindre $C = \{\pi_n \in A\}$ peut s'exprimer comme la réunion de tous les évènements de la forme $\{X_1 = x_1\} \cap \{X_2 = x_2\} \cap \dots \cap \{X_n = x_n\}$, où la suite (x_1, x_2, \dots, x_n)

parcourt l'ensemble (fini) A . Par conséquent, toute tribu qui rend tous les X_n mesurables contient nécessairement tous les cylindres, et donc aussi la tribu \mathfrak{F} .

3. Pour simplifier, notons la formule en question $\sum_A p_n(x_1, \dots, x_n)$. Soit $C = \{\pi_n \in A\} = \{\pi_m \in B\}$. Il s'agit de démontrer : $P\{\pi_n \in A\} = P\{\pi_m \in B\}$. En effet, l'égalité $\sum_A p_n(x_1, \dots, x_n) = \sum_B p_m(x_1, \dots, x_m)$ entraîne $A = B$, lorsque $n = m$, car π_n est surjectif. Si $m < n$, alors $\{\pi_n \in A\} = \{\pi_m \in B\}$ entraîne $A = \pi_n(\pi_m^{-1}(B)) = B \times S^{n-m}$. D'où $\sum_B p_m(x_1, \dots, x_m) = \sum_{B \times S} p_{m+1}(x_1, \dots, x_m, x_{m+1}) = \dots = \sum_{B \times S^{n-m}} p_n(x_1, \dots, x_m, \dots, x_n)$.
4. Si l'on se fixe n , la suite $\pi_n(C_m)$ ($m \geq 1$) est une suite décroissante de parties non vides de l'ensemble fini S^n . Il existe donc un indice $m(n)$ tel que $\pi_n(C_k) = \pi_n(C_{m(n)})$ pour tout $k \geq m(n)$. On a donc : $\bigcap_{m \geq 1} \pi_n(C_m) = \pi_n(C_{m(n)}) \neq \emptyset$. L'ensemble $\pi_1(C_{m(1)})$ est non vide. Il contient donc un élément s_1 . Supposons déjà obtenue la suite (s_1, \dots, s_n) de S^n telle que pour tout $k \leq n$ la sous-suite (s_1, \dots, s_k) appartienne à $\pi_k(C_{m(k)})$. En particulier, (s_1, \dots, s_n) est dans $\pi_n(C_{m(n)})$, donc aussi dans $\pi_n(C_{m(n+1)})$. On construit donc ainsi, par récurrence, un élément $\omega = (s_1, s_2, \dots)$ de Ω qui est tel que chaque suite initiale (s_1, s_2, \dots, s_n) appartient à $\bigcap_{m \geq 1} \pi_n(C_m)$ pour tout n . Prenons un entier $m \geq 1$. Alors C_m est un l -cylindre pour un certain entier l . Il en résulte que la suite (s_1, s_2, \dots, s_l) appartient à $\pi_l(C_m)$. La suite infinie ω est donc dans C_m . Comme ceci est vrai quel que soit m , il en résulte que l'intersection $\bigcap_m C_m$ est non vide.
5. D'abord P est évidemment positive, puisque les p_n le sont. Ensuite, en prenant $\Omega = \{\pi_1 \in S\}$, on a : $P(\Omega) = \sum_{x \in S} p_1(x) = 1$. Soient maintenant deux éléments C, D de l'algèbre \mathfrak{A} tels que $C \cap D = \emptyset$. Il existe alors un entier n tel que $A \subset S^n, B \subset S^n$ et $C = \pi_n^{-1}(A), D = \pi_n^{-1}(B)$. De là, $\pi_n^{-1}(A \cup B) = \pi_n^{-1}(A) \cup \pi_n^{-1}(B) = C \cup D$ et $\emptyset = \pi_n(C \cap D) = \pi_n(\pi_n^{-1}(A) \cap \pi_n^{-1}(B)) = \pi_n \pi_n^{-1}(A \cap B) = A \cap B$. On en tire $P(C \cup D) = \sum_{A \cup B} p_n(x_1, \dots, x_n) = \sum_A p_n(x_1, \dots, x_n) + \sum_B p_n(x_1, \dots, x_n) = P(C) + P(D)$. L'additivité de P est bien vérifiée. Pour démontrer la σ -additivité de P , on fait appel à l'exercice précédent. Si (C_m) est une suite décroissante de cylindres tendant vers \emptyset , alors il existe un entier m tel que $C_m = \emptyset$. Par conséquent, $C_m = C_{m+1} = \dots = \emptyset$. D'où $\lim_k P(C_k) = 0$, ce qui établit bien la σ -additivité.
6. En effet, toutes les conditions du théorème de prolongement d'une mesure de probabilité sur une algèbre sont réalisées. On peut donc prolonger la mesure à toute la tribu engendrée.
7. Prenons pour p_n la fonction $p_n(x_1, \dots, x_n) = p(x_1) \dots p(x_n)$. Les conditions (i), (ii) et (iii) de l'exercice 3 sont trivialement vérifiées. Le résultat de l'exercice 6 affirme qu'il existe une et une seule mesure

de probabilité P sur (Ω, \mathfrak{F}) telle que $P\{X_1 = x_1, \dots, X_n = x_n\} = p(x_1) \dots p(x_n)$. De plus, $P\{X_n = x_n\} = \sum P\{X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = x_n\}$, où la somme est sur l'ensemble de toutes les suites $(x_1, \dots, x_{n-1}) \in S^{n-1}$, soit $P\{X_n = x_n\} = \sum p(x_1) \dots p(x_{n-1})p(x_n)$, somme calculée sur le même ensemble. De là, $P\{X_n = x_n\} = p(x_n) \prod_{1 \leq i \leq n-1} (\sum_{x_i \in S} p(x_i)) = p(x_n) \cdot 1 = p(x_n)$. Ainsi la condition (ii) du présent exercice est bien vérifiée. La condition (i) l'est aussi, puisque $P\{X_1 = x_1, \dots, X_n = x_n\} = p(x_1) \dots p(x_n) = P\{X_1 = x_1\} \dots P\{X_n = x_n\}$.

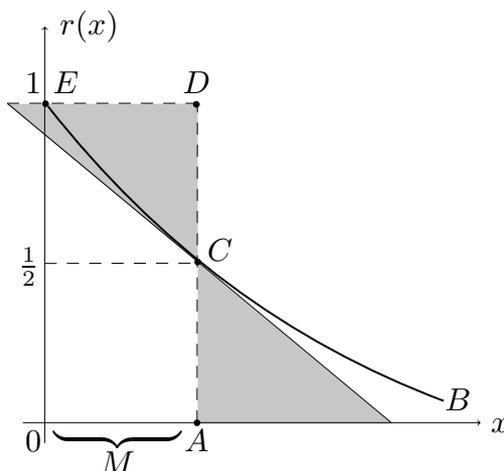
8. La fonction p_1 est donnée et pour $n \geq 2$ posons $p_n(x_1, \dots, x_n) = p_1(x_1)q_2(x_1, x_2) \dots q_n(x_1, \dots, x_n)$, quantité qui est égale à $\prod_{2 \leq i \leq n} P\{X_i = x_i | X_{i-1} = x_{i-1}, \dots, X_1 = x_1\}P\{X_1 = x_1\}$, par conséquent à $P\{X_1 = x_1, \dots, X_n = x_n\}$, si cette mesure de probabilité P existe. Or les trois conditions (i), (ii) et (iii) de l'exercice 3 sont trivialement vérifiées pour cette suite des (p_n) juste définie. Il existe donc une et une seule mesure de probabilité P sur (Ω, \mathfrak{F}) telle que $P\{X_1 = x_1, \dots, X_n = x_n\} = p_1(x_1)q_2(x_1, x_2) \dots q_n(x_1, \dots, x_n)$. Cette loi convient, car d'abord $P\{X_1 = x_1\} = p_1(x_1)$ par définition; ensuite
- $$\begin{aligned} P\{X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_1 = x_1\} \\ = \frac{P\{X_1 = x_1, \dots, X_n = x_n\}}{P\{X_1 = x_1, \dots, X_{n-1} = x_{n-1}\}} = \frac{p_1 q_2 \dots q_n}{p_1 q_2 \dots q_{n-1}} = q_n. \end{aligned}$$
9. C'est un cas particulier (important) de l'exercice précédent avec $q_n(x_1, \dots, x_n) = p_{x_{n-1}, x_n}$ et $p_1(x) = p_x$.

Chapitre 11

1. a) Notons les variables X et Y au lieu de X_1 et X_2 et formons les évènements $A_{n,j} = \{(j-1)/2^n \leq X < j/2^n\}$ et $B_{n,k} = \{(k-1)/2^n \leq X < k/2^n\}$ ($j, k = 1, 2, \dots, n2^n$), ainsi que $A_{n, n2^n+1} = \{n \leq X\}$ et $B_{n, n2^n+1} = \{n \leq Y\}$. Puisque X et Y sont supposés être indépendantes, chaque $A_{n,j}$ est indépendant de chaque $B_{n,k}$ pour tout $j, k = 1, 2, \dots, n2^n + 1$. Formons les variables aléatoires simples positives $X_n = \sum_{j=1}^{n2^n+1} ((j-1)/2^n) I_{A_{n,j}}$ et $Y_n = \sum_{k=1}^{n2^n+1} ((k-1)/2^n) I_{B_{n,k}}$. Pour tout couple (j, k) tel que $1 \leq j, k \leq n2^n + 1$, on a $P\{X_n = (j-1)/2^n, Y_n = (k-1)/2^n\} = P\{X_n = (j-1)/2^n\}P\{Y_n = (k-1)/2^n\} = P(A_{n,j}B_{n,k}) = P(A_{n,j})P(B_{n,k}) = P\{X_n = (j-1)/2^n\}P\{Y_n = (k-1)/2^n\}$. Ainsi X_n et Y_n sont indépendantes, d'où $\mathbb{E}[X_n Y_n] = \mathbb{E}[X_n]\mathbb{E}[Y_n]$. D'autre part, on a aussi $XY = \sup_n X_n Y_n$ et comme $(X_n Y_n)$ est une suite croissante de variables aléatoires simples positives, on conclut que $\mathbb{E}[XY] = \sup_n \mathbb{E}[X_n Y_n] = \sup_n \mathbb{E}[X_n]\mathbb{E}[Y_n] = \mathbb{E}[X]\mathbb{E}[Y]$.
- b) D'après la Proposition 6.2 du chap. 6, les variables X^+ et X^- sont indépendantes des variables Y^+ et Y^- . Ces quatre variables sont par

ailleurs positives et à espérance mathématique finie. D'après a) on en tire : $\mathbb{E}[X]\mathbb{E}[Y] = (\mathbb{E}[X^+] - \mathbb{E}[X^-])(\mathbb{E}[Y^+] - \mathbb{E}[Y^-]) = \mathbb{E}[X^+Y^+] - \mathbb{E}[X^+Y^-] - \mathbb{E}[X^-Y^+] + \mathbb{E}[X^-Y^-] = \mathbb{E}[(X^+ - X^-)(Y^+ - Y^-)] = \mathbb{E}[XY]$.

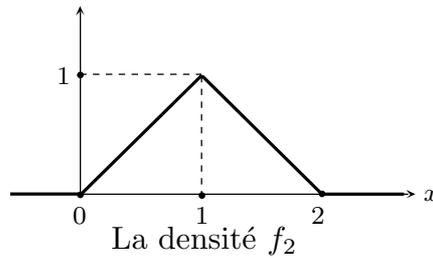
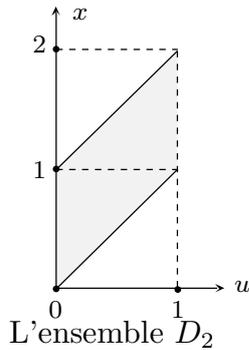
2. Par définition de la densité (voir la définition précédant la Proposition 4.2), pour tout ensemble borélien B , on a : $\mathbb{P}\{Y \in B\} = \mathbb{P}\{h \circ X \in B\} = \mathbb{P}_X\{h \in B\} = \int_{h^{-1}(B)} f_X(x) dx = \int_{S_X \cap h^{-1}(B)} f_X(x) dx$. Si $B \cap h(S_X) = \emptyset$, alors $S_X \cap h^{-1}(B) = \emptyset$ et donc $\mathbb{P}\{Y \in B\} = 0$. Ainsi pour tout borélien B on a : $\mathbb{P}\{Y \in B\} = \mathbb{P}\{Y \in B \cap h(S_X)\}$; ce qui prouve que le support de Y est contenu dans $h(S_X)$. Enfin, en appliquant la formule ci-dessus à $B = \{y\}$, on obtient : $\pi_Y(y) = \mathbb{P}\{Y = y\} = \int_{h^{-1}(\{y\})} f_X(x) dx$.
3. Pour la solution, voir la démonstration du Théorème 1.1 du chapitre 15 sur les changements de variables.
4. Il suffit de faire la démonstration pour $\mathbb{E}[X] < +\infty$. Il est clair que $f(x) \downarrow 0$, lorsque $x \rightarrow +\infty$ et que la fonction $r(x) = \mathbb{P}\{X > x\}$ est, pour $x \geq 0$, continûment dérivable, strictement décroissante et *convexe*. D'autre part, $\mathbb{E}[X] = \int_0^{+\infty} r(x) dx$ et M est l'unique nombre vérifiant $r(M) = 1/2$. C'est aussi l'aire du rectangle $(OADE)$. Menons en C la tangente au graphe de $r(\cdot)$; les triangles hachurés ont même aire. Il en résulte que l'aire curviligne (ABC) est supérieure à l'aire curviligne (CDE) . En ajoutant à chacune de ces aires curvilignes l'aire curviligne $(OACE)$, il vient $\mathbb{E}[X] = \text{aire curviligne } (OBE) > \text{aire du rectangle } (OADE) = M$.



