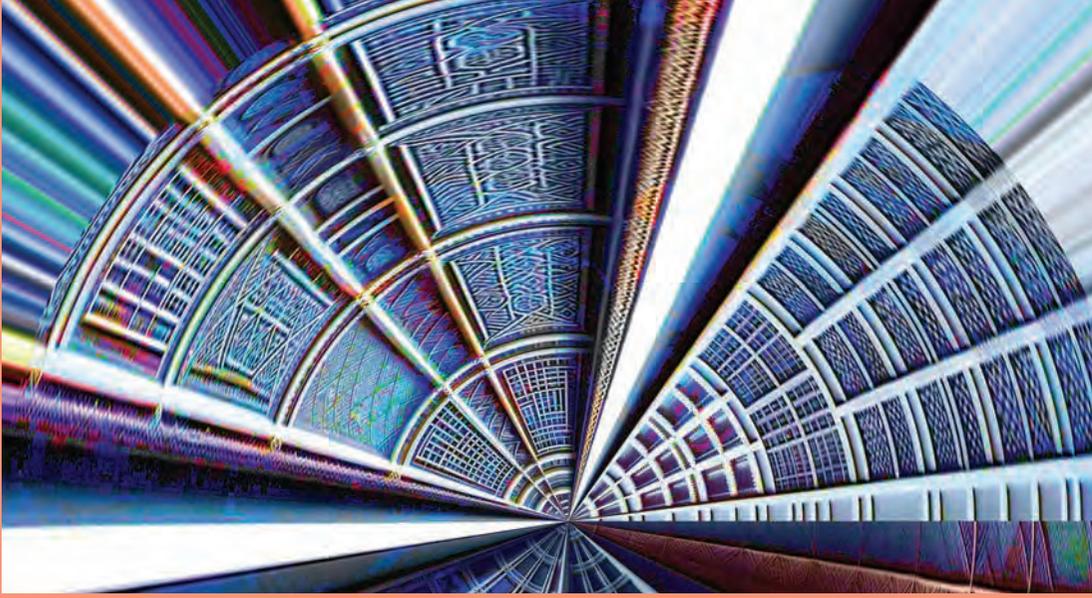


Les cours



Silvère Bonnabel, Michel Schmitt

# Probabilités



Presses des Mines



Silvère Bonnabel et Michel Schmitt, *Cours de probabilités*, Presses des MINES, collection  
Mathématiques et informatique

© Presses des MINES - TRANSVALOR, 2023  
60, boulevard Saint-Michel - 75272 Paris Cedex 06 - France  
[presses@mines-paristech.fr](mailto:presses@mines-paristech.fr)  
[www.pressesdesmines.com](http://www.pressesdesmines.com)  
ISBN : 978-2-38542-292-9

© Photo de couverture: Pixabay  
Dépôt légal: 2023  
Achevé d'imprimer en 2023 (Paris)

Cette publication a bénéficié du soutien de l'Institut Carnot M.I.N.E.S.

Tous droits de reproduction, de traduction, d'adaptation et d'exécution réservés pour  
tous les pays.

# Cours de probabilités

Collection Mathématiques et informatique

Dans la même collection

- Odile Bellenguez  
*À la recherche du temps de travail perdu*
- Hélène Coullon  
*Logiciel et industrie du futur*  
par : Gérard Dray et Sylvie Ranwez  
*Extraction, modélisation, gestion de connaissances dans  
le domaine de la santé et du risque*
- Philippe Fuchs  
*Théorie de la réalité virtuelle*
- Philippe Fuchs  
*Les casques de réalité virtuelle et de jeux vidéo*
- Georges Matheron,  
*Estimating and choosing*
- M. Schmitt, J. Mattioli  
*Morphologie mathématique*
- J.-M. Martins da Cruz  
*Spam*
- X. Deshen, P. Montesinos  
*Proceedings of the third International Workshop on  
Image Analysis*
- S. Desprès, M. Crampes  
*IC 2010*
- C. Lurgeau  
*Le siècle de la voiture intelligente*
- Dir. P. Fuchs, Coord. générale G. Moreau,  
Coord. du vol., S. Donikian,  
*Le traité de la réalité virtuelle vol.5*  
*Les humains virtuels*
- Dir. P. Fuchs,  
Coord. générale G. Moreau, coord. du vol. B.  
Arnaldi et P. Guitton  
*Le traité de la réalité virtuelle vol.4*  
*Les applications de la réalité virtuelle*
- Dir. P. Fuchs,  
Coord. générale G. Moreau, coord. du vol. G.  
Moreau et J. Tisseau  
*Le traité de la réalité virtuelle vol.3*  
*Outils et modèles informatiques*
- Dir. P. Fuchs,  
Coord. générale G. Moreau, coord. du vol.  
J.-M. Burkhardt et S. Coquillart  
*Le traité de la réalité virtuelle vol.2*  
*L'interfaçage, l'immersion et l'interaction en  
environnement virtuel*
- Dir. P. Fuchs,  
Coord. générale G. Moreau, coord. du vol. A.  
Berthoz et J.L. Vercher  
*Le traité de la réalité virtuelle vol.1*  
*L'homme et l'environnement virtuel*
- John Cagnol & Jean-Paul Zolesio (eds)  
*Information Processing*  
*Recent Mathematical Advances in Optimization and  
Control*
- F. Goulette  
*Modélisation 3D automatique*  
Outils de géométrie différentielle

Silvère Bonnabel et Michel Schmitt

# Cours de probabilités





# Préface

Le présent livre est la version rédigée sous forme d'un ouvrage de référence du support du cours de tronc commun de probabilités que nous avons enseigné de 2017 à 2020 aux étudiants de première année à Mines Paris-PSL (anciennement École des Mines de Paris, puis Mines ParisTech). Les probabilités sont souvent difficiles à assimiler par les étudiants, car elles nécessitent une manière de penser à part dans l'univers des sciences "dures". Curieusement, cet enseignement a été très bien perçu, au point qu'il a été classé dans les premiers par les auditeurs parmi les cours de sciences. Les élèves nous ont encouragé à reprendre le "poly" pour en assurer une diffusion au-delà du cercle de Mines Paris.

Ce livre est donc destiné avant tout aux étudiants dans les premières années d'école d'ingénieur ou de master, ainsi qu'à tous les curieux de la théorie moderne et de la pratique des probabilités. Bien que cela ne soit pas nécessaire, une certaine connaissance des probabilités discrètes, typiquement des jeux de cartes ou de dés, permettra d'appréhender plus aisément les concepts que nous développons. En effet, depuis une dizaine d'années, les classes préparatoires aux grandes écoles incluent les probabilités discrètes aux concours d'entrée des écoles d'ingénieur. Les élèves possèdent alors des notions de base à leur intégration.

La théorie des probabilités est désormais considérée comme un sous-domaine des mathématiques, et dans une certaine mesure plus précisément des mathématiques appliquées. Les domaines d'application sont incroyablement vastes, et l'enseignement de cette discipline est incontournable dans toute école d'ingénieur. Aussi, on trouve un grand nombre d'ouvrages de référence et de cours sur le sujet. La théorie moderne des probabilités étant assez difficile, puisqu'elle s'appuie fortement sur la théorie de la mesure, tout cours de probabilité implique un choix quant au niveau de théorie qu'on est prêt à exposer.

Cet ouvrage est tourné vers les applications des probabilités pour l'ingénieur. Il contient des pointeurs vers diverses applications, souvent rassemblés en fin de chapitre sous la forme de compléments facultatifs. Nous invitons toutefois le lecteur à prendre connaissance de ces compléments, qui lui permettent de voir l'intérêt des notions enseignées, et de parfaire sa culture scientifique. Notre livre ne fait qu'effleurer les outils pourtant fondamentaux que sont les transformées (Fourier, Laplace) de variables aléatoires, et essaie de réduire autant que possible les aspects liés à la théorie de la mesure abstraite. Étant donné

que les variables aléatoires discrètes sont vues en classes préparatoires, nous leur accordons une place secondaire dans le présent ouvrage, plus centré sur les variables aléatoires à valeurs dans un espace continu, les formules étant essentiellement les mêmes, en substituant aux sommes discrètes des intégrales.

Pour les lecteurs plus intéressés par les aspects mathématiques, ainsi que ceux qui souhaitent aller plus loin, nous recommandons la lecture de “Probability : Theory and Examples” de Rick Durrett, disponible en ligne. Nous indiquons également que toute la première partie de cet ouvrage s’inspire librement du cours de Probabilités de Francis Maisonneuve, cours dispensé avant nous à l’École des Mines, disponible aux Presses des Mines. Pour approfondir certaines applications, nous recommandons également le polycopié de Ramon Van Handel intitulé “Stochastic calculus and stochastic control” que nous trouvons passionnant et pédagogique (cours acm217 de Princeton). Finalement, nous devons aussi beaucoup au livre “Stochastic processes in physics and chemistry” de N.G. Van Kampen.

