Walter APPEL

Agrégation externe Mathématiques

Mathématiques pour l'agrégation externe

Probabilités

ÉLÉMENTS DE COURS

- Applications transversales
- Près de 120 exercices corrigés

deboeck SUPÉRIEUR

Walter Appel

Mathématiques pour l'agrégation externe

Probabilités

deboeck SUPÉRIEUR

Chez le même éditeur (extrait du catalogue)

Agrégation de mathématiques :

Rombaldi J.-É., *Mathématiques pour l'agrégation externe. Analyse*
Rombaldi J.-É., *Mathématiques pour l'agrégation. Analyse et probabilités*
Rombaldi J.-É., *Mathématiques pour l'agrégation. Algèbre et géométrie* – 2e édition
Rombaldi J.-É., *Exercices et problèmes corrigés pour l'agrégation de mathématiques*
Rombaldi J.-É., *Leçons d'oral pour l'agrégation de mathématiques. Première épreuve : les exposés*
Rombaldi J.-É., *Leçons d'oral pour l'agrégation de mathématiques. Seconde épreuve : les exercices*

Capes de mathématiques :

Darracq M.-C. & Rombaldi J.-É., *Mathématiques pour le Capes. Analyse*
Darracq M.-C. & Rombaldi J.-É., *Mathématiques pour le Capes. Algèbre et géométrie*
Darracq M.-C. & Rombaldi J.-É., *Mathématiques pour le Capes. Probabilités*

Pour toute information sur notre fonds et les nouveautés dans votre domaine de spécialisation, consultez notre site web :
www.deboecksuperieur.com

En couverture : Coupe d'un nautile © Bereta/Getty Images
Maquette et mise en page de l'auteur
Maquette de couverture : Primo&Primo
Couverture : SCM, Toulouse

Dépôt légal :
Bibliothèque royale de Belgique : 2024/13647/077
Bibliothèque nationale, Paris : juillet 2024
ISBN : 978-2-8073-6166-9

Sommaire

Table des matières

Annexes

Notations

	Abréviations	
p.s.	presque sûre, presque sûrement	
p.p. (μ-p.p)	presque partout (relativement à la mesure μ)	
	Ensembles	
$\mathbb{K}$	$\mathbb{R}$ ou $\mathbb{C}$	
$\mathbb{P}$	ensemble des entiers premiers $\mathbb{P} = \{2, 3, 5, 7, 11, \dots\}$	
$\mathcal{P}(\Omega)$	ensemble des parties de Ω	
$\mathrm{A} \cap \mathrm{B}$	intersection de A et B	
$\mathrm{A} \cup \mathrm{B}$	réunion de A et B	
$\mathrm{A} \sqcup \mathrm{B}$	réunion disjointe de A et B ($\mathrm{A} \cap \mathrm{B} = \varnothing$, $\mathrm{A} \sqcup \mathrm{B} := \mathrm{A} \cup \mathrm{B}$)	
$\bigsqcup_n \mathrm{A}_n$	réunion disjointe	
$\mathrm{A} \setminus \mathrm{B}$	$= \mathrm{A} \cap \mathrm{B}^{\mathrm{c}}$	
$\mathrm{A} - \mathrm{B}$	$= \mathrm{A} \setminus \mathrm{B}$, étant entendu que $\mathrm{B} \subset \mathrm{A}$	
A^{c}	complémentaire de A (dans Ω)	
$\mathrm{A}_n \uparrow \mathrm{A}$, $\mathrm{A} = \lim\uparrow \mathrm{A}_n$	limite croissante	p. 23
$\mathrm{A}_n \downarrow \mathrm{A}$, $\mathrm{A} = \lim\downarrow \mathrm{A}_n$	limite décroissante	p. 23
$\mathrm{A}^* = \limsup \mathrm{A}_n$	$\bigcap_n \bigcup_{k \geqslant n} \mathrm{A}_k$	§ 1.3
$\mathrm{A}_* = \liminf \mathrm{A}_n$	$\bigcup_n \bigcap_{k \geqslant n} \mathrm{A}_k$	§ 1.3
$\mathfrak{M}_n(\mathbb{K})$	ensemble des matrices carrées d'ordre n	
$\mathfrak{M}_{n,1}(\mathbb{K})$	ensemble des vecteurs colonnes d'ordre n	
$\mathscr{C}^k(\mathrm{I}, \mathbb{K})$	ensemble des fonctions $f : \mathrm{I} \to \mathbb{K}$ de classe $\mathscr{C}^k$	
	Fonctions	
$f(x^+),\ f(x+0)$	limite à droite $\lim_{h \to 0^+} f(x+h)$	
$f(x^-),\ f(x-0)$	limite à gauche $\lim_{h \to 0^-} f(x+h)$	
$f(+\infty)$	limite en $+\infty$: $\lim_{x \to +\infty} f(x)$	
$f(-\infty)$	limite en $-\infty$: $\lim_{x \to -\infty} f(x)$	
$f * g$	produit de convolution	§ A.10.d
$\|f\|_\infty^{\mathrm{K}}$	$\sup_{x \in \mathrm{K}} \big\lvert f(x) \big\rvert$	

	Probabilités	
Ω	univers, espace des épreuves	
$\mathfrak{T}$, $\mathfrak{F}$, $\mathfrak{A}$	tribus (« T » et « F » gothiques) ou algèbres (« A »)	
ω	réalisation du hasard (élément de Ω)	
$\mathbf{1}_{\mathrm{A}}$	indicatrice de A	§ 3.1.e
$\mathbf{P}$, $\mathbf{Q}$	probabilités	
$\mathbf{P}_{\mathrm{A}}$	probabilité conditionnée par A	§ 2.1.a
$\{\mathrm{X} \in \mathrm{A}\}$	événement lié à une variable aléatoire	
$\mathbf{P} \otimes \mathbf{Q}$	probabilité produit	§ 7.1.b
$\mathrm{C}^{\mathbf{X}}$	matrice de covariance du vecteur $\mathbf{X}$	§ 7.2
$\mathbf{Cov}(\mathrm{X}, \mathrm{Y})$	covariance du couple (X, Y)	
$\mathbf{E}(\mathrm{X})$	espérance de la variable X	
$\mathbf{P}(\mathrm{A}), \mathbf{P}\{\mathrm{X} \in \mathrm{B}\}$	probabilité	
$\mathbf{V}(\mathrm{X})$	variance de la variable X	
$\boldsymbol{\sigma}_{\mathrm{X}}$	écart-type de X	
$\mathfrak{N}(x)$	fonction de répartition de la loi $\mathscr{N}(0\,;1)$	§ 8.1.e
$\mathfrak{n}(x)$	densité de probabilité de la loi $\mathscr{N}(0\,;1)$	§ 8.14
$\mathrm{X} \wedge \mathrm{Y}$, $\mathrm{X} \wedge m$	$\min(\mathrm{X}, \mathrm{Y})$, $\min(\mathrm{X}, m)$	
X'	variable centrée $\mathrm{X}' = \mathrm{X} - \mathbf{E}(\mathrm{X})$	
X^*	variable centrée réduite $\mathrm{X}^* = (\mathrm{X} - \mathbf{E}\mathrm{X})/\boldsymbol{\sigma}_{\mathrm{X}}$	
X^+	$\max(\mathrm{X}, 0)$	p. 165
X^-	$\max(-\mathrm{X}, 0)$	p. 165
$\mathrm{X} \overset{\text{ps}}{=} \mathrm{Y}$	égalité presque sûre : $\mathbf{P}\{\mathrm{X} = \mathrm{Y}\} = 1$	
$\mathrm{X} \perp\!\!\!\perp \mathrm{Y}$	X et Y sont indépendantes	§ 7.1.b
$\mathrm{X}_n \xrightarrow{\text{ps}} \mathrm{X}$	convergence presque sûre	§ 13.1
$\mathrm{X}_n \xrightarrow{\mathbf{P}} \mathrm{X}$	convergence en probabilité	§ 13.2
$\mathrm{X}_n \xrightarrow{\mathcal{L}} \mathrm{X}$	convergence en loi (convergence étroite)	§ 13.4
$\mathrm{X}_n \xrightarrow{\mathrm{L}^p} \mathrm{X}$	convergence dans L^p	§ 13.16
$\mathrm{X} \sim \mathscr{L}$	X suit la loi $\mathscr{L}$	
$\mathscr{B}(p)$	loi de Bernoulli de paramètre $p \in [0\,;1]$	§ 3.2.b
$\mathscr{B}(n\,;p)$	loi binomiale de paramètres n et p	§ 3.2.d,
$\mathscr{C}(m\,;a)$	loi de Cauchy de paramètres m et a	§ 8.1.h
$\mathscr{E}(\alpha)$	loi exponentielle de paramètre α	§ 8.1.d
$\mathscr{G}(p)$	loi géométrique de paramètre p	§ 3.2.e
$\mathscr{N}(m\,;\sigma^2)$	loi normale de moyenne m et de variance σ^2	§ 8.1.e
$\mathscr{P}(\lambda)$	loi de Poisson de paramètre λ	§ 3.2.g
$\mathfrak{B}$, $\mathfrak{B}(\mathbb{R})$, $\mathfrak{B}(\mathbb{R}^n)$	tribu des boréliens de $\mathbb{R}$, de $\mathbb{R}^n$	§ B.3.a
$\mathfrak{T}$, $\mathfrak{E}$, ...	tribus	
$\mathfrak{D}_{\mathrm{X}}$	système complet induit par une variable discrète	p. 65
$\sigma(\mathrm{X})$	tribu engendrée par X	§ 7.1.b
$\sigma(\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n)$	tribu engendrée par $\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n$	§ 14.1
$\mathfrak{Q}_n$	$\sigma(\mathrm{X}_n, \mathrm{X}_{n+1}, \mathrm{X}_{n+2}, \ldots)$	§ 14.1
δ_a	mesure de Dirac en a	
$\mathcal{L}^0(\Omega)$	espace des variables aléatoires	
$\mathcal{L}^0_+(\Omega)$	espace des variables aléatoires positives	
$\mathcal{L}^0_{\mathrm{b}}(\Omega)$	espace des variables aléatoires bornées	
$\mathcal{L}^1$, L^1	espaces des variables intégrables : $\mathbf{E}\lvert\mathrm{X}\rvert < +\infty$	§ 6.2.b
$\mathcal{L}^2, \mathrm{L}^2$	espaces des variables de carre intégrables : $\mathbf{E}\lvert\mathrm{X}\rvert^2 < +\infty$	§ 6.2.c