5. a) La densité commune aux trois variables aléatoires est $f(x) = I_{[0,1]}(x)$. La densité de $X_1 + X_2$ est le produit de convolution de f par elle-même : $f_2(x) = (f * f)(x) = \int_{-\infty}^{\infty} f(u)f(x - u) du$. La fonction à intégrer est strictement positive sur $D_2 = \{(u, x) : 0 < u < 1, 0 < x - u < 1\} = \{(u, x) : \max(0, x - 1) < u < \min(1, x)\}$. On

$$\text{a donc : } f_2(x) = \begin{cases} \int_0^x du = x, & \text{pour } 0 \leq x \leq 1; \\ \int_{x-1}^1 du = 1 - (x-1) = 2-x, & \text{pour } 1 \leq x \leq 2. \end{cases}$$

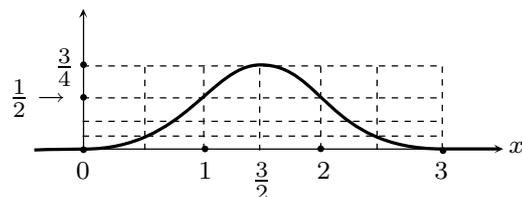
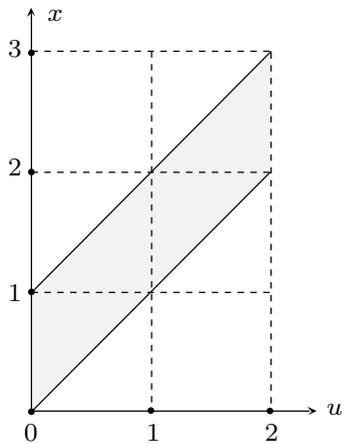
$$\text{D'où finalement } f_2(x) = \begin{cases} 1 - |1-x|, & \text{si } 0 \leq x \leq 2, \\ 0, & \text{sinon.} \end{cases}$$



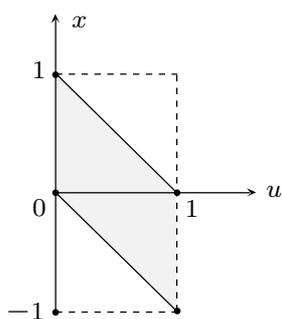
- b) La densité de $X_1 + X_2 + X_3$ est donnée par $f_3(x) = (f_2 * f)(x) = \int_{-\infty}^{\infty} f_2(u)f(x-u)du$. La fonction à intégrer est strictement positive sur $D_3 = \{(u, x) : 0 < u < 2, 0 < x-u < 1\} = \{(u, x) : \max(0, x-1) < u < \min(2, x)\}$. Trois cas à considérer : (1) pour $0 \leq x \leq 1$, on a $f_3(x) = \int_0^x u du = x^2/2$; (2) pour $1 \leq x \leq 2$, on a $f_3(x) = \int_{x-1}^x f_2(u)f(x-u)du = \int_{x-1}^1 u du + \int_1^x (2-u) du = \frac{1}{2}(1 - (x-1)^2) + 2(x-1) - \frac{1}{2}(x^2 - 1) = -(x - \frac{3}{2})^2 + \frac{3}{4}$; (3) pour $2 \leq x \leq 3$, on a : $f_3(x) = \int_{x-1}^2 (2-u) du = \int_0^{3-x} v dv = \frac{1}{2}(x-3)^2$.

$$\text{D'où finalement } f_3(x) = \begin{cases} x^2/2, & \text{pour } 0 \leq x \leq 1; \\ -(x - \frac{3}{2})^2 + \frac{3}{4}, & \text{pour } 1 \leq x \leq 2; \\ (x-3)^2/2, & \text{pour } 2 \leq x \leq 3; \\ 0, & \text{ailleurs.} \end{cases}$$

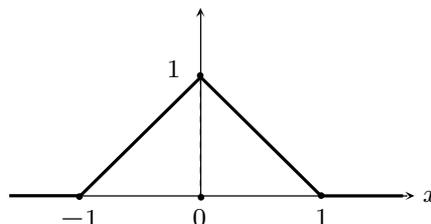
On constate, que contrairement à f_2 , la fonction f_3 est continûment dérivable.



- c) La densité de $X_1 - X_2$ est paire et donnée par $g(x) = \int_{-\infty}^{\infty} f(u)f(x+u) du$. La fonction à intégrer est strictement positive sur $D = \{(u, x) : 0 < u < 1, 0 < x + u < 1\} = \{(u, x) : \max(0, -x) < u < \min(1, 1 - x)\}$. De là, pour $0 \leq x \leq 1$, on a $g(x) = \int_0^{1-x} du = 1 - x$ et pour $-1 \leq x \leq 0$, $g(x) = \int_{-x}^1 du = 1 + x$, et finalement $g(u) = \begin{cases} 1 - |x|, & \text{si } |x| \leq 1, \\ 0, & \text{sinon.} \end{cases}$



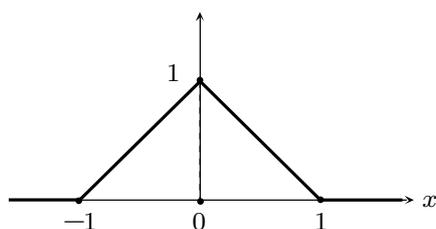
L'ensemble D



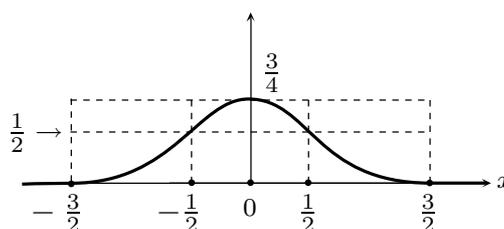
La densité g

6. a) Utilisons les notations de l'exercice précédent. Comme $Y_1 + Y_2 = (X_1 + X_2) - 1$, la variable $Y_1 + Y_2$ a pour densité $g_2(y) = f_2(y + 1) = 1 - |y|$, si $|y| \leq 1$, et 0 sinon.
 b) La variable $Y_1 + Y_2 + Y_3 = (X_1 + X_2 + X_3) - \frac{3}{2}$ a pour densité

$$g_2(y) = f_3(y + \frac{3}{2}) = \begin{cases} \frac{1}{2}(y + \frac{3}{2})^2, & \text{si } -\frac{3}{2} \leq y \leq -\frac{1}{2}, \\ -y^2 + \frac{3}{4}, & \text{si } -\frac{1}{2} \leq y \leq \frac{1}{2}, \\ \frac{1}{2}(y - \frac{3}{2})^2, & \text{si } \frac{1}{2} \leq y \leq \frac{3}{2}, \\ 0, & \text{ailleurs.} \end{cases}$$



La densité g_2



La densité g_3

- c) La variable $Y_1 - Y_2 = X_1 - X_2$ a la même densité que $Y_1 + Y_2$. On notera que le graphe de g_3 est le recollement de cinq tronçons de courbes différentes. Ce recollement est continûment différentiable. On ne sera pas surpris, au vu du graphe, que la loi de la somme $Y_1 + \dots + Y_n$, convenablement normée, converge en loi vers la normale $\mathcal{N}(0, 1)$ (cf. chap. 18).
7. Pour $0 \leq x \leq 1$ et $n = 1$ on retrouve : $f_1(x) = 1$. Supposons $n \geq 1$ et procédons par récurrence sur n . D'abord pour $0 \leq x \leq n + 1$ on a :

$$\begin{aligned} f_{n+1}(x) &= \int_{-\infty}^{+\infty} f_1(x-u) f_n(u) du = \int_{x-1}^x f_n(u) du \\ &= \sum_{k=0}^{n-1} (-1)^k \binom{n}{k} U(x, k), \text{ où } U(x, k) = \int_{x-1}^x \frac{((u-k)^+)^{n-1}}{(n-1)!} du. \end{aligned}$$

L'évaluation de cette intégrale donne :

$$\int_{x-1}^x \frac{((u-k)^+)^{n-1}}{(n-1)!} du = \begin{cases} 0, & \text{si } x \leq k; \\ \frac{(x-k)^n}{n!}, & \text{si } x-1 \leq k < x; \\ \frac{(x-k)^n}{n!} - \frac{(x-k-1)^n}{n!}, & \text{si } k < x-1. \end{cases}$$

Soit l l'unique entier positif défini par : $x-1 \leq l < x$. Alors

$$\begin{aligned} f_{n+1}(x) &= \sum_{k=0}^{n-1} (-1)^k \binom{n}{k} U(x, k) = \sum_{k=0}^l (-1)^k \binom{n}{k} U(x, k) \\ &= \sum_{k=0}^{l-1} (-1)^k \binom{n}{k} \left(\frac{(x-k)^n}{n!} - \frac{(x-k-1)^n}{n!} \right) + (-1)^l \binom{n}{l} \frac{(x-l)^n}{n!} \\ &= \frac{x^n}{n!} + \sum_{k=1}^l (-1)^k \binom{n}{k} \frac{(x-k)^n}{n!} + \sum_{k=1}^l (-1)^k \binom{n}{k-1} \frac{(x-k)^n}{n!} \\ &= \frac{x^n}{n!} + \sum_{k=1}^l (-1)^k \binom{n+1}{k} \frac{(x-k)^n}{n!} = \sum_{k=0}^l (-1)^k \binom{n+1}{k} \frac{(x-k)^n}{n!} \\ &= \sum_{k=0}^n (-1)^k \binom{n+1}{k} \frac{((x-k)^+)^n}{n!}. \end{aligned}$$

Chapitre 12

1. a) Comme $P\{X \in A, Y \in B\} = \int_{x \in A, y \in B} d\mu(x, y)$, puis

$$\mathbb{E}[Q_{(\cdot)}(A) \cdot I_{\{Y \in B\}}] = \int_{\mathbb{R}} Q_y(A) \cdot I_{\{Y \in B\}}(y) dP_Y(y) = \int_{y \in B} Q_y(A) dP_Y(y),$$

et aussi

$$Q_y(A) = \int I_A(x) dQ_y(x) = \int_{x \in A} dQ_y(x),$$

on peut récrire l'identité initiale comme

$$\int_{x \in A, y \in B} d\mu(x, y) = \int_{y \in B} \left(\int_{x \in A} dQ_y(x) \right) dP_Y(y).$$

b) Les conditions (1) et (2) sont banalement vérifiées. Pour l'identité (3) on

$$\text{a : } \mathbb{E}[Q_{(\cdot)}(A) \cdot I_{\{Y \in B\}}] = \sum_{j \in J} Q_{y_j}(A) I_{\{y_j \in B\}} P\{Y = y_j\} = \sum_{j \in J} P\{X \in$$

$$A|Y = y_j\} I_{\{y_j \in B\}} P\{Y = y_j\} = \sum_{j \in J} P\{X \in A, Y = y_j\} I_{\{y_j \in B\}} = P\{X \in A, Y \in B\}.$$

- c) La propriété (1) est banale; pour (2), le fait que $y \mapsto Q_y(A) = \int_A f_{X|Y}(x|y) dx$ soit mesurable résulte du théorème de Fubini. Enfin, la condition (3) réécrite dans sa forme intégrale s'exprime comme :

$$\int_{x \in A, y \in B} f_{X,Y}(x, y) dx dy = \int_{y \in B} \left(\int_{x \in A} f_{X|Y}(x|y) dx \right) f_Y(y) dy,$$

une identité qui a été vérifiée en (3.3).

2. a) $f_X(x) = e^{-x} I_{[0, +\infty[}(x)$, $f_Y(y) = e^{-y} I_{[0, +\infty[}(y)$.
 b) On a, pour tout $(x, y) \in \mathbb{R}^2$, l'identité $f(x, y) = f_X(x)f_Y(y)$.
3. Les variables X et Y ne sont pas indépendantes : on vérifie que $f_X(x) = 2e^{-2x} I_{[0, +\infty[}(x)$ et $f_Y(y) = 2e^{-y}(1 - e^{-y}) I_{[0, +\infty[}(y)$, de sorte que l'on n'a pas $f(x, y) = f_X(x)f_Y(y)$. [$f(x, y)$ n'est qu'en apparence le produit d'une fonction de x par une fonction de y .]
4. La condition est évidemment nécessaire. Montrons qu'elle est suffisante. En désignant par $f_X(x)$, $f_Y(y)$ les densités marginales de X , Y , il vient $f_X(x) = \int_{\mathbb{R}} f(x, y) dy = g(x)(\int_{\mathbb{R}} h(y) dy)$, et aussi $f_Y(y) = \int_{\mathbb{R}} f(x, y) dx = (\int_{\mathbb{R}} g(x) dx)h(y)$; d'où $1 = \int_{\mathbb{R}} f_Y(y) dy = (\int_{\mathbb{R}} g(x) dx)(\int_{\mathbb{R}} h(y) dy)$ et donc $f_X(x)f_Y(y) = g(x)h(y) = f(x, y)$.
5. a)

$$f_X(x) = \begin{cases} \frac{1}{\pi r^2} \int_{-\sqrt{r^2-x^2}}^{+\sqrt{r^2-x^2}} dt = \frac{2}{\pi} \frac{\sqrt{r^2-x^2}}{r^2}, & \text{si } |x| \leq r; \\ 0, & \text{sinon.} \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{2}{\pi} \frac{\sqrt{r^2-y^2}}{r^2}, & \text{si } |y| \leq r; \\ 0, & \text{sinon.} \end{cases}$$

$$\mathbb{E}[X] = \mathbb{E}[Y] = 0.$$

- b) X et Y ne sont pas indépendantes.
- c) $\text{Cov}(X, Y) = 0$; en effet, $\mathbb{E}[XY] = 0$ (par symétrie). Conclusion : le couple (X, Y) est non corrélé et pourtant il est non indépendant.
- d) $G(u) = P\{U \leq u\} = P\{X^2 + Y^2 \leq u\}$. Si $u \leq 0$, alors $G(u) = 0$. Si $0 < u \leq r^2$, on a $G(u) = P\{\sqrt{X^2 + Y^2} \leq \sqrt{u}\} = \pi(\sqrt{u})^2/(\pi r^2) = u/r^2$. Si $u > r^2$, on a $G(u) = 1$. Il en résulte, en dérivant, $g(u) = (1/r^2) I_{[0, r^2]}(u)$.
- e) $\mathbb{E}[U] = \int_{\mathbb{R}} ug(u) du = \frac{1}{r^2} \int_0^{r^2} u du = \frac{r^2}{2}$. Or $\mathbb{E}[U] = \mathbb{E}[X^2] + \mathbb{E}[Y^2]$ et comme $\mathbb{E}[X^2] = \mathbb{E}[Y^2]$, on a : $\mathbb{E}[X^2] = \mathbb{E}[Y^2] = r^2/4$. Comme $\mathbb{E}[X] = \mathbb{E}[Y] = 0$, on en déduit : $\text{Var } X = \text{Var } Y = r^2/4$.
- f) $f_{Y|X}(y|x) = \begin{cases} \frac{f_{X,Y}(x, y)}{f_X(x)}, & \text{si } f_X(x) > 0; \\ \text{densité arbitraire,} & \text{sinon.} \end{cases}$

Or $f_X(x) > 0$, si et seulement si $|x| < r$, d'où

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{2\sqrt{r^2-x^2}}, & \text{si } |x| < r, |y| \leq \sqrt{r^2-x^2}; \\ \text{densité arbitraire,} & \text{sinon.} \end{cases}$$

$$\mathbb{E}[Y^2 | X = x] = \int_{\mathbb{R}} y^2 f_{Y|X}(y|x) dy = \frac{r^2 - x^2}{3};$$

$$\mathbb{E}[X^2 + Y^2 | X = x] = x^2 + \frac{r^2 - x^2}{3} = \frac{r^2 + 2x^2}{3};$$

$$\mathbb{E}[X^2 + Y^2 | X] = \frac{r^2 + 2X^2}{3}.$$

g) $\mathbb{P}\{L \leq a\} = \mathbb{P}\{\sqrt{X^2 + Y^2} \leq a\} = \mathbb{P}\{X^2 + Y^2 \leq a^2\} = G(a^2) = a^2/r^2.$

$$\mathbb{P}\{\min(L_1, \dots, L_n) > a\} = \mathbb{P}\{L_1 > a\} \dots \mathbb{P}\{L_n > a\} = \left(1 - \frac{a^2}{r^2}\right)^n.$$

$\mathbb{P}\{\min(L_1, \dots, L_n) \leq a\} = 1 - \left(1 - \frac{a^2}{r^2}\right)^n$. C'est la probabilité pour qu'un tir au moins touche le disque $(0, a)$ de centre 0 et de rayon a .

