Ibrahima Dramé

L3
M1
Écoles
d'ingénieurs

Théorie des probabilités

Fondements, méthodes analytiques et simulation numérique

Cours complet et exercices corrigés



ellipses

Références sciences

Théorie des probabilités

Fondements, méthodes analytiques et simulation numérique

Ibrahima Dramé

ellipses

Collection Références sciences

dirigée par Paul de Laboulaye
paul.delaboulaye@editions-ellipses.fr

Retrouvez tous les livres de la collection et des extraits sur www.editions-ellipses.fr

ISBN 9782340-121003
Dépôt légal : octobre 2026
©Ellipses Édition Marketing S.A.
8/10 rue la Quintinie 75015 Paris

DANGER
LE PHOTOCOPILLAGE TUE LE LIVRE

Le Code de la propriété intellectuelle et artistique n'autorisant, aux termes des alinéas 2 et 3 de l'article L. 122-5, d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale, ou partielle, faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause, est illicite » (alinéa 1er de l'article L. 122-4).
Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

www.editions-ellipses.fr

Préface

L'époque où le calcul de probabilités était considéré par beaucoup de mathématiciens comme une spécialité mineure, et où on pouvait suivre un cursus universitaire complet en mathématiques sans apprendre les bases des probabilités, est maintenant bien révolue.

D'une part, à peu près tous les domaines des mathématiques les plus fondamentales, que ce soient la théorie des nombres, l'algèbre, l'analyse, la géométrie ou la logique, ont des interactions importantes avec les probabilités, et d'autre part dans les applications des mathématiques, la prise en compte des phénomènes aléatoires devient incontournable. L'algorithmique utilise de plus en plus de démarches utilisant l'aléatoire, le traitement statistique d'énormes quantités de données, et les progrès de l'intelligence artificielle, basée pour une large part sur l'apprentissage, sont construits sur les outils du calcul des probabilités et des processus stochastiques.

Ces évolutions rendent l'apprentissage du calcul des probabilités de plus en plus essentielle pour tous les mathématiciens, mais aussi pour de nombreux économistes, des spécialistes tant de la finance que des sciences physiques et de la vie. Et les étudiants sont de plus en plus nombreux à vouloir apprendre les mathématiques de l'aléatoire : calcul des probabilités et statistique.

C'est donc une bonne nouvelle lorsque un manuel de grande qualité, comme celui d'Ibrahima Dramé, vient enrichir la bibliothèque accessible aux étudiants.

Encore faut-il que le nouveau manuel apporte du neuf. Dans cet ouvrage, les bases du calcul des probabilités sont présentées de façon progressive, rigoureuse et systématique. Je voudrais saluer trois aspects de ce manuel, qui je pense contribueront à son succès. D'une part la liste des lois de probabilités, tant discrètes que continues, est particulièrement fournie. D'autre part ce manuel contient un grand nombre d'exercices, de difficultés variées, dont les corrections sont données en fin d'ouvrage. Enfin un chapitre entier est consacré à la simulation des variables aléatoires, et à la mise en oeuvre de la méthode de Monte Carlo. Je pense que c'est une excellente idée d'inclure un tel chapitre dans un ouvrage de base sur les probabilités. Simuler des variables aléatoires aide à approfondir la compréhension des objets étudiés dans un cours de probabilités.

Je formule le voeu que ce manuel aide de nombreux étudiants, en France, en Afrique et dans tous les pays francophones, à assimiler les bases de la théorie des probabilités.

Marseille, le 4 avril 2026
Étienne Pardoux

Avant-propos

Le mot **Hasard** est un mot d'origine arabe : *az-zahr*, le dé. Il est apparu en français pour signifier tout d'abord un jeu de dés, puis plus généralement un évènement non prévisible, et par extension le mode d'apparition de ce type d'évènement. Il peut sembler irréaliste, voire prétentieux, de chercher à quantifier le hasard, tant il semble échapper par nature à toute mesure. C'est pourtant ce qui a conduit à la notion de **Probabilité**. Nous allons dans ce livre introduire ce concept mathématique, dont la puissance permettra de modéliser d'innombrables situations où le hasard intervient, dépassant ainsi largement le cadre restreint des jeux de dés et tirages de cartes.

Les probabilités influencent notre quotidien bien plus qu'on ne le soupçonne. Elles aident à répondre à des questions concrètes : combien de temps vais-je attendre chez le boulanger ? Mon humeur dépend-elle plus de la météo ou de mon alimentation ? Les sondages reflètent-ils fidèlement l'opinion d'une population ? Derrière ces interrogations se cachent des phénomènes aléatoires que les mathématiciens ont étudiés et modélisés à l'aide de lois probabilistes, comme la loi normale ou la loi de Poisson. Ces outils, bien au-delà de leur intérêt théorique, sont également largement utilisés dans la vie professionnelle. Aujourd'hui, les probabilités constituent une branche des mathématiques à la fois dynamique et en constante évolution. L'intuition probabiliste est devenue un outil central dans de nombreuses disciplines, et elle forme le socle conceptuel des statistiques modernes.

La théorie des probabilités occupe une place centrale dans les mathématiques contemporaines, tant pour ses fondements conceptuels que pour ses nombreuses applications. Elle fournit un cadre rigoureux pour modéliser et analyser des phénomènes aléatoires, et intervient de manière transversale dans un large éventail de disciplines : physique (mécanique quantique, physique statistique), biologie (génétique, épidémiologie), climatologie, écologie, intelligence artificielle, traitement du signal, finance quantitative, économie, ou encore sociologie. Cette transversalité témoigne de la pertinence et de la puissance des outils probabilistes, aussi bien pour décrire l'incertitude que pour guider la prise de décision dans des contextes complexes.

Dans ce contexte, ce livre propose une introduction rigoureuse et progressive à la théorie des probabilités, en mettant l'accent sur les fondements théoriques nécessaires à une utilisation pertinente et éclairée des outils probabilistes. Il s'adresse à un public de niveau troisième année de licence ou première année de master en mathématiques, ainsi qu'aux élèves d'écoles d'ingénieurs. La lecture de cet ouvrage suppose une bonne maîtrise des notions fondamentales d'analyse réelle, d'algèbre linéaire et de calcul différentiel, telles qu'elles sont enseignées en première et deuxième années de licence dans les universités.

Le présent ouvrage propose une introduction structurée aux principaux outils de la théorie des probabilités. Il est conçu pour approfondir la compréhension des outils probabilistes usuels, tant du point de vue théorique qu'en vue d'applications concrètes. Le chapitre 1 est consacré aux fondements axiomatiques de la probabilité. On y introduit les espaces probabilisés au sens de Kolmogorov, les variables aléatoires, ainsi que les principales lois discrètes et continues. Cette première étape vise à établir le cadre mathématique rigoureux dans lequel s'inscrit l'ensemble de l'analyse probabiliste. Le chapitre 2 traite des fonctions caractéristiques, outils puissants permettant de caractériser entièrement la loi d'une variable aléatoire ou d'un vecteur aléatoire, et de faciliter l'étude des opérations sur les lois (sommes, convolutions, limites). Ce chapitre introduit également la notion de vecteur gaussien, centrale dans l'étude des lois multivariées, en particulier dans le contexte de la statistique et des processus stochastiques. Le chapitre 3 est consacré aux différentes notions de convergence des variables aléatoires : convergence en loi, en probabilité, presque sûre, en moyenne quadratique. Ces modes de convergence sont comparés, illustrés, et mis en œuvre dans le cadre des principaux théorèmes limites (loi des grands nombres, théorème central limite). Le chapitre 4 développe la théorie des probabilités conditionnelles et de l'espérance conditionnelle, notions fondamentales pour l'étude des dépendances partielles et de l'information partielle. Ce chapitre prépare à la compréhension des structures conditionnelles plus complexes comme les martingales ou les chaînes de Markov, et sert de pont vers la statistique bayésienne et les modèles hiérarchiques. Le chapitre 5 est consacré à la simulation des variables aléatoires, abordée du point de vue algorithmique et numérique. On y présente les techniques classiques de simulation (méthode de l'inversion, méthode de rejet, méthode de composition), ainsi que les fondements de la simulation de Monte Carlo, largement utilisée dans les méthodes numériques et l'estimation probabiliste. Enfin, le chapitre 6 rassemble des corrections détaillées des exercices, permettant au lecteur de consolider sa compréhension des concepts étudiés tout au long de l'ouvrage.

Ce parcours progressif vise à doter le lecteur d'un socle solide en probabilités, en insistant sur l'articulation entre les résultats théoriques, les outils calculatoires et les perspectives d'application. L'accent est mis sur la rigueur mathématique, sans négliger l'intuition et les motivations sous-jacentes à chaque notion introduite.

Chaque chapitre du livre se conclut par des exercices permettant de contrôler l'acquisition des notions essentielles qui y ont été introduites. Chaque exercice est rédigé sous forme d'un énoncé détaillé, destiné à guider le lecteur et à favoriser son autonomie. Les corrigés proposés sont toujours complets et commentés lorsque cela est nécessaire, en privilégiant les solutions méthodiques et raisonnables aux approches « astucieuses » ou « miraculeuses ». L'expérience montre en effet qu'un corrigé trop « brillant » peut décourager l'étudiant, qui se sent alors incapable de reproduire une telle performance, sans pour autant acquérir les étapes raisonnées qui mènent à la solution lorsqu'on dispose d'une bonne compréhension des concepts. L'expérience montre également la vertu du contre-exemple : il en est donc fait ici un large usage. Par ailleurs, lorsqu'un exercice mobilise un résultat d'algèbre ou de topologie, la démonstration est incluse si elle est courte, ou bien une référence bibliographique précise est indiquée. Certains exercices sont volontairement élémentaires, afin de faciliter l'adaptation de l'étudiant au niveau de la troisième année de licence. Les solutions détaillées donnent des exemples de rédaction possibles. Faire de nombreux exercices est sans doute le meilleur moyen de parvenir à la

compréhension de certaines notions parfois délicates. Rappelons cette maxime chinoise : « J'entends et j'oublie. Je vois et je retiens. Je fais et je comprends ».

Cet ouvrage est le fruit d'un long travail pédagogique, développé tout d'abord à l'université Gaston Berger de Saint Louis du Sénégal pour des étudiants en master mathématiques et informatiques pendant l'année académique 2013-2014 et poursuivi à Aix Marseille université, à l'université de Cheikh Anta Diop de Dakar, à l'école nationale de la statistique et de l'analyse économique Pierre Ndiaye (ENSAE de DAKAR).

Ce livre a commencé à exister comme polycopié de divers cours. Ceci dit je remercie les lecteurs qui ont déjà fait part de leurs commentaires et par avance ceux voudront bien nous communiquer leurs remarques et suggestions sur les éventuelles fautes d'orthographe, coquilles ou sur le contenu et l'organisation de l'ouvrage.

Je remercie également mes étudiants qui, année après année, m'ont (plus ou moins) écoutée raconter ce cours, ont peiné sur tous ces exercices et problèmes et m'ont poussée à l'améliorer. Mes remerciements vont enfin à feu Galaye DIA, Professeur titulaire des universités de classe exceptionnelle, Ahmadou Bamba SOW, Professeur titulaire en mathématiques Appliquées et Etienne PARDOUX, Professeur émérite à Aix Marseille Université de m'avoir donner le goût et l'amour inconditionnel de la théorie des probabilités. Je vous remercie pour vos conseils et votre soutien ; vous êtes vraiment appréciés. Merci d'être une source d'inspiration et de sagesse. Votre gentillesse et votre patience ont fait toute la différence. Merci, professeurs.

Je souhaite adresser un remerciement tout particulier au professeur Etienne PARDOUX, au professeur Ahmadou Bamba SOW et au professeur Papa NGOM, Professeur titulaire en mathématiques appliquées, pour avoir pris le temps de lire attentivement cet ouvrage et m'avoir suggéré des remarques essentielles ainsi que des corrections d'une grande rigueur. Leur lecture minutieuse et la pertinence de leurs conseils ont contribué de manière déterminante à renforcer la qualité scientifique et la clarté de ce livre.

Je tiens à exprimer ma profonde gratitude à Étienne PARDOUX, qui m'a fait l'honneur de rédiger la préface de cet ouvrage. Son soutien, sa bienveillance et la confiance qu'il m'a accordée tout au long de ce projet ont compté énormément dans son aboutissement.

Je souhaite remercier très sincèrement mon ami Abdou Salam SOUMARE, dont les encouragements constants ont été déterminants dans l'aboutissement de cet ouvrage. Sa confiance et sa présence bienveillante m'ont souvent redonné l'élan nécessaire pour ne pas me décourager et poursuivre ce projet jusqu'au bout.

A la mémoire de ma mère Adja Amy Thiam

Table des matières

1 Bases de la théorie des probabilités **13**

1.1 Préliminaires . . . 14
1.1.1 Notations et rappels sur les ensembles . . . 14
1.1.2 Tribu sur un ensemble . . . 21
1.1.3 Classes monotones . . . 24
1.1.4 Applications mesurables et mesures . . . 26
1.1.5 Compléments : les théorèmes de convergence . . . 30
1.2 Structures de base . . . 34
1.3 La notion de probabilité . . . 36
1.3.1 Espace de probabilité . . . 36
1.3.2 Probabilité conditionnelle . . . 38
1.3.3 Indépendance . . . 39
1.4 Variable aléatoire . . . 40
1.4.1 Généralités . . . 40
1.4.2 Loi d'une variable aléatoire . . . 41
1.4.3 Indépendance de variables aléatoires . . . 45
1.4.4 Espaces L^p et Moments d'une variable aléatoire . . . 47
1.5 Variable aléatoire discrète . . . 49
1.5.1 Définition et exemples . . . 49
1.5.2 Moments d'une variable aléatoire discrète . . . 50
1.5.3 Fonction génératrice . . . 51
1.5.4 Lois discrètes usuelles . . . 53
1.6 Vecteur aléatoire discret . . . 56
1.6.1 Loi conjointe . . . 56
1.6.2 Lois marginales . . . 58
1.6.3 Loi conditionnelle . . . 60
1.6.4 Moments . . . 60
1.6.5 Indépendance . . . 62
1.6.6 Espérance et matrice de covariance . . . 64
1.7 Variable aléatoire continue . . . 66
1.7.1 Définition et exemples . . . 66
1.7.2 Moments d'une variable aléatoire continue . . . 68
1.7.3 Fonction génératrice des moments . . . 70
1.7.4 Lois continues usuelles . . . 73
1.8 Vecteur aléatoire réel . . . 84

1.8.1 Fonction de répartition et densité . . . 84
1.8.2 Lois marginales . . . 86
1.8.3 Indépendance . . . 88
1.9 Changement de variables . . . 89
1.9.1 Cas univarié (variable aléatoire) . . . 90
1.9.2 Cas multivarié (vecteurs aléatoires) . . . 93
1.10 Convolution et somme de deux variables aléatoires . . . 97
1.10.1 Cas continu . . . 97
1.10.2 Cas discret . . . 98
1.10.3 Cas mixte . . . 100
1.11 Inégalités classiques . . . 101
1.12 Résumé sur les lois usuelles . . . 106
1.13 Exercices . . . 108

2 Fonction caractéristique et vecteurs gaussiens 113
2.1 Fonction caractéristique . . . 114
2.1.1 Définition et premières propriétés . . . 114
2.1.2 Quelques exemples classiques . . . 116
2.1.3 Propriété fondamentale de la fonction caractéristique . . . 117
2.1.4 La transformée de Laplace . . . 120
2.2 Vecteurs gaussiens . . . 123
2.2.1 Rappel sur les variables aléatoires gaussiennes . . . 123
2.2.2 Vecteur gaussien : définition et premières propriétés . . . 125
2.2.3 Caractérisation des vecteurs gaussiens . . . 126
2.2.4 Théorème de Cochran . . . 132
2.3 Exercices . . . 134

