MATHÉMATIQUES EN DEVENIR

Mathématiques en devenir

101. — Jacques Faraut. *Analyse sur les groupes de Lie. Une introduction.* Deuxième édition revue et augmentée

102. — Patrice Tauvel. *Corps commutatifs et théorie de Galois*

103. — Jean Saint Raymond. *Topologie, calcul diff. et variable complexe*

104. — Clément de Seguins Pazzis. *Invitation aux formes quadratiques*

105. — Bruno Ingrao. *Coniques projectives, affines et métriques*

106. — Wolfgang Bertram. *Calcul différentiel topologique élémentaire*

107. — Henri Lombardi & Claude Quitté. *Algèbre commutative. Méthodes constructives. Modules projectifs de type fini.*

108. — Frédéric Testard. *Analyse mathématique. La maîtrise de l'implicite*

109. — Grégory Berhuy. *Modules : théorie, pratique… et un peu d'arithmétique.* Deuxième édition

110. — Bernard Candelpergher. *Théorie des probabilités. Une introduction élémentaire (Nouveau tirage bonifié)*

111. — Philippe Caldero et Jérôme Germoni. *Histoires hédonistes de groupes et de géométries. Tome premier*

112. — Gema-Maria Díaz-Toca, Henri Lombardi & Claude Quitté. *Modules sur les anneaux commutatifs*

113. — Philippe Caldero et Jérôme Germoni. *Histoires hédonistes de groupes et de géométries. Tome second* – encores

114. — Alain Debreil. *Groupes finis et treillis de leurs sous-groupes*

115. — François Rouvière. *Initiation à la géométrie de Riemann*

116. — Nikolaï Nikolski. *Matrices et opérateurs de Toeplitz*

117. — Philippe Caldero et Jérôme Germoni. *Nouvelles histoires hédonistes de groupes et de géométries. Tome premier (Nouveau tirage)*

118. — Martine et Hervé Queffélec. *Analyse complexe et applications.* Nouvelle édition revue et bonifiée

119. — Alain Debreil, Jean-Denis Eiden, Rached Mneimné et Tuong-Huy NGuyen. *Formes quadratiques et géométrie*

120. — Christian Leruste.*Topologie algébrique–Une introduction, et au-delà*

121. — Grégory Berhuy. *Algèbre : le grand combat.* Deuxième édition

122. — Philippe Caldero et Jérôme Germoni. *Nouvelles histoires hédonistes de groupes et de géométries. Tome second*

123. — Charles-Michel Marle. *Géométrie symplectique et géométrie de Poisson*

124. — Pascal Boyer. *Petit compagnon des nombres et de leurs applications*

125. — Laurent Le Floch, Frédéric Testard. *Probabilités 1 – Le hasard* est *la nécessité*

126. — Laurent Le Floch, Frédéric Testard. *Probabilités 2 – Le hasard* est *la nécessité*

Laurent Le Floch & Frédéric Testard

Probabilités 2

Le hasard *est* la nécessité

Calvage & Mounet

Géomètre de formation, Laurent Le Floch est maître de conférences à l'université de La Rochelle. Il a partagé une grande partie des enseignements de probabilités de la licence de mathématiques de La Rochelle avec Frédéric Testard de 2010 à 2018. laurent.lefloch@univ-lr.fr

Frédéric Testard (1961-2018) a été coauteur avec R. Mneimné de la célèbre « Introduction à la théorie des groupes de Lie classiques », avec L. Salem et C. Salem du grand succès de librairie « Les plus belles formules mathématiques »(traduit en finnois, anglais, japonais, chinois et arabe) et auteur de « Analyse mathématique – La maîtrise de l'implicite ».