6. a) En désignant par $f(x_1, x_2)$ la densité conjointe de M et par $f_{X_1}(x_1)$ la densité marginale de X_1 , la densité de X_2 liée par $\{X_1 = x_1\}$ est donnée par

$$f_{X_2|X_1}(x_2|x_1) = \frac{f(x_1, x_2)}{f_{X_1}(x_1)} = \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2} \frac{(x_2 - \rho x_1)^2}{1-\rho^2}\right).$$

C'est la densité de la loi $\mathcal{N}(\rho x_1, \sqrt{1-\rho^2})$.

- b) On a $\mathbb{E}[X_2 | X_1] = \rho X_1$, d'où le résultat en vertu du Corollaire 2 du Théorème 5.6.

7. Nous donnerons trois exemples.

- a) Prenons $f(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x_1^2 + x_2^2)\right) + a g(x_1, x_2)$, où

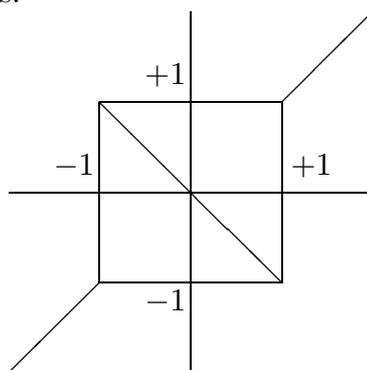
$$g(x_1, x_2) = \begin{cases} x_1 x_2, & \text{si } |x_1| \leq 1, |x_2| \leq 1, \\ 0, & \text{sinon,} \end{cases} \text{ et où } a \text{ est une constante}$$

positive telle que $f(x_1, x_2)$ soit strictement positive. On vérifie que $f(x_1, x_2)$ est une densité de probabilité, évidemment non normale, et que ses densités marginales ont les densités de $\mathcal{N}(0, 1)$.

- b) Soit (X, Z) un couple de variables aléatoires indépendantes, où X suit la loi normale $\mathcal{N}(0, 1)$ et Z la loi $\frac{1}{2}(\varepsilon_{-1} + \varepsilon_{+1})$. Posons $Y = XZ$; alors le couple (X, Y) fournit un exemple; en effet, X et Y suivent la loi $\mathcal{N}(0, 1)$. C'est le cas, par définition, pour X . Pour Y on le voit comme suit : $\mathbb{P}\{Y \leq y\} = \mathbb{P}\{Y \leq y | Z = +1\} \frac{1}{2} + \mathbb{P}\{Y \leq y | Z = -1\} \frac{1}{2} = \frac{1}{2}(\mathbb{P}\{X \leq y | Z = +1\} + \mathbb{P}\{-X \leq y | Z = -1\})$, une expression qui est égale à $\frac{1}{2}(\mathbb{P}\{X \leq y\} + \mathbb{P}\{-X \leq y\})$, puisque X et Z sont indépendants, soit encore $\mathbb{P}\{X \leq y\}$, puisque X est symétrique. En revanche, le couple (X, Y) ne suit pas une loi normale à deux dimensions. Il suffit pour le voir de remarquer que $X + Y = X(1 + Z)$ est égal à 0 avec probabilité $\frac{1}{2}$

et à $2X$ avec probabilité $\frac{1}{2}$. On voit, en outre, que la somme des deux variables aléatoires normales X, Y n'est pas normale.

- c) Considérons une loi normale à deux dimensions, centrée, mais dégénérée dont le support est la première bissectrice. Faisons tourner de 90° le segment pondéré de cette bissectrice dont les extrémités sont les points $(-1, -1)$ et $(+1, +1)$, tout en laissant l'autre segment en place. On obtient ainsi une loi à deux dimensions, centrée, qui n'est plus normale. Or ses lois marginales, qui sont les mêmes que les lois marginales initiales, sont normales.



8. Désignons par $h(x, y)$ la densité conjointe de (X, Y) et par $f(x)$ (resp. $g(y)$) la densité marginale de X (resp. de Y). On a par hypothèse $h(x, y) > 0$ et $f(x) > 0, g(y) > 0$ pour tout x et tout y . Bornons-nous à démontrer b) \Rightarrow a). Si b) est vérifié, on a $f = g$ et $h(x, y) = f(x)f(y) = \varphi(x^2 + y^2)$. En supposant φ dérivable, il vient $f'(x)f(y) = 2x\varphi'(x^2 + y^2)$ et $f(x)f'(y) = 2y\varphi'(x^2 + y^2)$; d'où $\frac{f'(x)}{2xf(x)} = \frac{f'(y)}{2yf(y)} = c$, c'est-à-dire $f'(x) - 2cx f(x) = 0$ et donc $f(x) = ke^{cx^2}$.
9. a) La loi conjointe de (X, Y) est donnée, d'une part, par la loi de X , qui est la loi uniforme sur $\{1, 2, \dots, 6\}$ (on a donc $\mathbb{E}[X] = 7/2$), d'autre part, par la loi conditionnelle de Y liée par X . Or la loi de Y liée par $\{X = x\}$ ($x = 1, 2, \dots, 6$) est la loi binomiale $B(x, 1/2)$.
- b) On a $\mathbb{E}[Y | X = x] = x/2$, d'où $\mathbb{E}[Y | X] = X/2$; c'est une variable aléatoire. On a ensuite $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y | X]] = \mathbb{E}[X/2] = 7/4$.
10. On a $\mathbb{E}[N] = \mathbb{E}[\mathbb{E}[N | X]] = \mathbb{E}[N | X = 1]p + \mathbb{E}[N | X = 0](1 - p)$. Or $\mathbb{E}[N | X = 1] = 1$ et $\mathbb{E}[N | X = 0] = 1 + \mathbb{E}[N]$; d'où $\mathbb{E}[N] = p + (1 + \mathbb{E}[N])(1 - p) = 1 + (1 - p)\mathbb{E}[N]$, d'où $\mathbb{E}[N] = 1/p$.
11. a) $a_{r,s} = \mathbb{E}[\mathbb{E}[N_{r,s} | X]] = \mathbb{E}[N_{r,s} | X=1](r/(r+s)) + \mathbb{E}[N_{r,s} | X=0](s/(r+s)) = 1 \times (r/(r+s)) + (1 + \mathbb{E}[N_{r,s-1}])(s/(r+s)) = (r/(r+s)) + (1 + a_{r,s-1})(s/(r+s))$. D'où $a_{r,s} = 1 + (s/(r+s))a_{r,s-1}$. Les conditions initiales $a_{r,0} = 1$ ($r \geq 1$) sont évidentes.
- b) La démonstration se fait par récurrence sur s , l'entier r étant fixé.
12. 1) $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y | X]] = \sum_{k \geq 0} \mathbb{P}\{X = k\} \mathbb{E}[Y | X = k] = \sum_{k \geq 0} \mathbb{P}\{X = k\} (k/2) = \mathbb{E}[X]/2$.

2) La loi conjointe de $(X - Y, Y)$ est donnée par $P\{X - Y = k, Y = l\} = P\{X = k + l, Y = l\} = P\{X = k + l\}P\{Y = l | X = k + l\} = P\{X = k + l\}/(k + l + 1)$ ($k, l \geq 0$); d'où, en posant $u_n = P\{X = n\}/(n + 1)$ ($n \geq 0$), l'expression $P\{X - Y = k, Y = l\} = u_{k+l}$. On voit que cette loi conjointe est symétrique en k, l ; il en résulte que les lois marginales de $X - Y$ et Y sont identiques. De façon précise, $P\{X - Y = k\} = \sum_{l \geq 0} u_{k+l} = \sum_{i \geq k} u_i$ et $P\{Y = l\} = \sum_{k \geq 0} u_{k+l} = \sum_{j \geq l} u_j$.

b \Rightarrow a : Si b) est vérifiée, on a $u_l = P\{Y = l\} - P\{Y = l + 1\} = q^l p - q^{l+1} p = q^l p^2$ ($l \geq 0$); d'où $u_{k+l} = q^{k+l} p^2 = (q^k p)(q^l p)$, c'est-à-dire $P\{X - Y = k, Y = l\} = P\{X - Y = k\}P\{Y = l\}$.

a \Rightarrow b : Si a) est vérifiée, on a $u_{k+l} = \left(\sum_{i \geq k} u_i\right)\left(\sum_{j \geq l} u_j\right)$. En faisant

$l = 0$ et en posant $p = \sum_{j \geq 0} u_j$, qui est dans $]0, 1[$, il vient $u_k = p \sum_{i \geq k} u_i$,

$u_k - u_{k+1} = pu_k$, $qu_k = u_{k+1}$ ($q = 1 - p$). Puisque $u_0 = p^2$, on obtient $u_k = q^k p^2$ ($k \geq 0$), d'où $P\{Y = l\} = \sum_{j \geq l} u_j = q^l p$ ($l \geq 0$).

13. a) On peut prendre $g(a, b) = \mathbb{E}[X | a \leq X \leq b]$. Or la fonction de survie conditionnelle pour tout $x \geq 0$ est donnée par

$$P\{X > x | a \leq X \leq b\} = \frac{P\{X > x, a \leq X \leq b\}}{P\{a \leq X \leq b\}} = \begin{cases} 1, & \text{si } 0 \leq x < a; \\ \frac{P\{X > x\} - P\{X > b\}}{P\{X > a\} - P\{X > b\}} = \frac{e^{-\lambda x} - e^{-\lambda b}}{e^{-\lambda a} - e^{-\lambda b}}, & \text{si } a \leq x \leq b; \\ 0, & \text{si } x > b. \end{cases}$$

$$\text{D'où } g(a, b) = \int_0^a 1 dx + \int_a^b \frac{e^{-\lambda x} - e^{-\lambda b}}{e^{-\lambda a} - e^{-\lambda b}} dx = \frac{1}{\lambda} + \frac{ae^{-\lambda a} - be^{-\lambda b}}{e^{-\lambda a} - e^{-\lambda b}}.$$

b) On a $\lim_{b \rightarrow \infty} g(a, b) = \mathbb{E}[X | X > a] = \frac{1}{\lambda} + a = \mathbb{E}[X] + a$, un résultat qu'on peut prévoir *a priori*, en raison de la propriété d'absence de mémoire de la loi exponentielle.

c) On a $g(a, a + \varepsilon) = \frac{1}{\lambda} + \left(-\frac{1}{\lambda} + a + o(\varepsilon)\right)$. D'où $\lim_{\varepsilon \rightarrow 0^+} g(a, a + \varepsilon) = a$.

Chapitre 13

1. Se rappeler la définition de l'aire d'un parallélogramme et faire plusieurs croquis annexes.

2. $g(u) = \mathbb{E}[e^{u(X-p)}] = pe^{qu} + qe^{-pu}$ ($u \in \mathbb{R}$). Ainsi $g(u)$ admet le développement suivant en série entière, valable pour tout réel u :

$$g(u) = p \sum_{k \geq 0} \frac{(qu)^k}{k!} + q \sum_{k \geq 0} \frac{(-pu)^k}{k!} = \sum_{k \geq 0} (pq^k + q(-p)^k) \frac{u^k}{k!}.$$

Le moment d'ordre $k \geq 1$ est le coefficient de $u^k/k!$ dans ce développement; d'où $\mathbb{E}[(X - p)^k] = pq^k + q(-p)^k$ ($k \geq 1$)

$$\begin{aligned} k = 1 : & \quad 0; \\ k = 2 : & \quad pq^2 + qp^2 = pq(p + q) = pq; \\ k = 3 : & \quad pq^3 - qp^3 = pq(q^2 - p^2) = pq(q - p); \\ k = 4 : & \quad pq^4 + qp^4 = pq(q^3 + p^3) = pq(1 - 3pq). \end{aligned}$$

3. a) $g'_1(u) = -\mathbb{E}[Xe^{-uX}]$, d'où, en multipliant par e^{-yu} ($u \geq 0$),
 $g'_1(u)e^{-yu} = -\mathbb{E}[Xe^{-u(X+y)}]$. En intégrant, on obtient

$$\int_0^\infty g'_1(u)e^{-yu} du = \mathbb{E}\left[\frac{X}{X+y}\right]. \quad (1)$$

En conditionnant par rapport à Y , il vient

$$\begin{aligned} \mathbb{E}\left[\frac{X}{X+Y}\right] &= \mathbb{E}\left[\mathbb{E}\left[\frac{X}{X+Y} \mid Y\right]\right] = \int_0^\infty \mathbb{E}\left[\frac{X}{X+Y} \mid Y=y\right] d\mu(y) \\ &= \int_0^\infty \mathbb{E}\left[\frac{X}{X+y}\right] d\mu(y); \end{aligned}$$

d'où, en remplaçant $\mathbb{E}[(X/X+y)]$ par sa valeur trouvée dans (1)

$$\mathbb{E}\left[\frac{X}{X+Y}\right] = - \int_0^\infty d\mu(y) \left(\int_0^\infty g'_1(u)e^{-yu} du \right);$$

enfin, en utilisant le théorème de Fubini

$$\mathbb{E}\left[\frac{X}{X+Y}\right] = - \int_0^\infty g'_1(u) \left(\int_0^\infty e^{-yu} d\mu(y) \right) du = - \int_0^\infty g'_1(u)g_2(u) du.$$

b) Supposons X, Y indépendants et de même loi; d'où $g_1 = g_2 = g$. Alors

$$\mathbb{E}\left[\frac{X}{X+Y}\right] = - \int_0^\infty g'(u)g(u) du = -\frac{1}{2} \int_0^\infty (g^2(u))' du = \frac{1}{2} [1 - \lim_{u \rightarrow \infty} g^2(u)].$$

Or $\mathbb{E}\left[\frac{X}{X+Y}\right] = \frac{1}{2}$, puisque $1 = \mathbb{E}\left[\frac{X+Y}{X+Y}\right]$ et que $\mathbb{E}\left[\frac{X}{X+Y}\right] = \mathbb{E}\left[\frac{Y}{X+Y}\right]$; d'où $\lim_{u \rightarrow +\infty} g^2(u) = 0$.