3 Convergence des variables aléatoires 139
3.1 Rappel . . . 140
3.1.1 Lemme de Borel-Cantelli et Loi du zéro-un de Borel . . . 140
3.1.2 Convergence monotone, lemme de Fatou et convergence dominée 142
3.2 Convergence des variables aléatoires . . . 144
3.2.1 Divers modes de convergence . . . 144
3.2.2 Critères de convergence presque sûre . . . 146
3.2.3 Liens entre les différents types de convergence . . . 149
3.3 Lois des grands nombres . . . 155
3.3.1 Inégalité de Kolmogorov . . . 155
3.3.2 Loi faible des grands nombres . . . 157
3.3.3 Loi forte des grands nombres . . . 158
3.3.4 Applications de la loi des grands nombres . . . 161
3.4 Convergence en loi . . . 162
3.5 Le théorème de la limite centrale . . . 168
3.6 Uniforme intégrabilité . . . 171
3.6.1 Définition et premières propriétés . . . 171
3.6.2 Critères d'uniforme intégrabilité . . . 172
3.6.3 Lien avec les modes de convergence . . . 173
3.7 Exercices . . . 176

4 Espérances et probabilités conditionnelles **181**
4.1 Rappels . . . 182
4.2 Espérance conditionnelle . . . 184
4.2.1 Linéarité de l'espérance conditionnelle : justification préalable . 184
4.2.2 Conditionnement par rapport à une sous-tribu (existence et unicité) 185
4.2.3 Conditionnement par rapport à un événement . . . 188
4.2.4 Propriétés de l'espérance conditionnelle . . . 189
4.2.5 Conditionnement par rapport à une variable aléatoire . . . 194
4.3 Probabilité conditionnelle . . . 196
4.4 Détermination pratique de l'espérance conditionnelle . . . 198
4.4.1 Cas discret . . . 198
4.4.2 Cas absolument continu . . . 199
4.4.3 Cas pratique : calcul de probabilité par conditionnement . . . 200
4.5 Conditionnement des vecteurs gaussiens . . . 202
4.6 Variance conditionnelle . . . 205
4.7 Exercices . . . 207

5 Simulation aléatoire et méthodes de Monte Carlo avec R **211**
5.1 Simulation de variables aléatoires . . . 212
5.1.1 Méthode d'inversion . . . 213
5.1.2 Méthode par transformation . . . 217
5.1.3 Méthode de rejet ou méthode d'acceptation-rejet . . . 219
5.1.4 Méthode de composition ou de mélange . . . 222
5.1.5 Travaux pratiques : lois classiques . . . 225
5.2 Simulation de vecteurs aléatoires . . . 230
5.2.1 Cas des vecteurs gaussiens . . . 231
5.2.2 Travaux pratiques : vecteurs gaussiens . . . 232
5.3 Intégration et estimation par Monte Carlo . . . 234
5.3.1 Intégration numérique . . . 235
5.3.2 Description et convergence de la méthode de Monte Carlo . . . 236
5.3.3 Vitesse de convergence et intervalles de confiance . . . 239
5.3.4 Techniques de réduction de variance . . . 241
5.4 Exercices . . . 265

6 Corrections des exercices **271**
6.1 Corrections des exercices proposés dans le cours . . . 271
6.1.1 Exercices du chapitre 1 . . . 271
6.1.2 Exercices du chapitre 2 . . . 278
6.1.3 Exercices du chapitre 3 . . . 283
6.1.4 Exercices du chapitre 4 . . . 286
6.1.5 Exercices du chapitre 5 . . . 294
6.2 Corrections des exercices proposés en fin de chapitre . . . 303
6.2.1 Corrigés des exercices du chapitre 1 . . . 303
6.2.2 Corrigés des exercices du chapitre 2 . . . 321
6.2.3 Corrigés des exercices du chapitre 3 . . . 335
6.2.4 Corrigés des exercices du chapitre 4 . . . 353
6.2.5 Corrigés des exercices du chapitre 5 . . . 373

Bibliographie **397**

Index **399**

Chapitre 1

Bases de la théorie des probabilités



Galerie de mathématiciens [1]

L'histoire des probabilités est intimement liée à celle des jeux de hasard. Bien que quelques calculs probabilistes apparaissent dès le Moyen Âge dans des contextes précis, ce n'est qu'au $XVII^{e}$ siècle que la théorie des probabilités commence véritablement à se constituer. Pendant près de deux siècles, elle se développe sans véritable formalisme,

1. Sources : portraits issus de Wikimedia Commons pour Beppo-Levi, Borel, Cauchy, Fubini, Laplace, Lebesgue, Markov, Poisson, Fatou et Pascal. Kolmogorov : SurovIA (Wikimedia Commons), CC BY-SA 4.0. Lévy : Konrad Jacobs (Wikimedia Commons), CC BY-SA 2.0 DE.

principalement autour du célèbre problème des partis, de problèmes d'urnes ou encore d'autres questions issues des jeux. Ce n'est qu'au XX^e siècle qu'apparaît la théorie moderne des probabilités, fondée sur la théorie de la mesure et celle de l'intégration. Depuis lors, cette théorie s'est largement diversifiée et a trouvé de nombreuses applications. Enfin, les difficultés liées aux échanges scientifiques, à la publication des ouvrages et à leur transmission à certaines époques rendent parfois délicate l'attribution de certains résultats : ainsi, la question de la paternité de la théorie des probabilités demeure en partie ouverte.

1.1 Préliminaires

1.1.1 Notations et rappels sur les ensembles

Dans cette partie, nous rappelons quelques notions fondamentales sur les ensembles, qui seront utiles tout au long de ce livre.

Considérons un ensemble E, c'est-à-dire une collection d'objets appelés *éléments* (ou *points*) de E. L'appartenance d'un point x à l'ensemble E est notée $x \in E$, tandis que $x \notin E$ signifie que x n'appartient pas à E. Un *sous-ensemble* de E est un ensemble dont tous les éléments appartiennent à E. On note $F \subset E$ lorsque F est un sous-ensemble de E (on dit aussi que F est inclus dans E).

Rappelons les opérations élémentaires sur les parties d'un ensemble. Soient A et B deux sous-ensembles de E.

• $\rightsquigarrow$ **Intersection :** $A \cap B$ est l'ensemble des points appartenant à la fois à A et B.

• $\rightsquigarrow$ **Réunion :** $A \cup B$ est l'ensemble des points appartenant à au moins l'un des deux ensembles A ou B.

• $\rightsquigarrow$ **Complémentaire :** le complémentaire de A dans E est l'ensemble des points de E qui n'appartiennent pas à A. On le note A^c ou $\bar{A}$. Si $A \subset B$, la différence ensembliste $B \backslash A$ désigne les éléments de B qui n'appartiennent pas à A ; on a alors $B \backslash A = B \cap A^c$.

• $\rightsquigarrow$ **Différence symétrique :** $A \Delta B$ est l'ensemble des points appartenant à A ou à B, mais pas aux deux. On a :

$$A \Delta B = (A \backslash B) \cup (B \backslash A) = (A \cup B) \backslash (A \cap B).$$

• $\rightsquigarrow$ **Egalité de deux ensembles :** pour montrer que $A = B$, on raisonne souvent par *double inclusion* : d'une part, montrer que $x \in A \implies x \in B$; et d'autre part que $x \in B \implies x \in A$.

• $\rightsquigarrow$ **Ensemble des parties :** on note $\mathscr{P}(E)$ l'ensemble de tous les sous-ensembles de E (aussi appelé *ensemble des parties de* E).

• $\rightsquigarrow$ **Ensemble vide :** c'est l'ensemble ne contenant aucun élément ; on le note $\emptyset$.

• $\rightsquigarrow$ **Ensembles disjoints :** deux ensembles A et B sont dits disjoints si $A \cap B = \emptyset$.

• $\rightsquigarrow$ **Propriétés usuelles :** étant donnés trois sous-ensembles $A, B, C \subset E$ (ou encore trois événements d'un espace probabilisé), les propriétés suivantes sont vérifiées :

$$A \cup B = B \cup A \qquad \text{(commutativité de la réunion)},$$
$$A \cap B = B \cap A \qquad \text{(commutativité de l'intersection)},$$

$$A\cup(B\cup C)=(A\cup B)\cup C \qquad \text{(associativité de la réunion)},$$
$$A\cap(B\cap C)=(A\cap B)\cap C \qquad \text{(associativité de l'intersection)},$$
$$A\cap(B\cup C)=(A\cap B)\cup(A\cap C) \qquad \text{(distributivité } \cap \text{ sur } \cup),$$
$$A\cup(B\cap C)=(A\cup B)\cap(A\cup C) \qquad \text{(distributivité } \cup \text{ sur } \cap),$$
$$(A\cup B)^c=A^c\cap B^c, \qquad (A\cap B)^c=A^c\cup B^c.$$

• $\rightsquigarrow$ **Lois de De Morgan :** [2] soit $(A_i)_{i\in I}$ une famille de sous-ensembles indexée par un ensemble quelconque I. Alors :

$$\left(\bigcup_{i\in I}A_i\right)\cap B=\bigcup_{i\in I}(A_i\cap B), \qquad \left(\bigcap_{i\in I}A_i\right)\cup B=\bigcap_{i\in I}(A_i\cup B),$$
$$\left(\bigcup_{i\in I}A_i\right)^c=\bigcap_{i\in I}A_i^c, \qquad \left(\bigcap_{i\in I}A_i\right)^c=\bigcup_{i\in I}A_i^c.$$

• $\rightsquigarrow$ **Ensembles usuels :**

$$\mathbb{N}=\{0,1,2,\dots\} \qquad \text{(entiers naturels)},$$
$$\mathbb{N}^*=\{1,2,3,\dots\} \qquad \text{(entiers naturels non nuls)},$$
$$\mathbb{Z}=\{\dots,-2,-1,0,1,2,\dots\} \qquad \text{(entiers relatifs)},$$
$$\mathbb{Q}=\left\{\tfrac{p}{q}\,:\,p\in\mathbb{Z},q\in\mathbb{N}^*,q\neq 0\right\} \qquad \text{(rationnels)},$$
$$\mathbb{R}=]-\infty,+\infty[\qquad \text{(réels)},$$
$$\mathbb{R}_+=[0,+\infty[=\{x\in\mathbb{R}:x\geq 0\}, \qquad \mathbb{R}_+^*=]0,+\infty[,$$
$$\mathbb{R}_-=]-\infty,0]=\{x\in\mathbb{R}:x\leq 0\}, \qquad \mathbb{R}_-^*=]-\infty,0[,$$
$$\overline{\mathbb{R}}=[-\infty,+\infty] \qquad \text{(droite réelle achevée)},$$
$$\mathbb{R}^d=\text{espace euclidien réel de dimension } d \quad (\mathbb{R}^1=\mathbb{R}),$$
$$\mathbb{C}=\{x+iy:x,y\in\mathbb{R}\} \qquad \text{(nombres complexes)}.$$

• $\rightsquigarrow$ **Cardinal d'un ensemble :** un sous-ensemble $A\subset E$ est dit *fini* si A est vide ou bien s'il existe un entier $n\geq 1$ et une bijection entre A et l'ensemble $\{1,2,\dots,n\}$. L'entier n est alors appelé le *nombre d'éléments de* A et est noté

$$\mathrm{Card}(A)=|A|=\#(A)=n.$$

Un ensemble est *infini* lorsqu'il n'est pas fini.

• $\rightsquigarrow$ **Ensemble dénombrable :** on dit qu'un ensemble E est *dénombrable* s'il existe une bijection entre E et $\mathbb{N}$, c'est-à-dire s'il est possible d'énumérer ses éléments sous la forme d'une suite $(x_n)_{n\in\mathbb{N}}$, avec $x_n\neq x_m$ dès que $n\neq m$. Exemples d'ensembles dénombrables :

$\mathbb{N}$, $\mathbb{N}^*$, $\mathbb{Z}$, $\mathbb{Q}$, {entiers pairs}, $\{n^2:n\in\mathbb{N}\}$, toute suite strictement croissante d'entiers.

2. Les lois de De Morgan sont des identités logiques formulées par le mathématicien britannique Augustus De Morgan (1806–1871).

En revanche, l'ensemble des réels $\mathbb{R}$, ainsi que tout intervalle $[a,b]$ avec $a < b$, ne sont pas dénombrables (résultat de Cantor).

Voici quelques propriétés fondamentales des ensembles dénombrables :

- Toute partie d'un ensemble dénombrable est finie ou dénombrable.
- La réunion d'une famille finie ou dénombrable d'ensembles finis ou dénombrables est elle-même finie ou dénombrable.
- Si A n'est pas fini ni dénombrable (c'est-à-dire s'il est *non dénombrable*), alors pour tout $B \subset A$ qui est fini ou dénombrable, l'ensemble $A \backslash B$ reste non dénombrable.

Définition 1.1.1 *Un ensemble E est dit* au plus dénombrable *s'il est fini ou dénombrable.*

Remarque 1.1.2 *Un ensemble non dénombrable est nécessairement infini, mais tous les ensembles infinis ne sont pas dénombrables. L'infinité des entiers naturels est plus "petite" que celle des réels :*

$$|\mathbb{N}| < |\mathbb{R}|,$$

au sens des cardinaux, ce qui illustre qu'il existe différents "niveaux" d'infini.

• $\rightsquigarrow$ **Produit cartésien d'ensembles :** si $E_1, ..., E_n$ sont des ensembles, on note $E_1 \times \cdots \times E_n$ l'ensemble des n-uplets $(x_1, \cdots, x_n)$ tels que $x_1 \in E_1$, $x_2 \in E_2$,...,$x_n \in E_n$.

• $\rightsquigarrow$ **Bijection, image directe et image réciproque :** soient E et F deux ensembles, et $f : E \longrightarrow F$ une application.

Définition 1.1.3 (**Injection, surjection, bijection**)

- **Injection :** *f est injective si deux éléments distincts de E ont des images distinctes dans F :*

  $$\forall x, x' \in E, \quad (x \neq x') \Longrightarrow (f(x) \neq f(x')).$$

  Équivalemment : $(f(x) = f(x')) \implies (x = x')$.

- **Surjection :** *f est surjective si tout élément de F possède au moins un antécédent dans E :*

  $$\forall y \in F, \ \exists x \in E, \quad f(x) = y.$$

- **Bijection :** *f est bijective si elle est à la fois injective et surjective, c'est-à-dire si chaque élément de F possède un unique antécédent dans E :*

  $$\forall y \in F, \ \exists! \ x \in E, \quad f(x) = y.$$

Remarque 1.1.4 *Si $f : E \to F$ est injective, en restreignant son ensemble d'arrivée à*

$$f(E) := \{y \in F : \exists x \in E, f(x) = y\},$$

la nouvelle application $f : E \to f(E)$ devient une bijection. Cette opération préserve l'injectivité et rend f surjective.