Le présent livre est divisé en onze chapitres. Chaque chapitre contient des exercices, dont la plupart sont corrigés en fin de chapitre. On trouvera également des exercices en fin de chapitre, qui étaient destinés à être traités par petits groupes d’étudiantes et étudiants, et qui seront pour la plupart corrigés dans un chapitre dédié à la fin de l’ouvrage. Le résumé du cours qui apparaît au début du présent ouvrage est assez informel. Il sert à avoir un aperçu synthétique de l’ensemble du livre, auquel le lecteur pourra se référer à chaque fois qu’il se sent un peu “perdu” dans les diverses notions. Son utilité a été plusieurs fois soulignée par les élèves. Un ensemble de résumés techniques précis, ou aide-mémoire, est également donné dans chaque chapitre. Ce sont ces résumés qui sont généralement utilisés au moment des examens.

Nos remerciements vont à nos collègues, Sébastien Boisgérault, Arnaud de La Fortelle, Francis Maisonneuve et Julien Reygner, pour leurs précieux retours et conseils (les deux derniers ont fait une relecture très attentive). Le premier auteur exprime sa gratitude à Francis Maisonneuve, qui lui a appris la théorie de la mesure et les probabilités lors de son cursus ingénieur, et dont les ouvrages aux presses de l’École des Mines ont été la principale source bibliographique pour le présent livre.

Silvère Bonnabel et Michel Schmitt  
Paris, février 2023

# Table des matières

<b>1 Pourquoi un formalisme ?</b>	<b>15</b>
1.1 Quelques limites . . . . .	16
1.1.1 L'équiprobabilité . . . . .	16
1.1.2 Passage à la limite . . . . .	19
1.2 La théorie incontournable de la mesure . . . . .	22
1.2.1 Jeu de pile ou face infini : notion d'événement . . . . .	22
1.2.2 Retour sur l'additivité : le paradoxe de Banach-Tarski . . . . .	24
1.2.3 La notion de tribu . . . . .	25
1.2.4 Dimension finie – dimension infinie . . . . .	25
1.3 Un peu de philosophie . . . . .	26
1.3.1 Les phénomènes chaotiques . . . . .	27
1.3.2 Simulations de Monte-Carlo . . . . .	30
1.4 Probabilités et information . . . . .	31
1.4.1 Probabilité des causes . . . . .	32
1.4.2 Estimation bayésienne de paramètre . . . . .	33
1.4.3 Simulations conditionnelles . . . . .	34
1.5 Il n'y a pas que les probabilités dans la vie . . . . .	38
1.6 Exercice . . . . .	38
1.6.1 Ruine nécessaire du joueur à pile ou face . . . . .	38
<b>2 Que contient ce livre ?</b>	<b>39</b>
2.1 Variables aléatoires réelles . . . . .	39
2.2 Les grands théorèmes des suites aléatoires . . . . .	41
2.3 Simulations de Monte-Carlo . . . . .	42
2.4 Variables aléatoires vectorielles . . . . .	42
2.5 Variables aléatoires à densité jointe . . . . .	43
2.6 Conditionnement . . . . .	44
2.6.1 Lois conditionnelles dans les cas discret et à densité . . . . .	44
2.6.2 Théorie générale de l'espérance conditionnelle . . . . .	45
2.6.3 Espérance conditionnelle et randomisation . . . . .	46
2.7 Régression . . . . .	46

<b>I</b>	<b>Phénomènes et variables aléatoires</b>	<b>49</b>
<b>3</b>	<b>Axiomes des probabilités</b>	<b>51</b>
3.1	Espace probabilisé . . . . .	51
3.2	Quelques remarques et explications . . . . .	52
3.3	Propriétés de la probabilité . . . . .	54
3.4	Probabilité conditionnelle, indépendance . . . . .	56
3.4.1	Probabilité conditionnelle . . . . .	56
3.4.2	Formule de Bayes . . . . .	57
3.4.3	Indépendance . . . . .	57
3.5	Ce qu'il faut retenir . . . . .	58
3.6	Exercices . . . . .	60
3.6.1	Pile ou face . . . . .	60
3.6.2	L'achat d'appartement . . . . .	60
3.6.3	Problème de Monty-Hall . . . . .	61
3.6.4	Probabilité des causes (Sally Clark) . . . . .	61
3.6.5	Indépendance conditionnelle et réseaux bayésiens . . . . .	63
<b>4</b>	<b>Variables aléatoires réelles</b>	<b>65</b>
4.1	Définition et outils fondamentaux . . . . .	65
4.1.1	Espérance d'une variable aléatoire . . . . .	66
4.1.2	Moments d'une variable aléatoire réelle . . . . .	68
4.1.3	Fonction caractéristique . . . . .	68
4.1.4	Fonction de répartition et lois . . . . .	69
4.2	Variables aléatoires discrètes et à densité . . . . .	71
4.2.1	Variables discrètes . . . . .	71
4.2.2	Variables à densité . . . . .	72
4.3	Quelques lois usuelles . . . . .	73
4.3.1	Lois discrètes et continues . . . . .	73
4.3.2	La loi gaussienne . . . . .	75
4.4	Indépendance des variables aléatoires . . . . .	78
4.5	Correction des exercices du chapitre . . . . .	80
4.6	Ce qu'il faut retenir . . . . .	82
4.7	Exercices . . . . .	85
4.7.1	Applications directes du cours . . . . .	85
4.7.2	Inégalité de Jensen . . . . .	85
4.7.3	Une borne inférieure utile . . . . .	85
4.7.4	Crues centennales de la Seine . . . . .	86
<b>5</b>	<b>Suite de variables aléatoires</b>	<b>87</b>
5.1	Convergence des suites de VAR . . . . .	87
5.2	Suite de VAR indépendantes . . . . .	90
5.2.1	Lois des grands nombres . . . . .	90
5.2.2	Théorème de la limite centrale . . . . .	92
5.3	Compléments facultatifs . . . . .	94
5.3.1	Illustration du théorème de la limite centrale dans un cas simple . . . . .	94
5.3.2	Correction des exercices du chapitre . . . . .	96

5.4	Ce qu'il faut retenir . . . . .	97
5.5	Exercices . . . . .	99
5.5.1	Convergences . . . . .	99
5.5.2	<i>Surbooking</i> . . . . .	99
5.5.3	Bien-fondé du $\chi^2$ . . . . .	100
5.5.4	Appels à un standard téléphonique . . . . .	100
5.5.5	Distribution du capital et th. de Glivenko-Cantelli . . . . .	102
<b>6</b>	<b>Vecteurs aléatoires</b>	<b>105</b>
6.1	Vecteurs aléatoires généraux . . . . .	105
6.2	Cas particulier des vecteurs à densité . . . . .	107
6.3	Fonction caractéristique . . . . .	109
6.4	Indépendance pour les vecteurs aléatoires . . . . .	110
6.5	Addition des VAR à densité . . . . .	111
6.6	Transformation des densités . . . . .	112
6.7	Gaussiennes multivariées à densité . . . . .	115
6.7.1	Loi du $\chi^2$ . . . . .	116
6.8	Compléments facultatifs . . . . .	117
6.8.1	Application en théorie de l'information : canaux gaussiens	117
6.9	Correction des exercices du chapitre . . . . .	118
6.10	Ce qu'il faut retenir . . . . .	120
6.11	Exercices . . . . .	122
6.11.1	Diffusion d'un faisceau de particules . . . . .	122
6.11.2	Transformation linéaire des gaussiennes . . . . .	122
6.11.3	Loi des vitesses dans un gaz de Maxwell (1860) . . . . .	122
<b>7</b>	<b>Simulations de Monte-Carlo</b>	<b>125</b>
7.1	Approximation d'intégrales . . . . .	127
7.2	Simulations des variables aléatoires . . . . .	129
7.3	Échantillonnage préférentiel . . . . .	130
7.4	Correction des exercices du chapitre . . . . .	132
7.5	Compléments facultatifs . . . . .	133
7.6	Ce qu'il faut retenir . . . . .	134
7.7	Exercices . . . . .	136
7.7.1	Simulation d'une variable aléatoire réelle . . . . .	136
7.7.2	Calcul probabiliste de $\pi$ . . . . .	136
7.7.3	Monte-Carlo et malédiction de la grande dimension . . . . .	136
7.7.4	Méthodes Monte-Carlo pour l'approximation de lois . . . . .	137
7.8	Sujet de travaux pratiques . . . . .	138
<b>II</b>	<b>Conditionnement</b>	<b>141</b>
<b>8</b>	<b>Conditionnement des variables à densité</b>	<b>143</b>
8.1	Conditionnement dans le cas discret . . . . .	143
8.2	Conditionnement dans le cas à densité . . . . .	144
8.2.1	Définition . . . . .	144
8.2.2	Illustration graphique . . . . .	144