4. Partons de la définition de la fonction gamma

$$\Gamma(p) = \int_0^\infty e^{-x} x^{p-1} dx \quad (p > 0)$$

et faisons le changement de variable $x = su$ ($s > 0$)

$$\Gamma(p) = s^p \int_0^\infty e^{-su} u^{p-1} du; \quad \frac{1}{s^p} = \frac{1}{\Gamma(p)} \int_0^\infty e^{-su} u^{p-1} du.$$

Prenons pour s une variable aléatoire X à valeurs *strictement positives* et posons $g(u) = \mathbb{E}[e^{-uX}]$; il vient

$$\frac{1}{X^p} = \frac{1}{\Gamma(p)} \int_0^\infty e^{-uX} u^{p-1} du,$$

d'où, en utilisant le théorème de Fubini, l'identité suivante dans $[0, +\infty[$

$$\mathbb{E}\left[\frac{1}{X^p}\right] = \frac{1}{\Gamma(p)} \int_0^\infty g(u)u^{p-1} du. \quad \square$$

5. Désignons par m_1 l'espérance mathématique de X et par μ_2, μ_3, μ_4 ses moments centrés d'ordre 2, 3, 4. Posons $X - m_1 = Y$, d'où $g(u) = e^{um_1}g_Y(u)$, $h(u) = \text{Log } g(u) = um_1 + \text{Log } g_Y(u)$. Or

$$g_Y(u) = 1 + \mu_2 \frac{u^2}{2!} + \mu_3 \frac{u^3}{3!} + \mu_4 \frac{u^4}{4!} + o(|u|^4) = 1 + \lambda(u);$$

$$\begin{aligned} \text{d'où } h(u) &= um_1 + \text{Log}(1 + \lambda(u)) = um_1 + \lambda(u) - \frac{\lambda^2(u)}{2!} + o(|u|^4) \\ &= um_1 + \mu_2 \frac{u^2}{2!} + \mu_3 \frac{u^3}{3!} + (\mu_4 - 3\mu_2^2) \frac{u^4}{4!} + o(|u|^4). \end{aligned}$$

On constate que $h'(0) = m_1$, $h''(0) = \mu_2$, $h'''(0) = \mu_3$, $h^{(4)}(0) = \mu_4 - 3\mu_2^2$. Ce sont les quatre premiers cumulants de X .

6. a) $\mathbb{E}[X^n] = \int_1^\infty x^n f(x) dx = a \int_1^\infty \frac{dx}{x^{(a+1)-n}}$. L'intégrale au dernier membre est convergente, si et seulement si $(a+1) - n > 1$, c'est-à-dire $n < a$, et dans ce cas elle vaut $a/(a-n)$.
- b) $g(u) = \mathbb{E}[e^{uX}] = \int_1^\infty e^{uX} f(x) dx = a \int_1^\infty \frac{e^{ux}}{x^{a+1}} dx$. Cette intégrale est définie, si et seulement si $u \in]-\infty, 0]$. Or $]-\infty, 0]$ n'est pas un voisinage ouvert de $u = 0$; la fonction g n'est donc pas une fonction génératrice des moments. D'ailleurs, si elle en était une, la variable aléatoire X admettrait des moments de tous les ordres, ce qui, d'après a), n'est pas le cas. En revanche, X admet une fonction caractéristique tout comme toute variable aléatoire.

7. $g(u) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{u|x|} e^{-x^2/2} dx = \frac{2}{\sqrt{2\pi}} \int_0^\infty e^{ux-x^2/2} dx$. Or $ux - \frac{x^2}{2} = -\frac{(x-u)^2}{2} + \frac{u^2}{2}$; d'où $g(u) = 2e^{u^2/2} \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-(x-u)^2/2} dx$. En faisant $x - u = t$, on obtient : $g(u) = 2e^{u^2/2} \frac{1}{\sqrt{2\pi}} \int_{-u}^\infty e^{-t^2/2} dt = 2e^{u^2/2} \frac{1}{\sqrt{2\pi}} \left(\int_{-u}^0 e^{-t^2/2} dt + \int_0^\infty e^{-t^2/2} dt \right)$, soit

$$g(u) = 2e^{u^2/2} \left(\frac{1}{2} - \frac{1}{\sqrt{2\pi}} \int_0^{-u} e^{-t^2/2} dt \right) = e^{u^2/2} (1 - 2\Phi(-u)), \text{ où l'on a posé } \Phi(u) = \frac{1}{\sqrt{2\pi}} \int_0^u e^{-t^2/2} dt.$$

8. On a $g(u) = \mathbb{E}[e^{uXY}] = \frac{1}{2\pi} \int_{\mathbb{R}^2} e^{uxy} e^{-x^2/2} e^{-y^2/2} dx dy$; d'où, en

intégrant à x constant,

$$\begin{aligned} g(u) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \left(\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{uxy - y^2/2} dy \right) e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{u^2 x^2/2} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(1-u^2)x^2/2} dx \\ &= \frac{1}{\sqrt{1-u^2}} \quad (|u| < 1). \end{aligned}$$

9. La variable $\Delta = X_1 X_4 - X_2 X_3$ est la somme des deux variables aléatoires *indépendantes* $X_1 X_4$ et $-X_2 X_3$. Or $\mathcal{L}(-X_2 X_3) = \mathcal{L}(X_2 X_3)$. Donc $\mathcal{L}(\Delta)$ est la loi de la somme des deux variables aléatoires indépendantes identiquement distribuées $X_1 X_4$ et $X_2 X_3$. D'après l'exercice précédent, on en tire : $g_{\Delta}(u) = \frac{1}{\sqrt{1-u^2}} \cdot \frac{1}{\sqrt{1-u^2}} = \frac{1}{1-u^2}$ ($|u| < 1$).
10. a) On a : $\sum_{k \geq 0} \alpha_k(\varphi)^k = e^{-\lambda} \sum_{k \geq 0} \frac{(\lambda\varphi)^k}{k!} = e^{-\lambda} e^{\lambda\varphi} = e^{\lambda(\varphi-1)}$. Si $\varphi = e^{it}$, alors $\sum_{k \geq 0} \alpha_k(\varphi)^k = \sum_{k \geq 0} \alpha_k e^{ikt}$, qui est la fonction caractéristique de la loi $\mathcal{P}(\lambda)$.
- b) On a : $\sum_{k \geq 0} \alpha_k(\varphi)^k = p \sum_{k \geq 0} (q\varphi)^k = \frac{p}{1-q\varphi} = \frac{1-q}{1-q\varphi} = \frac{\lambda-1}{\lambda-\varphi}$, où l'on a posé $1/q = \lambda > 1$. Dans le cas particulier où $\varphi(t) = e^{it}$, alors $\sum_{k \geq 0} \alpha_k(\varphi)^k = \sum_{k \geq 0} \alpha_k e^{ikt}$, qui est la fonction caractéristique de la loi géométrique $\sum_k p q^k \varepsilon_k$ ($k \geq 0$).
11. a) Prenons $\varphi_{\lambda}(t) = \varphi(\lambda t)$ ($\lambda \in [0, 1]$) et $f(\lambda) = I_{[0,1]}(\lambda)$. Alors $\int_0^1 \varphi_{\lambda}(t) d\lambda = \int_0^1 \varphi(\lambda t) d\lambda = (1/t) \int_0^t \varphi(u) du = \Phi(t)$ est une fonction caractéristique.
- b) Pour tout $\lambda \in [0, +\infty[$ la fonction $\varphi_{\lambda}(t) = e^{-\lambda^2 t^2}$ est une fonction caractéristique. De plus, pour $\alpha > -1$ la fonction $f(\lambda) = \lambda^{\alpha} e^{-\lambda^2} / I_{\alpha}$, où $I_{\alpha} = \int_0^{\infty} \lambda^{\alpha} e^{-\lambda^2} d\lambda < \infty$, est une densité de probabilité sur $[0, +\infty[$. Enfin, $\int_0^{\infty} e^{-\lambda^2 t^2} (\lambda^{\alpha} e^{-\lambda^2} / I_{\alpha}) d\lambda = (1/I_{\alpha}) \int_0^{\infty} \lambda^{\alpha} e^{-\lambda^2(1+t^2)} d\lambda$, une expression qui par le changement de variable $\lambda\sqrt{1+t^2} = u$ vaut $(1/I_{\alpha})(1/(1+t^2)^{(\alpha+1)/2}) \int_0^{\infty} u^{\alpha} e^{-u^2} du = 1/(1+t^2)^{(\alpha+1)/2}$ ou encore $1/(1+t^2)^{\gamma}$ [en posant $\gamma = (\alpha+1)/2 > 0$] et qui est bien une fonction caractéristique.
12. On a $\varphi_X(t) = \overline{\varphi_{-Y+Z}(t)} = \overline{\varphi_{-Y}(t)\varphi_Z(t)} = \overline{\varphi_Y(t)}\overline{\varphi_Z(t)}$ et aussi $\overline{\varphi_Y(t)} = \overline{\varphi_X(t)\varphi_Z(t)}$ en échangeant les rôles de X et Y . On en déduit $|\varphi_Z(t)|^2 = 1$, d'où $|\varphi_Z(t)| = 1$ et $\varphi_Z(t) = e^{ict}$ avec $c \in \mathbb{R}$.
13. Notons $\varphi(u_1, u_2)$ la fonction caractéristique du couple (Y_1, Y_2) . On a

$$\begin{aligned}\varphi(u_1, u_2) &= \mathbb{E}[e^{i(u_1 Y_1 + u_2 Y_2)}] = \mathbb{E}[e^{i(u_1 + u_2)X + u_1 X_1 + u_2 X_2}] \\ &= \varphi_X(u_1 + u_2) \varphi_{X_1}(u_1) \varphi_{X_2}(u_2). \text{ D'où l'on tire}\end{aligned}$$

$\varphi(u_1, 0) \varphi(0, u_2) = \varphi_X(u_1) \varphi_X(u_2) \varphi_{X_1}(u_1) \varphi_{X_2}(u_2)$. Or le couple (Y_1, Y_2) est indépendant, si et seulement si $\varphi(u_1, u_2) = \varphi(u_1, 0) \varphi(0, u_2)$, une relation qui se réécrit $\varphi_X(u_1 + u_2) = \varphi_X(u_1) \varphi_X(u_2)$. C'est l'équation fonctionnelle de l'exponentielle. D'où $\varphi_X(u) = e^{cu}$, avec $c = i\alpha$ et α réel, puisque $\varphi_X(u)$ est une fonction caractéristique. Par conséquent, $X = \alpha$, presque sûrement.

14. Pour $u \geq 0$, $x > 0$, on a, d'après l'inégalité de Markov, $P\{X \geq x\} = P\{e^{uX} \geq e^{ux}\} \leq e^{-ux} \mathbb{E}[e^{uX}]$, d'où le résultat en prenant l'infimum au dernier membre.

Chapitre 14

- Supposons que r soit continue à droite et vérifie l'identité. D'abord $r(0) = r(0)^2$, d'où $r(0) = 0$ ou 1 . Si $r(0) = 0$, l'identité entraîne que r est identiquement nulle. Supposons $r(0) = 1$; on a alors $r(1) = r(1/2)^2 \geq 0$. Si $r(1) = 0$, l'équation $r(1) = r\left(\frac{1}{n} + \dots + \frac{1}{n}\right) = r^n\left(\frac{1}{n}\right)$ ($n \geq 1$) entraîne que $r(1/n) = 0$ pour tout $n \geq 1$. Comme r est continue à droite, il vient $r(0) = 0$, contrairement à l'hypothèse. Par conséquent, $r(1) > 0$ et $r(1/n) = (r(1))^{1/n}$ pour tout $n \geq 1$ et aussi $r(m/n) = (r(1))^{m/n}$. Comme, d'une part, tout nombre réel $x \geq 0$ est limite à droite d'une suite de nombre rationnels (q_n) et, d'autre part, r est continue à droite, on a : $r(x) = r(\lim_n q_n) = \lim_n r(q_n) = \lim_n (r(1))^{q_n} = (r(1))^x$. On en déduit : $r(x) = e^{\alpha x}$ avec $\alpha = \text{Log } r(1)$.
- Pour $x \geq 1$ on a : $P\{X > x\} = P\{e^Y > x\} = P\{Y > \text{Log } x\} = e^{-\lambda \text{Log } x} = x^{-\lambda}$ et $f(x) = -(d/dx)P\{X > x\} = \lambda/x^{\lambda+1}$.
 - $\mathbb{E}[X] = \int_0^\infty r(x) dx = \int_0^1 dx + \int_1^\infty x^{-\lambda} dx = \lambda/(\lambda - 1)$, si $\lambda > 1$. Si $\lambda \leq 1$, l'intégrale ne converge pas.
 - On a : $P\{X^k > x\} = P\{X > \sqrt[k]{x}\} = x^{-\lambda/k}$, si $x \geq 1$, et 1 si $0 \leq x < 1$. D'où $\mathbb{E}[X^k] = \lambda/(\lambda - k)$, si $\lambda > k$ et $\mathbb{E}[X^k] = +\infty$, si $\lambda \leq k$. Une variable de Pareto n'admet pas des moments de tous les ordres, donc pas non plus de fonction génératrice des moments.
- Pour $x \geq 0$ on a $r(x) = P\{X \geq x\} = P\{Y \geq x^\alpha\} = e^{-\lambda x^\alpha}$, d'où $f(x) = -r'(x)$. Enfin, $\mathbb{E}[X] = \int_0^{+\infty} r(x) dx = \int_0^{+\infty} e^{-\lambda x^\alpha} dx = (1/\lambda^{1/\alpha})(1/\alpha) \int_0^{+\infty} e^{-u} u^{(1/\alpha)-u} du = (1/\lambda^{1/\alpha})\Gamma(1 + (1/\alpha))$, par le changement de variable $\lambda x^\alpha = u$.
- En effet, $F(x) = P\{X \leq x\} = P\{-\text{Log}(e^Y - 1) \leq x\} = P\{\text{Log}(e^Y - 1) \geq -x\} = P\{e^Y \geq 1 + e^{-x}\} = P\{Y \geq \text{Log}(1 + e^{-x})\} = \exp(-\text{Log}(1 + e^{-x})) = 1/(1 + e^{-x})$. Par dérivation, on obtient l'expression de $f(x)$. Enfin, la dérivée seconde de $\text{Log } f(x)$ est égale à $-1/(1 + \text{ch } x) < 0$.