Mathematics Subject Classification (1991) – Primary :

05-01 Combinatorics (Instructional exposition)
28-01 Measure and Integration (Instructional exposition)
28-A05 Measurable sets
28-A25 Integration with respect to measures
28-A35 Measures and integrals in product spaces
60-01 Probability Theory and Stochastic Processes (Instructional exposition)
60-A05 Foundations of probability theory (General questions)
60-A10 Foundations of probability theory (Probabilistic measure theory)
60-B10 Foundations of probability theory (Convergence of probability measures)
60-E05 Distribution theory (General theory)
60-E05 Distribution theory (Characteristic functions)
60-F05 Central limit and other weak theorems
60-F15 Strong theorems
60-G05 Foundations of stochastic processes
60-G42 Martingales with discrete parameter
60-G42 Sums of independent random variables ; random walks

ISBN 978-2-9163-5274-9

9 782916 352749

⊗ Imprimé sur papier permanent

À Sylvie, Marie et Paul
À Dominique

Avant-propos

Le hasard est le plus grand romancier du monde ;
pour être fécond, il n'y a qu'à l'étudier.
Honoré de Balzac

Ce livre est destiné aux élèves de lycée (pour une partie du premier tome), aux étudiants de licence ou de master de mathématiques (en particulier les candidats aux concours du CAPES ou de l'agrégation), aux élèves de CPGE, aux enseignants en mathématiques et plus généralement à toute personne intéressée par la théorie des probabilités. Le premier tome de ce livre correspond à l'approche « élémentaire » des probabilités telle qu'elle est classiquement enseignée jusqu'en deuxième année de licence de mathématiques. Le deuxième tome correspond à l'approche moderne des probabilités, initiée par Andreï Kolmogorov, qui s'appuie sur la théorie de la mesure.

Dans les discussions qui ont précédé l'écriture du livre, avant même de définir précisément son contenu, nous avions retenu trois grands principes :

• une approche spiralée dans la présentation des différents concepts probabilistes (variables aléatoires, indépendance, conditionnement) : ces concepts reviennent régulièrement dans des cadres différents (discret fini, discret infini, continu, abstrait) ;

• des énoncés d'exercices volontairement très détaillés. Ils permettent au lecteur d'approfondir sa compréhension des notions rencontrées. Les très nombreux exercices (plus de 700 au total) font partie intégrante du livre ;

• un recours important à l'outil informatique. Nous avons fait le choix d'utiliser principalement le langage Python et plus occasionnellement le tableur. L'outil informatique est un outil incontournable pour la simulation d'expériences aléatoires. Il nous faut cependant signaler qu'il ne s'agit pas d'un livre d'algorithmique. Nous avons en particulier systématiquement privilégié l'aspect « naturel » des programmes au détriment parfois de leur « efficacité ».

Les cinq premières parties du livre correspondent à l'approche spiralée évoquée précédemment. Ces parties correspondent à la nécessité d'introduire des modèles mathématiques de plus en plus sophistiqués pour résoudre des problèmes eux-mêmes de plus en plus complexes (souvent motivés par la solution de problèmes plus simples traités auparavant). Les derniers chapitres (en général les deux derniers) de chacune de ces parties constituent un petit *vade-mecum* d'outils nécessaires à la description ou l'étude de ces modèles.

La notion de modèle mathématique associé à une expérience aléatoire joue un rôle important dans les premières années d'apprentissage des probabilités. Dans la théorie mathématique des probabilités, la modélisation consiste initialement à construire un ensemble Ω, appelé l'*univers probabiliste* et contenant tous les résultats possibles d'une expérience aléatoire, et à le munir d'une probabilité P, qui associe à certaines parties de Ω appelées *événements* leur probabilité, en respectant un certain nombre de règles d'additivité. Quand l'ensemble Ω est fini (comme c'était, voilà quelques années, systématiquement le cas au lycée), il est naturel de munir toutes les parties de Ω d'une probabilité. Dans ce contexte, les parties à un élément sont appelées *événements élémentaires*, et l'*équiprobabilité* correspond au cas où tous les événements élémentaires ont la même probabilité. On calcule alors les probabilités en dénombrant cas favorables et cas possibles. Dans ce cadre (développé dans les parties 1 et 2 du livre), la théorie des probabilités s'identifie très fortement à l'analyse combinatoire.