8.2.3	Intuition et caractérisation . . . . .	146
8.2.4	Formule de Bayes pour les densités . . . . .	148
8.2.5	Formule des probabilités totales . . . . .	148
8.3	Espérance conditionnelle pour les variables à densité . . . . .	149
8.4	Conditionnement pour les gaussiennes multivariées . . . . .	150
8.5	Compléments facultatifs . . . . .	151
8.5.1	Une application en statistiques : le théorème de Rao-Blackwell . . . . .	151
8.6	Correction des exercices du chapitre . . . . .	152
8.7	Ce qu'il faut retenir . . . . .	154
8.8	Exercices . . . . .	155
8.8.1	Calcul de densités conditionnelles . . . . .	155
8.8.2	Densité conditionnelle à un événement . . . . .	155
8.8.3	Désintégration radioactive . . . . .	156
8.8.4	Problème : estimation bayésienne multi-hypothèses . . . . .	156
<b>9</b>	<b>Régression et espérance conditionnelle</b> . . . . .	<b>159</b>
9.1	Préambule : espérance conditionnelle dans $L^1$ . . . . .	160
9.2	L'espérance conditionnelle dans $L^2$ . . . . .	160
9.2.1	Formule d'Huygens . . . . .	160
9.2.2	Espérance conditionnelle pour un couple à densité . . . . .	161
9.2.3	Rappels de résultats d'analyse fonctionnelle . . . . .	162
9.2.4	Définition alternative de l'espérance conditionnelle . . . . .	163
9.3	Régression . . . . .	165
9.3.1	Régression fonctionnelle . . . . .	165
9.3.2	Régression linéaire . . . . .	166
9.4	Théorème de randomisation . . . . .	166
9.4.1	Exercices corrigés . . . . .	168
9.4.2	Exercices corrigés autour d'une notion générale . . . . .	169
9.5	Lien avec la régression en statistiques . . . . .	170
9.6	Correction des exercices du chapitre . . . . .	172
9.7	Ce qu'il faut retenir . . . . .	174
9.8	Exercices . . . . .	175
9.8.1	Fragmentations aléatoires . . . . .	175
9.8.2	Modèles de Markov Cachés (HMM) . . . . .	176
9.8.3	Régression logistique pour la classification binaire . . . . .	177
<b>10</b>	<b>Vecteurs gaussiens</b> . . . . .	<b>181</b>
10.1	Définition générale . . . . .	181
10.1.1	Cas où $\Sigma$ est inversible . . . . .	182
10.1.2	Cas où $\Sigma$ n'est pas inversible . . . . .	183
10.2	Régression dans le cas gaussien . . . . .	183
10.3	Compléments facultatifs . . . . .	184
10.3.1	Application pour l'ingénieur : le filtre de Kalman . . . . .	184
10.4	Ce qu'il faut retenir . . . . .	188
10.5	Exercice . . . . .	190

---

<b>11 Marches aléatoires</b>	<b>191</b>
11.1 Marches aléatoires et mouvement brownien . . . . .	191
11.2 Marches aléatoires et fortune aux jeux . . . . .	193
11.3 Marches aléatoires confinées et convergence vers une distribu- tion d'équilibre . . . . .	193
11.4 Correction des exercices du chapitre . . . . .	194
11.5 Exercices . . . . .	194
11.5.1 La formule d'Einstein - Smoluchowski . . . . .	194
11.5.2 Formule de Wald . . . . .	195
<b>Annexes</b>	<b>199</b>
Rappels sur la théorie de la mesure et de l'intégration . . . . .	204
Éléments de correction . . . . .	224
<b>Index</b>	<b>225</b>



# Chapitre 1

## Pourquoi un formalisme ?

L'objectif de ce chapitre introductif est de motiver le recours à une axiomatique moderne des probabilités due à Andreï Nikolaïevitch Kolmogorov<sup>1</sup>. Tant que l'on reste dans le cadre des probabilités discrètes, enseignées en France en classes préparatoires aux grandes écoles, les choses sont conceptuellement simples. En fait, dans ce modèle discret, tout est calculable, du moins en théorie, car il suffit d'énumérer toutes les configurations que peut prendre un système, la probabilité que le système vérifie une certaine propriété étant alors définie selon la règle

$$\text{Probabilité} = \frac{\text{nombre de cas favorables}}{\text{nombre total de cas}}$$

Les jeux de hasard se conforment bien à cette modélisation, car ils peuvent être reproduits un grand nombre de fois, et la probabilité apparaît alors comme la limite, lorsque le nombre d'expériences tend vers l'infini, du ratio du nombre de fois où le système est dans l'état désiré divisé par le nombre d'expériences<sup>2</sup>. C'est ce qu'a fait le Chevalier de Méré au 17<sup>e</sup> siècle lorsqu'il pariait sur différents types de jeux et essayait de rendre ces paris équitables. En fait, c'étaient les fluctuations qui pouvaient rendre riches ou ruinés les joueurs<sup>3</sup>.

D'un point de vue théorique, en probabilités discrètes, l'outil à notre disposition est le dénombrement. Calculer la probabilité d'un événement (par exemple obtenir un total de 10 en jetant deux dés) revient à compter de combien de manières différentes on peut obtenir 10 : 6 et 4, 5 et 5 ou 4 et 6, donc 3 parmi les 36 possibilités, soit une probabilité de  $3 \div 36 = 1 \div 12$ . Cependant, les calculs

---

1. La lecture de ce chapitre peut être omise et doit être vue comme une motivation pour poursuivre la lecture de ce livre.

2. On peut raffiner un peu ce modèle en supposant que tous les états n'ont pas la même probabilité, mais cela ne change pas fondamentalement ce qui suit. Au lieu de compter, il faut sommer ces probabilités élémentaires.

3. Ainsi, dans un jeu de pile ou face où chaque joueur a une somme  $S$  fixée à l'avance, l'un des deux joueurs sera ruiné avec probabilité 1, du moment que l'on attend assez longtemps (voir l'exercice 1.6.1 en fin de chapitre).

deviennent, dans la pratique, souvent complexes, parce qu'ils mettent jeu des nombres très grands que l'on ne peut pas manipuler directement<sup>4</sup>.

Cette approche est cependant très féconde et permet d'appréhender des notions probabilistes telles que les moments (espérance, variance...) et surtout le conditionnement, une manière d'introduire des connaissances nouvelles sur le système que l'on étudie, par la célèbre "formule de Bayes". Nous allons voir, cependant, que des questions très naturelles ne peuvent être résolues par ce formalisme et qu'une théorie plus complète est nécessaire. C'est l'objet de ce chapitre.

## 1.1 Quelques limites

Examinons plus en détails deux notions issues de ce qui précède, qui nous paraissent évidentes et qui génèrent des paradoxes si l'on n'y prend pas garde :

- la notion d'équiprobabilité ;
- le passage à la limite lorsque l'on renouvelle une expérience aléatoire simple un grand nombre de fois.

### 1.1.1 L'équiprobabilité

La notion d'équiprobabilité est profondément ancrée dans l'expérience subjective. Cependant, elle ne va pas de soi. Commençons par un exemple simple, tout en restant dans le cas discret.

#### Un exemple simple

Considérons les familles composées de deux enfants. On supposera les naissances de garçons et de filles équiprobables, ainsi que l'indépendance des naissances successives. Quelle est la probabilité d'avoir deux filles ?

- La probabilité d'avoir deux filles dans une famille est  $1/4$ .
- Si l'on sait que l'aînée est une fille, alors cette probabilité devient  $1/2$ .
- Si l'on suppose que l'on sait que la famille contient au moins une fille, la probabilité que la famille soit composée de deux filles est à présent  $1/3$ .
- Enfin, si l'on suppose que l'on rencontre une fille d'une famille de deux enfants, quelle est la probabilité que sa famille ait deux filles ?  $1/2$ .

Pour les trois premiers cas, il suffit de modéliser la famille par un 2-uple, le premier terme étant le sexe de l'aîné(e), le second celui du cadet(te). La différence du dernier cas avec les trois premiers provient du fait que lorsque

---

4. Le dénombrement fait souvent appel aux nombres de combinaisons. Ainsi  $\binom{n}{p}$  est le quotient du produit de  $p$  nombres décroissants à partir de  $n$  divisé par  $p!$ . Or  $p!$  excède très rapidement la précision de représentation des entiers en machines, pour peu que  $p$  soit de l'ordre de quelques dizaines.

l'on rencontre une fille, on ne sait pas si c'est l'aînée ou la cadette. On sur-représente donc les familles de deux filles<sup>5</sup>. Ainsi, la notion d'équiprobabilité ne va pas de soi, car elle dépend de la modélisation que l'on fait.

### Un nombre au hasard

Avant d'aborder un exemple plus complexe, le paradoxe de Bertrand, nous allons avoir besoin de généraliser légèrement les probabilités discrètes.

Qu'est ce qu'un nombre réel au hasard dans l'intervalle  $[0, 1]$  ? La probabilité qu'il prenne une valeur particulière, disons 0,5 est nulle. Si elle ne l'était pas, en sommant sur une toutes les valeurs entre 0 et 1, qui ont la même probabilité, on obtiendrait l'infini<sup>6</sup>. Mais étant nulle, sommer des 0 ne donne toujours que 0<sup>7</sup>. On contourne cette difficulté en posant que la probabilité que le nombre aléatoire tombe dans un intervalle  $[\alpha, \beta]$  sera  $\beta - \alpha$ , la longueur de l'intervalle. On peut facilement généraliser ce concept dans le plan (ou un espace de dimension finie), en posant que la probabilité pour qu'un point au hasard dans un ensemble  $A$  tombe dans  $B \subset A$  est égale à la surface de  $B$  divisée par la surface de  $A$ . Sans le savoir, nous sommes en train de définir une mesure.

### Le paradoxe de Bertrand

Le paradoxe de Bertrand est le suivant : en considérant un cercle de rayon unité, quelle est la probabilité qu'une corde de ce cercle ait une longueur supérieure à  $\sqrt{3}$  ? Ce nombre est choisi parce géométriquement, pour qu'une corde donnée ait une longueur supérieure à  $\sqrt{3}$ , il faut et il suffit qu'elle coupe le cercle de rayon un demi et de même centre (voir FIGURE 1.1).