Définition 1.1.5 - *Soit A un sous-ensemble de E. On appelle* **image directe** *de A par f l'ensemble :*

$$f(A) = \{y \in F \ : \ \exists\, x \in A,\ f(x) = y\}.$$

$f(A)$ est donc l'ensemble des éléments de F qui sont l'image d'un élément de A. Autrement dit, $f(A)$ est l'ensemble des éléments de F possédant un antécédent dans A. On peut écrire tout aussi bien : $f(A) = \{f(x) \ : \ x \in A\}$.
- Soit B un sous-ensemble de F. On appelle **image réciproque** *de B par f l'ensemble :*

$$f^{-1}(B) = \{x \in E \ : \ f(x) \in B\}.$$

$f^{-1}(B)$ est l'ensemble des éléments x de E dont l'image par f appartient à B. Autrement dit : $f^{-1}(B)$ est l'ensemble des antécédents par f des éléments de B. On a donc pour tout élément x de E :

$$x \in f^{-1}(B) \Longleftrightarrow f(x) \in B.$$

- Si l'on considère de plus une application $g : F \to G$, alors l'image réciproque d'une partie C de G par la composée $g \circ f$ est :

$$(g \circ f)^{-1}(C) = f^{-1}(g^{-1}(C)).$$

Remarque 1.1.6 - *Si $A \subset E$, alors $f(A) \subset F$, et si $B \subset F$, alors $f^{-1}(B) \subset E$.*
- Il faut noter que

$$f^{-1}(C_F B) = C_E\left(f^{-1}(B)\right),$$

où $C_F B$ désigne le complémentaire de B dans F et $C_E\left(f^{-1}(B)\right)$ désigne le complémentaire de $\left(f^{-1}(B)\right)$ dans E.
- L'image directe d'un singleton $f(\{x\}) = \{f(x)\}$ est un singleton. Par contre l'image réciproque d'un singleton $f^{-1}(\{y\})$ dépend de f. Cela peut être un singleton, un ensemble à plusieurs éléments ; mais cela peut-être E tout entier (si f est une fonction constante) ou même l'ensemble vide (si aucune image par f ne vaut y).

Propriétés 1.1.7 (**Lois de Hausdorff**) *Soit I un ensemble quelconque. Soient $(A_i)_{i\in I}$ une famille non vide de parties de E et $(B_i)_{i\in I}$ une famille non vide de parties de F. Alors nous avons les propriétés suivantes :*

$$\star\ f\left(\bigcup_{i\in I} A_i\right) = \bigcup_{i\in I} f(A_i) \qquad \star\ f\left(\bigcap_{i\in I} A_i\right) \subset \bigcap_{i\in I} f(A_i)$$

$$\star\ f^{-1}\left(\bigcup_{i\in I} B_i\right) = \bigcup_{i\in I} f^{-1}(B_i) \qquad \star\ f^{-1}\left(\bigcap_{i\in I} B_i\right) = \bigcap_{i\in I} f^{-1}(B_i)$$

De plus, pour toutes parties B_1 et B_2 de F, nous avons : $f^{-1}(B_1 \backslash B_2) = f^{-1}(B_1) \backslash f^{-1}(B_2)$.

Proposition 1.1.8 (Dénombrabilité par image surjective) *Soient A et B deux ensembles, et $f : A \to B$ une surjection.*

- *Si A est dénombrable, alors B est au plus dénombrable.*
- *Si A est fini, alors B est fini et $Card(B) \leq Card(A)$.*

Remarque 1.1.9 *Il est possible que A soit infini et B fini. Par exemple, l'application constante $f : \mathbb{N} \to \{0\}$ définie par $f(x) = 0$ est une surjection.*

Preuve : Supposons que A soit dénombrable. Définissons sur A la relation d'équivalence

$$x \sim x' \quad \Longleftrightarrow \quad f(x) = f(x').$$

Les classes d'équivalence forment une partition de A. Dans chaque classe, choisissons un représentant particulier [3]. Soit A_0 l'ensemble de tous ces représentants. Par construction, si $x, x' \in A_0$ sont distincts, alors x et x' appartiennent à des classes disjointes, donc $f(x) \neq f(x')$. Ainsi, la restriction de f à A_0 est injective. De plus, puisque f est surjective, tout $y \in B$ possède au moins un antécédent $x \in A$. Si c est la classe de x, alors le représentant de c dans A_0 a aussi y pour image par f. Ainsi, la restriction de f à A_0 est également surjective, et constitue donc une bijection de A_0 sur B. Puisque $A_0 \subset A$ et que A est dénombrable, A_0 est au plus dénombrable. Il en découle que B est au plus dénombrable. Le cas où A est fini se traite de manière analogue, en remarquant que toute partie d'un ensemble fini est finie, ce qui donne $Card(B) \leq Card(A)$. ■

• $\rightsquigarrow$ **Fonction indicatrice :** la fonction indicatrice joue un rôle fondamental dans la théorie des ensembles. Soit A une partie d'un ensemble E. On définit la fonction $\mathbf{1}_A : E \longrightarrow \{0,1\}$ par

$$\mathbf{1}_A(x) = \begin{cases} 1, & \text{si } x \in A, \\ 0, & \text{si } x \notin A. \end{cases}$$

La fonction $\mathbf{1}_A$ est appelée *fonction indicatrice* ou *fonction caractéristique* de A. Elle est parfois notée χ_A.

Exercice 1.1.10 *Soit E un ensemble. Soient A et B deux parties de E et $(A_n)_{n\in\mathbb{N}}$ une suite de parties de E.*
1) *Montrer que* $\mathbf{1}_{A^c} = 1 - \mathbf{1}_A$. 2) *Montrer que* $\mathbf{1}_{A\cap B} = \mathbf{1}_A \times \mathbf{1}_B = \inf(\mathbf{1}_A, \mathbf{1}_B)$.
3) *Montrer que* $\mathbf{1}_{A\setminus B} = \mathbf{1}_A - \mathbf{1}_A\mathbf{1}_B$. 4) *Montrer que* $\mathbf{1}_{A\cup B} = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_{A\cap B} = \sup(\mathbf{1}_A, \mathbf{1}_B)$.
5) *Montrer que si A et B sont disjoints, alors* $\mathbf{1}_{A\cup B} = \mathbf{1}_A + \mathbf{1}_B$. *En déduire que si les $(A_n)_{n\in\mathbb{N}}$ sont deux à deux disjoints, alors on a* $\sum_{n\in\mathbb{N}} \mathbf{1}_{A_n} = \mathbf{1}_{\bigcup_{n\in\mathbb{N}} A_n}$.
6) *Montrer que* $A \subseteq B \Longleftrightarrow \mathbf{1}_A \leq \mathbf{1}_B$.

On va maintenant introduire la notion de "limite" pour une suite $(A_n)_{n\geq 1}$ de parties de E.

Définition 1.1.11 *On dit qu'une suite $(A_n)_{n\geq 1}$ de parties de E converge (ou tend) vers la partie A, et on écrit $\lim_{n\to+\infty} A_n = A$, si pour tout $x \in A$ (resp. $x \notin A$), on a $x \in A_n$ (resp. $x \notin A_n$) pour tout n assez grand. En termes de quantificateurs, cela s'écrit :*

$$\forall x \in A,\ \exists\ n_0,\ \forall n \geq n_0,\ x \in A_n,$$
$$\forall x \notin A,\ \exists\ n_0,\ \forall n \geq n_0,\ x \notin A_n.$$

Il est facile de vérifier que cette définition revient à dire que la suite des fonctions indicatrices $(\mathbf{1}_{A_n})_{n\geq 1}$ converge simplement vers la fonction indicatrice $\mathbf{1}_A$. Autrement dit

$$\lim_{n\to+\infty} A_n = A \Longleftrightarrow \lim_{n\to+\infty} \mathbf{1}_{A_n}(x) = \mathbf{1}_A(x), \quad \text{pour tout } x \in E.$$

3. Par exemple, en fixant une numérotation de A par les entiers $(k \mapsto x_k)$, on peut choisir dans chaque classe l'élément d'indice minimal. Cela permet d'éviter l'invocation de l'axiome du choix.

Définition 1.1.12 *On dit qu'une suite $(A_n)_{n\geq 1}$ de parties de E est croissante (resp. décroissante) si $A_n \subset A_{n+1}$ (resp. $A_{n+1} \subset A_n$) pour tout n. La limite d'une suite croissante ou décroissante de parties est toujours définie. Autrement dit :*
- si $(A_n)_{n\geq 1}$ est croissante alors $\lim_{n\to+\infty} A_n = \bigcup_{n\geq 1} A_n$.
- Si $(A_n)_{n\geq 1}$ est décroissante alors $\lim_{n\to+\infty} A_n = \bigcap_{n\geq 1} A_n$.
On dit qu'une suite $(A_n)_{n\geq 1}$ est monotone si elle est croissante ou décroissante.

Exemple 1.1.13 *Considérons le cas où $E = \mathbb{R}$. On définit*

$$A_n = \left(\frac{1}{n}, 1\right] \quad et \quad B_n = \left(1 - \frac{1}{n}, 2 + \frac{1}{n}\right) \quad pour\ tout\ n \geq 2.$$

On vérifie facilement que la suite $(A_n)_{n\geq 2}$ est **croissante** *(au sens de l'inclusion). Ainsi, $\lim_{n\to+\infty} A_n = \bigcup_{n\geq 2} A_n = (0,1]$. De même, la suite $(B_n)_{n\geq 2}$ est* **décroissante**. *On a donc $\lim_{n\to+\infty} B_n = \bigcap_{n\geq 2} B_n = [1,2]$. Enfin, définissons $C_1 = [0,1]$, $C_2 = [0,2]$, $C_3 = \left[0, \frac{1}{3}\right]$, et, pour tout $n \geq 2$, $C_{2n} = [0, 2n]$, $C_{2n+1} = \left[0, \frac{1}{2n+1}\right]$. Dans ce cas, on voit que $(C_n)_{n\geq 1}$* **n'est pas monotone**.

Exercice 1.1.14 *On définit $A_n = (-\frac{1}{n}, 1]$, $B_n = [\frac{1}{n}, 1]$ et $C_n = [-n, n]$ pour tout $n \geq 1$. Vérifier que $\lim_{n\to+\infty} A_n = [0,1]$, $\lim_{n\to+\infty} B_n = (0,1]$ et $\lim_{n\to+\infty} C_n = \mathbb{R}$.*

Comme pour les suites de nombres, la limite d'une suite de parties n'est pas toujours définie. Autrement dit, il existe évidemment des suites $(A_n)_{n\geq 1}$ de parties qui ne convergent pas. Cependant, on peut toujours définir deux ensembles associés, appelés **limite supérieure** et **limite inférieure**.

Définition 1.1.15 *Soit $(A_n)_{n\geq 1}$ une suite de parties d'un ensemble E. On définit :*

$$\begin{aligned}
\limsup_n A_n = \overline{\lim} A_n &= \lim_{n\to\infty} \downarrow \left(\bigcup_{p\geq n} A_p\right) = \bigcap_{n\geq 1} \bigcup_{p\geq n} A_p = \left\{\omega \in E : \sum_{n=1}^{\infty} \mathbf{1}_{A_n}(\omega) = +\infty\right\} \\
&= \left\{\omega \in E : A_n(\omega) \text{ se produit infiniment souvent}\right\}, \\
\liminf_n A_n = \underline{\lim} A_n &= \lim_{n\to\infty} \uparrow \left(\bigcap_{p\geq n} A_p\right) = \bigcup_{n\geq 1} \bigcap_{p\geq n} A_p = \left\{\omega \in E : \sum_{n=1}^{\infty} \mathbf{1}_{A_n^c}(\omega) < +\infty\right\},
\end{aligned}$$

où $\lim\downarrow$ désigne la limite d'une suite décroissante et $\lim\uparrow$ celle d'une suite croissante.

Remarquons que la $\limsup A_n$ est l'ensemble des ω qui appartiennent à une infinité de A_n. La $\liminf A_n$ est l'ensemble des ω qui appartiennent à tous les A_n sauf un nombre fini. De plus, il est clair que

$$\liminf_n A_n \subset \limsup_n A_n,$$

et que la suite $(A_n)_{n\geq 1}$ admet une limite si et seulement si $\liminf_n A_n = \limsup_n A_n$, auquel cas

$$\lim_{n\to\infty} A_n = \liminf_n A_n = \limsup_n A_n.$$

Après avoir introduit les notions de $\limsup$ et $\liminf$ pour une suite de parties, il est naturel de se demander comment ces limites se comportent lorsque l'on distingue les termes pairs et impairs de la suite. De manière plus générale, on peut relier les $\limsup$ et $\liminf$ d'une suite arbitraire (A_n) à ceux de ses sous-suites paires et impaires.

Proposition 1.1.16 *Soit $(A_n)_{n\in\mathbb{N}}$ une suite de parties de E. Alors :*

$$(a) \quad \limsup_n A_n = (\limsup_n A_{2n}) \cup (\limsup_n A_{2n+1}),$$
$$(b) \quad \liminf_n A_n = (\liminf_n A_{2n}) \cap (\liminf_n A_{2n+1}).$$

Preuve : On sait que $x \in \limsup_n A_n$ ssi x appartient à une infinité de A_n. Autrement dit :

$$\begin{aligned} x \in \limsup_n A_n &\iff x \in A_n \text{ pour une infinité de } n \\ &\iff x \in A_{2n} \text{ pour une infinité de } n \text{ ou } x \in A_{2n+1} \text{ pour une infinité de } n \\ &\iff x \in (\limsup_n A_{2n}) \cup (\limsup_n A_{2n+1}). \end{aligned}$$

La preuve pour $\liminf$ repose sur une double inclusion, en s'appuyant sur le fait que $x \in \liminf_n A_n$ équivaut à dire que x est dans tous les A_k à partir d'un certain rang. ■

Exemple 1.1.17 *Considérons la suite $\Delta_n = [0, 2+(-1)^n]$ de parties de $\mathbb{R}$, avec $n \in \mathbb{N}$. On a $\liminf_n \Delta_n = [0,1]$ et $\limsup_n \Delta_n = [0,3]$. Par conséquent, $(\Delta_n)_{n\in\mathbb{N}}$ ne converge pas.*

Remarque 1.1.18 - *Si $(A_n)_{n\geq 1}$ est croissante ($A_n \subset A_{n+1}$), alors*

$$\limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n = \bigcup_{n\geq 1} A_n.$$

- Si $(A_n)_{n\geq 1}$ est décroissante ($A_{n+1} \subset A_n$), alors $\limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n = \bigcap_{n\geq 1} A_n$.

Exemple 1.1.19 *Déterminons les limites supérieure et inférieure des suites de parties de $\mathbb{R}$:*

$$A_n = \left[0, 1 + \tfrac{(-1)^n}{n}\right], \qquad B_n = \begin{cases} [0,n] & \text{si } n \text{ est pair}, \\ [n, +\infty[& \text{si } n \text{ est impair}. \end{cases}$$

Solution. - Pour (A_n) : on observe que (A_{2n}) est décroissante et (A_{2n+1}) croissante avec $A_{2n+1} \subset A_{2n}$. Ainsi, $\limsup_n A_n = [0,1]$, et $\liminf_n A_n = [0,1[$, d'où la suite (A_n) ne converge pas. - Pour (B_n) : on a (B_{2n}) croissante, (B_{2n+1}) décroissante. Ainsi, $\limsup_n B_n = \mathbb{R}_+$, et $\liminf_n B_n = \emptyset$. Donc (B_n) ne converge pas.

Exercice 1.1.20 1. *Déterminer les limites supérieure et inférieure de la suite $(A_n)_{n\geq 1}$ définie par :*

$$A_{2n-1} =]-2-\tfrac{1}{n}, 1], \qquad A_{2n} = [-1, 2+\tfrac{1}{n^2}[.$$

2. *Existe-t-il une suite $(B_n)_{n\geq 1}$ telle que $\limsup_n B_n = [-1,2]$ et $\liminf_n B_n = [-2,1]$?*

1.1.2 Tribu sur un ensemble

Dans cette section Ω désignera un ensemble non vide.