Introduction

Les dés aux minutes comptées,
les dés inaptes à étreindre,
parce qu'ils sont naissance et vieillesse.

René Char, *Partage formel* xxxviii.

Que trouve-t-on dans ce livre ?

Ce livre couvre en principe le programme de mathématiques de l'Agrégation externe

- pour les épreuves d'écrits,
- pour les leçons d'oral,

c'est-à-dire le programme commun à l'ensemble des candidats et candidates, indépendamment de l'option choisie.

S'il ne couvre pas l'ensemble des connaissances nécessaires à l'épreuve de modélisation de l'option A, il en traite certaines (vecteurs gaussiens, intervalles de confiance, convergence de suites de variables aléatoires).

On y trouvera entre autres :

- L'essentiel du cours : les théorèmes et la plupart des démonstrations, ou bien des renvois bibliographiques⟨1⟩.
- Quelques exercices d'applications, dont le but est double : s'entraîner à maîtriser les notions, et illustrer un plan de cours. Le niveau de difficulté des exercices est gradué en :

  ▷ (○) : exercice d'application directe.
  ▷ (⋆) , (⋆⋆) : exercice simple / plus délicat.
  ▷ (☕) , (☕☕) : exercice difficile, café recommandé / nécessaire

  Parfois, des indications suivent la liste d'exercice. Tous ceux de difficulté au moins une étoile sont corrigés.
- Des exemples d'applications, sortant parfois du cadre de la théorie des probabilités et débordant sur d'autres leçons — voire donnant quelques rares applications à la physique (après tout, la théorie des probabilités est née pour modéliser le monde réel, avant de devenir une discipline autonome).
- De nombreuses figures qu'il est bon d'avoir en tête au moment de présenter une leçon (pas seulement à l'Agrégation, mais dans votre futur métier de professeur).
- Quelques compléments, plus délicats, notés (∗), et qui peuvent être omis en première lecture.

⟨1⟩. Nous ne saurions trop conseiller de lire les grands ouvrages classiques, et ne pas se limiter à des manuels de cours !

Choix pédagogiques

Il existe de nombreuses manières de présenter un cours. Nous n'avons pas choisi d'être synthétiques mais, au contraire, de proposer plusieurs grilles de lecture. Par exemple, beaucoup de résultats sur les variables discrètes sont montrés avec des outils spécifiques à ces variables, bien qu'il eût été possible de les faire découler de théorèmes plus généraux.

Il faut cependant bien garder en tête que, lorsque vous présentez un développement, vous choisissez également le niveau auquel vous voulez le présenter. Si la théorie de la mesure vous met mal à l'aise, vous pouvez très bien présenter une belle leçon de probabilité dans laquelle la mesure ne joue qu'un rôle minime ou transparent⟨2⟩, en vous focalisant sur les variables discrètes qui offrent de très belles propriétés et des exemples d'applications nombreux. Dans tous les cas, *adaptez le niveau de votre leçon au niveau des questions que vous êtes prêts à entendre.*

Aussi, on ne s'étonnera pas de voir cohabiter des résultats très simples et d'autres plus délicats.

Liens internes

Le symbole « ✍ » indique le numéro d'un exercice qu'il est suggéré de chercher à ce stade de la lecture.

Le symbole « ✐ » indique un lien vers un autre chapitre au programme de l'Agrégation — que ce soit en Analyse, en Algèbre ou en Géométrie. Ce lien apparaît en petit, sous la forme suivante, généralement en début de paragraphe :

✐ *Arithmétique*

Ces différents liens sont également repris dans l'Index, sous l'entrée « **En lien avec…** ». Vous pouvez donc parcourir ce morceau d'Index lorsque vous recherchez, mettons, un exemple d'application à intégrer dans une leçon sur « Fonction définie par une intégrale dépendant d'un paramètre. Exemples et applications. »

Citations du jury

Vous trouverez régulièrement des indications de la forme

✍ **Rapport du jury** ✍

« Loi d'une variable aléatoire » *Les candidats aguerris pourront aborder le problème de la caractérisation de la loi par les moments.*

qui vous indiquent pourquoi tel ou tel complément de cours a été écrit ici. Bien évidemment, certaines remarques du jury sont des *suggestions* : ne vous sentez pas obligés de traiter les suggestions les plus difficiles si vous manquez d'assurance : il vaut mieux présenter un résultat modeste mais solidement établi. En revanche, une bonne maîtrise d'un sujet délicat sera évidemment très appréciée.

« Thèmes »

Une trentaine de « Thèmes » sont intégrés à l'ouvrage. Ce ne sont en aucun cas des développements prêts à l'emploi⟨3⟩ mais des suggestions, des digressions, d'agréables promenades (aléatoires bien sûr) que vous pouvez intégrer à des leçons, des exemples, des exercices, des développements.

Une liste des thèmes est présentée page 595.

⟨2⟩. Le premier tome du célèbre ouvrage de W. Feller *An introduction to probability theory and its applications* [13] s'en passe totalement.

⟨3⟩. Dont il faut toujours se méfier, car ils lassent les jurys au moment où, au contraire, il faut les réveiller avec énergie, bonne humeur, enthousiasme et, bien entendu, de solides connaissances.

Remerciements

Dans le monde entier, en tous lieux, à toute heure, une voix universelle n'implore que la Fortune ; on ne nomme qu'elle, [...] on lui impute la perte et le gain. [...] Et tel est sur nous l'empire du sort, qu'il n'y a plus d'autre divinité que ce même Sort, qui rend incertaine l'existence de Dieu.

Pline l'Ancien, *Histoire naturelle*

Je tiens à remercier vivement :

- Jean-Étienne Rombaldi pour m'avoir associé à son projet de cours de mathématiques pour l'Agrégation, et pour avoir bien voulu relire certains chapitres.
- Alain Luguet des éditions de Boeck, pour m'avoir fait confiance, et pour son travail éditorial soigné.
- Florence Palatin pour ses suggestions pertinentes et ses encouragements.
- Mes élèves des années passées et futures, sources constantes d'inspiration et de motivation !
- L'environnement chaleureux du Lycée du Parc (LdP), qui est pour beaucoup dans ce livre !
- De nombreuses discussions avec mes collègues et amis (notamment Denis Choimet, Nicolas Maillard, Jean-Pierre Barani, Kirone Mallick entre autres) m'ont apporté des éclairages nouveaux sur de nombreux sujets.
- Vincent Clapiès, Jérôme Demange, Olivier Dodane, Stéphane Gonnord, Denis Jourdan, Cécile Le Goff (LdP), ainsi que Bastien Marmeth et Marc Rezzouk pour avoir accepté de relire, dans des délais très serrés, des pans entiers de cet ouvrage.
- Une mention spéciale pour Marie Peronnier, pour sa relecture intégrale, chapitre après chapitre, et ses patientes questions !

J'accueillerai avec gratitude toute remarque, demande, suggestion pour un *errata* ou une éventuelle prochaine édition, à l'adresse `walter.appel@laposte.net`

Chapitre 1

Espaces probabilisés

Dans ce chapitre

Depuis les travaux de Kolmogorov⟨1⟩, il est d'usage de modéliser la notion intuitive de probabilité à l'aide de quelques axiomes, et d'une interprétation des différents objets mis en jeu.