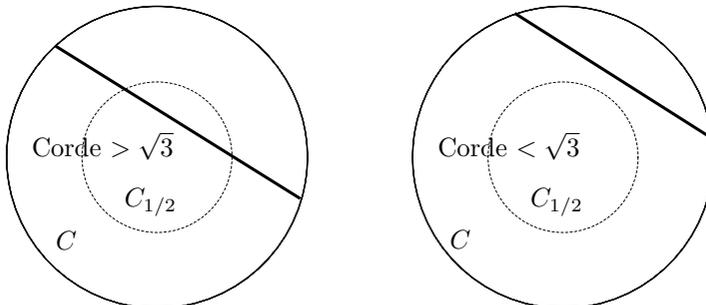


FIGURE 1.1 – Pour qu'une corde ait une longueur supérieure à  $\sqrt{3}$ , elle doit couper le cercle de rayon  $1/2$ .

5. Pour modéliser l'expérience, on peut par exemple utiliser un 3-uple, en ajoutant en troisième terme l'enfant que l'on rencontre (aîné ou cadet).

6. La phrase qui précède est une énormité mathématique, car on ne fait pas de somme sur un ensemble d'indices continu. On pourrait se limiter à faire la somme sur les rationnels, car ils sont en nombre dénombrable. La conclusion reste cependant la même : la probabilité du nombre 0,5 doit être nulle.

7. Ce type de constat motive l'analyse dite non standard, branche des mathématiques permettant de manipuler des infiniments petits (non nuls), mais qui débouche sur d'autres problèmes, notamment l'utilisation de l'axiome du choix.

Voici trois solutions intuitives.

**Solution 1 :** Par symétrie en rotation, on peut considérer les cordes horizontales seulement. Pour avoir une longueur supérieure à  $\sqrt{3}$ , il faut et il suffit qu'elle soit à une distance inférieure à  $1/2$  du centre du cercle. Donc on a naturellement une chance sur 2.

**Solution 2 :** Par symétrie en rotation, on peut considérer les cordes qui sont issues d'un point donné du cercle. Si l'on prend alors pour référence la tangente au cercle en ce point, toute corde est repérée par l'angle qu'elle fait par rapport à cette tangente, compris entre  $0$  et  $\pi$  (la corde d'angle  $\pi/2$  passe par le centre du cercle). Pour que la corde ait une longueur supérieure à  $\sqrt{3}$ , il faut et suffit que cet angle soit compris entre  $\pi/3$  et  $2\pi/3$ . Donc une chance sur 3 (voir FIGURE 1.2).

**Solution 3 :** Une corde peut être repérée par la projection du centre du cercle sur celle-ci. La corde a une longueur supérieure à  $\sqrt{3}$  si et seulement si cette projection se trouve dans le cercle de rayon  $1/2$  et de même centre. La surface du petit cercle est le quart de la surface du cercle de départ. Donc une chance sur 4.

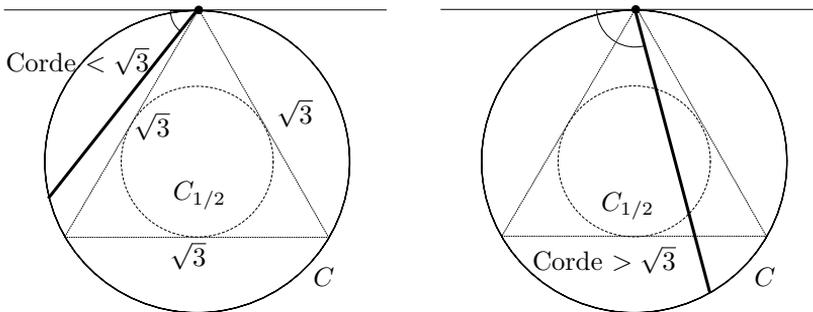


FIGURE 1.2 – Illustration de la solution 2 : à gauche, l'angle de la corde avec la tangente est inférieur à  $\pi/3$ . Cette corde a une longueur inférieure à  $\sqrt{3}$ . À droite, l'angle de la corde avec la tangente est compris entre  $\pi/3$  et  $2\pi/3$ . Cette corde a une longueur supérieure à  $\sqrt{3}$ .

De ce qui précède, on voit que de manière naturelle, la probabilité recherchée est alternativement  $1/2$ ,  $1/3$  puis  $1/4$ . Y a-t-il une erreur de calcul ? Non. En fait, c'est que la notion de "droite au hasard" n'admet pas de réponse naturelle et chacune des méthodes proposées plus haut, fondée sur des intuitions différentes fournit des statistiques différentes<sup>8</sup>. Pour l'illustrer, prenons comme référence de la corde, la projection du centre du cercle sur celle-ci. Les trois simulations présentées à la FIGURE 1.3 montrent trois répartitions autour du centre du cercle assez différentes, très dense à gauche, moins dense au milieu et équirépartie à droite. Ainsi, les trois façons de paramétrer une droite sont différentes. On voit, par exemple, que dans la solution 1 les cordes proches du

8. Dans un espace continu la notion d'équiprobabilité - comme nous l'avons plus haut étendue - dépend du système de coordonnées qu'on choisit.

centre sont beaucoup plus représentées que dans les autres approches, menant à la plus grande des trois probabilités. Si l'on veut lever le paradoxe, il faut changer de paramétrage des droites de manière “propre”, c'est-à-dire écrire une formule de changement de variables, comme dans le cas du changement de variable dans les intégrales et faire intervenir le jacobien de la transformation.

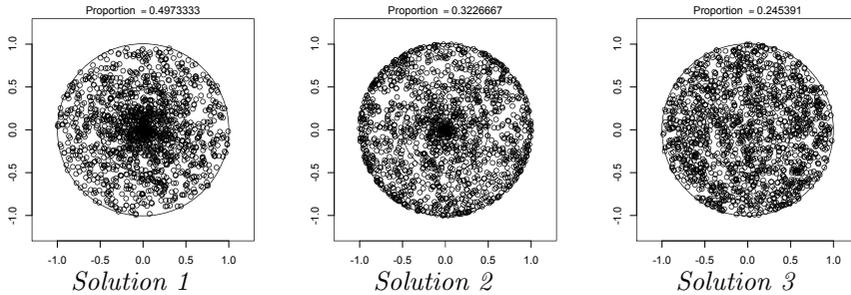


FIGURE 1.3 – Les trois solutions au problème de Bertrand. Chaque point représente la projection du centre du cercle sur la corde. 1000 cordes aléatoires sont représentées sur chaque figure, selon les solutions décrites plus haut.

### 1.1.2 Passage à la limite

Nous nous tournons maintenant vers un autre type de difficulté, celui du passage à la limite, qui semble pourtant si naturel. Prenons le jeu de pile ou face. Pile, représente une gain de 1, face, une perte de 1. Répétons l'expérience un grand nombre de fois et examinons le gain cumulé (qui peut être négatif). Intuitivement, il devrait être proche de 0 si la pièce n'est pas truquée.

La FIGURE 1.4 montre plusieurs réalisations. Ce que l'on constate est que, bien loin de fluctuer autour de zéro, le gain est de plus en plus erratique et peut prendre des valeurs de plus en plus grandes, dans un sens ou dans l'autre. En particulier, on peut montrer que la probabilité de dépasser une valeur (absolue) donnée tend vers 1, du moment que l'on attend assez longtemps.

Regardons maintenant la distribution des gains en fonction du nombre de lancers de la pièce. Il s'agit d'un loi binômiale : la probabilité d'observer  $p$  piles parmi  $n$  lancers vaut  $P(p) = \binom{n}{p} / 2^n$ . Ainsi, la probabilité de chaque gain particulier tend vers 0 lorsque le nombre de lancers tend vers l'infini. Si l'on passe à la limite, tout devient nul, et on n'a donc plus une probabilité (FIGURE 1.5). En fait, nous avons bien une mesure de masse 1 pour tout  $n$ . Simplement, cette masse s'évanouit à l'infini<sup>9</sup>.

9. Ceci illustre bien le cours d'intégration : l'intégrale de la limite (ici 0) est toujours inférieure ou égale à la limite de l'intégrale (ici 1). C'est le lemme de Fatou. Comme il n'y a pas de chapeau intégrable, le théorème de convergence dominée de l'intégrale de Lebesgue ne s'applique pas. Voir le chapitre consacré aux rappels d'intégration en fin de cet ouvrage.

Regardons plus en détails comment s'opère cette convergence vers 0, en renormalisant les axes du graphique. Le principe est de garder le même facteur de contraction sur l'axe des  $x$  et de dilatation sur l'axe des  $y$ . Le fait de prendre le même facteur de contraction et de dilatation préserve l'élément  $dx dy$ .

On peut tâtonner un peu, mais regardons deux cas éclairants.

D'abord, contractons l'axe des  $x$  d'un facteur  $n$ , et dilatoons alors celui des  $y$  du même facteur  $n$ . On observe que la distribution se resserre autour de 0 pour terminer en masse de Dirac. Il s'agit de la loi de :

$$\frac{1}{n} \sum_{i=1}^n X_i \quad (1.1)$$

où  $X_i$  représente la variable aléatoire du  $n^{\text{ème}}$  lancer de pièce (FIGURE 1.6). Il s'agit de la loi (faible) des grands nombres. Nous verrons d'ailleurs un résultat beaucoup plus fort dans la suite de ce livre, "la loi forte des grands nombres", qui affirme que ce résultat reste vrai pour presque toute réalisation d'une suite de lancers de pièce.