Lorsque l'ensemble Ω est infini mais dénombrable (par exemple $\mathbb{N}$), on peut encore en général calculer la probabilité de toutes les parties de Ω, mais il est impossible cette fois qu'il y ait équiprobabilité. Ce cadre est développé dans la troisième partie du livre.

Dans le cas (très fréquent en pratique) où $\Omega = \mathbb{R}$ (ou $\mathbb{R}^n$), des « obstructions techniques » liées à la *théorie de la mesure* font que l'on ne peut pas en général calculer la probabilité de toutes les parties de Ω. On se limite à calculer les probabilités de certaines parties (notamment les intervalles dans le cas réel, ou les « pavés » — produits cartésiens d'intervalles — dans le cas multidimensionnel). Ce sont ces parties que l'on appelle alors des événements, et l'ensemble de tous les événements, que nous noterons $\mathcal{T}$ est appelé une *tribu*, ou une *σ-algèbre*. La probabilité est alors une *mesure* P, définie sur $\mathcal{T}$, et de « masse totale » égale à 1. Le vocabulaire employé ici est celui de la théorie de la mesure. Le cas des variables à densité est developpé dans la quatrième partie du livre (la dernière du premier tome). Le cadre le plus général constitue la cinquième partie du livre (la première du deuxième tome). Nous indiquons ci-dessous les problèmes que peut poser le choix de Ω, dans les divers cadres mentionnés ci-dessus.

Dans l'approche classique, on choisit un univers Ω lié aux observations d'une expérience aléatoire. Cette approche soulève quelques difficultés, dont nous énumérons quelques-unes ci-après.

• Elle pose un problème lié à la subjectivité de l'observateur. Le choix de Ω peut dépendre du choix de l'observateur, de sa plus ou moins grande naïveté, mais aussi du choix des observations qu'il souhaite réaliser. Dans ce cadre, on peut concevoir que, à partir d'un même modèle expérimental, un changement de questionnement amène à modifier le choix de Ω.

• Elle pose un problème lié à « l'instabilité » éventuelle de Ω. Si l'expérience consiste en une suite de jeux non bornée d'avance (par exemple, on joue à pile ou face jusqu'au moment où l'on obtient dix piles consécutifs), la description de certains événements peut nécessiter que l'on « change de Ω à chaque étape », sauf si l'on se résout à introduire un univers Ω infini tenant compte de l'ensemble de l'expérience (et de sa durée potentiellement infinie), mais ceci pose — au minimum — d'autres problèmes, techniques en particulier.

Dans l'approche moderne (qui s'appuie sur la théorie de la mesure), l'univers Ω est un ensemble abstrait, en général inconnu (sinon par son nom). Une variable aléatoire X est une application de Ω dans $\mathbb{R}$ (par exemple), mais la variable ω n'est en général pas nommée, et l'écriture $P(X \in [0,1])$ représente la mesure de l'ensemble des éléments ω de Ω tels que $X(\omega)$ appartienne à $[0,1]$. Dans cette approche, on ne cesse de raisonner sur des parties de Ω, de leur appliquer le langage ensembliste, sans jamais référer à ces parties comme à des ensembles mais en les décrivant souvent par un discours non formalisé et événementiel.

Cette situation est unique dans le cadre du travail mathématique : c'est à peu près le seul cas où la variable dont dépendent les fonctions n'a aucune importance (d'où le fait qu'elle ne soit pas nommée en général), ceci alors même que les fonctions subissent la plupart des traitements de l'analyse classique, comme notamment les estimations d'intégrales qui permettent d'obtenir espérance et variance.