Rappelons tout d'abord que, dans la vie réelle, un résultat est réputé *aléatoire* lorsqu'il est le résultat d'une expérience dont on ne sait pas, pour un certain nombre

⟨1⟩. Les parents d'Andreï Nikolaïevitch Kolmogorov (1903–1987) ne l'éduquèrent pas, sa mère étant morte en couches et son père, un agriculteur, étant exilé pour raisons politiques. Il grandit donc dans l'affection de sa tante, puis devint conducteur de chemin de fer tout en poursuivant ses études, avant d'aller suivre des cours de physique et de mathématiques à l'Université de Moscou. Dans son temps libre, il écrit un traité sur la mécanique newtonienne. Ses travaux portèrent ensuite sur des sujets très divers, comme les séries de Fourier, la logique, l'axiomatisation des probabilités, les processus markoviens et la théorie de l'information. En physique, il contribua de manière décisive à la théorie de la turbulence, à l'étude du chaos et des systèmes dynamiques et à la théorie hamiltonienne (la « théorie KAM »).

de raisons, prédire avec certitude le résultat. Si cette expérience est reproductible, aussi précisément que possible, et qu'après un grand nombre de fois (noté n) un certain résultat R est réalisé r fois, alors on dira que la probabilité que ce résultat ce réalise est $\approx r/n$ (nombre de cas favorables sur nombre de cas total).

Toute formalisation des probabilités devra impérativement rendre compte de ce résultat.

Dans la théorie de Kolmogorov, ce résultat (et ses généralisations) sera justifié par le théorème appelé *Loi des grands nombres* (chapitre 16).

L'idée de Kolmogorov est de considérer tous les résultats possibles du hasard, toutes les réalisations possibles et imaginables, un peu comme dans les théories des univers multiples d'Everett⟨2⟩ ou tout simplement comme dans les mondes parallèles chers à la Science-Fiction classique. L'ensemble de tous ces mondes possibles est généralement appelé Ω.

Un observateur est simplement un habitant d'un de ces mondes parallèles, et il n'a accès qu'aux renseignements de ce monde (noté génériquement ω). Cet ω est tout simplement ce que le Destin a réservé à l'observateur (dans d'autres univers parallèles, le destin est un peu différent, ou l'est beaucoup).

Les différentes expériences possibles sont modélisées par des variables aléatoires (et des réalisations multiples d'une « même » expérience, comme lancer un dé, par une suite de variables aléatoires, mais toujours évaluées en l'unique ω connu de notre observateur).

1.1 Axiomes des probabilités

1.1.a Rappels de vocabulaire ensembliste

Soit Ω un ensemble.

- Si A et B sont des parties de Ω, on note $\mathrm{A} \smallsetminus \mathrm{B}$ l'ensemble des éléments de A qui ne sont pas dans B :

$$\mathrm{A} \smallsetminus \mathrm{B} = \mathrm{A} \cap \mathrm{B}^{\mathrm{c}} = \{x \in \mathrm{A} \; ; \; x \notin \mathrm{B}\}\,.$$

- Si B est inclus dans A, on notera $\mathrm{A} - \mathrm{B}$ l'ensemble $\mathrm{A} \smallsetminus \mathrm{B}$.
- Si A et B sont des parties disjointes, on note $\mathrm{A} \sqcup \mathrm{B}$ leur réunion disjointe :

$$\mathrm{A} \sqcup \mathrm{B} := \mathrm{A} \cup \mathrm{B} \qquad \text{quand } \mathrm{A} \cap \mathrm{B} = \varnothing.$$

  Si $(\mathrm{A}_i)_{i\in\mathrm{I}}$ est une famille de parties deux à deux disjointes, on note $\bigsqcup_i \mathrm{A}_i$ la réunion disjointe $\bigcup_i \mathrm{A}_i$.

⟨2⟩. Le mathématicien et physicien Hugh Everett (1930–1982) est connu pour avoir proposé une interprétation originale de la mécanique quantique, allant à l'encontre de l'interprétation de l'École de Copenhague, au cours de sa thèse de doctorat qu'il a préparé sous la direction de John Archibald Wheeler. Dans cette interprétation, l'existence de mondes multiples évite notamment le problème la réduction du paquet d'ondes qui introduisait une notion d'aléatoire peu satisfaisante.

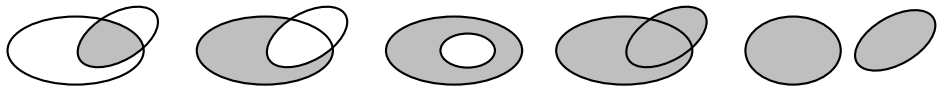


Figure 1.1 — *Opérations ensemblistes.*

- On étend les notations précédentes aux familles vides : si $\mathrm{I} = \varnothing$ est l'ensemble vide,
$$\bigcup_{i\in\mathrm{I}} \mathrm{A}_i := \varnothing \qquad \bigsqcup_{i\in\mathrm{I}} \mathrm{A}_i := \varnothing \qquad \bigcap_{i\in\mathrm{I}} \mathrm{A}_i := \Omega.$$
- On munit l'ensemble des parties de Ω de l'ordre partiel induit par la relation d'appartenance. Une suite $(\mathrm{A}_n)_{n\geqslant 0}$ est **croissante** si $\mathrm{A}_n \subset \mathrm{A}_{n+1}$ pour tout $n \in \mathbb{N}$.
- Si $(\mathrm{A}_n)_{n\geqslant 0}$ est une suite de parties de Ω, alors on définit sa **borne supérieure** et sa **borne inférieure** comme
$$\sup_n \mathrm{A}_n := \bigcup_{n=0}^{\infty} \mathrm{A}_n \qquad \text{et} \qquad \inf_n \mathrm{A}_n := \bigcap_{n=0}^{\infty} \mathrm{A}_n.$$
- Si $(\mathrm{A}_n)_{n\geqslant 0}$ est une suite *croissante* de parties de Ω (c'est-à-dire que $\mathrm{A}_n \subset \mathrm{A}_{n+1}$ pour tout entier n), sa borne supérieure est également appelée **limite** de $(\mathrm{A}_n)_{n\geqslant 0}$ et notée $\lim \uparrow \mathrm{A}_n$:
$$\lim_n \uparrow \mathrm{A}_n := \bigcup_{n=0}^{\infty} \mathrm{A}_n.$$
Si l'on note $\mathrm{A} := \lim \uparrow \mathrm{A}_n$, on écrit alors $\mathrm{A}_n \uparrow \mathrm{A}$.
- De même, si $(\mathrm{A}_n)_{n\geqslant 0}$ est *décroissante*, on écrit
$$\mathrm{A}_n \downarrow \mathrm{A} \qquad \text{avec} \qquad \mathrm{A} := \lim_n \downarrow \mathrm{A}_n := \bigcap_{n=0}^{\infty} \mathrm{A}_n.$$

1.1.b Tribus

Pour des raisons techniques⟨3⟩, dans un premier temps, mais également par commodité dans certaines parties de la théorie, nous devons nous contenter de définir une mesure non pas sur l'ensemble des parties de Ω, mais seulement sur certaines d'entre elles.

Afin de pouvoir écrire des formules de type
$$\mathbf{P}(\mathrm{A}^{\mathrm{c}}) = 1 - \mathbf{P}(\mathrm{A}) \qquad \text{ou} \qquad \mathbf{P}\left(\bigsqcup_n \mathrm{A}_n\right) = \sum_n \mathbf{P}(\mathrm{A}_n),$$
nous demanderons que l'ensemble de ces parties soit stable par passage au complémentaire par réunion dénombrable.

⟨3⟩. Liées à des difficultés lors de l'utilisation de l'Axiome du choix dans sa version la plus générale. Voir section B.1 page 570 de l'annexe B, et en particulier l'encadré 30 page 571.

Définition 1.1 (Tribu) — Une **tribu** sur Ω est une partie $\mathfrak{T}$ de l'ensemble $\mathcal{P}(\Omega)$ des parties de Ω, vérifiant les propriétés suivantes :

i) $\Omega \in \mathfrak{T}$;

ii) pour tout A de $\mathfrak{T}$, son complémentaire A^{c} est dans $\mathfrak{T}$;

iii) $\mathfrak{T}$ est stable par réunions (finies ou) dénombrables : si $(\mathrm{A}_n)_{n\geqslant 0}$ est une suite d'éléments de $\mathfrak{T}$, alors

$$\bigcup_{n=0}^{\infty} \mathrm{A}_n \in \mathfrak{T}.$$

Une tribu est également appelée **σ-algèbre**, la lettre σ rappelant que les réunions peuvent être dénombrables.