Définition 1.1.21 *Une classe $\mathscr{F}$ de parties de Ω est appelée une* **tribu** *(ou σ-algèbre de Boole) si :*

(i) $\Omega \in \mathscr{F}$,

(ii) $A \in \mathscr{F} \Longrightarrow A^c \in \mathscr{F}$ *("stabilité par passage au complémentaire")*

(iii) $A_1, A_2, \cdots \in \mathscr{F} \Longrightarrow \bigcup_{n\in\mathbb{N}^*} A_n \in \mathscr{F}$ *("stabilité par réunion dénombrable").*

De cette définition, on déduit immédiatement les propriétés élémentaires suivantes :

- $\emptyset \in \mathscr{F}$.
- $A_1, \ldots, A_n \in \mathscr{F} \implies A_1 \cup \cdots \cup A_n \in \mathscr{F}$ (stabilité par réunion finie).
- $A_1, \ldots, A_n \in \mathscr{F} \implies A_1 \cap \cdots \cap A_n \in \mathscr{F}$ (stabilité par intersection finie).
- $A, B \in \mathscr{F} \implies B \backslash A = B \cap A^c \in \mathscr{F}$ et $A \Delta B = (A \cup B) \backslash (A \cap B) \in \mathscr{F}$.

Exemple 1.1.22 $\star$ *La plus petite tribu sur Ω est $\mathscr{F} = \{\emptyset, \Omega\}$ (appelée* tribu grossière *ou* tribu triviale*).*

$\star$ *La plus grande tribu est l'ensemble $\mathscr{P}(\Omega)$, appelé* tribu discrète *ou* tribu complète.

$\star$ *Soit $\Omega = \{a, b, c\}$. La classe $\mathscr{F} = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$ est une tribu sur Ω.*

$\star$ *Soit $\Omega = \{1, 2, 3, 4, 5, 6\}$. On peut définir plusieurs tribus sur Ω : $\mathscr{F} = \mathscr{P}(\Omega)$, la tribu complète (maximale pour l'inclusion) ; $\mathscr{F}_0 = \{\emptyset, \Omega\}$, la tribu triviale ;*
$\mathscr{F}_1 = \{\emptyset, \{1\}, \{2, 3, 4, 5, 6\}, \Omega\}$; $\mathscr{F}_2 = \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\}$; etc.

Remarque 1.1.23 *Si $\mathscr{G} \subset \mathscr{F}$ et $A \in \mathscr{G}$, alors bien sûr $A \in \mathscr{F}$. En revanche, l'implication réciproque $A \in \mathscr{F}$ et $B \subset A \implies B \in \mathscr{F}$ est en général fausse dès que $\mathscr{F}$ n'est pas l'ensemble de toutes les parties de A.*

Exercice 1.1.24 (**Tribu trace**) *Soit $\mathscr{F}$ une tribu sur Ω et $F \subset \Omega$.*
Montrer que $\mathscr{F}_{|_F} = \{A \cap F \ : \ A \in \mathscr{F}\}$ est une tribu sur F.

Proposition 1.1.25 *Soit I un ensemble quelconque d'indices et soit $(\mathscr{F}_i)_{i\in I}$ une famille de tribu sur Ω. Alors $\bigcap_{i\in I} \mathscr{F}_i$ est une tribu sur Ω.*

Preuve : On sait que $\forall i \in I, \ \Omega \in \mathscr{F}_i \Longrightarrow \Omega \in \bigcap_{i\in I} \mathscr{F}_i$. Soit $A \in \bigcap_{i\in I} \mathscr{F}_i$. Alors $A \in \mathscr{F}_i$ pour tout $i \in I$. Et puisque $\mathscr{F}_i$ est une tribu, alors $A^c \in \mathscr{F}_i$ pour tout $i \in I$. Ce qui veut dire que $A^c \in \bigcap_{i\in I} \mathscr{F}_i$. Par raisonnement analogue, on obtient la réunion dénombrable.■

Exercice 1.1.26 *Donner un exemple de deux tribus $\mathscr{F}_1$ et $\mathscr{F}_2$ telles que $\mathscr{F}_1 \cup \mathscr{F}_2$ n'est pas une tribu.*

Des classes particulières de parties de Ω jouent un rôle fondamental car elles portent « en germe » toute la tribu.

Proposition 1.1.27 *Soit $\mathscr{A}$ une famille de parties de Ω. Il existe une plus petite tribu (au sens de l'inclusion) sur Ω contenant $\mathscr{A}$, c'est l'intersection de toutes les tribus contenant $\mathscr{A}$. On l'appelle* **tribu engendrée** *par $\mathscr{A}$ et on la note $\sigma_\Omega(\mathscr{A})$ (ou simplement $\sigma(\mathscr{A})$ quand il n'y a pas de risque de confusion). On dit aussi que $\mathscr{A}$ est un sous-ensemble générateur de la tribu $\sigma(\mathscr{A})$.*

Preuve : Soit $\mathscr{E}$ l'ensemble de toutes les tribus $\mathscr{M}$ sur E contenant $\mathscr{A}$. L'ensemble $\mathscr{E}$ n'est pas vide puisqu'il contient la tribu $\mathscr{P}(\Omega)$.

Posons $\mathscr{F} = \bigcap_{\mathscr{M}\in\mathscr{E}} \mathscr{M} = \{A \subset \Omega, \forall \mathscr{M} \in \mathscr{E},\ A \in \mathscr{M}\}$. Il est clair que $\mathscr{A} \subset \mathscr{F}$. De plus, $\mathscr{F}$ est une tribu comme intersection (quelconque) de tribus et, par définition, elle est contenue dans toute tribu contenant $\mathscr{A}$. ■

Exemple 1.1.28 - *Si $\mathscr{A} = \{\emptyset\}$ ou bien si $\mathscr{A} = \{\Omega\}$, alors $\sigma(\mathscr{A}) = \{\emptyset, \Omega\}$.*

- *Si $\mathscr{A} = \{A\}$ ou bien si $\mathscr{A} = \{A^c\}$ où $A \neq \emptyset$ et $A \neq \Omega$, alors $\sigma(\mathscr{A}) = \{\emptyset, A, A^c, \Omega\}$.*
- *Reprenons l'exemple 1.1.22. Si $\mathscr{A}_1 = \{\{1\}\}$, alors $\sigma(\mathscr{A}_1) = \mathscr{F}_1$. Si $\mathscr{A}_2 = \{\{1,3,5\}\}$, alors $\sigma(\mathscr{A}_2) = \mathscr{F}_2$. Et si $\mathscr{A} = \{\{1,3,5\},\{1,2,3\}\}$, alors $\sigma(\mathscr{A}) =$? (cf. exercice).*

Remarque 1.1.29 *Soient $\mathscr{A}_1$ et $\mathscr{A}_2 \in \mathscr{P}(\Omega)$. Alors nous avons :*

- *$\mathscr{A}_1 \subset \sigma(\mathscr{A}_1)$.*
- *Si $\mathscr{A}_1 \subset \mathscr{A}_2$ alors $\sigma(\mathscr{A}_1) \subset \sigma(\mathscr{A}_2)$.*
- *Si $\mathscr{F}$ est une tribu contenant $\mathscr{A}_1$, alors $\sigma(\mathscr{A}_1) \subset \mathscr{F}$.*
- *Si $\mathscr{F}$ est une tribu sur Ω, alors $\sigma(\mathscr{F}) = \mathscr{F}$.*

Proposition 1.1.30 *Soient E et F deux ensembles, $f : E \to F$ une application et $\mathscr{F}$ une tribu sur F. Alors $f^{-1}(\mathscr{F}) := \{f^{-1}(A) \ :\ A \in \mathscr{F}\}$ est une tribu sur E, appelée* **tribu image réciproque** *de $\mathscr{F}$ par f.*

Preuve : ($\star$) Comme $F \in \mathscr{F}$ et $f^{-1}(F) = E$, on a bien $E \in f^{-1}(\mathscr{F})$. ($\star$) Soit $A \in f^{-1}(\mathscr{F})$. Il existe $B \in \mathscr{F}$ tel que $A = f^{-1}(B)$. D'après la propriété de compatibilité des images réciproques avec le complémentaire (cf. remarque 1.1.6), on a $A^c = \left(f^{-1}(B)\right)^c = f^{-1}(B^c)$. Or $B^c \in \mathscr{F}$, donc $A^c \in f^{-1}(\mathscr{F})$. ($\star$) Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'éléments de $f^{-1}(\mathscr{F})$. Pour tout n, il existe $B_n \in \mathscr{F}$ tel que $A_n = f^{-1}(B_n)$. Comme $\bigcup_{n\in\mathbb{N}} B_n \in \mathscr{F}$ et que l'image réciproque commute à l'union dénombrable (cf. propriétés 1.1.7), on obtient :

$$\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} f^{-1}(B_n) = f^{-1}\left(\bigcup_{n\in\mathbb{N}} B_n\right) \in f^{-1}(\mathscr{F}).$$

Ainsi, $f^{-1}(\mathscr{F})$ est une tribu sur E. ■

Exercice 1.1.31 *Soient E et F deux ensembles, f une application de E dans F et $\mathscr{F}$ une tribu sur E. Montrer sur un exemple que $f(\mathscr{F}) = \{f(A) \ :\ A \in \mathscr{F}\}$ n'est pas une tribu sur F.*

Proposition 1.1.32 *Soient E et F deux ensembles, $f : E \to F$ une application et $\mathscr{F}$ une tribu sur E. Alors $\mathscr{F}_f := \{B \subset F \ :\ f^{-1}(B) \in \mathscr{F}\}$ est une tribu sur F, appelée* **tribu induite** *de $\mathscr{F}$ par f.*

Preuve : ($\star$) Comme $f^{-1}(F) = E \in \mathscr{F}$, on a $F \in \mathscr{F}_f$. ($\star$) Soit $B \in \mathscr{F}_f$. Alors $f^{-1}(B) \in \mathscr{F}$. Comme $\mathscr{F}$ est une tribu, $(f^{-1}(B))^c \in \mathscr{F}$. Or $(f^{-1}(B))^c = f^{-1}(B^c)$, donc $B^c \in \mathscr{F}_f$. ($\star$) Soit $(B_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\mathscr{F}_f$. Alors pour tout n, $f^{-1}(B_n) \in \mathscr{F}$. Comme $\mathscr{F}$ est stable par unions dénombrables, $\bigcup_{n\in\mathbb{N}} f^{-1}(B_n) \in \mathscr{F}$. Mais $\bigcup_{n\in\mathbb{N}} f^{-1}(B_n) = f^{-1}(\bigcup_{n\in\mathbb{N}} B_n)$, donc $\bigcup_{n\in\mathbb{N}} B_n \in \mathscr{F}_f$. Ainsi $\mathscr{F}_f$ est bien une tribu sur F. ■

Proposition 1.1.33 *Soient E et F deux ensembles, f une application de E dans F et $\mathscr{E}$ un sous-ensemble de $\mathscr{P}(F)$. Alors, nous avons $\sigma(f^{-1}(\mathscr{E})) = f^{-1}(\sigma(\mathscr{E}))$. Ce résultat est parfois appelé le* **lemme du transport**.

Preuve : Nous devons montrer les deux inclusions. **(i) Inclusion $\sigma(f^{-1}(\mathscr{E})) \subset f^{-1}(\sigma(\mathscr{E}))$.** Comme $\mathscr{E} \subset \sigma(\mathscr{E})$, on a aussitôt $f^{-1}(\mathscr{E}) \subset f^{-1}(\sigma(\mathscr{E}))$. De plus, d'après la proposition 1.1.30, $f^{-1}(\sigma(\mathscr{E}))$ est une tribu sur E. Elle contient donc la tribu engendrée par $f^{-1}(\mathscr{E})$, c'est-à-dire $\sigma(f^{-1}(\mathscr{E})) \subset f^{-1}(\sigma(\mathscr{E}))$.
(ii) Inclusion réciproque $f^{-1}(\sigma(\mathscr{E})) \subset \sigma(f^{-1}(\mathscr{E}))$. Considérons la tribu induite de $\sigma(f^{-1}(\mathscr{E}))$ par f : $\mathscr{F}_f := \{B \subset F : f^{-1}(B) \in \sigma(f^{-1}(\mathscr{E}))\}$. Par définition, $\mathscr{E} \subset \mathscr{F}_f$. Comme $\mathscr{F}_f$ est une tribu, elle contient la tribu engendrée par $\mathscr{E}$, donc $\sigma(\mathscr{E}) \subset \mathscr{F}_f$. En appliquant f^{-1}, on obtient $f^{-1}(\sigma(\mathscr{E})) \subset f^{-1}(\mathscr{F}_f)$. Or, par définition même de $\mathscr{F}_f$, on a $f^{-1}(\mathscr{F}_f) \subset \sigma(f^{-1}(\mathscr{E}))$. On en déduit bien $f^{-1}(\sigma(\mathscr{E})) \subset \sigma(f^{-1}(\mathscr{E}))$. En combinant (i) et (ii), on conclut que $\sigma(f^{-1}(\mathscr{E})) = f^{-1}(\sigma(\mathscr{E}))$. ■

Rappelons une classe de tribus très importante dans la théorie des probabilités : *la tribu borélienne* (également appelée *tribu de Borel* ou *tribu des boréliens*).

Définition 1.1.34 *Soit Ω un espace métrique*[4] *et $\mathscr{O}$ la famille des ouverts de Ω. On appelle* **tribu de Borel** *ou* **tribu borélienne**, *et on la note $\mathscr{B}(\Omega)$, la tribu engendrée par la famille $\mathscr{O}$. Autrement dit : $\mathscr{B}(\Omega) = \sigma(\mathscr{O})$. En d'autres termes, la tribu borélienne est la plus petite tribu sur Ω contenant tous les ouverts de Ω. On appelle* **borélien** *de Ω tout élément de $\mathscr{B}(\Omega)$.*

Proposition 1.1.35 *Soit $\mathscr{F}$ la famille des fermés de Ω. Alors la tribu borélienne $\mathscr{B}(\Omega)$ est également engendrée par $\mathscr{F}$, c'est-à-dire : $\mathscr{B}(\Omega) = \sigma(\mathscr{F})$.*

Preuve : (i) $\sigma(\mathscr{F}) \subset \mathscr{B}(\Omega)$: Soit $A \in \mathscr{F}$. Alors A^c est un ouvert, donc $A^c \in \mathscr{O} \subset \mathscr{B}(\Omega)$. Comme $\mathscr{B}(\Omega)$ est une tribu, on a aussi $A \in \mathscr{B}(\Omega)$. Ainsi $\mathscr{F} \subset \mathscr{B}(\Omega)$, ce qui entraîne $\sigma(\mathscr{F}) \subset \mathscr{B}(\Omega)$. **(ii) $\mathscr{B}(\Omega) \subset \sigma(\mathscr{F})$:** Soit $O \in \mathscr{O}$ un ouvert. Alors $O^c \in \mathscr{F} \subset \sigma(\mathscr{F})$. Comme $\sigma(\mathscr{F})$ est une tribu, on a aussi $O \in \sigma(\mathscr{F})$. D'où l'inclusion $\mathscr{O} \subset \sigma(\mathscr{F})$, et par conséquent $\mathscr{B}(\Omega) = \sigma(\mathscr{O}) \subset \sigma(\mathscr{F})$. ■

On peut maintenant décrire la tribu borélienne dans le cas particulier de $\Omega = \mathbb{R}$. Dans la topologie usuelle de $\mathbb{R}$, un ensemble O est ouvert si

$$\forall x \in O,\ \exists a < b \in \mathbb{R} \ \text{ tels que } \ x \in]a,b[\subset O.$$

De plus, tout ouvert $O \subset \mathbb{R}$ peut s'écrire comme une union dénombrable d'intervalles ouverts deux à deux disjoints, c'est-à-dire :

$$O = \bigcup_{(\rho,r)\in I}]\rho - r, \rho + r[, \quad I = \left\{(\rho,r) \in \mathbb{Q} \times \mathbb{Q}_+^* \ :]\rho - r, \rho + r[\subset O\right\}.$$

Proposition 1.1.36 *(i) La tribu de Borel $\mathscr{B}(\mathbb{R})$ est la tribu sur $\mathbb{R}$ engendrée par chacune des familles suivantes :*

(a) la famille des intervalles ouverts bornés : $\{\,]a,b[: a < b,\ a,b \in \mathbb{R}\,\}$;

(b) la famille des intervalles compacts : $\{\,[a,b] : a < b,\ a,b \in \mathbb{R}\,\}$;

4. Plus généralement, on peut travailler avec un espace topologique.

(c) la famille des intervalles semi-ouverts de la forme : $\{[a,b[: a<b,\ a,b\in\mathbb{R}\}$ *;*

(d) la famille des intervalles semi-ouverts de la forme : $\{]a,b] : a<b,\ a,b\in\mathbb{R}\}$ *;*

(e) la famille des intervalles de la forme : $\{]a,+\infty[: a\in\mathbb{R}\}$ *;*

(f) la famille des intervalles de la forme : $\{[a,+\infty[: a\in\mathbb{R}\}$ *;*

(g) la famille des intervalles de la forme : $\{]-\infty,a[: a\in\mathbb{R}\}$ *;*

(h) la famille des intervalles de la forme : $\{]-\infty,a] : a\in\mathbb{R}\}$.