5. Pour $0 \leq x \leq 1$ on a $P\{Y > x\} = (P\{X > x\})^n = (1-x)^n$, d'où $P\{Y \leq x\} = 1 - P\{Y > x\} = 1 - (1-x)^n$. La densité de Y est donc donnée par $f(x) = n(1-x)^{n-1} = nx^{1-1}(1-x)^{n-1}$, qui est la densité de probabilité de la loi $B(1, n)$. De même, $P\{X \leq x\} = x^n$, ce qui donne pour densité : $f(x) = nx^{n-1} = nx^{n-1}(1-x)^{1-1}$, soit la densité de la loi $B(n, 1)$.
6. a) $e_0 = \exp\left[\frac{1}{2}\left(\text{Log } 2 + \frac{\Gamma'(1/2)}{\Gamma(1/2)}\right)\right]$. En effet, on sait que : $\mathbb{E}[|X|^r] = \frac{2^{r/2}}{\sqrt{\pi}}\Gamma\left(\frac{r+1}{2}\right)$ ($r > -1$); d'où, puisque $\sqrt{\pi} = \Gamma(1/2)$, $\text{Log } e_r = \frac{1}{r}\left[\frac{r}{2}\text{Log } 2 + \text{Log } \frac{\Gamma(1/2+r/2)}{\Gamma(1/2)}\right] = \frac{\text{Log } 2}{2} + \frac{1}{r}\text{Log } \frac{\Gamma(1/2+r/2)}{\Gamma(1/2)}$. Or $\text{Log } \frac{\Gamma(1/2+r/2)}{\Gamma(1/2)} = \frac{\Gamma(1/2) + (r/2)\Gamma'(1/2) + o(r)}{\Gamma(1/2)} = 1 + \frac{r}{2}\frac{\Gamma'(1/2)}{\Gamma(1/2)} + o(r)$; d'où $\text{Log } \frac{\Gamma(1/2+r/2)}{\Gamma(1/2)} = \frac{r}{2}\frac{\Gamma'(1/2)}{\Gamma(1/2)} + o(r)$; le résultat s'en déduit.
- b) $e_0 = \lambda^{-1}e^{-\gamma}$, où γ est la constante d'Euler. En effet, on sait que $\mathbb{E}[X^r] = \Gamma(r+1)/\lambda^r$ ($r > -1$), d'où $\text{Log } e_r = \frac{1}{r}[-r \log \lambda + \text{Log } \Gamma(1+r)] = -\text{Log } \lambda + \frac{1}{r}\text{Log } \Gamma(1+r)$. Or $\Gamma(1+r) = \Gamma(1) + r\Gamma'(1) + o(r)$. On sait que $\Gamma'(1) = -\gamma$; d'où $\Gamma(1+r) = 1 - \gamma r + o(r)$, $\text{Log } \Gamma(1+r) = -\gamma r + o(r)$ et le résultat.
- c) Il suffit d'observer que $\mathbb{E}[|X|^r] < +\infty$ pour tout $r \in [0, 1[$.
- d) Prendre $X = e^Y$, où Y est une variable aléatoire de Cauchy $\mathcal{C}(0, 1)$. On voit que $\mathbb{E}[X^r]$ n'est finie pour aucune valeur $r > 0$.

7. On a pour tout $x > 0$

$$\frac{1}{\rho(x)} = \frac{r(x)}{f(x)} = \frac{\int_x^\infty e^{-\lambda t}(\lambda t)^{p-1} dt}{e^{-\lambda x}(\lambda x)^{p-1}} = \int_x^\infty e^{-\lambda(t-x)}\left(\frac{t}{x}\right)^{p-1} dt;$$

d'où, en faisant le changement de variable $t - x = u$

$$\frac{1}{\rho(x)} = \int_0^\infty e^{-\lambda u}\left(1 + \frac{u}{x}\right)^{p-1} du;$$

d'où le résultat.

8. $\mathbb{E}[X^r] = \frac{\lambda^p}{\Gamma(p)} \int_0^\infty e^{-\lambda x} x^{p+r-1} dx$; en faisant le changement de variable $\lambda x = u$ on obtient : $\mathbb{E}[X^r] = \frac{1}{\lambda^r} \frac{1}{\Gamma(p)} \int_0^\infty e^{-u} u^{p+r-1} du$. L'intégrale du second membre est convergente, si et seulement si $p+r > 0$ et dans ce cas vaut $\Gamma(p+r)$.
9. a) $g(u) = \mathbb{E}[e^{u|X|}] = \frac{1}{2} \int_{\mathbb{R}} e^{(u-1)|x|} dx = \int_0^\infty e^{(u-1)x} dx$. Cette intégrale est convergente pour $u-1 < 0$; en faisant le changement de variable $(u-1)x = y$, il vient $g(u) = 1/(1-u)$ ($u < 1$); la loi de $|X|$ est $\mathcal{E}(1)$.

b) Soit g la fonction génératrice de $\mathcal{E}(1)$: $g(u) = \frac{1}{1-u}$ ($u < 1$). Alors

$$g_{X_1+X_2}(u) = g_{X_1}(u)g_{X_2}(u) = \left(\frac{1}{1-u}\right)^2 \quad (u < 1) \quad (\text{loi } \Gamma(2, 1));$$

$$g_{X_1-X_2}(u) = g_{X_1}(u)g_{-X_2}(u) = g_{X_1}(u)g_{X_2}(-u) = \frac{1}{1-u} \cdot \frac{1}{1+u} = \frac{1}{1-u^2} \quad (|u| < 1) \quad (\text{première loi de Laplace});$$

$$g_{|X_1-X_2|}(u) = \frac{1}{1-u} \quad (u < 1) \quad (\text{loi } \mathcal{E}(1)).$$

10. a) Désignons par μ la loi de X ; il vient $\mathbb{P}\{Y > X\} = \int_0^\infty \mathbb{P}\{Y > X \mid X = x\} d\mu(x) = \int_0^\infty \mathbb{P}\{Y > x \mid X = x\} d\mu(x)$ d'où, puisque X, Y sont indépendants, $\mathbb{P}\{Y > X\} = \int_0^\infty \mathbb{P}\{Y > x\} d\mu(x) = \int_0^\infty e^{-\lambda x} d\mu(x) = \mathbb{E}[e^{-\lambda X}] = L(\lambda)$.

b) D'après a) on a, puisque les X_k sont indépendantes,

$$\mathbb{P}\{Y > X_1 + \dots + X_n\} = L_{X_1+\dots+X_n}(\lambda) = L_{X_1}(\lambda) \dots L_{X_n}(\lambda) = \mathbb{P}\{Y > X_1\} \dots \mathbb{P}\{Y > X_n\}.$$

c) On a $\Omega = \{M = X_1\} + \dots + \{M = X_n\}$, d'où $\{M > \sum_{k=1}^n X_k - M\} = \{X_1 > \sum_{k \neq 1} X_k\} \cup \dots \cup \{X_n > \sum_{k \neq n} X_k\}$ et $\mathbb{P}\{M > \sum_{k=1}^n X_k - M\} = n\mathbb{P}\{X_1 > \sum_{k \neq 1} X_k\}$.

D'après b), puisque $\mathbb{P}\{X_1 > X_2\} = 1/2$, on en tire :

$$\mathbb{P}\{M > \sum_{k=1}^n X_k - M\} = n[\mathbb{P}\{X_1 > X_2\}]^{n-1} = n\left(\frac{1}{2}\right)^{n-1}.$$

11. $e_r = [\mathbb{E}[X^r]]^{1/r} = (1+r)^{-1/r}$ ($r > 0$); $e_0 = \lim_{r \downarrow 0} e_r = e^{-1}$.

13. a) Le support de X est $[-1, +1]$; pour $x \in [-1, +1]$, sa densité est donc $f(x) = \int_{\mathbb{R}} I_{[0,1]}(u) I_{[0,1]}(x+u) du$. La fonction à intégrer est strictement positive pour $0 < u < 1$ et $0 < x+u < 1$, soit $f(x) = (1-|x|) I_{[-1,+1]}(x)$. Pour obtenir la fonction caractéristique, on opère ainsi : $\varphi_{X_1}(t) = (e^{it} - 1)/(it)$, $\varphi_X(t) = \varphi_{X_1}(t) \varphi_{X_1}(-t) = 2(1 - \cos t)/t^2 = (\sin(t/2)/(t/2))^2$. La variable aléatoire $2X$ a pour densité et fonction caractéristique $f_{2X}(x) = (1/2)f_X(x/2) = (1/2)(1 - (|x|/2))I_{[-2,+2]}(x)$ et $\varphi_{2X}(t) = \varphi_X(2t) = (\sin t/t)^2$.

b) $\varphi_{Y_1}(t) = (\sin t)/t$, $\varphi_Y(t) = \varphi_{Y_1}(t)^2 = ((\sin t)/t)^2$; c'est la fonction caractéristique de $2X$.

Chapitre 15

1. Utiliser le Théorème 1.1. Une solution plus élégante consiste à faire usage de la dernière application du présent chapitre.
2. Utiliser le Théorème 1.1.
3. Utiliser la formule de l'exemple 3 (loi du rapport).

4. Faisons le changement de variable $u = x + y$, $v = x/(x + y)$, qui établit une bijection de $]0, +\infty[\times]0, +\infty[$ sur $]0, +\infty[\times]0, 1[$. On a : $x = uv$, $y = u(1 - v)$, $D(x, y)/D(u, v) = \begin{vmatrix} v & u \\ 1 - v & -u \end{vmatrix} = -u$. La densité conjointe de (X, Y) est donnée par

$$f(x, y) = \frac{\lambda^r}{\Gamma(r)} e^{-\lambda x} x^{r-1} \cdot \frac{\lambda^s}{\Gamma(s)} e^{-\lambda y} y^{s-1} = \frac{\lambda^{r+s}}{\Gamma(r)\Gamma(s)} e^{-\lambda(x+y)} x^{r-1} y^{s-1};$$

la densité conjointe de (U, V) est donc $g(u, v) = f(uv, u(1 - v))u$, soit

$$\begin{aligned} g(u, v) &= \frac{\lambda^{r+s}}{\Gamma(r)\Gamma(s)} (uv)^{r-1} (u(1 - v))^{s-1} e^{-\lambda(uv+u(1-v))} u \\ &= \frac{\lambda^{r+s}}{\Gamma(r+s)} u^{r+s-1} e^{-\lambda u} \cdot \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} v^{r-1} (1 - v)^{s-1}. \end{aligned}$$

Par conséquent, $g(u, v)$ est le produit de la densité de la loi $\Gamma(r + s, \lambda)$ par la densité de la loi Beta(r, s).

5. a) $f(x) = \int_{\mathbb{R}} I_{[0,1]}(v)g(x/v)(1/|v|) dv = \int_0^1 (1/v)g(x/v) dv$.
 b) La variable aléatoire X prend ses valeurs dans $[0, +\infty[$. Supposons $x > 0$ et faisons le changement de variable $x/v = u$. Il vient $f(x) = \int_x^\infty (g(u)/u) du$. De là f est dérivable et $f'(x) = -g(x)/x$ ($x > 0$). Comme, par hypothèse, $g(x) > 0$ pour tout $x > 0$, on voit que $f'(x) < 0$ pour tout $x > 0$. Il en résulte que f est strictement décroissante sur $]0, +\infty[$ et admet donc un mode et un seul en $x = 0$.
 c) La variable aléatoire X prend cette fois-ci ses valeurs dans \mathbb{R} . Reprenons la représentation de f donnée en a) et effectuons toujours le changement de variable $x/v = u$. Nous distinguons deux cas :
 (1) pour $x > 0$, on obtient $f(x) = \int_x^\infty (g(u)/u) du$ ($x > 0$);
 (2) pour $x < 0$, on obtient $f(x) = \int_x^{-\infty} (g(u)/u) du$ ($x < 0$).
 Dans les deux cas, on voit que f est dérivable et que $f'(x) = -g(x)/x$ ($x \neq 0$). Par hypothèse, $g(x) > 0$ pour tout $x \in \mathbb{R}$. Pour $x < 0$ on a donc $f'(x) < 0$ et pour $x > 0$, $f'(x) > 0$. Il en résulte que f est strictement décroissante sur $]0, \infty[$, strictement croissante sur $] -\infty, 0[$. Elle admet donc un mode et un seul en $x = 0$.
 d) La relation $xf'(x) + g(x) = 0$ donne $f'(x) = -(x/\sqrt{2\pi})e^{-x^2/2}$, d'où $f(x) = (1/\sqrt{2\pi})e^{-x^2/2}$.
6. a) Faisons le changement de variable $\begin{pmatrix} u \\ v \end{pmatrix} = A \begin{pmatrix} x \\ y \end{pmatrix}$, $\begin{pmatrix} x \\ y \end{pmatrix} = A^{-1} \begin{pmatrix} u \\ v \end{pmatrix}$, $(D(x, y)/D(u, v)) = \det A^{-1} = \pm 1$. D'où, avec les mêmes notations qu'en 3)

$$g(u, v) = f(x(u, v), y(u, v)) |\det A^{-1}| = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2(u, v) + y^2(u, v))\right).$$

Or, puisque A est orthogonale, on a $x^2 + y^2 = u^2 + v^2$ et donc

$$g(u, v) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(u^2 + v^2)\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) \times \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}v^2\right).$$

- b) On sait que $T = Y/X$ suit la loi de Cauchy $\mathcal{C}(0, 1)$. On peut donc écrire $Z = \frac{a + bT}{c + dT} = \frac{aX + bY}{cX + dY}$. C'est le rapport de deux variables aléatoires indépendantes dont chacune suit la loi $\mathcal{N}(0, 1)$, c'est donc une variable aléatoire qui suit la loi de Cauchy $\mathcal{C}(0, 1)$.

7. a) La densité conjointe de (X, Y) est $f(x, y) = (1/(2\pi))e^{-(x^2+y^2)/2}$. Faisons le changement de variable $\begin{cases} u = 2x \\ v = x - y \end{cases}$ qui établit une bijection de \mathbb{R}^2 sur \mathbb{R}^2 : $\begin{cases} x = u/2 \\ y = (u/2) - v \end{cases}$, $\frac{D(x, y)}{D(u, v)} = \begin{vmatrix} 1/2 & 0 \\ 1/2 & -1 \end{vmatrix} = -\frac{1}{2}$. La densité conjointe $g(u, v)$ de (U, V) est donc égale à :

$$g(u, v) = f\left(\frac{u}{2}, \frac{u}{2} - v\right) \frac{1}{2} = \frac{1}{2} \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left(\left(\frac{u}{2}\right)^2 + \left(\frac{u}{2} - v\right)^2\right)\right).$$

Les densités marginales s'en déduisent :

$$g(u, \cdot) = \int_{\mathbb{R}} g(u, v) dv = \frac{1}{2} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{u^2}{4}\right) \quad (\text{loi } \mathcal{N}(0, \sigma = 2));$$

$$g(\cdot, v) = \int_{\mathbb{R}} g(u, v) du = \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{v^2}{2}\right) \quad (\text{loi } \mathcal{N}(0, \sigma = \sqrt{2})).$$

- b) La densité conditionnelle cherchée est :

$$g_{U|V}(u, 0) = \frac{g(u, 0)}{g(\cdot, 0)} = \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{u^2}{2}\right) \quad (\text{loi } \mathcal{N}(0, \sigma = \sqrt{2})).$$

- c) Posons $W = X + Y$, $V = X - Y$. En faisant le changement de variable $\begin{cases} w = x + y \\ v = x - y \end{cases}$, un calcul analogue au précédent fournit

$$g_{W|V}(w, 0) = \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{w^2}{2}\right) \quad (\text{loi } \mathcal{N}(0, \sigma = \sqrt{2})).$$

- d) On constate effectivement que $\mathcal{L}(2X | X - Y = 0) = \mathcal{L}(X + Y | X - Y = 0) = \mathcal{L}(X + Y)$. La raison en est que, dans les hypothèses de l'énoncé, les variables aléatoires $X + Y$ et $X - Y$ sont *indépendantes* et que, en outre, conditionnellement à $X - Y = 0$, on a $X + Y = 2X$.

8. Par récurrence.

9. Il suffit de montrer b). Soit A un borélien de \mathbb{R} . On a : $P\{Y \in A\} = \frac{1}{2}P\{X \in A\} + \frac{1}{2}P\{(1/X) \in A\}$. Or d'après a), $P\{(1/X) \in A\} = P\{X \in A\}$, d'où $P\{Y \in A\} = P\{X \in A\}$.