Ce changement de statut de l'ensemble Ω est à l'origine de nombreuses difficultés d'apprentissage. Certaines situations permettent une approche mixte : du type théorie de la mesure avec un univers Ω explicite permettant de tout calculer. Cette approche, d'un certain point de vue fondamentale si l'on veut se convaincre que les objets dont parle l'approche moderne ont effectivement une existence mathématique (comme l'existence d'un modèle pour le jeu de pile ou face infini par exemple), est néanmoins peu utilisée en pratique. Elle conduit en général à des calculs en fait inutilement compliqués, précisément parce qu'elle donne une identité aux éléments abstraits de la théorie, détournant ainsi l'attention des concepts essentiels. On peut penser qu'il s'agit de la pire manière de faire des probabilités, mais que le passage par cette étape peut constituer une phase intéressante dans l'apprentissage, en motivant les approches plus abstraites indiquées plus haut.

Donnons un exemple. On choisit $\Omega = [0,1[$, $\mathcal{T}$ est la tribu borélienne (plus petite tribu contenant tous les intervalles de $[0,1]$, complétée pour contenir aussi toutes les parties contenues dans un événement de mesure de Lebesgue nulle), et P la mesure de Lebesgue, qui à un intervalle associe sa longueur. On définit les variables aléatoires $X_1, X_2, \ldots$ de la manière suivante :

- ▷ $X_1(t) = 1$ si $t \in [0, 1/2[$, $X_1(t) = 0$ sinon ;
- ▷ $X_2(t) = 1$ si $t \in [0, 1/4[$ ou si $t \in [1/2, 3/4[$, $X_2(t) = 0$ sinon ;

et ainsi de suite. Chaque X_{i+1} vaut 1 sur la "première moitié" des 2^i intervalles associés à X_i et 0 sur la deuxième moitié.

La famille $(X_1, X_2, \ldots)$ constitue un modèle du jeu de pile ou face infini (où l'on suppose la pièce équilibrée et l'indépendance des résultats). Dans ce modèle, le choix initial de t détermine tout le déroulement du jeu (jusqu'à l'infini). On peut voir là au moins deux difficultés.

• La signification de ce choix de t peut poser des difficultés variées, liées au manque de compatibilité entre l'idée du hasard qui préside au résultat de chaque partie et celle au contraire d'une forme de prédestination…

• D'un point de vue didactique et épistémologique, le dénuement du modèle rend difficile à accepter l'idée de hasard associée au jeu.

Toutes ces remarques indiquent les difficultés que pose le choix d'un modèle quantitatif quand il s'agit de théorie des probabilités. Pourtant ces modèles ont du succès, notamment dans leurs applications. Ce succès tient au fait que, une fois établi et assimilé un dictionnaire entre « *vision naïve des probabilités* » et « *propriétés mathématiques du modèle* », la capacité prédictive des modèles probabilistes devient aussi bonne que celle des modèles mathématiques de la physique, par exemple.

Cette capacité prédictive des modèles mathématiques des probabilités à travers des lois du hasard, quantitatives et asymptotiques, c'est l'objet de la sixième partie du livre. Nous parlons ici de prédiction quantitative. Examinons un premier énoncé quantitatif :

« *Si on lance un dé équilibré, chaque face a 1 chance sur 6, ou encore une probabilité 1/6, d'apparaître* ».

Cet énoncé quantitatif a un sens précis dans le cadre d'un modèle. Il n'en a aucun, sinon l'expression de notre croyance en une symétrie géométrique a priori que l'expérience aléatoire va se faire un plaisir de briser, dans le cadre du jeu de dé lui-même. Dans l'expérience consistant à lancer un dé et à observer le numéro écrit sur la face qui se trouve sur le dessus, il n'y a rien qui soit égal à 1/6.

On donnera plus de sens à ce nombre 1/6 en répétant 1000 fois l'expérience et en déterminant la fréquence d'apparition de chaque face. Cette fréquence représente maintenant une donnée quantitative que l'on peut comparer à 1/6, et l'expérience prouve alors que les six fréquences obtenues au bout des 1000 lancers sont en général assez voisines de 1/6. Et si on lance 10 000 fois au lieu de 1000, ce sera encore plus vrai…

C'est ce type de constatation qui conduit à dire que les lois mathématiques du hasard sont forcément des lois asymptotiques, et que ce sont ces lois qui ont une force prédictive.