(ii) *La tribu de Borel* $\mathscr{B}(\overline{\mathbb{R}})$ *est la tribu sur* $\overline{\mathbb{R}}$ *engendrée par l'une des quatre familles suivantes : (a)* $\{[a,+\infty] : a\in\mathbb{R}\}$ *; (b)* $\{]a,+\infty] : a\in\mathbb{R}\}$ *; (c)* $\{[-\infty,a[: a\in\mathbb{R}\}$ *; (d)* $\{[-\infty,a] : a\in\mathbb{R}\}$.

Preuve : Preuve du point (i) **:** $\mathscr{B}(\mathbb{R})$ (a) **Intervalles ouverts bornés** $]a,b[$ **:** Tout ouvert de $\mathbb{R}$ peut s'écrire comme une réunion dénombrable d'intervalles ouverts bornés. Ainsi, la tribu engendrée par les intervalles ouverts bornés contient tous les ouverts de $\mathbb{R}$, et donc est égale à la tribu borélienne $\mathscr{B}(\mathbb{R})$. (b) **Intervalles compacts** $[a,b]$ **:** On a $[a,b]\subset\mathbb{R}$ fermé, donc inclus dans $\mathscr{B}(\mathbb{R})$. Inversement, tout intervalle ouvert borné s'écrit comme

$$]a,b[=\bigcup_{n\geq 1}\left[a+\frac{1}{n},b-\frac{1}{n}\right]\in\sigma(\{[a,b]\}).\quad\text{Ainsi}\quad\sigma(\{[a,b]\})=\mathscr{B}(\mathbb{R}).$$

(c) **Intervalles semi-ouverts** $[a,b[$ **:** On peut écrire $]a,b[=\bigcup_{n\geq 1}[a+\frac{1}{n},b[$, donc tout ouvert est engendré par les intervalles de type $[a,b[$. (d) **Intervalles semi-ouverts** $]a,b]$ **:** De même, $]a,b]=\bigcap_{n\geq 1}]a,b+\frac{1}{n}[$, donc la tribu engendrée par ces intervalles contient tous les ouverts. (e) **Intervalles de type** $]a,+\infty[$ **:** Tout intervalle semi-ouvert $]a,b]$ s'écrit comme $]a,b]=]a,+\infty[\cap(\mathbb{R}\backslash]b,+\infty[)$, donc la tribu engendrée par $]a,+\infty[$ contient tous les intervalles semi-ouverts et donc tous les ouverts.

(f) **Les autres familles** $([a,+\infty[,]-\infty,a[,[-\infty,a])$ **:** On procède de la même manière par complément et intersection avec $\mathbb{R}$, montrant que chaque famille engendre la tribu borélienne.

Preuve du point (ii) **:** $\mathscr{B}(\overline{\mathbb{R}})$. La démonstration est analogue. On considère les intervalles englobant $\pm\infty$, et on utilise le même raisonnement par unions, intersections et compléments pour montrer que chaque famille génère la tribu borélienne de $\overline{\mathbb{R}}$. ■

Remarque 1.1.37 *Pour un ensemble fini* Ω*, on choisit souvent* $\mathscr{F}=\mathscr{P}(\Omega)$ *et pour* $\Omega\subset\mathbb{R}$ *ou* $\Omega\subset\mathbb{R}^n$*, on choisit souvent* $\mathscr{F}=\mathscr{B}(\Omega)$.

Définition 1.1.38 *Soit* $\mathscr{F}$ *une tribu sur* Ω*. Les éléments de* $\mathscr{F}$ *sont dits* $\mathscr{F}$**-mesurables** *(ou simplement mesurables lorsqu'il n'y a pas de risque de confusion). Le couple* $(\Omega,\mathscr{F})$ *est appelé un* **espace mesurable**. *Si* $\mathscr{G}\subset\mathscr{F}$ *est également une tribu, on dit que* $\mathscr{G}$ *est une* **sous-tribu** *de* $\mathscr{F}$.

1.1.3 Classes monotones

Soit Ω un ensemble non vide et $\mathscr{P}(\Omega)$ l'ensemble de ses parties. Pour montrer qu'une certaine propriété est vérifiée pour une tribu $\mathscr{F}$, il est souvent plus commode de l'établir

d'abord pour une famille $\mathscr{L}$ de parties de Ω, plus grande que $\mathscr{F}$, puis de restreindre le résultat à $\mathscr{F}$. Un outil fondamental dans cette perspective est le **lemme de la classe monotone**.

Définition 1.1.39 *Une famille $\mathscr{L} \subset \mathscr{P}(\Omega)$ est appelée* **classe monotone** *(ou encore* **λ-système***) si les conditions suivantes sont satisfaites :*

(i) $\Omega \in \mathscr{L}$;

(ii) *$\mathscr{L}$ est stable par différence propre : pour tous $A, B \in \mathscr{L}$ tels que $A \subset B$, on a $B \setminus A \in \mathscr{L}$;*

(iii) *$\mathscr{L}$ est stable par union dénombrable croissante : si $(A_n)_{n\in\mathbb{N}}$ est une suite croissante d'éléments de $\mathscr{L}$, alors $\bigcup_{n\in\mathbb{N}} A_n \in \mathscr{L}$.*

Remarque 1.1.40 *Une classe monotone n'est pas nécessairement une tribu. L'intersection de classes monotones est encore une classe monotone. Si l'on ajoute la stabilité par intersections finies (ou par unions finies), on obtient une tribu.*

Définition 1.1.41 *Une famille $\mathscr{C} \subset \mathscr{P}(\Omega)$ est un* **π-système** *si $\Omega \in \mathscr{C}$ et si elle est stable par intersection finie, c'est-à-dire : $A, B \in \mathscr{C} \;\Rightarrow\; A \cap B \in \mathscr{C}$.*

Exemple 1.1.42 *Les familles suivantes sont des π-systèmes :*

- *la classe des intervalles : $\mathscr{C}_1 = \{]-\infty, x] \;:\; x \in \mathbb{R}\}$;*
- *la classe des singletons : $\mathscr{C}_2 = \big\{\{x\} : x \in \Omega\big\} \cup \{\emptyset\}$;*
- *la classe des pavés : $\mathscr{C}_3 = \{A \times B \;:\; A, B \in \mathscr{P}(\Omega)\}$.*

Ces notions, introduites par Dynkin[5], sont également appelées **systèmes de Dynkin**. Nous allons d'abord établir un résultat intermédiaire, qui nous servira dans le résultat fondamental de cette section : le **lemme de la classe monotone**. Ce lemme, dû àSierpinski[6] et popularisé par Dynkin, permet de démontrer l'égalité entre deux mesures de probabilité de manière élégante et économique : de même que deux applications linéaires qui coïncident sur une base coïncident sur tout l'espace, deux probabilités qui coïncident sur un π-système coïncident sur la tribu engendrée par ce π-système.

Proposition 1.1.43 *Une classe monotone stable par intersection est une tribu. En particulier, tout λ-système qui est aussi un π-système est une tribu.*

Preuve : Soit $\mathscr{L}$ une classe monotone. Comme $\Omega \in \mathscr{L}$ et que $\mathscr{L}$ est stable par différence propre, elle est stable par complémentaire. Si, de plus, $\mathscr{L}$ est stable par intersection, alors par les lois de De Morgan elle est aussi stable par union :

$$A, B \in \mathscr{L} \;\Rightarrow\; A^c, B^c \in \mathscr{L} \;\Rightarrow\; A^c \cap B^c \in \mathscr{L} \;\Rightarrow\; (A^c \cap B^c)^c = A \cup B \in \mathscr{L}.$$

Une récurrence immédiate montre que $\mathscr{L}$ est stable par unions finies.

Soit alors $(A_n)_{n\geq 1} \subset \mathscr{L}$. En posant $B_n = \bigcup_{k=1}^{n} A_k$, on obtient une suite croissante $(B_n)_{n\geq 1}$ d'éléments de $\mathscr{L}$, donc $\bigcup_{n\geq 1} B_n \in \mathscr{L}$. Mais $\bigcup_{n\geq 1} B_n = \bigcup_{n\geq 1} A_n$, ce qui prouve la stabilité par unions dénombrables. Ainsi, $\mathscr{L}$ est une tribu. ■

5. Eugene B. Dynkin (1924–2014), mathématicien russe, spécialiste de théorie des probabilités et de processus stochastiques.

6. Waclaw Franciszek Sierpinski (1882-1969), mathématicien polonais, connu pour ses contributions à la théorie des ensembles, la théorie des nombres et la topologie.

Lemma 1.1.44 *(***Lemme de la classe monotone***) La plus petite classe monotone contenant un π-système $\mathscr{C}$ est exactement la tribu engendrée par $\mathscr{C}$.*

Remarque 1.1.45 *Dans certains ouvrages, ce résultat est appelé le* **théorème π-λ de Dynkin**.

Preuve : On note $\mathbb{L}$ l'ensemble des classes monotones $\mathscr{L}$ telles que $\mathscr{C} \subset \mathscr{L}$. $\mathbb{L}$ n'est pas vide car $\mathscr{P}(\Omega) \in \mathbb{L}$. L'intersection de toutes les classes monotones de $\mathbb{L}$, notée $\mathscr{L}_{\mathscr{C}}$, est bien définie, contient $\mathscr{C}$ et reste une classe monotone. C'est donc le plus petit élément de $\mathbb{L}$. D'autre part, la tribu engendrée $\sigma(\mathscr{C})$ est une classe monotone contenant $\mathscr{C}$, donc $\mathscr{L}_{\mathscr{C}} \subset \sigma(\mathscr{C})$. Pour montrer l'égalité, il suffit de prouver que $\mathscr{L}_{\mathscr{C}}$ est une tribu. D'après la proposition 1.1.43, il reste à montrer que $\mathscr{L}_{\mathscr{C}}$ est stable par intersection. Soit $A \subset \Omega$ et posons $\mathscr{D}_A = \{B \in \mathscr{L}_{\mathscr{C}} : B \cap A \in \mathscr{L}_{\mathscr{C}}\}$. On vérifie que $\mathscr{D}_A$ est une classe monotone dès que $A \in \mathscr{L}_{\mathscr{C}}$. - Si $A \in \mathscr{C}$, alors par stabilité de $\mathscr{C}$ par intersection, on a $\mathscr{C} \subset \mathscr{D}_A$, donc $\mathscr{L}_{\mathscr{C}} \subset \mathscr{D}_A$. - Ainsi, pour tout $B \in \mathscr{L}_{\mathscr{C}}$, on obtient $B \cap A \in \mathscr{L}_{\mathscr{C}}$. Finalement, pour tous $B, B' \in \mathscr{L}_{\mathscr{C}}$, on a $B \cap B' \in \mathscr{L}_{\mathscr{C}}$. Donc $\mathscr{L}_{\mathscr{C}}$ est une tribu. Par minimalité de $\sigma(\mathscr{C})$, on en déduit $\sigma(\mathscr{C}) = \mathscr{L}_{\mathscr{C}}$. ■

1.1.4 Applications mesurables et mesures

Définition 1.1.46 (**Mesurabilité**) *Soient $(\Omega_1, \mathscr{F}_1)$ et $(\Omega_2, \mathscr{F}_2)$ deux espaces mesurables et soit $f : \Omega_1 \to \Omega_2$. On dit que f est $(\mathscr{F}_1, \mathscr{F}_2)$-mesurable si, pour tout $B \in \mathscr{F}_2$, on a $f^{-1}(B) \in \mathscr{F}_1$, autrement dit si $f^{-1}(\mathscr{F}_2) \subset \mathscr{F}_1$.*

Remarque 1.1.47 *S'il n'y a pas d'ambiguïté sur les tribus considérées, on dira simplement « mesurable » au lieu de « $(\mathscr{F}_1, \mathscr{F}_2)$-mesurable ».*

Exemple 1.1.48 *Un exemple fondamental : pour $A \in \mathscr{P}(\Omega)$, la fonction caractéristique (ou indicatrice) de A, définie sur Ω par $\mathbf{1}_A(x)$ est $(\mathscr{F}, \mathscr{B}(\mathbb{R}))$-mesurable si et seulement si $A \in \mathscr{F}$, car : $\mathbf{1}_A^{-1}(\mathscr{B}(\mathbb{R})) = \{\emptyset, \Omega, A, A^c\} = \sigma(\{A\})$. En effet, pour tout $B \subset \mathbb{R}$, on a*

$$\mathbf{1}_A^{-1}(B) = \begin{cases} \emptyset & \text{si } 0 \notin B \text{ et } 1 \notin B, \\ A & \text{si } 0 \notin B \text{ et } 1 \in B, \\ A^c & \text{si } 0 \in B \text{ et } 1 \notin B, \\ \Omega & \text{si } 0 \in B \text{ et } 1 \in B. \end{cases}$$

Ainsi, $\mathbf{1}_A$ est $(\mathscr{F}, \mathscr{B}(\mathbb{R}))$-mesurable pour toute tribu $\mathscr{F}$ contenant A.

En dehors de tels exemples élémentaires, il est généralement difficile (voire impossible) de tester directement l'appartenance de $f^{-1}(B)$ à $\mathscr{F}_1$ pour tout $B \in \mathscr{F}_2$, notamment en raison de l'absence de description explicite des éléments de $\mathscr{F}_2$ lorsque cette tribu est riche (par exemple $\mathscr{B}(\mathbb{R})$). La proposition suivante montre qu'il suffit en réalité de restreindre le test aux éléments B d'une famille génératrice de $\mathscr{F}_2$.