Remarque 1.2 On définit une **algèbre** en n'exigeant que la stabilité par réunion finie. Voir l'annexe B page 569.

Exemple 1.3 Soit Ω un ensemble. Alors $\{\varnothing, \Omega\}$ est une tribu (**tribu grossière**), ainsi que $\mathcal{P}(\Omega)$ (**tribu triviale**).

Exemple 1.4 Pour tout $n \in \mathbb{Z}$, notons $\mathrm{I}_n := [\,n\,;n+1\,[$. On construit une tribu $\mathfrak{T}$ sur $\mathbb{R}$ en considérant tous les ensembles de la forme $\mathrm{A}_{\mathrm{K}} := \bigsqcup_{n\in \mathrm{K}} \mathrm{I}_n$, où K est une partie quelconque de $\mathbb{Z}$. Alors $\mathfrak{T}$ est une tribu, dont on remarquera qu'elle ne contient pas d'élément plus petit qu'un intervalle $[\,n\,;n+1\,[$ (et notamment, pas de singleton).

Théorème 1.5 (Propriétés élémentaires des tribus) — *Soit $\mathfrak{T}$ une tribu de Ω.*

i) $\mathfrak{T}$ est stable par intersection dénombrable : si $(\mathrm{A}_n)_{n\geqslant 0}$ est une suite d'éléments de $\mathfrak{T}$, alors

$$\bigcap_{n=0}^{\infty} \mathrm{A}_n \in \mathfrak{T}.$$

ii) Ainsi, $\mathfrak{T}$ est stable par passage à la borne supérieure et borne inférieure.

iii) L'intersection d'une famille de tribus de Ω est encore une tribu.

Comme avec des nombreuses structures connues (espace vectoriels, groupes), on peut réduire et définir des sous-tribus, ou bien prendre une partie et considérer la plus petite tribu la contenant :

Définition 1.6 — Soit $\mathfrak{F}$ une tribu sur Ω. Une partie $\mathfrak{G}$ de $\mathcal{P}(\Omega)$ est une **sous-tribu** de $\mathfrak{F}$ si c'est une tribu et si $\mathfrak{G} \subset \mathfrak{F}$.

Définition 1.7 (Tribu engendrée) — Soit $\mathfrak{F} = (\mathrm{F}_i)_{i\in \mathrm{I}}$ un ensemble non vide de parties de Ω. Comme vu dans la proposition 1.5, l'intersection de toutes les tribus contenant $\mathfrak{F}$ est encore une tribu (et elle est non vide puisque $\mathcal{P}(\Omega)$ est bien une tribu contenant $\mathfrak{F}$). On l'appelle **tribu engendrée par $\mathfrak{F}$**, et on la note $\sigma(\mathfrak{F})$. C'est, par construction, la plus petite tribu contenant $\mathfrak{F}$.

Conséquence immédiate du fait que $\sigma(\mathfrak{A})$ est l'intersection de toutes les tribus contenant $\mathfrak{A}$:

■ **Proposition 1.8** — *Soit $\mathfrak{A}$ une partie de $\mathcal{P}(\Omega)$.*

i) Si $\mathfrak{A}$ est une tribu, alors $\sigma(\mathfrak{A}) = \mathfrak{A}$.

ii) Si $\mathfrak{A} \subset \mathfrak{G}$ et si $\mathfrak{G}$ est une tribu, alors $\sigma(\mathfrak{A}) \subset \mathfrak{G}$.

⟦ Démonstration : On suppose que $\mathfrak{A}$ est une tribu. On sait que $\sigma(\mathfrak{A})$ est l'intersection de toutes les tribus contenant $\mathfrak{A}$, elle est donc incluse dans toute tribu contenant $\mathfrak{A}$, et notamment dans $\mathfrak{A}$; ainsi, $\mathfrak{A} \subset \sigma(\mathfrak{A}) \subset \mathfrak{A}$ ce qui permet de conclure que $\sigma(\mathfrak{A}) = \mathfrak{A}$.

On suppose que $\mathfrak{G}$ est une tribu et que $\mathfrak{A}$ est une partie de $\mathcal{P}(\Omega)$ incluse dans $\mathfrak{G}$. Alors $\sigma(\mathfrak{A})$ est l'intersection d'un certain ensemble de tribus, au rang desquelles on trouve $\mathfrak{G}$, donc $\sigma(\mathfrak{A}) \subset \mathfrak{G}$. ⟧

On en déduit la propriété de croissance ci-dessous, élémentaire mais très utile.

■ **Proposition 1.9 (Croissance)** — *Si $\mathfrak{F}$ et $\mathfrak{G}$ sont des ensembles de parties de Ω et si $\mathfrak{F} \subset \mathfrak{G}$, alors $\sigma(\mathfrak{F}) \subset \sigma(\mathfrak{G})$.*

⟦ Démonstration : On sait déjà que $\mathfrak{F} \subset \mathfrak{G} \subset \sigma(\mathfrak{G})$; puisque $\sigma(\mathfrak{G})$ est une tribu, la proposition 1.8 permet alors de conclure que $\sigma(\mathfrak{F}) \subset \sigma(\mathfrak{G})$. ⟧

Définition 1.10 — On appelle tribu des **boréliens de $\mathbb{R}$**, et on note $\mathfrak{B} = \mathfrak{B}(\mathbb{R})$, la tribu engendrée par les intervalles de $\mathbb{R}$.

■ **Proposition 1.11 (Autres caractérisations des boréliens)** —

i) $\mathfrak{B}$ est la tribu engendrée par les intervalles de la forme $]-\infty\,;a]$, pour a parcourant $\mathbb{R}$;

ii) $\mathfrak{B}$ est la tribu engendrée par les ouverts de $\mathbb{R}$ (et donc également la tribu engendrée par les fermés).

1.1.c Espaces probabilisés

Définition 1.12 — On appelle **espace probabilisable** un couple $(\Omega, \mathfrak{T})$ où Ω est un ensemble quelconque, et $\mathfrak{T}$ une tribu sur Ω.

- L'ensemble Ω est appelé **univers** ou **espace des épreuves**, et ses éléments, génériquement notés ω, sont les **épreuves**, ou **aléas**, ou **réalisations du hasard**, ou encore **résultats d'une expérience**.
- Les éléments de la tribu $\mathfrak{T}$ sont appelés des **événements observables** ou, plus succinctement, **événements**.
- On dit que deux événements A et B sont **incompatibles** lorsqu'ils sont disjoints, c'est-à-dire $A \cap B = \varnothing$.

Remarque 1.13 (Caractérisation d'une propriété par un ensemble de vérité) Un événement est souvent décrit par une propriété L, pouvant être ou non vérifiée par les éléments ω ; on assimilera dans toute la suite la propriété L et l'*ensemble de vérité* de L, qui est l'événement $\{\omega \in \Omega\,;\ L(\omega)\}$.

Le cas le plus fréquent est le suivant : si $X : \Omega \to \mathbb{R}$ est une fonction et si A est une partie de $\mathbb{R}$, alors on note $\{X \in A\}$ l'ensemble $\{\omega \in \Omega\,;\ X(\omega) \in A\}$.

Définition 1.14 (Espaces probabilisés) — Soit $(\Omega, \mathfrak{T})$ un espace probabilisable. Une **probabilité** (ou encore : **mesure de probabilité**) sur cet espace est une application $\mathbf{P} : \mathfrak{T} \to \mathbb{R}$ vérifiant les propriétés suivantes :

⋆ **Axiome 1 : P** est à valeurs positives ;

⋆ **Axiome 2 :** $\mathbf{P}(\Omega) = 1$;

⋆ **Axiome 3 :** $\mathbf{P}$ est **σ-additive** : pour toute suite $(\mathrm{A}_n)_{n\geqslant 0}$ d'éléments deux à deux disjoints de $\mathfrak{T}$,

$$\mathbf{P}\Big(\bigsqcup_{n=0}^{\infty} \mathrm{A}_n\Big) = \sum_{n=0}^{\infty} \mathbf{P}(\mathrm{A}_n).$$

Le triplet $(\Omega, \mathfrak{T}, \mathbf{P})$ est appelé **espace probabilisé**.

Exemple 1.15 (Masse de Dirac) Soit Ω un ensemble quelconque ; on le munit de la tribu triviale $\mathfrak{T} = \mathcal{P}(\Omega)$. Fixons maintenant un élément $a \in \Omega$ particulier et posons, pour toute partie $\mathrm{A} \in \mathfrak{T}$,

$$\mathbf{P}(\mathrm{A}) := \begin{cases} 1 & \text{si } a \in \mathrm{A}, \\ 0 & \text{sinon.} \end{cases}$$

Alors $\mathbf{P}$ est une probabilité : on l'appelle la **masse de Dirac**, ou **mesure de Dirac**, centrée en a, et on la note habituellement δ_a.