La plus connue de ces lois asymptotiques est la loi des grands nombres. Cette loi possède deux variantes, connues sous le nom de loi faible et loi forte, que nous énonçons ci-dessous sans entrer dans les détails techniques.

Dans la suite, X_1, X_2, … désignent les résultats (quantitatifs) d'expériences aléatoires reproduites à l'identique et indépendamment. On note M_n la moyenne empirique
$$M_n = \frac{X_1 + \cdots + X_n}{n}\cdot$$

Loi forte des grands nombres.– Avec probabilité 1, la moyenne empirique tend vers l'espérance E(X). On dit aussi que la suite de terme général M_n converge presque sûrement vers E(X).

Loi faible des grands nombres.– Pour tout $\varepsilon > 0$, la probabilité pour que $|M_n - E(X)| > \varepsilon$ tend vers 0 quand n tend vers l'infini. On dit aussi que la suite de terme général M_n converge en probabilité vers $E(X)$.

Pour des variables aléatoires X_n de carré intégrable, la loi faible se démontre très facilement en utilisant l'additivité de la variance pour des variables indépendantes et l'inégalité de Bienaymé-Tchebycheff. La loi forte est plus difficile, même pour des variables de carré intégrable. Elle est vraie — mais la démonstration est plus difficile— pour des variables seulement supposées intégrables (en théorie des probabilités, c'est une hypothèse plus faible). Par ailleurs, il est facile de prouver que la convergence presque sûre implique la convergence en probabilités, et par conséquent que la loi forte implique la loi faible. Cette dernière est donc, elle aussi, vraie pour des variables seulement intégrables.

Ces deux lois ont des significations très différentes.

• La première n'a de sens que si l'on suppose que chaque « occurrence » ω du hasard détermine l'ensemble infini de tous les résultats des expériences réalisées. La loi forte dit alors que cette suite converge en moyenne de Césaro vers $E(X)$ sauf pour un ensemble de valeurs de ω dont la probabilité vaut 0. On retrouve les difficultés épistémologiques liées à l'infini mentionnées plus haut.

• La deuxième pose beaucoup moins de problèmes d'interprétation. Elle dit que, lorsque l'on réalise n fois l'expérience et que l'on considère tous les résultats possibles pour la moyenne empirique M_n, la plupart des valeurs prises par la variable M_n sont voisines de $E(X)$ quand n est grand.

Au lycée, les élèves se sont habitués à l'idée de stabilisation des fréquences mentionnée plus haut. Ils s'y sont habitués :

▷ dans un cadre d'expérimentation (jeux ou expériences de hasard réalisés dans la classe, le nombre d'élèves permettant d'obtenir un grand nombre de tirages) ;
▷ dans un cadre de simulation : calculatrices, tableurs et autres simulateurs ad hoc permettent de « *faire comme si* » en économisant du temps et sans doute un peu de désordre...

L'adéquation des résultats obtenus dans les deux cadres montre a priori uniquement que les simulateurs sont bien programmés (ceci étant d'ailleurs plus ou moins vrai d'une technologie à l'autre). En revanche, la stabilité des fréquences est un fait scientifique, de nature expérimentale, qui ne relève pas du modèle mathématique mais de ce que nous appellerons des « lois du hasard », auxquelles il ne semble pas plus inadéquat de croire qu'à la loi de la gravitation universelle, dans la mesure où dans les deux cas n'existe guère qu'une validation expérimentale, mais où celle-ci n'a jamais été mise en défaut. Les élèves débutants manquent sans doute d'outils et de connaissances pour se poser de telles questions, mais on peut considérer que l'adéquation entre les résultats expérimentaux (les lois du hasard) et les prévisions du modèle mathématique (la loi faible des grands nombres) constitue une validation raisonnable du modèle. Cette adéquation permet en particulier de parler de *probabilité a priori* (celle du modèle : *subjective et souvent géométrique*) et de *probabilité a posteriori* (celle de l'expérience : *objective et asymptotique*) et de les comparer.