Un deuxième type de renormalisation est plus intéressant, car il montre une régularité que l'on ne peut plus appréhender dans le cadre des probabilités discrètes. Ainsi, si on contracte l'axe des  $x$  d'un facteur  $\sqrt{n}$ , et que l'on dilate celui des  $y$  du même facteur on observe la convergence vers une courbe en cloche très régulière (FIGURE 1.7). De par la contraction de l'axe des abscisses, les valeurs de gain deviennent de plus en plus denses et se rapprochent d'une courbe continue. On peut montrer que cette courbe limite, appelée courbe de Gauss, a pour équation :

$$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad (1.2)$$

Ceci illustre bien que le cadre discret ne suffit plus et qu'un cadre continu s'impose. Nous avons besoin de représenter une probabilité sur un ensemble continu, tel  $\mathbb{R}$  ou  $\mathbb{R}^n$ . C'est là qu'intervient le cadre axiomatique de Kolmogorov, s'appuyant sur les notions mathématiques de théorie de la mesure et d'intégrale de Lebesgue, qui sont donc essentielles pour exprimer la loi limite de :

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i. \quad (1.3)$$

En termes d'outils théoriques, nous voyons par cet exemple que l'on aimerait pouvoir décrire les probabilités par des fonctions, ici, une gaussienne ou loi de Gauss. C'est ce que ferons en introduisant la notion de "densité".

Quelle est la généralité de ce que nous venons de faire ? Prenons cette fois-ci une pièce fortement pipée, par exemple, la probabilité d'obtenir pile est 0.1. Cette distribution est fortement dissymétrique. Si on compense le fait que la moyenne s'enfuit à  $-\infty$ , le même type de renormalisation fonctionne encore, et on converge toujours vers la courbe de Gauss (FIGURE 1.8).

Ce résultat est très général, et nous le verrons sous le terme de théorème de la limite centrale. C'est également ce théorème qui justifie l'importance accordée aux vecteurs gaussiens, omniprésents dans l'univers des probabilités.

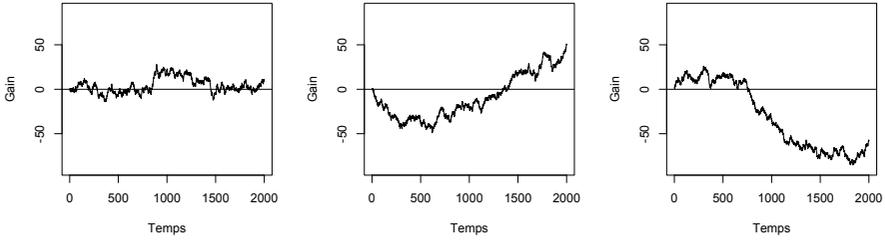


FIGURE 1.4 – Trois simulations d’un jeu de pile ou face. On remarquera que le gain ne tend pas vers 0 au cours du temps.

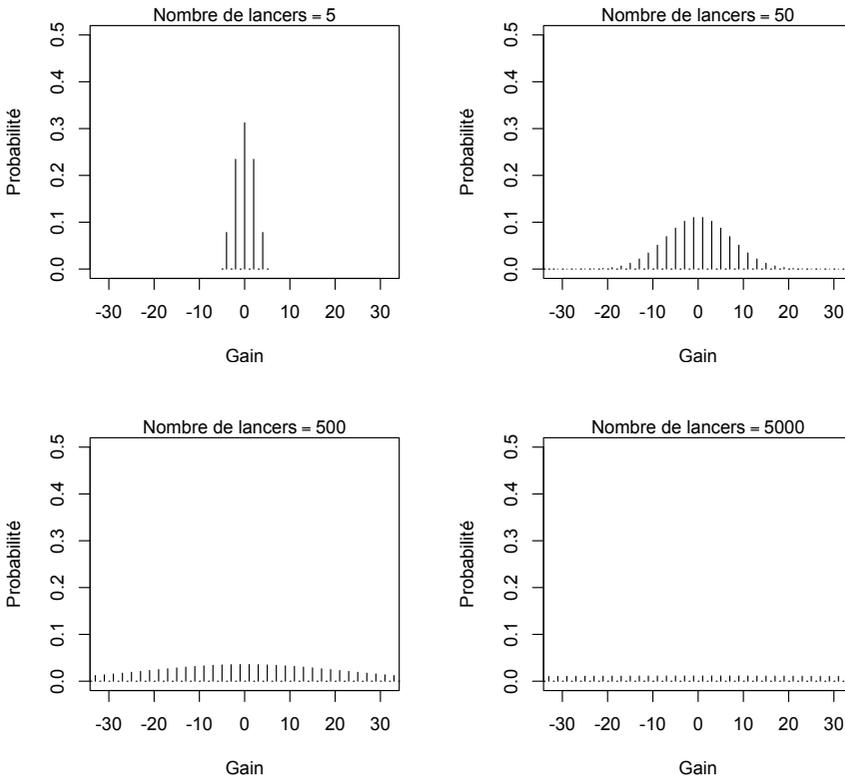


FIGURE 1.5 – Convergence vers 0 des probabilités des gains en fonction du nombre de lancers.

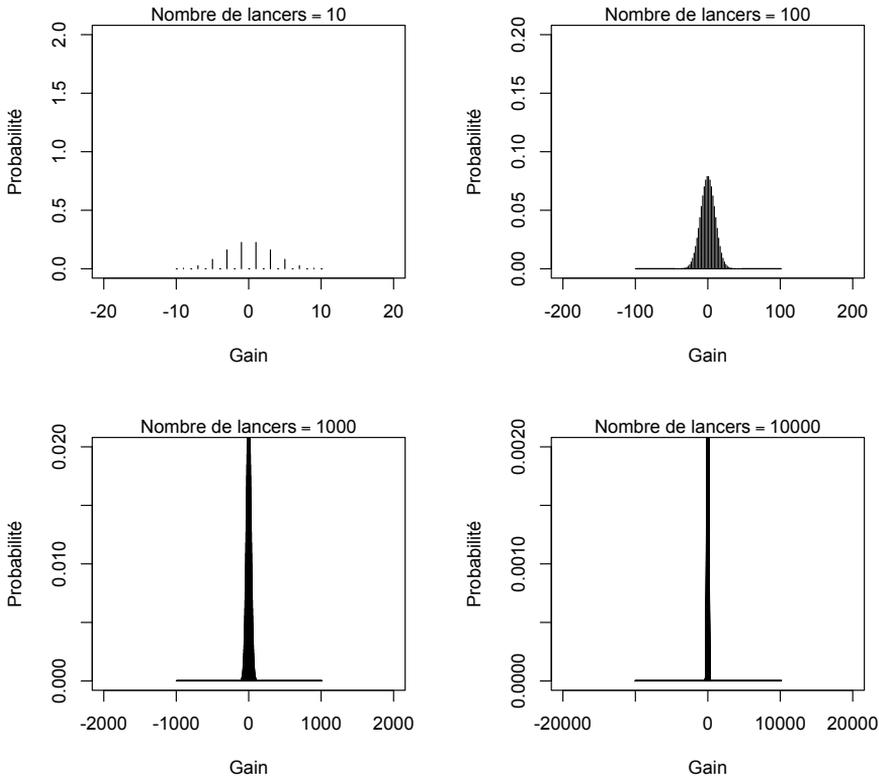


FIGURE 1.6 – Convergence vers une masse de Dirac en 0 lorsque l’on renormalise les axes par  $n$ . Le “pic central” tend vers l’infini.

## 1.2 La théorie incontournable de la mesure

Nous avons déjà progressé un peu, en voyant que les probabilités discrètes sont insuffisantes pour étudier certains cas limites. Cependant, nous avons écrit quelques énormités mathématiques, mais elle nous ont permis d’avancer. Ce paragraphe va nous montrer que nous devons prendre des précautions si nous voulons obtenir des résultats consistants.

### 1.2.1 Jeu de pile ou face infini : notion d’événement

Reprenons le jeu de pile ou face. Si nous voulons rester dans le cadre discret, il faut se restreindre à un nombre de lancers fini, disons  $n$ , puis on peut essayer de passer à la limite, mais ceci peut s’avérer risqué, voire impossible. Par exemple, il n’est pas aisé de répondre rigoureusement à la question : “toute suite de lancers contient une infinité de fois pile”. Intuitivement, c’est toujours le cas, et la probabilité de cet événement vaut bien 1, mais on ne peut se restreindre à des suites finies pour y répondre.