Exemple 1.16 (Une probabilité à densité) $\Omega = \mathbb{R}$ est muni de la tribu $\mathfrak{T} = \mathfrak{B}$ des boréliens. Soit $f : \mathbb{R} \to \mathbb{R}^+$ une fonction continue[⟨4⟩] et positive, intégrable et telle que $\int_{\mathbb{R}} f = 1$. Pour tout borélien A, on pose

$$\mathbf{P}(\mathrm{A}) := \int_{\mathrm{A}} f(t)\,\mathrm{d}t = \int_{-\infty}^{+\infty} \mathbf{1}_{\mathrm{A}}(t)\, f(t)\,\mathrm{d}t,$$

où $\mathbf{1}_{\mathrm{A}}$ est l'indicatrice de A, définie par $\mathbf{1}_{\mathrm{A}}(t) = 1$ si $t \in \mathrm{A}$ et $\mathbf{1}_{\mathrm{A}}(t) = 0$ sinon. Alors $\mathbf{P}$ est une probabilité, de **densité** f.

Des raisonnements « finis » permettent de démontrer les propriétés élémentaires suivantes :

■ **Proposition 1.17 (Propriétés « finie »)** — *Soit $(\Omega, \mathfrak{T}, \mathbf{P})$ un espace probabilisé. Soient* A, B *des éléments de* $\mathfrak{T}$.

i) $\mathbf{P}(\varnothing) = 0$;

ii) $\mathbf{P}$ *est à valeurs dans* $[0\,;1]$;

iii) ***(croissance)*** *: si* $\mathrm{A} \subset \mathrm{B}$, *alors* $\mathbf{P}(\mathrm{A}) \leqslant \mathbf{P}(\mathrm{B})$;

iv) Si A *et* B *sont disjoints, alors* $\mathbf{P}(\mathrm{A} \sqcup \mathrm{B}) = \mathbf{P}(\mathrm{A}) + \mathbf{P}(\mathrm{B})$;

v) Si $\mathrm{A} \subset \mathrm{B}$, *alors* $\mathbf{P}(\mathrm{B} - \mathrm{A}) = \mathbf{P}(\mathrm{B}) - \mathbf{P}(\mathrm{A})$;

vi) $\mathbf{P}(\mathrm{A} \smallsetminus \mathrm{B}) = \mathbf{P}(\mathrm{A}) - \mathbf{P}(\mathrm{A} \cap \mathrm{B})$;

vii) $\mathbf{P}(\mathrm{A}^c) = 1 - \mathbf{P}(\mathrm{A})$;

viii) $\mathbf{P}(\mathrm{A} \cup \mathrm{B}) = \mathbf{P}(\mathrm{A}) + \mathbf{P}(\mathrm{B}) - \mathbf{P}(\mathrm{A} \cap \mathrm{B})$.

ix) ***(additivité)*** *: si* $\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_n$ *sont des événements deux à deux disjoints, alors*

$$\mathbf{P}(\mathrm{A}_1 \sqcup \mathrm{A}_2 \sqcup \cdots \sqcup \mathrm{A}_n) = \mathbf{P}(\mathrm{A}_1) + \mathbf{P}(\mathrm{A}_2) + \cdots + \mathbf{P}(\mathrm{A}_n).$$

x) ***(sous-additivité)*** *: si* $\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_n$ *sont des événements quelconques,*

$$\mathbf{P}\left(\mathrm{A}_1 \cup \mathrm{A}_2 \cup \cdots \cup \mathrm{A}_n\right) \leqslant \mathbf{P}(\mathrm{A}_1) + \mathbf{P}(\mathrm{A}_2) + \cdots + \mathbf{P}(\mathrm{A}_n).$$

⟨4⟩. Ou du moins mesurable pour la tribu des boréliens, c'est-à-dire telle que, pour tout borélien B de $\mathbb{R}$, son image réciproque $f^{-1}(\mathrm{B})$ est encore un borélien.

⟦ Démonstration : *i*) : on pose $\mathrm{A}_0 := \Omega$ et $\mathrm{A}_n := \varnothing$ pour tout $n \geqslant 1$. La suite $(\mathrm{A}_n)_{n\geqslant 0}$ est composée d'éléments deux à deux disjoints, l'axiome 3 permet donc d'écrire

$$1 = \mathbf{P}(\Omega) = \mathbf{P}\Big(\bigsqcup_{n=0}^{\infty} \mathrm{A}_n \Big) = \mathbf{P}(\Omega) + \sum_{n=1}^{\infty} \mathbf{P}(\varnothing),$$

et donc $\mathbf{P}(\varnothing) = 0$.

iv) et *ix*) : soit n un entier non nul ; soient $\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_n$ des éléments de $\mathfrak{T}$. On pose $\mathrm{A}_k := \varnothing$ pour tout entier $k \geqslant n+1$; la propriété de σ-additivité et la propriété *i*) permettent de conclure.

iii) et *v*) : on écrit $\mathrm{B} = \mathrm{A} \sqcup (\mathrm{B} - \mathrm{A})$ donc $\mathbf{P}(\mathrm{B}) = \mathbf{P}(\mathrm{A}) + \mathbf{P}(\mathrm{B} - \mathrm{A})$ grâce à la propriété d'additivité.

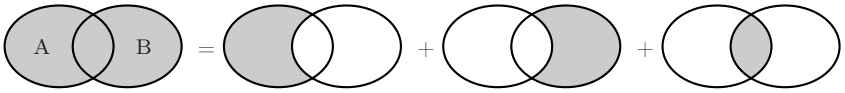
ii) et *vii*) : pour tout élément A de $\mathfrak{T}$, $\mathbf{P}(\mathrm{A}) + \mathbf{P}(\mathrm{A}^{\mathrm{c}}) = 1$.

vi) : découle de la formule précédente et de la relation $\mathrm{A} \smallsetminus \mathrm{B} = \mathrm{A} - (\mathrm{A} \cap \mathrm{B})$.

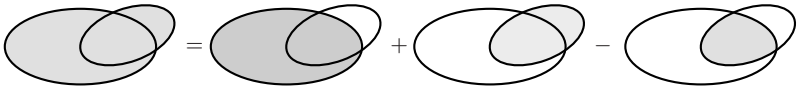
viii) : on écrit

$$\mathrm{A} \cup \mathrm{B} = \big[\mathrm{A} - (\mathrm{A} \cap \mathrm{B})\big] \sqcup \big[\mathrm{B} - (\mathrm{A} \cap \mathrm{B})\big] \sqcup \mathrm{A} \cap \mathrm{B}$$

et on applique les propriétés *iv*) et *v*).



Alternativement, on peut considérer la figure ci-dessous :



$\mathbf{P}(\mathrm{A} \cup \mathrm{B}) \quad = \quad \mathbf{P}(\mathrm{A}) \quad + \quad \mathbf{P}(\mathrm{B}) \quad - \quad \mathbf{P}(\mathrm{A} \cap \mathrm{B})$

x) : se démontre par récurrence sur n en utilisant la propriété *viii*). ⟧

Mais nous aurons surtout besoin de propriétés faisant intervenir une quantité infinie d'événements. C'est ici que les tribus et la notion de σ-additivité prennent leur sens.

1.1.d Continuité monotone d'une probabilité

Théorème 1.18 (Propriétés « dénombrables ») — *Soit $(\Omega, \mathfrak{T}, \mathbf{P})$ un espace probabilisé. Alors*

i) ***Continuité monotone croissante*** *: pour toute suite croissante $(\mathrm{B}_n)_{n\geqslant 0}$ d'éléments de $\mathfrak{T}$,*

$$\mathbf{P}\Big(\bigcup_{n=0}^{\infty} \mathrm{B}_n \Big) = \lim_{n\to\infty} \mathbf{P}(\mathrm{B}_n)$$

ce que l'on écrit aussi $\mathbf{P}\big(\lim_{n\to\infty} \uparrow \mathrm{B}_n \big) = \lim_{n\to\infty} \mathbf{P}(\mathrm{B}_n)$.

ii) ***Continuité monotone décroissante*** *: pour toute suite $(\mathrm{C}_n)_{n\geqslant 0}$ décroissante d'éléments de $\mathfrak{T}$,*

$$\mathbf{P}\Big(\bigcap_{n=0}^{\infty} \mathrm{C}_n \Big) = \lim_{n\to\infty} \mathbf{P}(\mathrm{C}_n).$$

ce que l'on écrit aussi $\mathbf{P}\big(\lim_{n\to\infty} \downarrow \mathrm{C}_n \big) = \lim_{n\to\infty} \mathbf{P}(\mathrm{C}_n)$.

⟦ Démonstration : Montrons la continuité croissante.