Ce type de questionnement a été retenu par les auteurs du programme de première et apparaît dans les commentaires de ce programme : « la simulation permet d'une part d'avoir des estimations de résultats impossibles à calculer explicitement et d'autre part, par la comparaison de résultats simulés à des résultats expérimentaux, de valider des modèles ».

Enfin, la septième partie du livre est consacrée à des chapitres plus « avancés » de la théorie des probabilités, comme les chaînes de Markov, les martingales ou les marches aléatoires, qui intéresseront principalement les étudiants de master. On y trouve également des chapitres concernant quelques thèmes chers aux auteurs comme la loi de Benford, des problèmes de dés ou encore quelques développements autour du problème des anniversaires.

Nos collègues de l'université de La Rochelle nous ont soutenus tout au long de notre projet. Nous pensons notamment à Laurence Cherfils, Fabienne Marotte et Jean-Marc Garnier. Qu'ils en soient remerciés. Nous remercions particulièrement nos collègues Gilles Bailly-Maitre, Noel Fraisseix (tous deux également soutien de la première heure) ainsi que Gregory Liorit qui ont relu quelques chapitres du livre et suggéré quelques modifications.

Un grand merci enfin aux éditions Calvage & Mounet, et plus particulièrement à Rached Mneimné, pour le soin apporté à l'aboutissement de ce livre.

À la mémoire de Frédéric Testard (1961-2018)

La décision d'écrire ce livre date de la fin de l'année 2015. Frédéric et moi partagions l'essentiel de nos enseignements depuis une dizaine d'années et l'idée d'écrire un livre ensemble date des débuts de cette période. Cependant, l'ampleur de la tâche m'a longtemps fait peur. C'était sans compter sur l'opiniâtreté et la force de persuasion de Frédéric. Il est le principal artisan de ce livre. Il en est à l'origine et à la conclusion. C'est lui encore qui rompait les périodes d'inactivité. Je me souviens du moment où il m'a décrit les grandes lignes du projet. Et, devant ma mine déconfite, pour me rassurer, il m'a dit qu'il y avait un millier de pages rédigées et disséminées dans son ordinateur. C'était devenu un sujet de plaisanterie entre nous : Frédéric mène mille à zéro. Durant ces presque trois années de rédaction, son envie d'expliquer des mathématiques n'a jamais faibli.

Table des matières

Probabilités 1 - P'têt ben qu'non...

Première partie - Des réponses aux problématiques issues des jeux et de la vie courante

Probabilités 2 - ... P'têt ben qu'oui

Cinquième partie - La grande unification

Sixième partie - Le hasard dompté

XXX. Divers modes de convergence

XXXI. Le théorème central-limite

XXXII. Loi forte des grands nombres

Septième partie - Pour aller plus loin

XXXIII. Les dés sont jetés

XXXIV. Des coïncidences pas si troublantes

Le hasard *est* la nécessité

Probabilités 2
... P'têt ben qu'oui

Cinquième partie

La grande unification

Chapitre XXIV

Espaces probabilisés abstraits

Comprendre c'est avant tout unifier.
Albert Camus

1. Panorama

À relire rapidement les points de vue développés dans les trois parties précédentes, on serait tenté de penser qu'il n'y a pas *une* mais *plusieurs* théories des probabilités, à utiliser chacune en fonction d'un contexte particulier. Des théories certes pas étanches comme nous l'avons vu à plusieurs reprises, mais possédant chacune des outils *ad hoc* et parfois guère transférables.

Nous allons ici nous attacher à montrer ce qui unifie ces points de vue. On ne peut qu'être frappé par la similarité des résultats : pour citer un simple exemple, des variables indépendantes ont une covariance nulle, que l'on soit dans le cadre fini, discret ou continu. Ceci s'applique même aux outils : la ressemblance entre les théorèmes de comparaison pour la convergence des séries ou des intégrales généralisées suggère l'existence d'un énoncé général les englobant.