Proposition 1.1.49 *Soient $(\Omega_1, \mathscr{F}_1)$ et $(\Omega_2, \mathscr{F}_2)$ deux espaces mesurables, et soit $\mathscr{S}$ une famille de parties de Ω_2 engendrant $\mathscr{F}_2$ (c'est-à-dire $\sigma(\mathscr{S}) = \mathscr{F}_2$). L'application $f : \Omega_1 \to \Omega_2$ est $(\mathscr{F}_1, \mathscr{F}_2)$-mesurable si, pour tout $B \in \mathscr{S}$, on a $f^{-1}(B) \in \mathscr{F}_1$, autrement dit si $f^{-1}(\mathscr{S}) \subset \mathscr{F}_1$.*

Preuve : Comme $\mathscr{F}_1$ est une tribu sur Ω_1, l'hypothèse $f^{-1}(\mathscr{S}) \subset \mathscr{F}_1$ entraîne, par minimalité, $\sigma(f^{-1}(\mathscr{S})) \subset \mathscr{F}_1$. Or, d'après la proposition 1.1.33, on a

$$\sigma(f^{-1}(\mathscr{S})) = f^{-1}(\sigma(\mathscr{S})).$$

Comme $\sigma(\mathscr{S}) = \mathscr{F}_2$, on obtient $f^{-1}(\mathscr{F}_2) \subset \mathscr{F}_1$, ce qui prouve la mesurabilité de f. ■

Exemple 1.1.50 *Soit $\Omega = \{0,1,2\}$, $\mathscr{F}_1 = \{\emptyset, \Omega\}$, $\mathscr{F}_2 = \mathscr{P}(\Omega)$, et $f : (\Omega, \mathscr{F}_1) \to (\Omega, \mathscr{F}_2)$ défini par $f(x) = x$. Alors f n'est pas mesurable, car l'image réciproque de $\{1\}$ (élément de $\mathscr{F}_2$) n'appartient pas à $\mathscr{F}_1$.*

Critère de mesurabilité des fonctions numériques

Nous examinons maintenant l'effet des opérations usuelles sur la mesurabilité des fonctions. Soit $f : \Omega_1 \to \Omega_2$. Dans tout ce qui suit, si $\Omega_2 = \mathbb{R}$ (resp. $\overline{\mathbb{R}}$), on notera, pour $a, b \in \mathbb{R} \cup \{+\infty\}$:

$$\{f < a\} = f^{-1}(]-\infty, a[) \quad (\text{resp. } f^{-1}([-\infty, a[)),$$

et de même pour $\{f \leq a\}, \{f > a\}, \{f \geq a\}, \{a < f < b\}, \{a \leq f < b\}, \{a < f \leq b\}$ et $\{a \leq f \leq b\}$.
Ainsi, si $\Omega_2 = \mathbb{R}$ ou $\overline{\mathbb{R}}$ muni de sa tribu de Borel, on déduit des propositions 1.1.36 et 1.1.49 un critère simple de mesurabilité.

Théorème 1.1.1 *(i) Une fonction $f : \Omega_1 \to \mathbb{R}$ est $(\mathscr{F}_1, \mathscr{B}(\mathbb{R}))$-mesurable si et seulement si l'une des huit conditions équivalentes suivantes est vérifiée :*

(a) pour tout $a \in \mathbb{R}$, $\{f < a\} \in \mathscr{F}_1$;
(b) pour tout $a \in \mathbb{R}$, $\{f \leq a\} \in \mathscr{F}_1$;
(c) pour tout $a \in \mathbb{R}$, $\{f > a\} \in \mathscr{F}_1$;
(d) pour tout $a \in \mathbb{R}$, $\{f \geq a\} \in \mathscr{F}_1$;
(e) pour tout $(a,b) \in \mathbb{R}^2$, $a \leq b$, $\{a \leq f < b\} \in \mathscr{F}_1$;
(f) pour tout $(a,b) \in \mathbb{R}^2$, $a \leq b$, $\{a \leq f \leq b\} \in \mathscr{F}_1$;
(g) pour tout $(a,b) \in \mathbb{R}^2$, $a \leq b$, $\{a < f \leq b\} \in \mathscr{F}_1$;
(h) pour tout $(a,b) \in \mathbb{R}^2$, $a \leq b$, $\{a < f < b\} \in \mathscr{F}_1$.

(ii) Une fonction $f : \Omega_1 \to \overline{\mathbb{R}}$ est $(\mathscr{F}_1, \mathscr{B}(\overline{\mathbb{R}}))$-mesurable si et seulement si l'une des quatre conditions équivalentes suivantes est vérifiée : (a) pour tout $a \in \mathbb{R}$, $\{f < a\} \in \mathscr{F}_1$; (b) pour tout $a \in \mathbb{R}$, $\{f \leq a\} \in \mathscr{F}_1$; (c) pour tout $a \in \mathbb{R}$, $\{f > a\} \in \mathscr{F}_1$; (d) pour tout $a \in \mathbb{R}$, $\{f \geq a\} \in \mathscr{F}_1$.

Définition 1.1.51 (Mesure) *Soit $(\Omega, \mathscr{F})$ un espace mesurable. On appelle* mesure *sur $(\Omega, \mathscr{F})$ une application $\mu : \mathscr{F} \to [0, +\infty]$ vérifiant :*

1. $\mu(\emptyset) = 0$;
2. *Pour toute suite $(A_n)_{n \geq 0}$ d'éléments de $\mathscr{F}$ deux à deux disjoints,*

$$\mu\left(\bigcup_{n \geq 0} A_n\right) = \sum_{n \geq 0} \mu(A_n).$$

Cette propriété est appelée σ-additivité *de μ et μ est dite* σ-additive. *De plus, on dit que le triplet $(\Omega, \mathscr{F}, \mu)$ est un* espace mesuré.

Proposition 1.1.52 (Continuité monotone des mesures) *Soit $(\Omega, \mathscr{F}, \mu)$ un espace mesuré.*

*(i) (***Continuité croissante***) Si $(A_n)_{n\geq 1}$ est une suite croissante d'éléments de $\mathscr{F}$, c'est-à-dire $A_n \subseteq A_{n+1}$, alors l'ensemble limite $A := \bigcup_{n=1}^{\infty} A_n$ appartient à $\mathscr{F}$ et*

$$\mu(A_n) \xrightarrow[n\to\infty]{} \mu(A).$$

*(ii) (***Continuité décroissante***) Si $(A_n)_{n\geq 1}$ est une suite décroissante d'éléments de $\mathscr{F}$, c'est-à-dire $A_n \supseteq A_{n+1}$, et si $\mu(A_1) < +\infty$, alors l'ensemble limite $A := \bigcap_{n=1}^{\infty} A_n$ appartient à $\mathscr{F}$ et*

$$\mu(A_n) \xrightarrow[n\to\infty]{} \mu(A).$$

Preuve : (i) *Continuité croissante.* Puisque la suite (A_n) est croissante, on peut écrire $A_1 = B_1,\ B_n := A_n \setminus A_{n-1}\ (n \geq 2)$. Les ensembles $(B_n)_{n\geq 1}$ sont deux à deux disjoints et satisfont $A = \bigcup_{n=1}^{\infty} A_n = \bigsqcup_{n=1}^{\infty} B_n$. La σ-additivité de μ donne alors

$$\mu(A) = \sum_{n=1}^{\infty} \mu(B_n), \quad \mu(A_k) = \sum_{n=1}^{k} \mu(B_n) \quad (k \geq 1).$$

Ainsi $(\mu(A_k))_{k\geq 1}$ est une suite croissante bornée par $\mu(A)$, et ses sommes partielles convergent vers la série ci-dessus. On obtient donc $\mu(A_k) \xrightarrow[k\to\infty]{} \mu(A)$. (ii) *Continuité décroissante.* Supposons maintenant $A_n \supseteq A_{n+1}$ et $\mu(A_1) < +\infty$, et posons $A := \bigcap_{n=1}^{\infty} A_n$. Considérons les ensembles $C_n := A_1 \setminus A_n$, $n \geq 1$. La suite (C_n) est croissante, et l'on a

$$A_1 \setminus A = \bigcup_{n=1}^{\infty} C_n.$$

En appliquant la continuité croissante à (C_n), on obtient $\mu(C_n) \xrightarrow[n\to\infty]{} \mu(A_1 \setminus A)$. Or, pour tout n, $\mu(C_n) = \mu(A_1) - \mu(A_n)$, expression bien définie puisque $\mu(A_1) < +\infty$. En passant à la limite, on déduit

$$\mu(A_1) - \mu(A_n) \xrightarrow[n\to\infty]{} \mu(A_1) - \mu(A),$$

ce qui équivaut à $\mu(A_n) \xrightarrow[n\to\infty]{} \mu(A)$. ■

Quelques propriétés supplémentaires donnent lieu à la définition suivante.

Définition 1.1.53 ⋆ *On appelle* **poids** *de μ la quantité $\mu(\Omega)$.*

- ⋆ *Une mesure μ est dite* **finie** *ou* **bornée** *si $\mu(\Omega) < +\infty$.*
- ⋆ *Une mesure μ est dite* **σ-finie** *s'il existe une suite $(A_n)_{n\geq 0}$ d'éléments de $\mathscr{F}$ telle que $\Omega = \bigcup_{n\geq 0} A_n$ et, pour tout $n \in \mathbb{N}$, $\mu(A_n) < +\infty$.*
- ⋆ *Une mesure μ sur $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ est dite* **borélienne** *si, pour tout borélien borné $B \subset \mathbb{R}^n$, on a $\mu(B) < +\infty$.*
- ⋆ *Une mesure μ est dite* **diffuse** *si, pour tout $a \in \Omega$, $\{a\} \in \mathscr{F}$ et $\mu(\{a\}) = 0$.*

Remarque 1.1.54 *Si $\mu(\Omega) = 1$, alors μ est appelée* **probabilité** *ou* **mesure de probabilité**, *et le couple $(\Omega, \mathscr{F})$ est appelé* **espace probabilisable**.

Exemple 1.1.55 • $\rightsquigarrow$ **Mesure de Dirac (ou masse de Dirac) :** *Soit $a \in \Omega$. La masse de Dirac en a est la mesure δ_a sur $(\Omega, \mathscr{P}(\Omega))$ définie par*

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

Par restriction, δ_a est aussi une mesure sur tout espace mesurable $(\Omega, \mathscr{F})$. Vérifions que δ_a est σ-additive. Soit $(A_n)_{n\geq 0}$ une suite d'éléments de $\mathscr{P}(\Omega)$ deux à deux disjoints :

Cas 1 : *$a \notin \bigcup_{n\in\mathbb{N}} A_n$. Alors $\delta_a(\bigcup A_n) = 0$. De plus, $a \notin A_n$ pour tout n, donc $\delta_a(A_n) = 0$ et $\sum_{n\in\mathbb{N}} \delta_a(A_n) = 0$.*

Cas 2 : *$a \in \bigcup_{n\in\mathbb{N}} A_n$. Alors $\delta_a(\bigcup A_n) = 1$. Comme les A_n sont disjoints, a appartient à un seul A_{n_0}, donc $\delta_a(A_{n_0}) = 1$ et $\delta_a(A_n) = 0$ pour $n \neq n_0$, d'où $\sum_{n\in\mathbb{N}} \delta_a(A_n) = 1$.*

Ainsi, dans les deux cas : $\delta_a(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} \delta_a(A_n)$, donc δ_a est bien une mesure.

Remarque 1.1.56 *δ_a indique si un ensemble contient le point a. Par exemple, nous avons : $\delta_0([-1,1]) = 1$, $\delta_0(]0,1]) = \delta_0(\mathbb{R}^*) = 0$, $\delta_2(\mathbb{N}) = 1$, $\delta_\pi(\mathbb{Q}) = 0$. Plus généralement, pour tout $A \in \mathscr{F}$, $\delta_a(A) = \mathbf{1}_A(a)$.*

• $\rightsquigarrow$ **Séries de mesures finies :** *Si $(\mu_k)_{k\in\mathbb{N}}$ est une famille de mesures finies sur $(\Omega, \mathscr{F})$ et $(\alpha_k)_{k\in\mathbb{N}}$ une suite de réels positifs, alors $\mu(A) = \sum_{k\in\mathbb{N}} \alpha_k \mu_k(A)$, $A \in \mathscr{F}$, définit une mesure sur $(\Omega, \mathscr{F})$. La vérification se fait directement en utilisant la σ-additivité de chaque μ_k et la sommation par paquets dans $\mathbb{R}_+$.*

Exercice 1.1.57 1. **Mesure de dénombrement (ou de comptage) :** *Soit D un ensemble dénombrable (par exemple $\mathbb{N}$, $\mathbb{Z}$, ou $\mathbb{Q}$). On définit $\mu(A) = \mathrm{Card}(A)$, $A \subset D$. Montrer que μ est une mesure sur $(D, \mathscr{P}(D))$.*

2. **Mesure image :** *Soit $(\Omega, \mathscr{F}, \mu)$ un espace mesuré, $(E, \mathscr{E})$ un espace mesurable et $f : \Omega \to E$ une application $(\mathscr{F}, \mathscr{E})$-mesurable. On définit $\mu^f(B) = \mu\big(f^{-1}(B)\big)$, $B \in \mathscr{E}$. Montrer que μ^f est une mesure sur $(E, \mathscr{E})$. (On appelle μ^f la* mesure image *de μ par f, parfois notée $\mu \circ f^{-1}$).*

On admet qu'il existe une unique mesure sur $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$, notée $\lambda^{(d)}$ et appelée *mesure de Lebesgue*, telle que pour tout pavé $]a_1, b_1] \times \cdots \times]a_d, b_d]$, $a_i < b_i$, on ait

$$\lambda^{(d)}\Big(]a_1, b_1] \times \cdots \times]a_d, b_d]\Big) = \prod_{i=1}^{d} (b_i - a_i).$$

La mesure de Lebesgue généralise la longueur $(d = 1)$, l'aire $(d = 2)$ et le volume $(d = 3)$. On note $\lambda := \lambda^{(1)}$ dans le cas $d = 1$. On a $\lambda^{(d)}(\mathbb{R}^d) = +\infty$, ce qui en fait une mesure non finie, contrairement aux probabilités qui sont des cas particuliers de mesures finies.

Exercice 1.1.58 *Vérifier que*

$$\mathbb{R} = \bigcup_{k\in\mathbb{Z}}]k, k+1]$$

et en déduire que $\lambda(\mathbb{R}) = +\infty$ en appliquant la σ-additivité de la mesure de Lebesgue.

1.1.5 Compléments : les théorèmes de convergence

L'objectif de ce livre n'est pas d'exposer en détail la théorie générale de l'intégration, mais d'en présenter les éléments nécessaires au calcul des probabilités. Nous nous appuierons pour cela sur les ouvrages de Marc Briance et Gilles Pagès [8], d'El Haj Lamri [31] ainsi que de Thierry Gallouët et Raphaèle Herbin [21], auxquels nous renvoyons le lecteur pour les preuves et des développements plus complets.

Dans ce paragraphe, nous introduisons les grands théorèmes d'interversion entre limite et intégrale, qui font toute la puissance de la théorie de l'intégration au sens de Lebesgue. Pour cela, nous rappelons d'abord la construction de l'intégrale de Lebesgue sur un espace mesuré quelconque $(\Omega, \mathscr{F}, \mu)$.

Dans toute la suite, nous adopterons la convention que 0 multiplié par une quantité infinie vaut 0, c'est-à-dire $0 \times \mu(A) = 0$ si $\mu(A) = +\infty$, $0 \times f(x) = 0$ si $f(x) = \pm\infty$, et plus généralement $0 \times a = 0$ pour $a = \pm\infty$.

Nous commençons par définir l'intégrale pour des fonctions très simples, de type indicatrices.

Définition 1.1.59 *Soit $A \in \mathscr{F}$. La fonction indicatrice de A, définie par $f(\omega) = \mathbf{1}_A(\omega)$, est mesurable. Son intégrale par rapport à μ est donnée par*

$$\int_\Omega \mathbf{1}_A \, d\mu := \mu(A).$$

En particulier : - si $\mathscr{F} = \mathscr{B}(\mathbb{R})$, en prenant μ la mesure de Lebesgue et $A =]a,b]$, on obtient

$$\int_\Omega \mathbf{1}_A \, d\mu = b - a,$$

c'est-à-dire la longueur de l'intervalle. Cette définition coïncide alors avec l'intégrale de Riemann, puisque

$$\int_{\mathbb{R}} \mathbf{1}_A(x)\, dx = \int_a^b 1 \, dx = b - a;$$

- si $\mathscr{F} = \mathscr{P}(\mathbb{N})$, en prenant μ la mesure de comptage sur $\mathbb{N}$, et $A \subset \mathbb{N}$, alors

$$\int_\Omega \mathbf{1}_A \, d\mu = \mathrm{Card}(A).$$

Ainsi, la définition de l'intégrale de Lebesgue permet de traiter de manière unifiée des exemples aussi différents que la mesure de Lebesgue et la mesure de comptage. Nous savons maintenant intégrer les fonctions indicatrices d'ensembles mesurables, ce qui est encore très limité ! Pour obtenir une théorie plus riche, nous souhaitons que l'intégrale soit *linéaire*, ce qui conduit naturellement à l'étendre aux fonctions étagées.