Soit $(B_n)_{n\geqslant 0}$ une suite croissante d'éléments de $\mathfrak{T}$. On pose $B_{-1} = \varnothing$ et $A_n = B_n \smallsetminus B_{n-1}$ pour tout n entier. Ainsi, les A_n sont deux à deux disjoints et

$$B_n = A_0 \sqcup A_1 \sqcup A_2 \sqcup \cdots \sqcup A_n$$

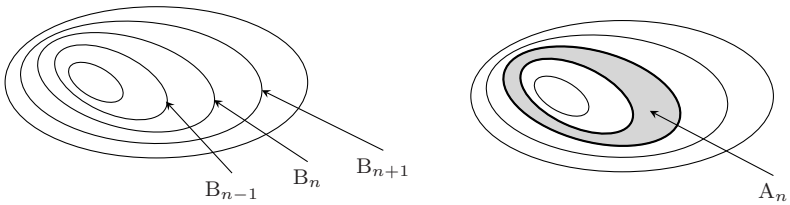
pour tout n, donc notamment $\mathbf{P}(B_n) = \mathbf{P}(A_0) + \mathbf{P}(A_1) + \cdots + \mathbf{P}(A_n)$. Vérifions par double inclusion que $\bigcup_{n=0}^{\infty} B_n = \bigsqcup_{n=0}^{\infty} A_n$.

- Soit $\omega \in \bigcup_{n=0}^{\infty} B_n$. Il existe un entier N tel que $\omega \in B_N = A_0 \sqcup \cdots \sqcup A_N$, et donc il existe un (unique) $k \in [\![0\,;N]\!]$ tel que $\omega \in A_k$. Ainsi, $\omega \in \bigsqcup_{n=0}^{\infty} A_n$.
- Soit $\omega \in \bigsqcup_{n=0}^{\infty} A_n$. Il existe un (unique) entier k tel que $\omega \in A_k$, et comme $A_k \subset B_k$, on a bien $\omega \in B_k \subset \bigcup_{n=0}^{\infty} B_n$.

Ainsi,

$$\mathbf{P}\Big(\bigcup_{n=0}^{\infty} B_n\Big) = \mathbf{P}\Big(\bigsqcup_{n=0}^{\infty} A_n\Big) = \sum_{n=0}^{\infty} \mathbf{P}(A_n) = \lim_{N\to\infty} \sum_{n=0}^{N} \mathbf{P}(A_n) = \lim_{N\to\infty} \mathbf{P}(B_N),$$

ce qu'il fallait démontrer.



Par passage au complémentaire, on obtient la continuité monotone décroissante. ⟧

Réciproquement, pour une fonction $\mathbf{P} : \mathfrak{T} \to \mathbb{R}^+$ additive et vérifiant $\mathbf{P}(\Omega) = 1$, ↛1.3 chacune des deux propriétés de continuité monotone est équivalente à la σ-additivité.

On pourra retenir ces résultats en utilisant les notations de la page 23 (« $C_n \downarrow C$ » pour indiquer que la suite $(C_n)_{n\geqslant 0}$ est décroissante et de limite C, et $B_n \uparrow B$ pour une suite croissante de limite B) :

σ-additivité			$\mathbf{P}\big(\bigsqcup_n A_n\big) = \sum_n \mathbf{P}(A_n)$
continuité croissante	si $B_n \uparrow B$,	alors	$\mathbf{P}(B_n) \uparrow \mathbf{P}(B)$
continuité décroissante	si $C_n \downarrow C$,	alors	$\mathbf{P}(C_n) \downarrow \mathbf{P}(C)$

En pratique, on utilisera souvent ce théorème sous la forme suivante :

Théorème 1.19 (Probabilité d'une réunion / d'une intersection) — *Soit $(\Omega, \mathfrak{T}, \mathbf{P})$ un espace probabilisé.*

i) Si $(A_n)_{n\geqslant 0}$ est une suite quelconque d'événements, la suite de terme général $B_n = \bigcup_{k=0}^{n} A_k$ est croissante, et

$$\mathbf{P}\Big(\bigcup_{n=0}^{\infty} A_n\Big) = \lim_{n\to\infty} \mathbf{P}\Big(\bigcup_{k=0}^{n} A_k\Big).$$

ii) Si $(\mathrm{A}_n)_{n\geqslant 0}$ *est une suite quelconque d'événements, la suite de terme général* $\mathrm{C}_n = \bigcap_{k=0}^{n} \mathrm{A}_k$ *est décroissante, et*

$$\mathbf{P}\Big(\bigcap_{n=0}^{\infty} \mathrm{A}_n\Big) = \lim_{n\to\infty} \mathbf{P}\Big(\bigcap_{k=0}^{n} \mathrm{A}_k\Big).$$

⟦ Démonstration : Il suffit de remarquer que $\bigcup_{k=0}^{n} \mathrm{A}_k = \bigcup_{k=0}^{n} \mathrm{B}_k$ et que $\bigcap_{k=0}^{n} \mathrm{A}_k = \bigcap_{k=0}^{n} \mathrm{C}_k$, par double inclusion. ⟧

Théorème 1.20 (Inégalité de Boole / propriété de σ-sous-additivité) — *Si* $(\mathrm{A}_n)_{n\geqslant 1}$ *est une suite d'événements quelconques,*

$$\mathbf{P}\left(\bigcup_{n=0}^{\infty} \mathrm{A}_n\right) \leqslant \sum_{n=0}^{\infty} \mathbf{P}(\mathrm{A}_n),$$

le deuxième membre de l'inégalité pouvant être fini ou égal à $+\infty$.

⟦ Démonstration : On sait que $\mathbf{P}(\mathrm{A}\cup\mathrm{B}) \leqslant \mathbf{P}(\mathrm{A}) + \mathbf{P}(\mathrm{B})$ pour tous événements A et B. On en déduit, par récurrence, que $\mathbf{P}(\mathrm{A}_0 \cup \cdots \cup \mathrm{A}_n) \leqslant \mathbf{P}(\mathrm{A}_0) + \cdots + \mathbf{P}(\mathrm{A}_n)$ pour tout $n \in \mathbb{N}$.

Si l'on note maintenant $\mathrm{B}_n = \mathrm{A}_0 \cup \cdots \cup \mathrm{A}_n$, la suite $(\mathrm{B}_n)_{n\geqslant 0}$ est croissante, et

$$\mathbf{P}(\mathrm{B}_n) = \mathbf{P}(\mathrm{A}_0 \cup \cdots \cup \mathrm{A}_n) \leqslant \mathbf{P}(\mathrm{A}_0) + \cdots + \mathbf{P}(\mathrm{A}_n)$$

pour tout $n \in \mathbb{N}$. De plus,

$$\mathbf{P}\Big(\bigcup_{n=0}^{\infty} \mathrm{A}_n\Big) = \mathbf{P}\Big(\bigcup_{n=0}^{\infty} \mathrm{B}_n\Big) = \mathbf{P}\big(\lim_{n\to\infty}\uparrow \mathrm{B}_n\big) = \lim_{n\to\infty} \mathbf{P}(\mathrm{B}_n)$$

par continuité monotone croissante. Enfin, l'inégalité précédente donne

$$\lim_{n\to\infty} \mathbf{P}(\mathrm{B}_n) \leqslant \sum_{n=0}^{\infty} \mathbf{P}(\mathrm{A}_n).$$ ⟧

⟦ Démonstration alternative : On peut également transformer la réunion des A_n en une réunion disjointe : pour cela, on note

$$\mathrm{A}'_0 = \mathrm{A}_0 \qquad \mathrm{A}'_n = \mathrm{A}_n \smallsetminus \bigcup_{k=0}^{n-1} \mathrm{A}_k$$

Par construction, les A'_n sont disjoints deux à deux, $\mathrm{A}'_n \subset \mathrm{A}_n$ et $\mathrm{A}_0 \cup \cdots \cup \mathrm{A}_n = \mathrm{A}'_0 \sqcup \cdots \sqcup \mathrm{A}'_n$. On peut alors utiliser la propriété de σ-additivité de la mesure de probabilité pour obtenir

$$\mathbf{P}\Big(\bigcup_{n=0}^{\infty} \mathrm{A}_n\Big) = \mathbf{P}\Big(\bigsqcup_{n=0}^{\infty} \mathrm{A}'_n\Big) = \sum_{n=0}^{\infty} \mathbf{P}(\mathrm{A}'_n) \leqslant \sum_{n=0}^{\infty} \mathbf{P}(\mathrm{A}_n).$$ ⟧

1.1.e Événements négligeables et presque sûrs

En théorie des probabilités, les événements négligeables et presque sûrs jouent un rôle central : ce sont en effet eux qui vont naturellement apparaître dans la plupart des théorèmes limites⟨5⟩.

⟨5⟩. Et par conséquent transformer l'étude de l'incertain, qui a historiquement fondé les probabilités, en théorie de la certitude.