C'est au début du XX$^{\text{ème}}$ siècle qu'a été élaborée cette unification, avec la création par Lebesgue [1] de la théorie de la mesure et de l'intégration qui porte son nom. Au prix d'un sérieux effort d'abstraction, ces théories fournissent un cadre pour l'intégrale ordinaire dans lequel beaucoup de théorèmes, notamment sur la convergence, se simplifient, mais aussi une écriture unifiée pour tous les contextes probabilistes.

Dans ce nouveau cadre, la variété des univers aléatoires décrits précédemment n'a plus lieu d'être : un univers « *générique* », encore noté Ω, les

1. Henri Lebesgue (1875-1941) est le créateur de la théorie de l'intégration étudiée dans cette cinquième partie du livre. Il présenta ses résultats essentiels lors de la soutenance de sa thèse, en 1902.

remplace tous. Cet univers, muni d'une mesure de probabilité P, est alors la source de variables aléatoires qui se chargeront, par leurs propriétés particulières, de prêter vie aux divers contextes — fini, discret, continu — que nous avons décrits jusqu'ici. L'outil fondamental que constitue le calcul des espérances est incarné dans ce cadre par la théorie de l'intégration, et des théorèmes aux preuves multiples tels que la linéarité de l'espérance, deviennent de simples propriétés de calcul intégral.

Enfin, il n'est pas désagréable d'observer que théorie des séries et des intégrales généralisées relevant elles aussi (mais cette fois dans un contexte non probabiliste) de cette même conception unifiée, les théorèmes de comparaison évoqués plus haut apparaissent effectivement comme des variations d'un même énoncé général sur l'intégrabilité au sens de Lebesgue.

2. Espaces probabilisés abstraits

Un espace probabilisé est un ensemble Ω, *a priori* quelconque — mais de préférence assez gros si on veut pouvoir décrire des phénomènes complexes, comme nous le verrons au chapitre suivant — dans lequel on a identifié certaines parties appelées des **événements**, et défini une manière de calculer la probabilité de ces événements, non pas constructive comme la donnée des p_i sur un univers fini par exemple, mais axiomatique : une probabilité doit satisfaire certaines règles de calcul (d'ailleurs sans grande surprise).

Nous décrivons ci-dessous les propriétés exigées tant des événements que de leurs probabilités : ces propriétés constituent l'axiomatique de la théorie abstraite des probabilités.

2.1. Les événements

Dans les contextes finis ou dénombrables, toute partie de l'univers aléatoire était un événement. Dans le contexte continu, nous nous sommes surtout intéressés à des événements tels que « $X \in [a,b]$ », ou des intersections ou réunions d'événements de ce type. Dans toutes ces situations, les événements étaient définis *après* l'expérience aléatoire, la conception — plus ou moins implicite — de l'univers aléatoire faisant en sorte que l'on puisse toujours calculer leur probabilité.

Le point de vue ici est à peu près opposé. On se donne d'abord Ω, puis les événements, et les variables aléatoires n'arrivent qu'après. C'est, comme indiqué plus haut, le choix d'un Ω suffisamment riche qui permettra d'étudier les variables qui nous intéressent (ce sera décrit au chapitre suivant, où nous prouverons que la théorie n'est pas construite sur du sable et que les objets dont nous décrivons les propriétés existent effectivement).

2.1.1. Définition. Soit Ω un ensemble et soit $\mathcal{A}$ une famille de parties de Ω.

On dit que $\mathcal{A}$ est une **tribu**, ou une **σ-algèbre** si elle vérifie les propriétés suivantes.

(a) La partie Ω appartient à $\mathcal{A}$.

(b) Si $A \in \mathcal{A}$, la partie A^C (complémentaire de A) appartient aussi à $\mathcal{A}$.

(c) Si les parties A_n ($n \in \mathbb{N}$) appartiennent à $\mathcal{A}$, leur réunion aussi.

On reformule souvent la propriété (c) en disant que $\mathcal{A}$ est « *stable par réunion dénombrable* », ou « *σ-stable* ».