10. La densité conjointe de (X, Y) est $f(x, y) = (1/(2\pi))e^{-(x^2+y^2)/2}$. Faisons le changement de variable $\begin{cases} u = xy \\ v = x/y \end{cases}$, qui établit une application

de \mathbb{R}^2 dans D , avec $D = \{(u, v) : u \geq 0, v \geq 0, \text{ ou } u \leq 0, v \leq 0\}$. Cette application n'est cependant *pas* une bijection de \mathbb{R}^2 sur D (les couples (x, y) et $(-x, -y)$ ont même image). Comme $x^2 = uv$ et $y^2 = u/v$, prenons $x = +\sqrt{uv}$, $y = +\sqrt{u/v}$, d'où $\frac{D(u, v)}{D(x, y)} = \left| \begin{array}{cc} y & x \\ 1/y & -x/y^2 \end{array} \right| = -2(x/y) = -2v$ et $\frac{D(x, y)}{D(u, v)} = -1/(2v)$. Comme tout élément (u, v) de D est l'image de *deux* éléments de \mathbb{R}^2 , on a

$$\begin{aligned} g(u, v) &= 2f(x(u, v), y(u, v)) \frac{1}{2|v|} = 2 \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left(uv + \frac{u}{v}\right)\right) \frac{1}{2|v|} \\ &= \frac{1}{2\pi} \exp\left(-\frac{u}{2}\left(v + \frac{1}{v}\right)\right) \frac{1}{|v|} \quad ((u, v) \in D). \end{aligned}$$

Densité marginale de U :

Pour $u \geq 0$: $g(u, \cdot) = \frac{1}{2\pi} \int_0^\infty \exp\left(-\frac{u}{2}\left(v + \frac{1}{v}\right)\right) \frac{dv}{v}$.

Pour $u \leq 0$: $g(u, \cdot) = \frac{1}{2\pi} \int_{-\infty}^0 \exp\left(-\frac{u}{2}\left(v + \frac{1}{v}\right)\right) \frac{dv}{-v}$
 $= \frac{1}{2\pi} \int_0^\infty \exp\left(\frac{u}{2}\left(t + \frac{1}{t}\right)\right) \frac{dt}{t}$.

D'où

$$g(u, \cdot) = \frac{1}{2\pi} \int_0^\infty \exp\left(-\frac{|u|}{2}\left(t + \frac{1}{t}\right)\right) \frac{dt}{t} \quad (u \in \mathbb{R}).$$

Densité marginale de V :

Pour $v \geq 0$: $g(\cdot, v) = \frac{1}{2\pi} \frac{1}{v} \int_0^\infty \exp\left(-\frac{u}{2}\left(v + \frac{1}{v}\right)\right) du$

soit en posant $\frac{u}{2}\left(v + \frac{1}{v}\right) = t$,

$$g(\cdot, v) = \frac{1}{\pi} \frac{1}{v} \int_0^\infty e^{-t} \frac{dt}{v + (1/v)} = \frac{1}{\pi} \frac{1}{1 + v^2}$$

Pour $v \leq 0$: $g(\cdot, v) = \frac{1}{2\pi} \frac{1}{-v} \int_{-\infty}^0 \exp\left(-\frac{u}{2}\left(v + \frac{1}{v}\right)\right) du$,

soit avec $\frac{u}{2}\left(v + \frac{1}{v}\right) = t$,

$$g(\cdot, v) = \frac{1}{\pi} \frac{1}{-v} \int_{+\infty}^0 \frac{e^{-t} dt}{v + (1/v)} = \frac{1}{\pi} \frac{1}{1 + v^2} \int_0^\infty e^{-t} dt.$$

D'où

$$g(\cdot, v) = \frac{1}{\pi} \frac{1}{1 + v^2} \quad (v \in \mathbb{R}).$$

12. a) $\mathcal{L}(U) = \mathcal{C}(0, 1)$; b) $\mathcal{L}(Z) = \mathcal{C}(0, 1)$.

Chapitre 16

1. a) Considérons l'inégalité suivante, valable pour tout $n \geq 1$ et tout $\varepsilon > 0$:

$$P\{|X_n + Y_n - X - Y| > 2\varepsilon\} \leq P\{|X_n - X| > \varepsilon\} + P\{|Y_n - Y| > \varepsilon\}.$$

Le résultat en découle en laissant ε fixe et en faisant tendre n vers l'infini.

b) Considérons l'inégalité suivante, valable pour tout $n \geq 1$ et tout $\varepsilon > 0$:

$$\begin{aligned} P\{|X_n Y_n - XY| > 3\varepsilon\} &\leq P\{|X_n - X| |Y| > \varepsilon\} \\ &\quad + P\{|Y_n - Y| |X| > \varepsilon\} + P\{|X_n - X| |Y_n - Y| > \varepsilon\}. \end{aligned}$$

Majorons le premier terme du second membre; on a pour tout $\varepsilon > 0$ et tout $A > 0$:

$$P\{|X_n - X| |Y| > \varepsilon\} \leq P\{|X_n - X| > \varepsilon/A\} + P\{|Y| > A\}.$$

On peut choisir A assez grand pour que $P\{|Y| > A\} < \eta$; le nombre A étant ainsi choisi, on peut prendre n assez grand pour que $P\{|X_n - X| > \varepsilon/A\} < \eta$, de sorte que $P\{|X_n - X| |Y| > \varepsilon\} < 2\eta$. Le deuxième terme du second membre se traite de la même façon. Enfin le troisième terme du second membre tend vers 0 lorsque n tend vers l'infini, puisque

$$P\{|X_n - X| |Y_n - Y| > \varepsilon\} \leq P\{|X_n - X| > \sqrt{\varepsilon}\} + P\{|Y_n - Y| > \sqrt{\varepsilon}\}.$$

2. a) On a, en vertu du Lemme 5.3, pour tout $\eta > 0$, l'inégalité

$$|F_{X_n + Y_n}(x) - F_{X_n}(x)| \leq F_{X_n}(x + \eta) - F_{X_n}(x - \eta) + P\{|Y_n| > \eta\}.$$

On conclut en prenant pour x , $x - \eta$, $x + \eta$ ($\eta > 0$) des points de continuité de la fonction de répartition F de X et en laissant tendre n vers l'infini.

b) On a, pour tout $A > 0$ et tout $\varepsilon > 0$

$$P\{|X_n Y_n| > A\varepsilon\} \leq P\{|X_n| > A\} + P\{|Y_n| > \varepsilon\}.$$

Puisque $X_n \xrightarrow{\mathcal{L}} X$ (où X est une variable aléatoire), le premier terme du second membre peut être rendu inférieur à η pour A , n assez grands. Puisque $Y_n \xrightarrow{p} 0$ le second terme tend vers 0 lorsque n tend vers l'infini. Ces deux points permettent de conclure.

3. a) La série de terme général $\mathbb{E}[|X_n|^{1/2}] = n^{-3/2}$ est convergente; donc $X_n \xrightarrow{p.s.} 0$, d'après le deuxième critère de convergence presque sûre.

b) $\mathbb{E}[X_n^2] = 1$; donc X_n ne tend pas vers 0 en moyenne quadratique.

4. Supposons que $X_n \xrightarrow{p} X$ et qu'il existe $C > 0$ tel que pour tout $n \geq 1$ on ait $P\{|X_n| \leq C\} = 1$. Il en résulte que $P\{|X| \leq C\} = 1$, puisque, pour tout $\varepsilon > 0$, on a $P\{|X| \leq C + \varepsilon\} = \lim_{n \rightarrow \infty} P\{|X_n| \leq C + \varepsilon\} = 1$. Posons $E_n(\varepsilon) = \{|X_n - X| > \varepsilon\}$. On a $|X_n - X|^r \leq \varepsilon^r I_{E_n^c(\varepsilon)} + (2C)^r I_{E_n(\varepsilon)}$, d'où, en prenant l'espérance mathématique et en faisant tendre n vers l'infini $\limsup \mathbb{E}[|X_n - X|^r] \leq \varepsilon^r$.
6. a) Avec $B = \frac{t + \text{Log } g(\varepsilon)}{\varepsilon}$ on a :
- $$g(\varepsilon) \geq \int_{\{X \geq B\}} e^{\varepsilon X} dP \geq e^{\varepsilon B} \int_{\{X \geq B\}} dP = e^t g(\varepsilon) P\{X \geq B\}.$$
7. a) Utiliser le deuxième critère de convergence presque sûre.
 b) Si $\sum_{n \geq 1} \mathbb{E}[X_n^2] < \infty$, alors $\mathbb{E}[X_n^2] \rightarrow 0$ lorsque n tend vers l'infini, d'où le résultat.
8. $\mathbb{E}\left[\left(\frac{X_n}{\mu_n} - 1\right)^2\right] = \text{Var} \frac{X_n}{\mu_n} = \frac{\sigma_n^2}{(\mu_n)^2} = \frac{O(1)}{|\mu_n|} \rightarrow 0$. Ainsi $\frac{X_n}{\mu_n} - 1$ tend vers 0 en moyenne quadratique, donc aussi en probabilité.
9. Puisque $X_n \xrightarrow{p} 0$, on peut extraire de la suite (X_n) une suite partielle (X_{n_k}) telle que $X_{n_k} \downarrow 0$ p.s. Or ceci entraîne $X_n \downarrow 0$ p.s., puisque la suite (X_n) est décroissante.
10. Pour tout $\varepsilon > 0$, on a $P\{X_n > \varepsilon\} \leq 1/n \rightarrow 0$, alors que, pour tout $n \geq 1$, on a $\mathbb{E}[X_n^2] = +\infty$, puisque l'intégrale $\int_0^{1/n} (1/x) dx$ est divergente.
11. a) $X_n \xrightarrow{p} 0$, car pour tout $\varepsilon > 0$, on a $P\{X_n > \varepsilon\} \leq P\{X_n > 0\} = 1/n \rightarrow 0$.
 b) Montrons que la suite (Y_n) ne converge *pas* vers 0 en probabilité. Remarquons, tout d'abord, que $\{Y_{2n} \geq 1/2\} \supset \bigcup_{n < k \leq 2n} \{X_k = k\}$, d'où $P\{Y_{2n} \geq 1/2\} \geq P\left(\bigcup_{n < k \leq 2n} \{X_k = k\}\right) = 1 - P\left(\bigcap_{n < k \leq 2n} \{X_k = 0\}\right) = 1 - \prod_{n < k \leq 2n} (1 - (1/k)) \geq 1 - \prod_{n < k \leq 2n} \exp(-1/k) = 1 - \exp\left(-\sum_{n < k \leq 2n} (1/k)\right) \geq 1 - e^{-1/2} > 0$. Ainsi $P\{Y_{2n} \geq 1/2\}$ ne tend *pas* vers 0 lorsque n tend vers l'infini ; la suite (Y_n) ($n \geq 1$) ne tend donc *pas* vers 0 en probabilité.
12. Le support de Z_n est $[0, n]$. Pour $n \geq x$ on a $P\{Z_n > x\} = P\{\min(U_1, \dots, U_n) > x/n\} = P\{U_1 > x/n, \dots, U_n > x/n\} = (P\{U > x/n\})^n = (1 - (x/n))^n$, qui tend vers e^{-x} lorsque n tend vers l'infini.
13. Le support de $e^{-\lambda X}$ est $]0, 1[$. Pour $0 < x < 1$ on a $P\{e^{-\lambda X} < x\} = P\{X > -(\text{Log } x/\lambda)\} = e^{-\lambda(-\text{Log } x/\lambda)} = x$; soit $\mathcal{L}(e^{-\lambda X}) = \mathcal{L}(U)$.
14. a) D'après les exercices 12 et 13, on a $\mathcal{L}(A_n) = \mathcal{L}(Z_n)$, où Z_n désigne la v.a. définie à l'exercice 12; d'où, d'après l'exercice 12, $A_n \xrightarrow{\mathcal{L}} Y$, où Y est une variable aléatoire exponentielle de paramètre 1.

- b) Comme $B_n = (A_n)^{1/\lambda}$, on a $B_n \xrightarrow{\mathcal{L}} Y^{1/\lambda}$. La loi limite est la loi de *Weibull* (cf. exercice 3 du chap. 14), dont la fonction de survie pour $x > 0$ est donnée par $P\{Y^{1/\lambda} > x\} = P\{Y > x^\lambda\} = e^{-x^\lambda}$.
- c) Comme $C_n = (A_n)^{-1/\lambda}$, on a $C_n \xrightarrow{\mathcal{L}} Y^{-1/\lambda}$. La loi limite est la loi de *Fréchet*, dont la fonction de répartition pour $x > 0$ est donnée par $P\{Y^{-1/\lambda} < x\} = P\{Y > x^{-\lambda}\} = e^{-(x^{-\lambda})}$.
- d) On a $D_n = -\text{Log } A_n$; d'où $D_n \xrightarrow{\mathcal{L}} -\text{Log } Y$. La loi limite est la loi de *Gumbel*, dont la fonction de répartition pour tout x est donnée par $P\{-\text{Log } Y < x\} = P\{Y > e^{-x}\} = e^{-(e^{-x})}$.
15. a) Pour $x > 0$, on a : $P\{Z_n \leq x\} = (P\{X \leq nx\})^n = (1 - P\{X > nx\})^n = (1 - o(1/(nx)))^n$. D'où $\text{Log } P\{Z_n \leq x\} = n \text{Log}(1 - o(1/(nx)))$, qui tend vers 0 lorsque n tend vers l'infini.
- b) Pour $x > 0$, on a : $P\{Z_n \leq x\} = (1 - P\{X > n^{1/\lambda}x\})^n = \left(1 - \frac{\alpha}{nx^\lambda} + o(1/n)\right)^n$, qui tend vers $e^{-\alpha x^{-\lambda}}$, lorsque n tend vers l'infini.
Exemples pour a) : $\mathcal{L}(X) = \mathcal{E}(\theta)$ ($\theta > 0$), $X = |Y|$, avec $\mathcal{L}(Y) = \mathcal{N}(0, 1)$.
Exemples pour b) : $X = e^Y$, avec $\mathcal{L}(Y) = \mathcal{E}(\theta)$ ($\theta > 0$) et $\alpha = 1$, $\lambda = \theta$; ou bien $X = |Y|$ avec $\mathcal{L}(Y) = \mathcal{C}(0, 1)$ et $\alpha = 2/\pi$, $\lambda = 1$; ou bien encore $X = Y - 1$, avec $\mathcal{L}(Y) = \text{Pareto}(1, 1)$ et $\alpha = 1$, $\lambda = 1$.