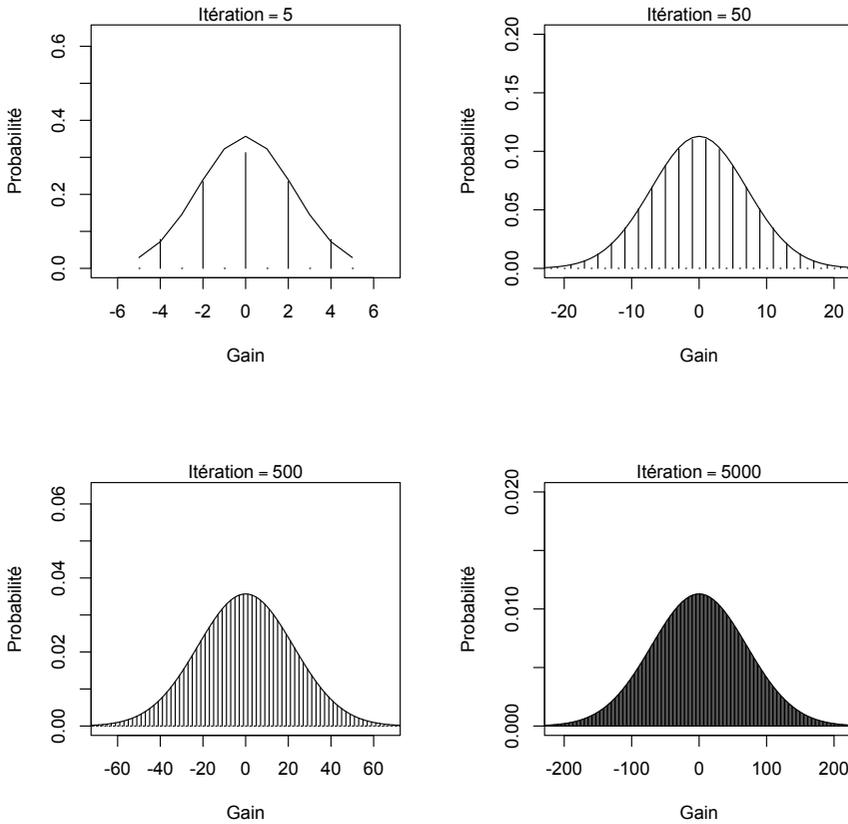


FIGURE 1.7 – Convergence vers une Gaussienne lorsque l'on renormalise les axes par  $\sqrt{n}$ .

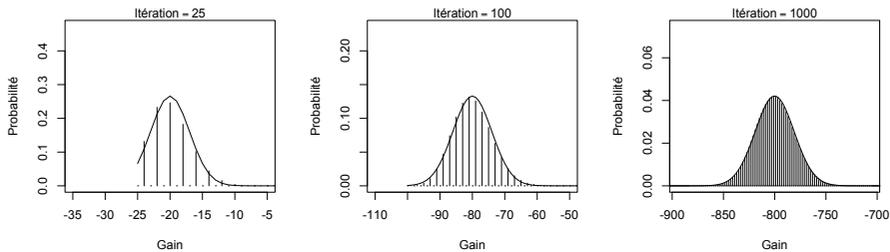


FIGURE 1.8 – Convergence vers une Gaussienne lorsque l'on renormalise les axes par  $\sqrt{n}$ , dans le cadre d'une pièce pipée (probabilité de tomber sur pile de 0.1). On observera que pour voir la courbe de Gauss, il est nécessaire de centrer le graphique sur la moyenne, soit  $n \times 0.8$ .

Si nous considérons l'ensemble de toutes les suites infinies de lancers, cet ensemble n'est pas dénombrable<sup>10</sup>. L'univers  $\Omega$  dans lequel on se place est l'ensemble de toutes les suites infinies de piles ou faces. Chaque suite particulière a naturellement une probabilité nulle puisqu'il y a une infinité (non dénombrable) de suites possibles. Cependant, la probabilité d'avoir pile au premier lancer est 0.5. Pour avancer, il faut donc pouvoir évaluer la probabilité de groupes de suites, les "mesurer". Mathématiquement, un groupe de suites sera un sous-ensemble de  $\Omega$ , que l'on appelle en probabilités un événement. Voici donc pourquoi la notion d'événement, qui est triviale dans le cas discret, – tout sous-ensemble de  $\Omega$  est un événement – revêt une importance particulière dans les cas plus généraux.

### 1.2.2 Retour sur l'additivité : le paradoxe de Banach-Tarski

En reprenant la notion précédente d'univers  $\Omega$ , peut-on définir la probabilité  $P$  pour tout sous-ensemble de  $\Omega$ ? La propriété minimale de  $P$  que l'on souhaite est la suivante :

$$\text{Pour tous } A \text{ et } B \text{ disjoints, } P(A \cup B) = P(A) + P(B) \quad (1.4)$$

C'est la propriété d'additivité que vérifient les probabilités discrètes lorsque l'on fait du dénombrement.

Malheureusement, les mathématiques ne sont pas très compatissantes. Il n'existe pas de fonction additive  $P : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$ , où  $\mathcal{P}(\Omega)$  est l'ensemble des parties de  $\Omega$ .

Un exemple est donné par le paradoxe de Banach-Tarski<sup>11</sup>. Dit de façon simple, il est possible de découper une sphère de rayon unité en un nombre fini de morceaux et de les recombinaer de telle sorte qu'ils s'assemblent sous forme de deux sphères de rayon unité. De manière formelle, plaçons dans  $\mathbb{R}^3$  et considérons une sphère de rayon unité  $S_1$ . Il existe un entier  $n$  et des ensembles  $(A_i)_{i=1}^n$  disjoints, vérifiant :

$$A_i \cap A_j = \emptyset \text{ pour } i \neq j \text{ et } \bigcup_{i=1}^n A_i = S_1, \quad (1.5)$$

et tels que par  $n$  isométries  $\mathcal{D}_i$  (rotations - translations) dans  $\mathbb{R}^3$ , donc sans modifier la forme ou la taille des  $A_i$ , ils puissent se recombinaer en deux sphères disjointes de rayon unité,  $S_1$  et  $S'_1$ .

$$\mathcal{D}_i(A_i) \cap \mathcal{D}_j(A_j) = \emptyset \text{ pour } i \neq j \text{ et } \bigcup_{i=1}^n \mathcal{D}_i(A_i) = S_1 \cup S'_1. \quad (1.6)$$

Cet exemple montre que l'on ne peut pas "mesurer" le volume de toute partie de  $\mathbb{R}^3$  sans obtenir une incohérence, car le "volume" (nous mettons le

10. On peut le prouver par le procédé diagonal de Cantor. Supposons que le nombre de suites soit dénombrable. On peut alors indexer chaque suite  $S$  par un entier  $i \in \mathbb{N}$ . Notons donc  $S_i = (s_i^j)_{j \in \mathbb{N}}$  la  $i^{\text{ème}}$  suite de lancers. On construit alors la suite  $S = (\bar{s}_j^j)_{j \in \mathbb{N}}$  (la barre indique que pile devient face et inversement) qui est différente de chacune des suites, donc qui n'a pas été énumérée dans les  $S_i$ .

11. Ce paradoxe qui en fait est un théorème a été démontré en 1924 par les brillants mathématiciens polonais Stefan Banach (1892-1945) et Alfred Tarski (1901-1983).

volume entre guillemets, car ce ne peut pas être un vrai volume...) de  $A_i$  est égal au volume de  $\mathcal{D}_i(A_i)$ . Donc la somme de  $i = 1$  à  $n$  des volumes des  $A_i$  égale la somme des volumes des  $\mathcal{D}_j(A_j)$  qui sont en nombre fini. Pourtant, le volume d'une sphère de rayon 1 est 2 fois plus petit que celui de deux sphères de rayon 1 disjointes.

Ainsi, dans l'espace, on ne peut mesurer le volume de n'importe quel ensemble. C'est pourquoi on se restreint à certains sous-ensembles, les sous-ensembles dits mesurables (les boréliens) qui garantissent une construction cohérente de la notion de volume.

En passant, les ensembles  $A_i$  construits dans le théorème de Tarski ne sont pas des boréliens. La démonstration du théorème de Banach-Tarski fait intervenir l'axiome du choix, nécessaire à la construction d'ensembles non mesurables. A noter également que ce théorème reste vrai dans  $\mathbb{R}^n$  pour  $n \geq 3$ . Il n'est cependant pas vrai en dimension 2.

### 1.2.3 La notion de tribu

Du paragraphe précédent, on conclut qu'il faut abandonner l'une des deux notions, soit l'additivité, soit le fait de pouvoir mesurer tout sous-ensemble de l'univers  $\Omega$ . La théorie de la mesure introduit la notion de "tribu", qui regroupe les ensembles que l'on se permet de mesurer et garantit alors l'additivité, plus précisément la  $\sigma$ -additivité, associée à une famille finie ou dénombrable.

La modélisation qui en découle est la suivante :

**Une expérience aléatoire** est modélisée par un triplet  $(\Omega, \mathcal{A}, \mathcal{P})$ , appelé *espace probabilisé* où :

1.  $\Omega$  représente tous les états possibles du *monde* ou *univers*,
2.  $\mathcal{A}$  est une tribu sur  $\Omega$ , représentant toutes ce dont on a le droit le parler de la probabilité, i.e. toutes les questions que l'on peut poser sur le monde,
3.  $P$  est une mesure positive de masse 1 sur l'espace mesurable  $(\Omega, \mathcal{A})$ .

**La tribu**  $\mathcal{A}$  est un ensemble de parties de l'univers  $\Omega$ , contenant  $\emptyset$  et  $\Omega$ , stable par intersection et union dénombrable. Chaque ensemble  $A$  de  $\mathcal{A}$  est appelé événement.