Les éléments de $\mathcal{A}$ sont appelés des **événements**. L'élément Ω est appelé **événement certain**.

Exemples

L'ensemble $\mathcal{A} = \mathcal{P}(\Omega)$ toutes les parties de Ω est évidemment une tribu. C'est la plus grosse possible. On l'appelle la **tribu discrète**.

L'ensemble $\mathcal{A} = \{\Omega, \emptyset\}$ est aussi une tribu (nous laissons le lecteur le vérifier). C'est la plus petite. On l'appelle la **tribu grossière**.

Les propriétés (a) à (c) ont quelques conséquences simples et utiles.

2.1.2. Théorème. *Soit Ω un ensemble et soit $\mathcal{A} \subset \mathcal{P}(\Omega)$ une tribu.*

(i) La partie $\emptyset$ (ensemble vide) appartient à $\mathcal{A}$. On l'appelle ***événement impossible****.*

(ii) Si les parties A_n ($n \in \mathbb{N}$) appartiennent à $\mathcal{A}$, leur intersection aussi.

En d'autres termes, $\mathcal{A}$ est aussi « *stable par intersection dénombrable* ».

Démonstration. (i) C'est une conséquence immédiate de (a) et (b).

(ii) On utilise l'égalité ensembliste :

$$\left(\bigcap_{n=0}^{+\infty} A_n\right) = \left(\bigcup_{n=0}^{+\infty} A_n^C\right)^C$$

et les points (a) et (c). $\square$

En pratique, on étudie souvent des réunions (ou intersections) finies d'événements : le théorème suivant indique qu'il s'agit là encore d'événements.

2.1.3. Théorème. *Toute tribu est stable par réunion ou intersection finies.*

Démonstration. Pour montrer que si $A_1, \ldots, A_n \in \mathcal{A}$, leur réunion

$$F = A_1 \cup \cdots \cup A_n$$

appartient à $\mathcal{A}$, on pose $A_{n+1} = A_{n+2} = \cdots = A_n$ et on observe que F est égal à la réunion dénombrable des A_i.

Pour une intersection, soit on reprend la démonstration du point (ii) du théorème précédent, soit on « complète » encore l'intersection avec des A_i égaux à A_n ou Ω. □

Nous laissons le lecteur vérifier la propriété suivante.

2.1.4. Proposition. *Si A et B sont des événements, $A \setminus B$ (éléments de A qui ne sont pas dans B) est un événement.*

Ces propriétés permettent de faire sur les événements les opérations habituelles : le fait qu'un événement ne se produise pas est encore un évenement (propriété (ii) de la définition), le fait que l'un des événements d'une famille dénombrable se produise est encore un événement (propriété (iii) : par exemple, le fait qu'une pièce « finisse par tomber sur pile » est la réunion dénombrable des événements « *la pièce tombe sur pile au $n^{\text{ème}}$ coup* »), le fait que tous se produisent aussi, etc.

La nécessité que Ω soit un événement peut ne pas apparaître aussi claire : mais elle est nécessaire si on veut qu'existe au moins un événement A (car alors A^C est un événement et $A \cup A^C = \Omega$ en est aussi un), mais aussi pour pouvoir écrire des phrases telles que « *la probabilité totale est égale à* 1 : il s'agit bien de la probabilité de l'événement certain Ω.

2.1.5. Définition. Espace probabilisable

On appelle **espace probabilisable** tout couple $(\Omega, \mathcal{A})$, où Ω est un ensemble et $\mathcal{A}$ une tribu de parties de Ω.

2.2. Définition axiomatique d'une probabilité

Une fois que l'on dispose d'un espace probabilisable, on peut associer aux événements leur probabilité.

2.2.1. Définition. Mesure de probabilité

Soit $(\Omega, \mathcal{A})$ un espace probabilisable. Une application $P : \mathcal{A} \to \mathbb{R}$ est une **mesure de probabilité** si elle vérifie les conditions suivantes.