Chapitre 17

1. a) On a $\mathbb{E}[Y_n] = m$, $\text{Var } Y_n = \sigma^2/n$. Pour calculer $\mathbb{E}[Z_n]$, on supposera $m = 0$ (en effet, Z_n ne dépend pas de m). Un calcul élémentaire montre que $\sum_{k=1}^n (X_k - Y_n)^2 = \sum_{k=1}^n X_k^2 - nY_n^2$, d'où $\mathbb{E}\left[\sum_{k=1}^n (X_k - Y_n)^2\right] = n\sigma^2 - \sigma^2 = (n-1)\sigma^2$ et $\mathbb{E}[Z_n] = \sigma^2$.
- b) $Z_n = (1/(n-1)) \sum_{k=1}^n X_k^2 - (n/(n-1))Y_n^2$. Or en utilisant la loi forte des grands nombres de Kolmogorov (Théorème 2.3), on obtient $(1/n) \sum_{k=1}^n X_k^2 \xrightarrow{p.s.} \mathbb{E}[X_1^2]$, puisque $\mathbb{E}[X_1^2] < \infty$ et $Y_n \xrightarrow{p.s.} \mathbb{E}[X_1]$, puisque $\mathbb{E}[|X_1|] < \infty$; d'où $Z_n \xrightarrow{p.s.} \mathbb{E}[X_1^2] - (\mathbb{E}[X_1])^2 = \sigma^2$.
2. Désignons par φ la fonction caractéristique commune aux X_n . Puisque X_n est dans L^1 et qu'elle est centrée, on a $\varphi'(0) = 0$. Or $\varphi_{Y_n}(t) = (\varphi(t/n))^n$. En posant $\psi = \text{Log } \varphi$, on obtient $\psi_{Y_n}(t) = n\psi(t/n) = t(\psi(t/n) - \psi(0))/(t/n)$. En faisant tendre n vers l'infini, il vient $\psi_{Y_n}(t) \rightarrow t\psi'(0) = t\varphi'(0)/\varphi(0) = 0$; d'où $\varphi_{Y_n}(t) \rightarrow 1$ et $Y_n \xrightarrow{\mathcal{L}} 0$. La valeur limite étant une constante, la dernière propriété est équivalente à $Y_n \xrightarrow{p} 0$.
3. a) Comme $X_n/n = Y_n - ((n-1)/n)Y_{n-1}$, on a $X_n/n \xrightarrow{p.s.} 0$ ($n \rightarrow \infty$). Il en résulte qu'avec une probabilité égale à 1, l'évènement $A_n = \{|X_n| \geq n\}$

ne se réalise que pour un nombre *fini* d'indices n . Les X_n étant indépendantes, les évènements A_n sont indépendants; il résulte alors du lemme de Borel-Cantelli que $\sum_{n \geq 1} P(A_n) < \infty$.

- b) Les X_n étant identiquement distribuées, il suffit de montrer que $\mathbb{E}[|X_1|] < \infty$. Or $\mathbb{E}[|X_1|] = \int_0^{+\infty} P\{|X_1| \geq x\} dx$. En posant $H_k = \{x \in \mathbb{R}^+ : k \leq x < k + 1\}$, on a $\mathbb{E}[|X_1|] = \sum_{k \geq 0} \int_{H_k} P\{|X_1| \geq x\} dx \leq \sum_{k \geq 0} P\{|X_1| \geq k\} \int_{H_k} dx = \sum_{k \geq 0} P\{|X_1| \geq k\} < +\infty$.
- c) Les X_n étant dans L^1 , il résulte de la loi forte des grands nombres de Kolmogorov que, presque sûrement, $Y = \mathbb{E}[X_1] = \text{constante}$.
4. Pour tout $\varepsilon > 0$ on a $P\{|S_n/n| > \varepsilon\} = P\{|S_n/\sqrt{n}| > \varepsilon\sqrt{n}\}$. Or, pour tout $N > 0$, il existe N_0 tel que pour tout $n \geq N_0$ on ait $\varepsilon\sqrt{n} > N$. Par suite $P\{|S_n/n| > \varepsilon\} \leq P\{|S_n/\sqrt{n}| > N\} \rightarrow \int_{|x| > N} d\mu(x)$, où μ est la loi de Y . Puisque N est arbitraire, il en résulte que $P\{|S_n/n| > \varepsilon\} \rightarrow 0$.
6. a) La densité $g_n(x, R)$ est égale à $V_{n-1}(\sqrt{R^2 - x^2})I_{[-R, +R]}(x)$, d'où

$$f_n(x, R) = \frac{V_{n-1}(\sqrt{R^2 - x^2})}{V_n(R)} = \frac{1}{\sqrt{\pi}} \frac{\Gamma(1 + n + 2)}{\Gamma(1 + (n - 1)/2)} \frac{1}{R} \left(1 - \frac{x^2}{R^2}\right)^{(n-1)/2} I_{[-R, +R]}(x),$$

soit, en prenant $R = \sqrt{n}$, la densité

$$f_n(x, \sqrt{n}) = \frac{1}{\sqrt{n}} \frac{\Gamma(1 + n/2)}{\Gamma(1 + (n - 1)/2)} \left(1 - \frac{x^2}{R^2}\right)^{(n-1)/2} I_{[-\sqrt{n}, +\sqrt{n}]}(x).$$

- b) En utilisant la formule de Stirling $\Gamma(1 + p) \sim (p/e)^p \sqrt{2\pi p}$ ($p \rightarrow \infty$), on obtient après un calcul simple

$$\frac{1}{\sqrt{n}} \frac{\Gamma(1 + n/2)}{\Gamma(1 + (n - 1)/2)} \rightarrow \frac{1}{\sqrt{2}} \quad (n \rightarrow \infty).$$

- c) Prenons x réel et l'entier n assez grand pour que $|x| < \sqrt{n}$. On voit que $\left(1 - \frac{x^2}{n}\right)^{(n-1)/2} \rightarrow e^{-x^2/2}$ lorsque n tend vers l'infini.

On sait, d'autre part, que pour de grandes valeurs de n , le volume de la boule $B_n(0, R)$ a tendance à se concentrer sur la périphérie; il suffit pour le voir de remarquer que pour tout $h > 0$ on a $\frac{V_n(R)}{V_n(R + h)} = \left(\frac{R}{R + h}\right)^n \rightarrow 0$, lorsque n tend vers l'infini. On peut donc s'attendre au résultat suivant :

Projetons la surface $A_n(R)$ de la boule $B_n(0, R)$ sur l'axe des x : on obtient une distribution de masses de densité $g_n^*(x, R)$, que l'on norme pour avoir une densité de *probabilité* $f_n^*(x, R) = \frac{g_n^*(x, R)}{A_n(R)}$. Pour $R =$

\sqrt{n} , la suite des densités $f_n^*(x, \sqrt{n})$ converge ponctuellement, lorsque n tend vers l'infini, vers la densité de la loi normale $\mathcal{N}(0, 1)$. Cette conjecture est effectivement vraie et un calcul analogue au précédent permet de la démontrer.

7. Il suffit de démontrer 3). Pour tout $n \geq 1$ posons $A_n = \{X_n = 0\}$. Alors $\{X_n \rightarrow 0\} = \liminf_{n \rightarrow \infty} A_n = (\limsup_{n \rightarrow \infty} A_n^c)^c$; d'où $P\{X_n \rightarrow 0\} = 1 \Leftrightarrow P(\limsup_{n \rightarrow \infty} A_n^c) = 0$. Or, les A_n étant indépendants, on a, en vertu du Lemme de Borel-Cantelli : $P(\limsup_{n \rightarrow \infty} A_n^c) = 0 \Leftrightarrow \sum_{n \geq 1} P(A_n^c) < \infty \Leftrightarrow \sum_{n \geq 1} u_n < +\infty$.

Chapitre 18

1. La fonction génératrice de X_λ est $g_0(u) = \exp(\lambda(e^u - 1))$; celle de $(X_\lambda - \lambda)/\sqrt{\lambda}$ est donc $g(u) = e^{-u\sqrt{\lambda}} \exp(\lambda(e^{u/\sqrt{\lambda}} - 1))$. D'où $\text{Log } g(u) = -u\sqrt{\lambda} - \lambda + \lambda e^{u/\sqrt{\lambda}} = -u\sqrt{\lambda} - \lambda + \lambda \left(1 + \frac{u}{\sqrt{\lambda}} + \frac{u^2}{2\lambda} + o\left(\frac{1}{\lambda}\right)\right) = \frac{u^2}{2} + o(1) \rightarrow \frac{u^2}{2}$; ainsi $g(u) \rightarrow e^{u^2/2}$.
2. La fonction génératrice de X_p est $g_0(u) = (\lambda/(\lambda - u))^p$ ($u < \lambda$); donc celle de $\frac{X_p - (p/\lambda)}{\sqrt{p}/\lambda}$ est $g(u) = e^{-u\sqrt{p}} g_0\left(\frac{\lambda}{\sqrt{p}}u\right) = e^{-u\sqrt{p}} \left(\frac{1}{1 - \frac{u}{\sqrt{p}}}\right)^p$. D'où $\text{Log } g(u) = -u\sqrt{p} - p \text{Log}\left(1 - \frac{u}{\sqrt{p}}\right) = -u\sqrt{p} - p\left(-\frac{u}{\sqrt{p}} - \frac{u^2}{2p} + o\left(\frac{1}{p}\right)\right) = \frac{u^2}{2} + o\left(\frac{1}{p}\right) \rightarrow \frac{u^2}{2}$; ainsi $g(u) \rightarrow e^{u^2/2}$.
3. a) $\mathbb{E}[S_n] = \sum_{k=1}^n \frac{1}{k} \sim \log n$, $C_n^2 = \text{Var } S_n = \sum_{k=1}^n \frac{1}{k} \left(1 - \frac{1}{k}\right) \sim \text{Log } n$.
- b) Posons $X'_n = X_n - 1/n$, $S'_n = \sum_{k=1}^n X'_k$. On a $\text{Var } S'_n = \text{Var } S_n = C_n^2$. La suite (X'_n) ($n \geq 1$) est une suite de variables aléatoires indépendantes, centrées, de L^2 . Elle vérifie, de plus, la condition de Liapounov pour $\delta = 1$; en effet,

$$\begin{aligned} \mathbb{E}[|X'_k|^3] &= \frac{1}{k} \left(1 - \frac{1}{k}\right)^3 + \left(1 - \frac{1}{k}\right) \left(\frac{1}{k}\right)^3 = \frac{1}{k} \left(1 - \frac{1}{k}\right) \left[\left(1 - \frac{1}{k}\right)^2 + \left(\frac{1}{k}\right)^2\right] \\ &\leq \frac{1}{k} \left(1 - \frac{1}{k}\right) \leq \frac{1}{k}, \end{aligned}$$

d'où

$$\frac{1}{(C_n)^3} \sum_{k=1}^n |X'_k|^3 \leq \frac{1}{(C_n)^3} \sum_{k=1}^n \frac{1}{k} \sim \frac{1}{\sqrt{\text{Log } n}} \rightarrow 0 \quad (n \rightarrow \infty).$$

Il en résulte que $S'_n/C_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$, d'où aussi $Y_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$.

4. Soit (X_k) ($k \geq 1$) une suite de variables aléatoires indépendantes de loi commune $p\varepsilon_1 + q\varepsilon_0$. Posons $S_n = \sum_{k=1}^n X_k$. La somme dont on calcule la limite est égale, vu que S_n suit la loi binomiale $B(n, p)$, à $P\{[n/2] + 1 \leq S_n \leq n\} = P\left\{\frac{[n/2] + 1 - np}{\sqrt{npq}} \leq \frac{S_n - np}{\sqrt{npq}} \leq \frac{n - np}{\sqrt{npq}}\right\}$, une expression qui pour n grand est équivalente à : $P\left\{\frac{(n/2) - np}{\sqrt{npq}} \leq \frac{S_n - np}{\sqrt{npq}} \leq \frac{nq}{\sqrt{npq}}\right\} = P\left\{-\frac{p - (1/2)}{\sqrt{pq}}\sqrt{n} \leq \frac{S_n - np}{\sqrt{npq}} \leq \sqrt{\frac{q}{p}}\sqrt{n}\right\}$. D'après le théorème central limit, cette dernière expression est équivalente à $\frac{1}{\sqrt{2\pi}} \int_{-(p-(1/2))\sqrt{n}/\sqrt{pq}}^{\sqrt{(q/p)}\sqrt{n}} e^{-x^2/2} dx$, qui, puisque $p > \frac{1}{2}$, tend vers $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2} dx = 1$.

Remarque. — Interprétons X_k comme la fonction indicatrice de la voix recueillie pour un candidat A par le $k^{\text{ième}}$ électeur. S'il y a n électeurs, S_n représente le nombre de voix recueillies par A . Le précédent résultat montre que si $p > \frac{1}{2}$, c'est-à-dire si la probabilité est *a priori* favorable au candidat A , alors la probabilité pour que la décision, prise à la majorité d'une seule voix, soit favorable à A , tend vers 1, lorsque le nombre des votants tend vers l'infini.

5. a) Désignons par F_n, F les fonctions de répartition de Y_n, Y . Pour tout point de continuité x de F et tout $n \geq 1$, on a $P\{Y_{N_n} \leq x\} = \sum_{k \geq 1} P\{Y_{N_n} \leq x, N_n = k\} = \sum_{k \geq 1} P\{Y_k \leq x, N_n = k\}$. En vertu de l'indépendance des suites $(Y_n), (N_n)$, on en déduit $P\{Y_{N_n} \leq x\} = \sum_{k \geq 1} P\{Y_k \leq x\}P\{N_n = k\} = \sum_{k \geq 1} F_k(x)P\{N_n = k\}$, d'où

$$P\{Y_{N_n} \leq x\} - F(x) = \sum_{k \geq 1} (F_k(x) - F(x))P\{N_n = k\} = \sum_{k=1}^M (F_k(x) - F(x))P\{N_n = k\} + \sum_{k > M} (F_k(x) - F(x))P\{N_n = k\} = J_1 + J_2.$$

Supposons à présent que $Y_n \xrightarrow{\mathcal{L}} Y$.

On peut choisir M assez grand pour que pour tout $k \geq M$ on ait $|F_k(x) - F(x)| < \varepsilon$, d'où $|J_2| < \varepsilon$.

Le nombre M ainsi fixé, faisons tendre n vers l'infini. Puisque par hypothèse $N_n \xrightarrow{p} +\infty$, on a, pour tout $k \geq 1$, $\lim_{n \rightarrow \infty} P\{N_n = k\} = 0$.

On peut donc choisir n_0 assez grand pour que pour tout $n > n_0$ on ait $|J_1| < \varepsilon$.

- b) C'est une conséquence directe de a), puisque d'après le Théorème «central limit», on a $Y_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$.