Définition 1.1.60 *Soit f une fonction étagée positive*[7]*, de décomposition canonique $f = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}$, avec les A_i mesurables disjoints. On appelle* intégrale de f par rapport à la mesure μ *la quantité*

$$\int_\Omega f\,d\mu := \sum_{i=1}^{n} a_i \int_\Omega \mathbf{1}_{A_i}\,d\mu = \sum_{i=1}^{n} a_i \mu(A_i).$$

Grâce à cette définition, on peut montrer que l'intégrale est un opérateur *additif, croissant et positivement homogène* sur l'ensemble des fonctions mesurables, étagées et positives ou nulles (voir [8], propriétés 7.1). Le résultat suivant permet d'étendre la définition aux fonctions mesurables à valeurs dans $\overline{\mathbb{R}}_+$.

Proposition 1.1.61 *Soit f une fonction mesurable de $(\Omega, \mathscr{F})$ dans $(\mathbb{K}, \mathscr{B}(\mathbb{K}))$ avec $\mathbb{K} = \mathbb{R}, \overline{\mathbb{R}}$ ou $\mathbb{C}$. Il existe une suite de fonctions étagées $(f_n)_{n\geq 1}$ telle que*

$$f(\omega) = \lim_{n\to+\infty} f_n(\omega), \quad \textit{pour tout } \omega \in \Omega.$$

Si $f \geq 0$, la suite $(f_n)_{n\geq 1}$ peut être choisie croissante et positive, c'est à dire $0 \leq f_n \leq f_{n+1}$ pour tout $n \in \mathbb{N}^*$.

Preuve : Voir [8], théorème 5.1. ■

Nous pouvons maintenant étendre la définition de l'intégrale aux fonctions positives. On note $\mathbf{M}^+$ l'ensemble des fonctions : $f : \Omega \longrightarrow \overline{\mathbb{R}}_+, \quad (\mathscr{F}, \mathscr{B}(\overline{\mathbb{R}}_+))$-mesurables.

Définition 1.1.62 *Soit $f \in \mathbf{M}^+$. On définit son intégrale par rapport à μ comme la quantité*

$$\int_\Omega f\,d\mu := \sup\left\{ \int_\Omega g\,d\mu \ : \ g \textit{ étagée, positive, et } g(\omega) \leq f(\omega),\ \forall \omega \in \Omega \right\}.$$

L'intégrale $\int_\Omega f\,d\mu$ prend ses valeurs dans $\overline{\mathbb{R}}_+$, et nous dirons que f est μ-intégrable sur Ω *(ou simplement* intégrable par rapport à μ *quand il n'y a pas de risque de confusion) si $\int_\Omega f\,d\mu < \infty$.*

Exemple 1.1.63 *La fonction $f = \mathbf{1}_\mathbb{R}$ est δ_0-intégrable sur $\mathbb{R}$ mais n'est pas λ-intégrable sur $\mathbb{R}$. En effet,*

$$\int_\mathbb{R} f\,d\lambda = \lambda(\mathbb{R}) = +\infty \quad \textit{et} \quad \int_\mathbb{R} f\,d\delta_0 = \delta_0(\mathbb{R}) = 1.$$

7. Une fonction mesurable $f : (\Omega, \mathscr{F}) \to (\mathbb{K}, \mathscr{B}(\mathbb{K}))$ avec $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$ est dite *étagée* si son ensemble de valeurs $f(\Omega)$ est fini. En notant $f(\Omega) = \{a_1, a_2, \ldots, a_n\}$ (valeurs toutes distinctes) et $A_i = f^{-1}(\{a_i\}) = \{f = a_i\}$, on a la décomposition canonique

$$f = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i},$$

où $(A_i)_{1\leq i\leq n}$ forme une partition finie de Ω. Autrement dit, f est étagée si f est une combinaison linéaire finie de fonctions caractéristiques mesurables.

Le développement de la théorie de l'intégration sur $\mathbf{M}^+$ repose sur un théorème fondamental de convergence des intégrales, propre aux suites croissantes d'éléments de $\mathbf{M}^+$. Ce théorème, connu sous le nom de *théorème de convergence monotone*, permet, sous des hypothèses minimales, de commuter « intégration » et « passage à la limite ». Ses nombreuses applications justifient pleinement l'intérêt de la théorie de l'intégrale de Lebesgue.

Théorème 1.1.2 Théorème de convergence monotone ou de Beppo Levi
Soit $(f_n)_{n\in\mathbb{N}}$ une suite croissante d'éléments de $\mathbf{M}^+$ (au sens où $0 \le f_n \le f_{n+1}$). Alors

$$f := \sup_{n\in\mathbb{N}} f_n = \lim_{n\to+\infty} f_n \in \mathbf{M}^+ \quad \text{et} \quad \int_\Omega f\,d\mu = \lim_{n\to+\infty}\int_\Omega f_n\,d\mu = \sup_{n\in\mathbb{N}}\int_\Omega f_n\,d\mu.$$

Preuve : Voir [8], théorème 7.1 ou bien [4], théorème II.2.1. ■

Corollaire 1.1.64 *Soit $(f_n)_{n\in\mathbb{N}}$ une suite croissante de $\mathbf{M}^+$ et soit $f = \sum_{n\in\mathbb{N}} f_n$. Alors*

$$\int_\Omega f\,d\mu = \sum_{n\in\mathbb{N}}\int_\Omega f_n\,d\mu.$$

Preuve : La suite $g_n = \sum_{m=0}^n f_m$ est croissante et converge simplement vers f. Le théorème 1.1.2 implique $\lim_{n\to+\infty}\int_\Omega g_n\,d\mu = \int_\Omega \lim_{n\to+\infty} g_n\,d\mu = \int_\Omega f\,d\mu$, ce qui donne le résultat. ■

En l'absence d'hypothèse de croissance, on dispose tout de même du *lemme de Fatou*, outil fondamental permettant entre autres de vérifier si la limite simple d'une suite de fonctions μ-intégrables reste μ-intégrable.

Théorème 1.1.3 (Lemme de Fatou) *Soit $(f_n)_{n\in\mathbb{N}}$ une suite quelconque d'éléments de $\mathbf{M}^+$. Alors*

$$0 \le \int_\Omega \liminf_{n\to+\infty} f_n\,d\mu \le \liminf_{n\to+\infty}\int_\Omega f_n\,d\mu \le \infty.$$

Preuve : Voir [8], théorème 8.2. ■

Exemple 1.1.65 (**Illustration du lemme de Fatou**) *Prenons $\Omega = \mathbb{R}$, $\mathscr{F} = \mathscr{B}(\mathbb{R})$ et $\mu = \lambda$ (mesure de Lebesgue). Considérons la suite de fonctions étagées*

$$f_n = \begin{cases} \mathbf{1}_{[0,1]} & \text{si } n \text{ est pair,} \\ \mathbf{1}_{]1,2]} & \text{si } n \text{ est impair.} \end{cases}$$

On a clairement $\liminf f_n = 0$ sur $\mathbb{R}$, par conséquent $\int_{\mathbb{R}} \liminf_{n\to+\infty} f_n\,d\lambda = 0$. Par contre, $\int_{\mathbb{R}} f_n\,d\lambda = 1$ pour tout n, donc $\liminf_{n\to+\infty}\int_{\mathbb{R}} f_n\,d\lambda = 1$. Ainsi, dans cet exemple,

$$0 = \int_{\mathbb{R}} \liminf_{n\to+\infty} f_n\,d\lambda < \liminf_{n\to+\infty}\int_{\mathbb{R}} f_n\,d\lambda = 1.$$

Nous étendons maintenant l'intégrale des fonctions positives aux fonctions de signe quelconque. Pour cela, si f est une fonction mesurable, on définit

$$f^+ = f \vee 0, \qquad f^- = -(f \wedge 0),$$

de sorte que $f = f^+ - f^-$ et $|f| = f^+ + f^-$. Les fonctions f^+ et f^- sont boréliennes si f l'est. On note $\mathbf{M}$ l'ensemble des fonctions mesurables $f : \Omega \to \overline{\mathbb{R}}$, $(\mathscr{F}, \mathscr{B}(\overline{\mathbb{R}}))$-mesurables. Notons que si $f \in \mathbf{M}$, alors $f^+,\ f^- \in \mathbf{M}^+$.

Définition 1.1.66 *Soit $f \in \mathbf{M}$. On dit que f est* μ-intégrable *si $\int_\Omega |f| d\mu < +\infty$. Dans ce cas, on appelle* intégrale de f par rapport à μ, *et on note $\int_\Omega f d\mu$, le réel défini par*

$$\int_\Omega f d\mu := \int_\Omega f^+ d\mu - \int_\Omega f^- d\mu.$$

Remarque 1.1.67 *Si f est μ-intégrable et $B \in \mathscr{F}$, alors*

$$\int_B f d\mu = \int_\Omega \mathbf{1}_B f d\mu.$$

De plus, grâce au théorème de convergence monotone, les propriétés essentielles de l'intégrale sur les fonctions positives (linéarité, monotonie) s'étendent à l'intégrale des fonctions de signe quelconque.

Nous présentons à présent un théorème fondamental qui permet de transférer une intégration définie sur Ω vers un autre espace mesuré plus simple. C'est le *théorème de transfert*, dont l'usage est omniprésent dans la suite. Autrement dit, il s'appuie sur la notion de *mesure image*. Rappelons que si f est une application mesurable de $(\Omega, \mathscr{F}, \mu)$ vers $(E, \mathscr{E})$, la mesure image μ^f est définie par : $\mu^f(B) = \mu(f^{-1}(B))$, $B \in \mathscr{E}$.

Theorem 1.1.68 (Théorème de transfert) *Soit $f : (\Omega, \mathscr{F}, \mu) \to (E, \mathscr{E})$ mesurable et soit $\phi : E \to \mathbb{R}$ borélienne.*

- *Si $\phi \geq 0$, alors*
$$\int_E \phi d\mu^f = \int_\Omega \phi \circ f d\mu.$$
- *Dans le cas général, ϕ est μ^f-intégrable si et seulement si $\phi \circ f$ est μ-intégrable, et dans ce cas,*
$$\int_E \phi d\mu^f = \int_\Omega \phi \circ f d\mu.$$

Preuve : Si $\phi = \mathbf{1}_B$ avec $B \in \mathscr{E}$, alors

$$\int_E \mathbf{1}_B d\mu^f = \mu^f(B) = \mu(f^{-1}(B)) = \int_\Omega \mathbf{1}_B \circ f d\mu,$$

donc la formule est vraie. Par linéarité, elle reste valable pour les fonctions étagées. Pour une fonction mesurable positive ϕ, on l'approche par une suite croissante de fonctions étagées ϕ_n. Alors $(\phi_n \circ f)$ est croissante vers $\phi \circ f$, et le théorème de Beppo-Levi (cf. théorème 1.1.2) donne

$$\int_E \phi d\mu^f = \lim_{n\to\infty} \int_E \phi_n d\mu^f = \lim_{n\to\infty} \int_\Omega \phi_n \circ f d\mu = \int_\Omega \phi \circ f d\mu.$$

Dans le cas général, on remarque que

$$\int_E |\phi| d\mu^f = \int_\Omega |\phi \circ f| d\mu.$$

Ainsi, ϕ est μ^f-intégrable si et seulement si $\phi \circ f$ est μ-intégrable, et la formule résulte de la décomposition $\phi = \phi^+ - \phi^-$. ■

Nous en venons maintenant à l'énoncé central de la théorie de l'intégration : le *théorème de convergence dominée de Lebesgue*, qui garantit le passage à la limite sous le signe intégral.

Theorem 1.1.69 Théorème de convergence dominée de Lebesgue
Soit $(f_n)_{n\geq 0}$ une suite de fonctions mesurables de $(\Omega,\mathscr{F},\mu)$ dans $\mathbb{R}$. Supposons que :

1. *Il existe une fonction f mesurable telle que $f_n \to f$ μ-presque partout*[8].
2. *Il existe une fonction $g \in L^1(\mu)$, $g \geq 0$, telle que $|f_n| \leq g$ μ-presque partout pour tout n.*

Alors $f \in L^1(\mu)$ et $\displaystyle\int_\Omega f\,d\mu = \lim_{n\to\infty}\int_\Omega f_n\,d\mu, \qquad \lim_{n\to\infty}\int_\Omega |f_n - f|\,d\mu = 0.$

Preuve : Voir [8], théorème 8.3. ■

Un corollaire classique de ce théorème est l'interversion entre « somme » et « intégrale ».

Theorem 1.1.70 *Soit $(f_n)_{n\geq 0}$ une suite de fonctions mesurables de $(\Omega,\mathscr{F},\mu)$ dans $\mathbb{R}$.*

1. *Si $f_n \geq 0$ pour tout n, alors* $\displaystyle\int_\Omega \Big(\sum_{n\geq 0} f_n\Big)\,d\mu = \sum_{n\geq 0}\int_\Omega f_n\,d\mu.$
2. *Si* $\displaystyle\sum_{n\geq 0}\int_\Omega |f_n|\,d\mu < +\infty$, *alors* $\displaystyle\sum_{n\geq 0} f_n$ *est définie μ-presque partout, elle est μ-intégrable, et*
$$\int_\Omega \Big(\sum_{n\geq 0} f_n\Big)\,d\mu = \sum_{n\geq 0}\int_\Omega f_n\,d\mu.$$

Preuve : Voir [8], théorème 8.4. ■

Remarque 1.1.71 *Les énoncés ci-dessus sont formulés pour des fonctions à valeurs réelles. Ils restent néanmoins valables pour des fonctions à valeurs complexes. En effet, si (f_n) est une suite de fonctions complexes et si $|f_n| \leq g$ avec $g \in L^1(\mu)$, on écrit*

$$f_n = u_n + iv_n \text{ et } f = u + iv$$

avec u_n, v_n, u, v réelles. Alors $|u_n| \leq |f_n| \leq g$ et $|v_n| \leq |f_n| \leq g$, donc on applique les théorèmes précédents aux suites réelles (u_n) et (v_n). Alternativement, on remarque que $\left|\int (f_n - f)\right| \leq \int |f_n - f|$ et l'on conclut en appliquant la version réelle au réel non négatif $|f_n - f|$.

1.2 Structures de base



Comme tout domaine scientifique, la théorie des probabilités repose sur un vocabulaire précis. Des mots du langage courant y prennent un sens rigoureux et parfois différent.

8. Soit $(f,g) \in \mathbf{M}^+ \times \mathbf{M}^+$. On dit qu'elles sont égales μ-presque partout lorsque $\{f \neq g\}$ est négligeable. C'est à dire $f = g$ presque partout $\iff \mu(\{x \in E : f(x) \neq g(x)\}) = 0$. On note $f = g$ μ *p.p.*

Le calcul des probabilités est la branche des mathématiques qui modélise les phénomènes aléatoires. *Modéliser*, c'est simplifier la réalité pour rendre possible le calcul et la prédiction.

La formulation moderne remonte à 1933, lorsque Kolmogorov en établit les fondements axiomatiques : un *événement* est un sous-ensemble d'un *univers* et une probabilité est une fonction qui associe à chaque événement un nombre entre 0 et 1, selon des règles précises. [9]

En résumé, les probabilités offrent un langage abstrait pour quantifier et analyser l'incertain. Leur origine est intimement liée aux jeux de hasard, qui restent aujourd'hui un outil pédagogique central :

- point de départ historique de la discipline ;
- modèles simples et reproductibles, mais assez riches pour éclairer des situations générales.