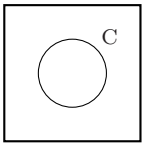
Définition 1.21 — On appelle **événement négligeable** tout élément N de $\mathfrak{T}$ de probabilité nulle, c'est-à-dire tel que $\mathbf{P}(\mathrm{N}) = 0$.
À l'opposé, un événement **presque sûr** est le complémentaire d'un événement négligeable, c'est-à-dire un élément Ω' de $\mathfrak{T}$ tel que $\mathbf{P}(\Omega') = 1$.

■ **Proposition 1.22** — *Si* N *est un événement négligeable, et si* A *est un événement tel que* $\mathrm{A} \subset \mathrm{N}$*, alors* A *est négligeable. Si des événements* $\mathrm{N}_1, \mathrm{N}_2, \ldots$*, en quantité dénombrable, sont négligeables, alors* $\mathrm{N}' := \bigcup_{k=1}^{\infty} \mathrm{N}_k$ *est négligeable.*

⟦ Démonstration : Par croissance, $0 \leqslant \mathbf{P}(\mathrm{A}) \leqslant \mathbf{P}(\mathrm{N}) = 0$, donc A est négligeable. Par propriété des tribus, N′ est un événement et, par σ-sous-additivité, $\mathbf{P}(\mathrm{N}') \leqslant \sum_k \mathbf{P}(\mathrm{N}_k) = 0$. ⟧

Bien évidemment, un événement *non vide* peut très bien être de probabilité nulle, ce qui veut dire également qu'un événement de probabilité nulle peut arriver.

Exemple 1.23 Voici, sans justification, quelques exemples d'événements, négligeables ou presque sûrs.



- On munit le carré $\Omega = [0\,;1] \times [0\,;1]$ de la mesure μ de Lebesgue sur $\mathbb{R}^2$, qui est alors une mesure de probabilité. Rappelons que pour toute partie A « suffisamment régulière » de Ω, $\mu(\mathrm{A})$ est l'aire de la partie A. Alors une courbe C (par exemple le cercle d'équation $x^2 - x + y^2 - y + \frac{7}{16} = 0$) est d'aire nulle : $\mathbf{P}(\mathrm{C}) = 0$.
- Sur $\Omega = [0\,;1]$, muni de la mesure de Lebesgue, la partie $\mathbb{Q} \cap [0\,;1]$ constituée uniquement des points rationnels, est de mesure nulle. En effet, c'est une réunion *dénombrable* (ce point est capital !) de singletons, tous de mesure nulle.
- On considère l'espace $\Omega = \{0,1\}^{\mathbb{N}^*}$ des suites $(x_n)_{n\geqslant 1}$ à valeurs dans $\{0,1\}$, décrivant le résultat d'une succession infinie de tirages à *pile* ou *face*, avec une pièce honnête. Alors l'ensemble A des tirages vérifiant la propriété suivante :

  $$\text{la suite des moyennes } \frac{x_1 + \cdots + x_n}{n} \text{ converge vers } \frac{1}{2}$$

  est de probabilité 1. (C'est un résultat qui découle de la loi forte des grands nombres, mais qu'on peut démontrer de manière relativement élémentaire, voir l'exercice 16.1 page 428.)

1.1.f Systèmes complets d'événements

Il est souvent très utile de découper l'univers Ω en événements disjoints, généralement caractérisés par des propriétés numériques⟨6⟩.

Dans la notion de partition, tous les éléments sont non vides ; en probabilité, comme il est parfois difficile de savoir *a priori* si certains éléments du découpage sont vides ou non, on met cette contrainte de côté, et on pose la définition suivante :

Définition 1.24 (Système complet d'événements) — Une famille $(\mathrm{A}_i)_{i\in\mathrm{I}}$, finie ou dénombrable, de parties de Ω est appelée **système complet d'événements** lorsque Ω est la réunion disjointe des A_i :

$$\Omega = \bigsqcup_{i\in\mathrm{I}} \mathrm{A}_i.$$

⟨6⟩. Typiquement, des valeurs ou des intervalles de valeurs prises par une variable aléatoire $\mathrm{X} : \Omega \to \mathbb{R}$. Voir la définition 3.4 page 65 d'un système complet d'événements associé à une variable aléatoire discrète.

Notamment, si $(\mathrm{A}_i)_{i\in\mathrm{I}}$ est un système complet d'événements, alors pour toute réalisation ω du hasard, *un et un seul* des événements A_i est réalisé.

Pour des raisons pratiques⟨7⟩, on s'autorisera à ne pas recouvrir Ω tout entier, mais à en oublier une partie négligeable.

Définition 1.25 (Système quasi-complet d'événements) — Une famille $(\mathrm{A}_i)_{i\in\mathrm{I}}$, finie ou dénombrable, d'événements est appelée **système quasi-complet d'événements** lorsque ses éléments sont incompatibles, ($\mathrm{A}_i \cap \mathrm{A}_j = \varnothing$ si $i \neq j$) et que leur réunion est presque sûre :

$$\mathbf{P}\left(\bigsqcup_{i\in\mathrm{I}} \mathrm{A}_i\right) = 1.$$

Ainsi, lorsque $(\mathrm{A}_i)_{i\in\mathrm{I}}$ est un système quasi-complet d'événements, il existe un événement négligeable N tel que

$$\Omega = \mathrm{N} \sqcup \bigsqcup_{i\in\mathrm{I}} \mathrm{A}_i.$$

Cette notion permet de découper tout événement B en

$$\mathrm{B} = (\mathrm{B}\cap\mathrm{N}) \sqcup \bigsqcup_{i\in\mathrm{I}} \mathrm{B}\cap\mathrm{A}_i$$

puis, en remarquant que $\mathbf{P}(\mathrm{B}\cap\mathrm{N}) \subset \mathbf{P}(\mathrm{N}) = 0$, on déduit la proposition suivante :

■ **Proposition 1.26** — *Si* $(\mathrm{A}_i)_{i\in\mathrm{I}}$ *est un système quasi-complet d'événements, et si* B *est un événement, alors*

$$\mathbf{P}(\mathrm{B}) = \sum_{i\in\mathrm{I}} \mathbf{P}(\mathrm{B}\cap\mathrm{A}_i).$$

1.1.g Formule du crible

La probabilité de la réunion de deux événements non disjoints est simple à calculer :

$$\mathbf{P}(\mathrm{A}\cup\mathrm{B}) = \mathbf{P}(\mathrm{A}) + \mathbf{P}(\mathrm{B}) - \mathbf{P}(\mathrm{A}\cap\mathrm{B})$$

Pour trois événements, la figure ci-dessous permet de se convaincre que

$$\mathbf{P}(\mathrm{A}\cup\mathrm{B}\cup\mathrm{C}) = \mathbf{P}(\mathrm{A})+\mathbf{P}(\mathrm{B})+\mathbf{P}(\mathrm{C})-\mathbf{P}(\mathrm{A}\cap\mathrm{B})-\mathbf{P}(\mathrm{B}\cap\mathrm{C})-\mathbf{P}(\mathrm{C}\cap\mathrm{A})+\mathbf{P}(\mathrm{A}\cap\mathrm{B}\cap\mathrm{C}).$$

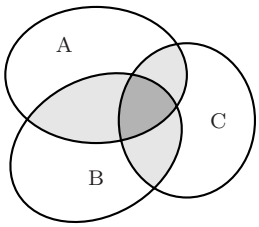


Figure 1.2 — *En gris foncé,* $\mathrm{A}\cap\mathrm{B}\cap\mathrm{C}$. *En gris (clair et foncé),* $\mathrm{A}\cap\mathrm{B}$, $\mathrm{A}\cap\mathrm{C}$ *et* $\mathrm{B}\cap\mathrm{C}$.

⟨7⟩. Notamment lors de l'étude de variables aléatoires qui prendront presque sûrement une valeur finie mais, exceptionnellement, la valeur $+\infty$.

Plus généralement, la probabilité d'une réunion est donnée par la **formule de Poincaré**, ou **du crible**.

Théorème 1.27 (Formule de Poincaré, ou du crible) — *Soit $n \geqslant 1$. Pour tous événements $\mathrm{A}_1, \dots, \mathrm{A}_n$,*

$$\mathbf{P}\left(\bigcup_{i=1}^{n} \mathrm{A}_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n} \mathbf{P}(\mathrm{A}_{i_1} \cap \mathrm{A}_{i_2} \cap \cdots \cap \mathrm{A}_{i_k}).$$

⟦ Démonstration : Par récurrence sur n. Notons $\mathscr{H}_n$ la propriété au rang n (valable pour tous événements $\mathrm{A}_1, \mathrm{A}_2, \dots, \mathrm{A}_n$).