INDEX

- σ -additive (mesure —) 15, 114
- σ -algèbre 8
- σ -corps 8
- σ -finie (mesure —) 114

- Abel (lemme d'—) 103
- absence de mémoire 184
- absolument continue (fonction de répartition —) 134, 143
- additivité
 - simple 23
 - (σ - —) 15, 114
- algèbre 7
 - (σ - —) 8
- incipitale 172, 220
- anniversaires 31
- application réciproque 43
- approche
 - de Gauss 237
 - de Laplace 239
- approximation (lemme d'—) 121
- arc sinus (lois —) 190
- arithmétique (densité —) 39
- arrangement
 - avec répétitions 29
 - sans répétition 30
- atome 66, 281, 282, 283

- Banach
 - (problème des boîtes d'allumettes de —) 76
- Bayes (formule de —) 55
- Beppo Levi (théorème de convergence monotone de —) 125
- Bernoulli 225
 - (loi de —) 68
 - (processus de —) 95
- Bernstein 234, 242, 273

- bêta
 - (fonction —) 74
 - (loi —) 189
- Bienaymé (inégalité de —) 90
- binomial (coefficient —) 20
- binomiale
 - (identité —) 20
 - (loi —) 68
 - (loi — négative) 74, 112
- Boole
 - (algèbre de —) 7
 - (inégalité de —) 17
- Borel
 - (loi (0, 1) de —) 233
 - (σ -corps de —) 8
- Borel-Cantelli
 - (lemmes de —) 231
- borélienne (tribu —) 9
- Bose-Einstein
 - (modèle de —) 42
- boule $B_n(0, R)$ 235

- caractéristique
 - (fonction —) 164
- Carathéodory 115
- cardinal 27
- cardinalité 27
- Cauchy (loi de —) 186
- central limit (théorème —) 241
- centrée
 - (variable aléatoire —) 86, 130
- centrée réduite
 - (variable aléatoire —) 86
- certain (évènement —) 3
- chaînes de Markov 127
- changement de
 - variables (formule du —) 198
- Chernoff (inégalité de —) 176
- Chi-deux (loi du —) 189

- Chu-Vandermonde
 (identité de —) 20
 classe monotone 11
 — engendrée 11
 classes
 indépendantes d'évènements 60
 coefficient
 binomial 20, 31
 de corrélation 89
 multinomial 34
 commutativité complète 82
 complet (système —) 54
 complète
 (commutativité —) 82
 (mesure —) 114
 complétée (mesure —) 116
 condition
 de Liapounov 247
 de Lindeberg 244
 conditionnelle
 (densité —) 145
 (espérance
 mathématique —) 145
 (loï de probabilité —) 144
 (mesure de probabilité —) 54
 (probabilité —) 54, 150
 conditionnement
 (formule du double —) 54
 conjointe
 (fonction de répartition —) 142
 (loï —) 141
 conjonction d'évènements 3
 continuité (théorème de —) 106
 contraire (évènement —) 3
 convergence
 dominée 126
 en loi 205
 en moyenne d'ordre r 209
 en probabilité 207
 étroite 205
 monotone 125
 presque sûre 209
 convolution 81, 99, 200
 corps
 (σ - —) 8
 de Borel 8
 corrélation (coefficient de —) 89
 courbe de régression 147
 covariance 88
 croissante
 (dénombrément de suites — s) 33
 (suite — d'évènements) 3
 cumulant 169

 Darmois (Georges) 240
 décisions judiciaires 70
 défaillance (taux de —) 192
 dégénérée
 (loï normale —) 152
 dénombrément 29
 densité
 arithmétique 39
 conditionnelle 145
 conjointe 143
 de probabilité 135
 densités
 (produit de convolution de —) 138
 dérangements (nombre de —) 39
 différence
 d'évènements 3
 propre d'évènements 3
 symétrique d'évènements 6
 diffuse
 (loï de probabilité —) 50, 135, 136
 Dirac (mesure de —) 25
 discrète
 (mesure de probabilité —) 25
 (variable aléatoire —) 67
 disjoints (évènements —) 3
 distance en variation 263
 droite achevée 72, 120
 Dynkin
 (système de —) 10, 58
 (système de — engendré) 11
 (système de — faible) 39

 écart
 d'ordre r 88
 moyen minimum 94

- écart médian 94
- écart-type 86
- engendré
 - (classe monotone — e) 11
 - (système de Dynkin —) 11
 - (tribu — e) 9
- ensemble
 - de toutes les parties 2, 29
 - fondamental 2
- entraîne
 - (l'évènement $A - B$) 3
- épreuve 1
- équirépartition 26
- équivalents (évènements —) 3
- espace
 - mesurable 10, 113
 - mesuré 114
 - probabilisable 10
 - probabilisé 15
 - (produit d' s — s probabilisés) 127
- espérance mathématique 82, 129
 - conditionnelle 145, 146
- étagée (variable aléatoire —) 68, 121
- Euler (formule
 - arithmétique d'—) 37
- évènement 1
 - certain 3
 - élémentaire 2
 - impossible 3
- exponentielle (loi —) 182
- extérieure (mesure —) 115

- factoriel (moment —) 87, 104
- factorielle montante 19
- Fatou
 - (inégalités de —) 23
 - (lemme de —) 125
- Fermi-Dirac (modèle de —) 42
- fiabilité (fonction de —) 182, 273
- fini (ensemble —) 26
- finie (mesure σ —) 114
- fonction
 - arithmétique d'Euler 37
 - caractéristique 164
 - caractéristique d'un couple 171
 - de fiabilité 182
 - de masse 49
 - de répartition 48
 - de répartition conjointe 142
 - de répartition d'un
 - vecteur aléatoire 52
 - de survie 71, 136, 182
 - de variables aléatoires 195
 - de vraisemblance 238
 - génératrice 100
 - génératrice des moments 159, 170
 - hypergéométrique 20
 - mesurable 44
- fondamental (ensemble —) 2
- formule
 - de Bayes 55
 - de la somme 27
 - de Poincaré 16, 37, 259
 - de Stirling 243
 - de Wald 96
 - du changement de variables 198
 - du produit 28
- fractions continues 269
- Fréchet (loi de —) 223, 320
- Fubini (théorème de —) 132

- gamma
 - (fonction —) 19, 187
 - (loi —) 187
- Gauss
 - (approche de —) 237
 - (identité de —) 78
 - (loi de Laplace —) 178
 - (loi de Laplace —
 - à deux dimensions) 151
- génératrice
 - (fonction —) 100
 - (fonction — des
 - moments) 159, 170
- géométrique
 - (loi —) 71
 - (moyenne —) 91, 192
- Gumbel (loi de —) 320

- Hardy-Littlewood
 (théorème de —) 254
- hypergéométrique
 (fonction —) 20
 (loi —) 69
- identité
 binomiale 20
 de Chu-Vandermonde 20
 multinomiale 34
- image réciproque 43
- impossible (événement —) 3
- inclusion-exclusion
 (principe d' —) 27
- incompatibles (événements —) 3
- indépendance 58, 80
 de classes d'événements 60
 de variables aléatoires 61
 mutuelle 59
- indicatrice 5, 46
- inégalité
 de Boole 17
 de Chernoff 176
 de Liapounov 93
 de Markov 90
 de Tchebychev 90
- injection 30
- injective (suite (n, p) —) 30
- inspection (paradoxe de l' —) 77
- intégrabilité 125
- intégrale
 de Lebesgue 133
 de Riemann 133
 de Stieltjes-Lebesgue 134
 d'une variable aléatoire 122, 123
 impropre de Riemann 133
- inverse d'une
 fonction de répartition 177, 217
- Khintchine
 (loi de —) 224
 (théorème de —) 176
 (théorème du log. itéré de —) 254
- Kolmogorov 15, 210, 230
- Laplace
 (approche de —) 239
 (première loi de —) 185
 (seconde loi de —) 186
 (transformation de —) 159
- Laplace-Gauss (loi de —) 178
- Lebesgue
 (intégrale de —) 133
 (mesure de —) 118
 (théorème de —) 137
 (théorème de convergence
 dominée de —) 126
- lemme
 d'Abel 103
 d'approximation 121
- Letta (Giorgio) 219, 277
- Lévy (Paul)
 (inverse des fonctions de
 répartition d'après —) 177, 217
 (théorème de — sur les
 fonctions caractéristiques) 219
 (théorème de Lindeberg —) 241
- Liapounov
 (inégalité de —) 93
 (théorème de —) 247
- limite d'une suite d'événements 4
- limite inférieure, supérieure
 d'une suite d'événements 4
- Lindeberg (théorème de —) 243
- Lindeberg-Lévy
 (théorème de —) 241
- linéarité
 de l'espérance mathématique 85
 de l'intégrale 125
- log-normale (loi —) 163, 181
- logarithme itéré (loi du —) 254
- logistique standard (loi —) 192
- loi
 bêta 189
 binomiale 68
 négative 74, 112
 conjointe 141
 de Bernoulli 68
 de Cauchy 186
 de Laplace-Gauss 178

- de l'Arc sinus 190
- de la somme 200
- de Fréchet 223, 320
- de Gumbel 320
- de Khintchine 224
- de Lorentz 187
- de Maxwell 194
- de Pareto 191
- de Pascal 112
- de Poisson 26, 72
- de probabilité conditionnelle 144
- de probabilité
 - d'une variable aléatoire 47
- de Rayleigh 194
- de Weibull 191, 320
- des grands nombres
 - faible 225
 - forte 225
- du logarithme itéré 254
- du produit 200
- du rapport 201
- exponentielle 182
- gamma 187
- géométrique 71
- hypergéométrique 69
- log-normale 181
- logistique standard 192
- normale 178
 - à deux dimensions 151, 154
- uniforme 177
 - sur $[a, b]$ 178
 - $(0, 1)$ de E. Borel 233
- Lorentz (loi de) 187
- marginales
 - (lois —) 80, 131
 - (mesures de probabilité —) 131
 - (variables aléatoires —) 80, 141
- Markov
 - (chaîne de —) 127
 - (inégalité de —) 90
- masse (fonction de —) 49
- matrice des
 - variances-covariances 151
- Maxwell (loi de —) 194
- Maxwell-Boltzmann
 - (modèle de —) 41
- médiane 93
- Méré (Chevalier de —) 40
- mesurable
 - (espace —) 10, 113
 - (fonction —) 44
- mesure
 - de Dirac 25
 - de Lebesgue 118, 132
 - de Lebesgue λ^n sur \mathbb{R}^n 119
 - de Stieltjes-Lebesgue 116
 - de Stieltjes-Lebesgue sur \mathbb{R}^n 119
 - extérieure 115
 - finie 114
 - marginale 131
 - produit 131
 - singulière 25
 - sur une algèbre 114
- mesuré (espace —) 114
- mesure de probabilité 15
 - conditionnelle 54
 - discrète 25, 67
 - induite par une
 - fonction de répartition 119
- méthode du
 - maximum de vraisemblance 237
- modèle
 - de Boltzmann-Maxwell 41
 - de Bose-Einstein 42
 - de Fermi-Dirac 42
- de Moivre-Laplace (théorème de) 241
- moment 86, 130
 - absolu 88, 130
 - factoriel 87, 104
 - (fonction génératrice des — s) 159
- monotone
 - (classe —) 11
 - (classe — engendrée) 11
 - (suite — d'évènements) 4
- monotonie
 - de l'espérance mathématique 85
 - de l'intégrale 125

- montante (factorielle —) 19
moyenne 86
 arithmétique 93, 97
 géométrique 91, 93, 97, 192
 harmonique 93, 97
multinomial
 (coefficient —) 34
 (identité — e) 34
 (lois — e) 65
mutuelle
 (indépendance —) 59, 61
négligeable 116
non corrélées 88
non dégénérée (loi normale —) 152
non-vieillesse 184
normale
 (lois —) 178
 (lois — à deux dimensions) 151
(n, p)-injective 30
numérotation 26
paradoxe de l'inspection 77
Pareto (lois de —) 191
Pascal
 (lois de —) 74, 112
 (triangle de —) 32
permutation 30
Poincaré (formule de —) 16, 37, 259
Poisson (lois de —) 26, 72
poissonisation 77
poissons 69
Polya (Georg) 240
polynômes de Bernstein 235
Pratelli (Luca) 230, 246, 277
première loi de Laplace 185
presque
 (propriété vraie — partout) 124
 (propriété vraie — sûrement) 129
principe
 d'inclusion-exclusion 27
 de réflexion 35
probabilisable (espace —) 10
probabilisé (espace —) 15
probabilité
 conditionnelle 150
 d'un événement 15
 (lois de —
 d'une variable aléatoire) 47
 (mesure de —) 15
 (mesure de — conditionnelle) 54
 (mesure de — discrète) 25
 (mesures de — marginales) 131
problème
 des boîtes d'allumettes 76
 des rencontres 38, 259
 du scrutin 37
 du tricheur 55
processus de Bernoulli 95
produit
 d'espaces probabilisés 127
 de convolution 81, 99, 137, 200
 des densités 138
 de mesures de probabilité 131
 (formule du —) 28
 (lois du —) 200
prolongement (théorème de —) 115
Rajchman (théorème de —) 229
rapport (lois du —) 201
Rayleigh (lois de —) 194
réciproque (application —) 43
réflexion (principe de —) 35
régression (courbe de —) 147
remise
 (tirage avec —) 64, 68
 (tirage sans —) 65, 69
rencontres (problème des —) 38, 259
répartition
 (fonction de —) 48
 (fonction de — conjointe) 142
 (fonction de — d'un
 vecteur aléatoire) 49
répétition
 (arrangement avec —) 29
 (arrangement sans — s) 30
réunion d'événements 3
Riemann (intégrale de —) 133

- Scheffé (théorème de —) 216
 scrutin (problème du —) 37
 seconde
 fonction caractéristique 169
 loi de Laplace 186
 simple (variable aléatoire —) 68, 121
 singleton 2
 singulière
 (mesure de probabilité —) 25
 Skorohod (théorème de —) 217
 somme
 aléatoire 106
 de variables aléatoires 105
 (formule de la —) 27
 (loï de la —) 200
 Stieltjes-Lebesgue
 (intégrale de —) 134
 (mesure de —) 116
 suites croissantes
 d'évènements 3
 (dénombrément de —) 33
 support d'une
 loi de probabilité 26, 50, 135
 survie (fonction de —) 71, 136, 182
 symétrisée
 (d'une variable aléatoire) 185
 système
 complet 54
 de Dynkin 10, 58
 engendré 11

 taux de défaillance 192
 Tchebychev 225
 (inégalité de —) 90
 test de Wilcoxon 249
 théorème
 « central limit » 240
 d'approximation
 de Weierstrass 234
 de Fubini 132
 de prolongement 115
 du transfert 83, 130, 142
 tirage avec remise 64, 68, 273
 tirage sans remise 65, 69

 transfert
 (théorème du —) 83, 130, 142
 triangle de Pascal 32
 tribu 8
 borélienne 9
 achevée 120
 de \mathbb{R}^n 10
 complétée 116
 des ensembles mesurables 118
 engendrée 9
 par une variable aléatoire 50
 tricheur (problème du —) 55
 triplet fondamental 1

 unicité (théorème d'—) pour les
 fonctions caractéristiques 166
 fonctions génératrices 101
 uniforme (loi —) 177

 variable aléatoire
 à deux dimensions 141
 à valeurs dans \mathbb{R}^n 46
 étagée 68, 121
 (loï de probabilité d'une —) 47
 réelle 46, 120
 simple 68, 121
 (tribu engendrée par une —) 50
 variables aléatoires
 indépendantes 61
 marginales 80, 141
 mutuellement indépendantes 61
 (somme de —) 105
 variance 130
 variancés-covariancés
 (matrice des —) 151
 vecteur aléatoire
 à deux dimensions 141
 (fonction de
 répartition d'un —) 52
 volume de la boule $B_n(0, R)$ 194
 vraisemblance (fonction de —) 238

 Wald (formule de —) 96
 Weibull (loi de —) 191, 320