Deux approches classiques motivent la notion de probabilité :

- l'approche *a priori*, fondée sur la symétrie et l'équité des situations (ex. lancer d'une pièce équilibrée) ;
- l'approche *fréquentiste*, qui évalue la probabilité comme la fréquence limite d'un événement lors de répétitions nombreuses.

Passons maintenant aux notions de base nécessaires pour formaliser rigoureusement ces idées.

Expérience aléatoire : c'est une expérience dont l'issue dépend du hasard. L'ensemble de toutes les issues possibles est l'*univers* Ω.

Exemple 1.2.1 *Lancer d'un dé :* $\Omega = \{1,2,3,4,5,6\}$. *Si l'on ne regarde que la parité :* $\Omega = \{\textit{pair}, \textit{impair}\}$. *Tirage d'une carte parmi* 32 *: selon l'observation,*

$$\Omega = \{7,8,9,10,V,D,R,As\}, \quad \Omega = \{\textit{trèfle}, \textit{carreau}, \textit{cœur}, \textit{pique}\}, \quad \Omega = \{\textit{rouge}, \textit{noir}\}.$$

L'univers peut être fini ou infini :

Exemple 1.2.2 - *On répète le lancer d'une pièce jusqu'à obtenir pile :* $\Omega = \mathbb{N}^*$*, un ensemble infini dénombrable.*

- *On mesure la durée de vie d'un atome :* $\Omega = \mathbb{R}_+$*, infini non dénombrable.*

Événement : un *événement* est un sous-ensemble de l'univers Ω. Parmi eux, on distingue :

- l'**événement impossible**, noté $\emptyset$;
- l'**événement certain**, qui coïncide avec tout l'univers Ω.

Comme ce sont des ensembles, les événements se combinent par les opérations usuelles :

$$A \cup B \text{ (réunion)}, \quad A \cap B \text{ (intersection)}, \quad A^c = \Omega \setminus A \text{ (complémentaire)}.$$

On dit que deux événements sont **incompatibles** lorsqu'ils ne peuvent pas se produire simultanément, c'est-à-dire lorsque $A \cap B = \emptyset$.

9. Historiquement, l'incertitude était attribuée au hasard ou au destin ; la théorie en propose une formalisation mathématique.

Exemple 1.2.3 *Au lancer d'un dé :* $A = \{2,4,6\}$ *(pair) et* $B = \{1,3,5\}$ *(impair) sont incompatibles.*

Ces notions constituent le socle du langage probabiliste. Nous pouvons maintenant introduire la notion centrale de *probabilité*, qui permet d'attribuer un poids numérique à chaque événement.

1.3 La notion de probabilité

Le terme *probabilité*, issu du latin *probabilitas*, désigne initialement l'opposé de la certitude. Il représente également le degré de plausibilité d'un événement par un nombre compris entre 0 et 1. La modélisation mathématique du calcul des probabilités repose sur un triplet $(\Omega, \mathscr{F}, \mathbb{P})$ que nous allons décrire.

1.3.1 Espace de probabilité

Un *espace de probabilité* est un triplet $(\Omega, \mathscr{F}, \mathbb{P})$ où :

- Ω est un ensemble appelé *univers*,
- $\mathscr{F}$ est une *tribu* (ou σ-algèbre) sur Ω, appelée la famille des événements observables,
- $\mathbb{P}$ est une *mesure de probabilité* sur $(\Omega, \mathscr{F})$.

Définition 1.3.1 *Une* probabilité *sur* $(\Omega, \mathscr{F})$ *est une mesure* $\mathbb{P} : \mathscr{F} \to [0,1]$ *telle que :*
1) $\mathbb{P}(\Omega) = 1$ *;*
2) *Pour toute famille* $(A_n)_{n\geq 1} \subset \mathscr{F}$ *d'événements deux à deux disjoints,*

$$\mathbb{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Cette propriété est appelée σ-additivité.

Ainsi, une probabilité est une mesure positive de masse finie égale à 1. Plusieurs propriétés fondamentales en découlent.

Proposition 1.3.2 *Soit* $(\Omega, \mathscr{F}, \mathbb{P})$ *un espace de probabilité. Les propriétés suivantes sont vérifiées :*
i) *La probabilité de l'ensemble vide est égale à zéro, c'est-à-dire* $\mathbb{P}(\emptyset) = 0$ *;*
ii) *Si* $(A_k)_{k\in\{1,\dots,n\}}$ *est une suite finie d'événements deux à deux disjoints de* $\mathscr{F}$*, alors*

$$\mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} \mathbb{P}(A_k);$$

iii) *La probabilité de l'événement complémentaire d'un événement quelconque* $A \in \mathscr{F}$ *s'obtient par :*

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A);$$

iv) Si un événement en implique un autre, sa probabilité est plus petite, c'est-à-dire :

$$A \subset B \implies \mathbb{P}(A) \leq \mathbb{P}(B);$$

v) La probabilité de l'union de deux événements s'obtient par la formule de Poincaré :

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B);$$

vi) Pour toute suite d'événements quelconque $(A_n)_{n\geq 1}$ de $\mathscr{F}$, non nécessairement disjoints, on a l'inégalité de Boole :

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

Cette propriété est appelée σ-sous-additivité.
vii) Si $(A_n)_{n\geq 1} \subset \mathscr{F}$, $A_n \subseteq A_{n+1}$ et $\bigcup_{n=1}^{\infty} A_n = A$, alors $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$;
viii) Si $(A_n)_{n\geq 1} \subset \mathscr{F}$, $A_n \supseteq A_{n+1}$ et $\bigcap_{n=1}^{\infty} A_n = A$, alors $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.

Preuve : *i*) Utilisons la σ-additivité avec $A_1 = \Omega$ et $A_i = \emptyset$ pour $i \geq 2$:

$$\mathbb{P}(\Omega) = \mathbb{P}(\Omega) + \sum_{i\geq 2} \mathbb{P}(\emptyset).$$

Comme $\mathbb{P}(\Omega) = 1$ et que $\mathbb{P}(\emptyset) \geq 0$, on en déduit que $\sum_{i\geq 2} \mathbb{P}(\emptyset) = 0$, d'où le résultat.
ii) Soient $A_1, \ldots, A_n$ des événements deux à deux disjoints. Pour $k > n$, posons $A_k = \emptyset$. On obtient une suite infinie $(A_k)_{k\geq 1}$ d'événements deux à deux disjoints. Par σ-additivité :

$$\mathbb{P}\Big(\bigcup_{k=1}^{n} A_k\Big) = \mathbb{P}\Big(\bigcup_{k\geq 1} A_k\Big) = \sum_{k=1}^{n} \mathbb{P}(A_k) + \sum_{k>n} \mathbb{P}(A_k) = \sum_{k=1}^{n} \mathbb{P}(A_k),$$

puisque $\mathbb{P}(A_k) = 0$ pour $k > n$.
iii) On a $\Omega = A \cup A^c$ (union disjointe). En appliquant *ii*), il vient immédiatement que $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
iv) Si $A \subset B$, alors $B = A \cup (B \cap A^c)$ (union disjointe). Par *ii*) et le fait que $\mathbb{P}(B \cap A^c) \geq 0$, on obtient $\mathbb{P}(A) \leq \mathbb{P}(B)$.
v) Décomposons A et B en ensembles mutuellement disjoints :

$$A = (A \cap B) \cup (A \cap B^c), \quad B = (A \cap B) \cup (A^c \cap B), \quad A \cup B = (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B).$$

En utilisant *ii*), la formule de Poincaré s'en déduit.
vi) Considérons la suite d'événements disjoints :

$$B_1 = A_1,\ B_2 = A_2 \cap A_1^c, \quad \ldots, \quad B_n = A_n \cap \bigcap_{k=1}^{n-1} A_k^c.$$

Alors $(B_n)_{n\geq 1}$ est une suite deux à deux disjoints, $B_n \subset A_n$, et $\bigcup_{n\geq 1} B_n = \bigcup_{n\geq 1} A_n$. Ainsi,

$$\mathbb{P}\Big(\bigcup_{n\geq 1} A_n\Big) = \mathbb{P}\Big(\bigcup_{n\geq 1} B_n\Big) = \sum_{n\geq 1} \mathbb{P}(B_n) \leq \sum_{n\geq 1} \mathbb{P}(A_n),$$

où l'inégalité résulte de $B_n \subset A_n$. Les points *vii*) et *viii*) résultent directement de la continuité croissante et décroissante des mesures (cf. proposition 1.1.52). Comme $\mathbb{P}$ est une mesure finie, ces propriétés s'appliquent immédiatement et donnent les limites annoncées. ■

Définition 1.3.3 (Événements négligeables et presque sûrs) *Soit $(\Omega, \mathscr{F}, \mathbb{P})$ un espace de probabilité.*

- *Un événement $A \in \mathscr{F}$ est* négligeable *si $\mathbb{P}(A) = 0$.*
- *Un événement $A \in \mathscr{F}$ est* presque sûr *(p.s.) si $\mathbb{P}(A) = 1$, c'est-à-dire si A^c est négligeable.*

Proposition 1.3.4 *Si $(A_n)_{n\geq 1} \subset \mathscr{F}$ est une famille d'événements négligeables, alors $\bigcup_{n\geq 1} A_n$ est négligeable.*

Preuve : On applique l'inégalité de Boole : $\mathbb{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n) = 0$. ■

1.3.2 Probabilité conditionnelle

Définition 1.3.5 *Soit $(\Omega, \mathscr{F}, \mathbb{P})$ un espace de probabilité. Soient A et B deux événements quelconques de $\mathscr{F}$ tels que $\mathbb{P}(B) \neq 0$. La probabilité conditionnelle de A sachant B, notée $\mathbb{P}(A/B)$ ou $\mathbb{P}_B(A)$, est définie par :*

$$\mathbb{P}(A/B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

On en déduit la relation suivante, appelée **formule de Bayes** :

$$\mathbb{P}(A/B) \times \mathbb{P}(B) = \mathbb{P}(B/A) \times \mathbb{P}(A).$$

Cette formule, établie par Thomas Bayes $(1702 - 1761)$, constitue la base de ce que l'on appelle la *théorie subjective des probabilités*. Par son caractère de réciprocité, elle permet de réévaluer, a posteriori, la validité d'une hypothèse posée a priori.

Proposition 1.3.6 (Formule des probabilités composées) *Soient $A_1, \ldots, A_n$ des événements issus d'une expérience aléatoire d'univers Ω, tels que $\mathbb{P}(A_1 \cap \cdots \cap A_{n-1}) > 0$. On a :*

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\,\mathbb{P}(A_2/A_1) \times \cdots \times \mathbb{P}\big(A_n/A_1 \cap \cdots \cap A_{n-1}\big).$$

Preuve : La preuve s'obtient aisément par récurrence. ■

Exemple 1.3.7 *Considérons une urne contenant 4 boules blanches et 6 boules rouges. On tire successivement 3 boules sans remise. Quelle est la probabilité d'obtenir 3 boules blanches ? On note A_i l'événement : « la i-ème boule tirée est blanche ». Alors :*

$$\mathbb{P}(A_1) = \frac{4}{10} = \frac{2}{5}, \quad \mathbb{P}(A_2/A_1) = \frac{3}{9} = \frac{1}{3}, \quad \mathbb{P}(A_3/A_1 \cap A_2) = \frac{2}{8} = \frac{1}{4}.$$

Par la formule des probabilités composées : $\mathbb{P}(A_1 \cap A_2 \cap A_3) = \frac{2}{5} \times \frac{1}{3} \times \frac{1}{4} = \frac{1}{30}$.

Remarque 1.3.8 *On obtient le même résultat par un raisonnement combinatoire. En effet :* $\mathbb{P}(B) = \frac{\text{Card}(B)}{\text{Card}(\Omega)} = \frac{A_4^3}{A_{10}^3} = \frac{1}{30}$, *où* B *désigne l'événement « tirer* 3 *boules blanches ».*

Définition 1.3.9 *On appelle* **système complet d'événements** *toute partition* $E_1, \ldots, E_n$ *de* Ω, *c'est-à-dire :* $\forall i \neq j, \quad E_i \cap E_j = \emptyset$, *et* $\bigcup_{i=1}^n E_i = \Omega$.

Proposition 1.3.10 (Formule des probabilités totales) *Soit* $E_1, \ldots, E_n$ *un système complet d'événements sur* $\mathscr{F}$ *tel que* $\mathbb{P}(E_i) > 0$ *pour tout* $i \in \{1, \ldots, n\}$. *Alors, pour tout* $A \in \mathscr{F}$ *:* $\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(E_i)\,\mathbb{P}(A/E_i) = \sum_{i=1}^n \mathbb{P}(A \cap E_i)$.

Preuve : On a, d'après la définition 1.3.9 : $A = A \cap \Omega = A \cap \bigcup_{i=1}^n E_i = \bigcup_{i=1}^n (A \cap E_i)$. Comme les E_i sont deux à deux disjoints : $\mathbb{P}(A) = \mathbb{P}\Big(\bigcup_{i=1}^n (A \cap E_i)\Big) = \sum_{i=1}^n \mathbb{P}(A \cap E_i)$. ■

On déduit alors une deuxième formulation de la formule de Bayes :

Proposition 1.3.11 (Formule de Bayes) *Soit* $A \in \mathscr{F}$ *tel que* $\mathbb{P}(A) \neq 0$. *Alors, pour tout* $i = 1, \ldots, n$, *on a :*

$$\mathbb{P}(E_i/A) = \frac{\mathbb{P}(E_i)\,\mathbb{P}(A/E_i)}{\sum_{j=1}^n \mathbb{P}(E_j)\,\mathbb{P}(A/E_j)}.$$

Preuve : Il suffit de combiner la première formule de Bayes avec la proposition 1.3.10.■

1.3.3 Indépendance

- **Indépendance d'évènements**

Définition 1.3.12 *Deux évènements* A *et* B *sont dits indépendants si et seulement si*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B).$$

Attention à ne pas confondre indépendance et incompatibilité. Dans le cas de l'incompatibilité, on a $A \cap B = \emptyset$ et donc $\mathbb{P}(A \cap B) = 0$. En particulier, si deux évènements A et B sont incompatibles (ou disjoints), alors : $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Définition 1.3.13 *Une famille finie d'événements* $\{A_1, \ldots, A_n\} \subset \mathscr{F}$ *est dite* indépendante *(ou* mutuellement indépendante*) si, pour tout sous-ensemble non vide* $I \subset \{1, \ldots, n\}$, *on a*

$$\mathbb{P}\Big(\bigcap_{i \in I} A_i\Big) = \prod_{i \in I} \mathbb{P}(A_i).$$

Remarque 1.3.14 *Le fait que les évènements* A_i *soient deux à deux indépendants n'implique pas que la famille* $\{A_1, \ldots, A_n\}$ *soit mutuellement indépendante, comme le montre l'exemple suivant.*

Exemple 1.3.15 *On considère* $\Omega = \{1,2,3,4\}$ *muni de la loi uniforme* $p(\omega) = 1/4$, *pour tout* $\omega \in \Omega$. *On définit les évènements :* $A = \{1,2\}$, $B = \{2,3\}$, *et* $C = \{1,3\}$. *On a immédiatement* $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}$, *et* $\mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = \frac{1}{4}$. *Ainsi,* A, B *et* C *sont deux à deux indépendants. En revanche,* $A \cap B \cap C = \emptyset$, *d'où*

$$\mathbb{P}(A \cap B \cap C) = 0 \neq \tfrac{1}{8} = \mathbb{P}(A) \times \mathbb{P}(B) \times \mathbb{P}(C).$$