**La probabilité**  $P$  est une mesure positive de masse 1 et par suite possède des propriétés suivantes :

1.  $P(\emptyset) = 0$ ,  $P(\Omega) = 1$  et  $\forall A \in \mathcal{A}$ ,  $P(A) \geq 0$ .
2. Pour toute suite d'événements  $(A_n)_{n \in \mathbb{N}}$  deux à deux disjoints on a

$$p\left(\biguplus_{n=0}^{\infty} A_n\right) = \sum_{n=0}^{\infty} P(A_n)$$

### 1.2.4 Dimension finie – dimension infinie

Dans  $\mathbb{R}^n$ , il n'y a qu'une façon raisonnable de construire une tribu. C'est celle des boréliens, la plus petite contenant les pavés.

En dimension infinie, c'est un autre débat. Essayons de définir une "fonction aléatoire", par exemple  $f : \mathbb{R} \rightarrow \mathbb{R}^3$ . Il s'agira typiquement d'une trajectoire d'un objet. Il nous faut construire une tribu sur  $(\mathbb{R}^3)^{\mathbb{R}}$ . Classiquement, cela se fait par le produit de tribus<sup>12</sup> sur  $\mathbb{R}^3$ . Cette construction ne permet d'appréhender ce qui se passe qu'en un nombre fini d'instants. Avec cette tribu, l'ensemble des trajectoires continues n'est pas mesurable...

Il existe cependant des alternatives, comme le montre la théorie des ensembles aléatoires développée dans les années 1960 à l'Ecole des Mines de Paris par Georges MATHERON<sup>13</sup>, connue sous le nom de "Géostatistique". Cette théorie réduit l'univers aux fermés topologiques<sup>14</sup> de  $\mathbb{R}^n$ , ce qui permet de considérer la tribu engendrée par les événements de la forme "un fermé donné coupe l'ensemble aléatoire". En particulier, "un compact donné est inclus dans l'ensemble aléatoire" est mesurable. Cette approche et celle par produit de tribus sont alors profondément distinctes. Sur les similitudes et différences entre ces deux approches, voir par exemple le chapitre X du livre *Morphologie Mathématique* édité aux Presses des Mines (2013).

Illustrons la différence entre les deux approches par "un point au hasard dans le carré unité". De manière intuitive, ce point au hasard est caractérisé par ses deux coordonnées qui sont deux variables aléatoires uniformes et indépendantes. Dans la première approche, un ensemble de points fini ou dénombrable a une probabilité nulle de contenir notre point aléatoire. On dira qu'il est indiscernable de l'ensemble vide. Dans la deuxième approche, la probabilité que le point aléatoire coupe un fermé particulier (en fait, y soit inclus) est égal à la surface de ce fermé, ce qui permet de le distinguer de l'ensemble vide ! Si l'on a plusieurs points aléatoires, on peut même compter le nombre de points<sup>15</sup>.

### 1.3 Un peu de philosophie

En fait, tout ce qui précède parle bien peu du hasard. On cherche plutôt à le mesurer pour arriver à faire des prévisions. Cependant, cela ne répond absolument pas à la question "ce phénomène est-il réellement aléatoire ?".

La physique nous fournit quelques rares cas.

**Désintégration de particule :** la demi-vie d'une particule radioactive est le temps au bout duquel la moitié d'une population de telles particules se sera désintégrée. Cependant, du point de vue d'une particule, le temps au bout duquel elle se désintégrera est considéré comme aléatoire.

12. Le produit de tribus sur  $(\Omega_i)_{i \in I}$ , où  $I$  est un ensemble d'indices quelconque, est la tribu engendrée par – la plus petite tribu contenant – les ensembles de la forme  $\Pi_{i \in I} A_i$ , où tous les  $A_i$  sont mesurables et seulement un nombre fini d'entre eux est différent de  $\Omega_i$ .

13. *Random Sets & Integral Geometry*, Wiley, 1974.

14. Ceci n'est pas très restrictif, car dans la pratique – ici dans les sciences de la terre – un ensemble et son adhérence jouent le même rôle.

15. Ceci découle directement du théorème de Choquet de caractérisation des fermés aléatoires. Un ensemble aléatoire  $X$  est caractérisé par sa fonctionnelle  $T(K) = P(K \cap X \neq \emptyset)$  où  $K$  parcourt l'ensemble des compacts de  $\mathbb{R}^n$ . Deux fonctionnelles différentes donnent naissance à deux ensembles aléatoires que l'on peut distinguer en termes aléatoires.

**Mécanique quantique** : le hasard est ici intrinsèque à la théorie, avec la notion de fonction d'onde. Ce comportement aléatoire n'est essentiellement perceptible qu'à l'échelle atomique.

Dans les autres cas, les probabilités sont plutôt une façon de modéliser notre méconnaissance fine d'un phénomène ou la volonté de ne pas entrer dans les détails. On imagine que si on avait plus de connaissances sur le phénomène en question (équation d'évolution, conditions initiales entre autres), on pourrait passer à des modèles déterministes et par suite parfaitement prévisibles. L'aléatoire serait ainsi évacué totalement.

La mécanique statistique fournit un exemple de cela. Si on considère une boîte contenant un grand nombre de molécules, ce qui est le cas pour une boîte de taille humaine, on ne peut imaginer suivre le mouvement de chaque molécule et leurs collisions mutuelles. La loi des grands nombres, que l'on verra dans la suite de ce cours, montre que l'on peut considérer l'ensemble comme un groupe avec des grandeurs macroscopiques telles que la pression, notion parfaitement déterministe, alors qu'elle correspond aux molécules qui frappent une paroi donnée, et cela, de manière "aléatoire".

### 1.3.1 Les phénomènes chaotiques

Regardons maintenant de plus près ce qui nous permettrait d'affirmer qu'un phénomène est intrinsèquement aléatoire. Considérons la série temporelle de la FIGURE 1.9. Intuitivement, il s'agit d'une suite de valeurs au hasard. Cependant, cette série a été générée par des techniques purement déterministes.

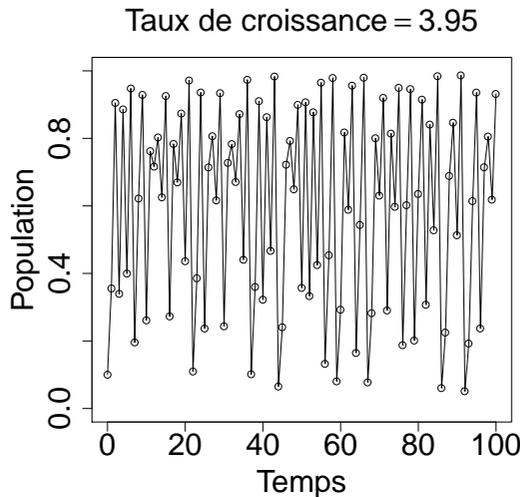


FIGURE 1.9 – Série temporelle présentant un caractère fortement aléatoire.

Il correspond à un exemple de dynamique de populations, connu sous le nom de "suite logistique", équivalent discret du "modèle dynamique de Verhulst". On considère le nombre d'individus  $u_n$  d'une population à des temps discrets

$n = 1, 2, 3, \dots$ . Le taux de croissance est  $\alpha$ . Pour  $\alpha$  supérieur à 1, la population croît de manière exponentielle. Si on considère que cette population doit se nourrir et cela, sur un territoire donné fini, cette croissance sera contrebalancée par le fait que, si la population est trop importante, le facteur  $\alpha$  va diminuer. Si on renormalise la population à 1, correspondant à une population que le territoire peut nourrir raisonnablement, on obtient l'équation d'évolution suivante, où le taux de croissance devient  $\alpha(1 - u_n)$  :

$$u_{n+1} = \alpha \cdot u_n (1 - u_n). \quad (1.7)$$

Les suites temporelles que l'on voit sur la FIGURE 1.10 sont obtenues pour différentes valeurs de  $\alpha$ . Les résultats peuvent surprendre : on passe d'un régime où la population croît jusqu'à une asymptote, à un régime où l'on oscille de part et d'autre de l'asymptote, puis à un régime à deux asymptotes, puis à un régime où l'on ne distingue plus de limite.

Si l'on examine les valeurs limites pour  $\alpha \in [1, 4]$ , on observe pour les petites valeurs de  $\alpha$  une convergence vers une valeur limite, solution de l'équation  $u_{n+1} = u_n$ . Puis on observe un dédoublement des valeurs limites, une "bifurcation" dans le vocabulaire des systèmes dynamiques, puis l'ensemble des valeurs limites devient dense dans un intervalle.

Cet exemple montre que des systèmes purement déterministes peuvent "mimer" des situations que l'on qualifierait d'aléatoires. En fait, ces systèmes partagent deux caractéristiques fondamentales :

- **la sensibilité aux conditions initiales** : même si les conditions initiales ( $u_0$  par exemple pour la suite logistique) sont connues avec une très grande précision, les trajectoires vont se mettre rapidement à diverger.
- **la récurrence** : pour toutes les conditions initiales (presque toutes, car il y a toujours des cycles et points fixes, mais ils sont "rares"), le système passe de manière dense et une infinité de fois dans toutes les parties de l'espace des phases, nous emmenant vers des propriétés statistiques.

Ces systèmes, appelés systèmes chaotiques on fait l'objet de nombreuses recherches et sont à la base des générateurs aléatoires que l'on trouve de manière simple dans tous les ordinateurs<sup>16</sup>. Ce sont ceux-là qui nous ont permis de simuler le gain dans le jeu de pile ou face. Ce sont encore eux qui permettent à une grande échelle de simuler des phénomènes aléatoires. Dans la suite du cours, nous verrons comment, à partir d'une distribution aléatoire générique, on peut simuler toutes sortes de lois de probabilités (changement de variable, inversion de la fonction de répartition, méthode d'acceptation et rejet, etc.).

Notons enfin que de nombreux phénomènes physiques sont modélisés de manière déterministe par des équations d'évolution qui présentent les caractéristiques du chaos déterministe, tels les systèmes à trois corps ou la notion de pression en physique statistique.

Si l'on veut revenir à la philosophie, nous sommes donc pris entre deux affirmations en apparence contradictoires :

---

16. Voir à ce sujet le livre pionnier de Donald KNUTH, *The Art of Computer Programming*, Volume 2, Addison Wesley, 1997 pour la troisième édition.

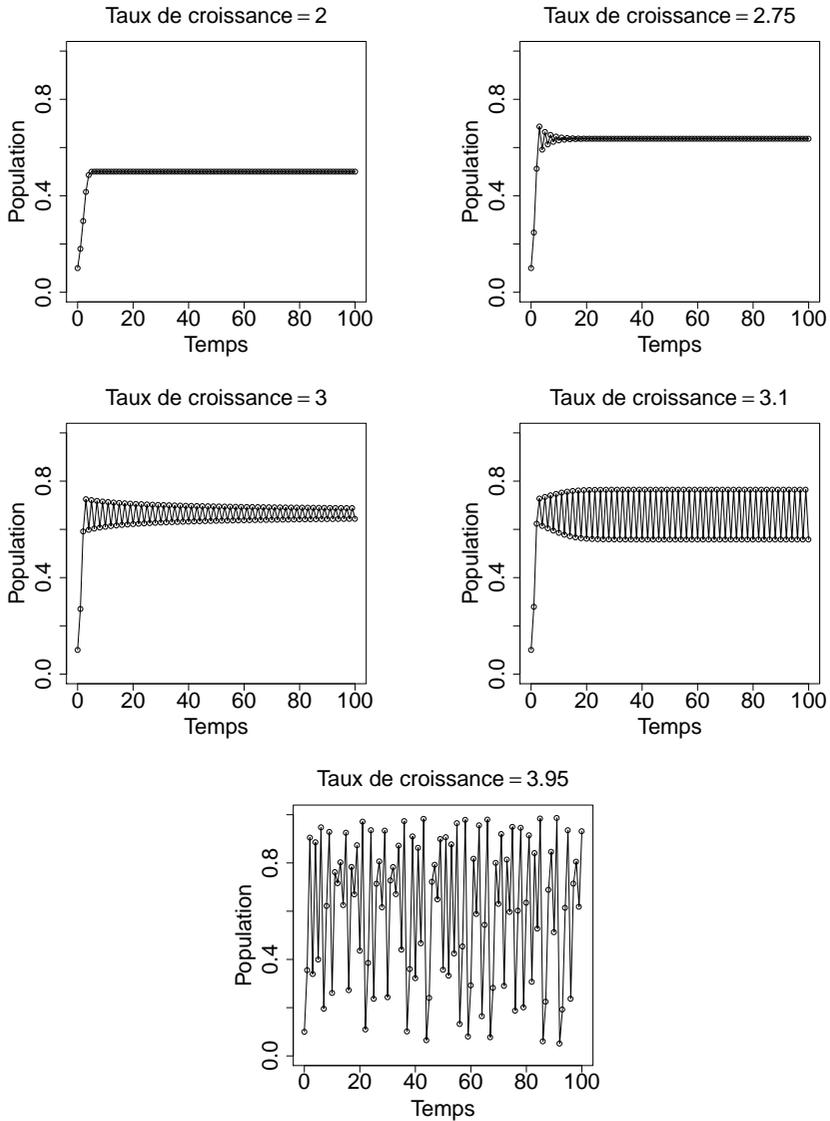


FIGURE 1.10 – Séries temporelles de la suite logistique pour différentes valeurs du taux de croissance. On observera le passage d’une asymptote à deux asymptotes, puis à un régime chaotique.

1. Le monde est purement aléatoire, et tout déterminisme qu'on y observe est le fait de la loi des grands nombres (on n'observe que la moyenne sur un très grand nombre d'échantillons).
2. Le monde est régi par des lois déterministes et tout aléatoire n'est que l'expression du chaos déterministe.

Choisir entre les deux n'est pas l'objet de ce livre. On se contentera du "modèle probabiliste" qui n'est qu'un modèle et qui se juge par la fécondité de son explication des phénomènes que l'on observe et par la qualité de ses prévisions. Ceci peut se résumer dans la citation de Georges MATHERON (Estimer et choisir) : *il n'y a pas de probabilités en soi, il n'y a que des modèles probabilistes.*

### 1.3.2 Simulations de Monte-Carlo

De manière étonnante, les méthodes probabilistes ont montré leur efficacité pour effectuer des calculs purement déterministes. En voici quelques exemples.

#### Mesure de surface (calcul d'intégrales)

Supposons que l'on veuille mesurer la surface d'une forme inscrite dans un carré, potentiellement grand. Si on connaît de manière analytique les contours de la forme, on peut espérer effectuer le calcul directement. Cependant, si ces équations sont trop complexes, ou si de telles équations n'existent pas, et que la seule question à laquelle on puisse répondre est : pour tout point dans le carré, on peut dire s'il appartient ou non à la forme, alors, la probabilité qu'un point tiré au hasard dans le carré tombe dans la forme est la surface de la forme divisée par la surface du carré. Pour arriver au résultat, on considère un grand nombre  $n$  de points aléatoires indépendants tombant dans le carré et on mesure la proportion de points tombés à l'intérieur de la forme, que l'on multiplie par la surface du carré. Cela nous donne un estimateur de la surface que l'on cherche, d'autant plus précis que le nombre de points est grand<sup>17</sup>.

En fait, dans l'exemple précédent, il s'agit d'estimer une intégrale au moyen d'une espérance, comme nous le verrons dans le chapitre dédié aux méthodes de Monte-Carlo. Le calcul d'intégrales par des méthodes probabilistes présente un intérêt fort dans les espaces de grande dimension. En effet, si l'on veut échantillonner le cube de côté unité en dimension, disons, 100, si l'on divise ne serait-ce que chaque axe en 2, nous obtenons  $2^{100}$  valeurs à renseigner, ce qui est énorme, et le calcul de surface est alors grossier. Simuler un point aléatoire dans ce même cube nécessite la génération de 100 valeurs uniformes sur  $[0, 1]$ , ce qui est immédiat.

#### Descente de gradient

Ajouter de l'aléatoire dans un algorithme de calcul numérique a également prouvé son efficacité, comme le montrent les exemples du recuit simulé ou de

---

17. La distribution de l'estimateur est une loi binômiale, dont le paramètre est précisément la probabilité que l'on cherche. On a donc au passage également un estimateur de la précision de la méthode, donné par la variance de la loi binômiale.

# P

ourquoi un nouveau livre d'introduction aux probabilités ?

Inspiré du cours de probabilités de première année du cycle ingénieur civil dispensé par les auteurs à l'École des mines de Paris, cet ouvrage se distingue par sa dimension concrète et son ancrage dans diverses applications du monde de l'ingénieur, illustrant ainsi la devise de l'École : Théorie et Pratique. L'évaluation anonyme annuelle de ce cours, plébiscité par les étudiants, a permis de recueillir de nombreux commentaires pour enrichir et améliorer cet ouvrage.

Sans sacrifier la rigueur propre à l'enseignement des mathématiques, les auteurs font preuve d'un grand effort de pédagogie, avec l'objectif de permettre aux esprits les moins abstraits d'appréhender les concepts probabilistes. Ils apportent des réponses claires à la question fréquemment posée : « mais à quoi ça sert » ?

L'ouvrage aborde notamment le sujet des simulations de Monte-Carlo et esquisse de nombreux liens avec les statistiques ou l'intelligence artificielle. Il propose également une leçon inaugurale inédite, motivant l'introduction du formalisme moderne des probabilités, ainsi que de nombreux exercices corrigés, largement tirés des expériences en recherche de ses auteurs.

**Silvère Bonnabel** est ingénieur civil des Mines de Paris, docteur en Mathématiques et Automatique, titulaire d'une habilitation à diriger les recherches en mathématiques. Professeur à Mines Paris - PSL, ses travaux se concentrent sur l'élaboration de méthodes mathématiques et leurs applications à des problèmes industriels, essentiellement dans le domaine du filtrage probabiliste, avec des applications industrielles au traitement du signal radar et à la navigation haute précision.

**Michel Schmitt** est polytechnicien et ingénieur du corps des Mines, docteur en morphologie mathématique et titulaire d'une habilitation à diriger les recherches. Il a rejoint Mines Paris - PSL pour y exercer successivement les fonctions de directeur du centre de géostatistique, puis directeur de la recherche. Ses centres d'intérêt scientifiques incluent l'analyse d'image par la morphologie mathématique, les ensembles aléatoires et les réseaux de neurones.

35 euros