Elle est vraie pour $n = 2$. Soit $n \geqslant 3$, et supposons la propriété vraie jusqu'à un rang $n-1$. Alors

$$\begin{aligned}
\mathbf{P}\left(\bigcup_{i=1}^{n} \mathrm{A}_i\right) &= \mathbf{P}\big((\mathrm{A}_1 \cup \mathrm{A}_2 \cup \cdots \cup \mathrm{A}_{n-1}) \cup \mathrm{A}_n\big) \\
&= \mathbf{P}(\mathrm{A}_1 \cup \mathrm{A}_2 \cup \cdots \cup \mathrm{A}_{n-1}) + \mathbf{P}(\mathrm{A}_n) - \mathbf{P}\big((\mathrm{A}_1 \cup \mathrm{A}_2 \cup \cdots \cup \mathrm{A}_{n-1}) \cap \mathrm{A}_n\big) \\
&= \mathbf{P}(\mathrm{A}_1 \cup \mathrm{A}_2 \cup \cdots \cup \mathrm{A}_{n-1}) + \mathbf{P}(\mathrm{A}_n) - \mathbf{P}\big((\mathrm{A}_1 \cap \mathrm{A}_n) \cup \cdots \cup (\mathrm{A}_{n-1} \cap \mathrm{A}_n)\big) \\
&= \sum_{k=1}^{n-1} (-1)^{k+1} \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n-1} \mathbf{P}(\mathrm{A}_{i_1} \cap \cdots \cap \mathrm{A}_{i_k}) \\
&\quad - \sum_{k=1}^{n-1} (-1)^{k+1} \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n-1} \mathbf{P}(\mathrm{A}_{i_1} \cap \cdots \mathrm{A}_{i_k} \cap \mathrm{A}_n) + \mathbf{P}(\mathrm{A}_n)
\end{aligned}$$

Il ne reste plus qu'à constater que toute suite strictement croissante $(i_1, i_2, \dots, i_k)$ de $[\![1\,;n]\!]$ s'écrit

- soit comme une suite strictement croissante de $[\![1\,;n-1]\!]$,
- soit comme une suite strictement croissante de $[\![1\,;n-1]\!]$ suivie de $i_k = n$,
- soit enfin comme le terme seul n (lorsque $k = 1$).

On obtient alors la formule de $\mathscr{H}_n$, ce qui achève la récurrence. ⟧

Thème 1 : Probabilité que deux entiers soient premiers entre eux

Arithmétique

Pour tout $n \geqslant 1$, on note r_n la probabilité que deux nombres, choisis indépendamment et aléatoirement selon la loi uniforme sur $[\![1\,;n]\!]$, soient premiers entre eux. On souhaite montrer que

$$\lim_{n\to\infty} r_n = \frac{1}{\zeta(2)} = \frac{6}{\pi^2}.$$

Notons

$$\mathrm{A}_n = \big\{(a,b) \in [\![1\,;n]\!]^2 \ ;\ a \wedge b = 1\big\}.$$

Si l'on note $p_1, p_2, \dots, p_k$ les nombres premiers inférieurs ou égaux à n, on note pour $1 \leqslant i \leqslant k$, $\mathrm{V}_i = \{a \in [\![1\,;n]\!] \ ;\ p_i \mid a\}$ et

$$\mathrm{U}_i = \big\{(a,b) \in [\![1\,;n]\!]^2 \ ;\ p_i \mid a \text{ et } p_i \mid b\big\} = \mathrm{V}_i \times \mathrm{V}_i$$

de sorte que $\mathrm{A}_n = [\![1\,;n]\!]^2 \setminus (\mathrm{U}_1 \cup \cdots \cup \mathrm{U}_k) = \bigcap_{i=1}^{k} \mathrm{U}_k^{\mathrm{c}}$.

Or, le nombre de multiples dans $[\![1\,;n]\!]$ d'un entier a est $\lfloor n/a \rfloor$, d'où l'on déduit que

$$\operatorname{Card} \mathrm{V}_i = \left\lfloor \frac{n}{p_i} \right\rfloor \qquad \text{et} \qquad \operatorname{Card} \mathrm{U}_i = \left\lfloor \frac{n}{p_i} \right\rfloor^2$$

Si i et j sont des entiers distincts, on a de même

$$\operatorname{Card} \mathrm{V}_i \cap \mathrm{V}_j = \left\lfloor \frac{n}{p_i\, p_j} \right\rfloor \qquad \text{et} \qquad \operatorname{Card} \mathrm{U}_i \cap \mathrm{U}_j = \left\lfloor \frac{n}{p_i\, p_j} \right\rfloor^2$$

et, plus généralement, pour toute partie J non vide de $[\![1\,;n]\!]$,

$$\operatorname{Card} \left(\bigcap_{i\in \mathrm{I}} \mathrm{U}_i \right) = \left\lfloor \frac{n}{\prod\limits_{i\in \mathrm{I}} p_i} \right\rfloor^2 .$$

On peut alors appliquer la formule du crible, en faisant apparaître la fonction arithmétique μ de Möbius, dont les propriétés sont rappelées en annexe (théorème A.64 page 566 de l'annexe A) :

$$\begin{aligned} \operatorname{Card} \mathrm{A}_n &= \sum_{k=0}^{n} (-1)^k \sum_{i_1<\cdots<i_k} \left\lfloor \frac{n}{\prod_j p_{i_j}} \right\rfloor^2 \\ &= \sum_{k=0}^{n} \sum_{i_1<\cdots<i_k} \mu(p_{i_1}\cdots p_{i_k}) \left\lfloor \frac{n}{\prod_j p_{i_j}} \right\rfloor^2 = \sum_{d=1}^{n} \mu(d) \left\lfloor \frac{n}{d} \right\rfloor^2 \end{aligned}$$

puisqu'à la dernière étape, on ne rajoute que des termes tels que $\mu(d) = 0$, et que l'on n'oublie que des termes vérifiant $\lfloor n/d \rfloor = 0$.
On a donc

$$r_n = \frac{\operatorname{Card} \mathrm{A}_n}{n^2} = \sum_{d=1}^{n} \frac{\mu(d)}{n^2} \left\lfloor \frac{n}{d} \right\rfloor^2 = \sum_{d=1}^{\infty} \psi_d(n)$$

où l'on a noté

$$\begin{aligned} \psi_d \colon \ \mathbb{N} &\longrightarrow \mathbb{R} \\ n &\longmapsto \frac{\mu(d)}{n^2} \left\lfloor \frac{n}{d} \right\rfloor^2 \end{aligned}$$

Or, d'une part,

$$\forall d \in \mathbb{N}^* \qquad \lim_{n\to\infty} \psi_d(n) = \frac{\mu(d)}{d^2},$$

et d'autre part, $\|\psi_d\|_\infty \leqslant \frac{1}{d^2}$ donc la série $\sum \psi_d$ converge normalement.
Le théorème de la double limite permet alors d'affirmer que

$$\lim_{n\to\infty} r_n = \sum_{d=1}^{\infty} \frac{\mu(d)}{d^2}$$

Enfin, le théorème A.64 montre que cette dernière quantité vaut $1/\zeta(2)$, ou encore

$$\lim_{n\to\infty} r_n = \frac{6}{\pi^2} \approx 0{,}608.$$

Remarque 1.28 Bien que l'on soit passé à la limite $n \to \infty$, il est important de voir qu'on n'a pas défini de probabilité sur $\mathbb{N}^* \times \mathbb{N}^*$. (On verra au chapitre 13 que la suite de mesures que l'on a définies sur $\mathbb{N}^* \times \mathbb{N}^*$ n'est pas **tendue** : sa limite n'est pas une mesure de probabilité.)

Remarque 1.29 Robert Matthews a mis a profit cette propriété pour calculer une valeur approchée de π en utilisant, pour générateurs de nombres entiers « aléatoires » des distances angulaires entre les 100 plus brillantes étoiles du ciel, obtenant ainsi 4950 distances ; il obtient une valeur de π à 0,5 % près [24].

1.2 Exemples d'espaces probabilisés

1.2.a Probabilités sur un ensemble fini ou dénombrable

Munir un univers fini $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ d'une probabilité revient à associer, à chaque élément ω_k, un réel $p_k \geqslant 0$, avec pour seule contrainte que $p_1 + \cdots + p_n = 1$.

Si A est une partie quelconque de Ω, on pose

$$\mathbf{P}(\mathrm{A}) := \sum_{k\ ;\ \omega_k \in \mathrm{A}} p_k$$

et cela suffit à définir une probabilité sur $\big(\Omega, \mathcal{P}(\Omega)\big)$ (elle répond bien aux exigences des trois axiomes).

Exemple 1.30 Dans le cas d'un lancer de dé, on peut choisir $\Omega = \{1, 2, 3, 4, 5, 6\}$ et poser $p_k = 1/6$ pour $k = 1, \ldots, 6$; cette situation modélise un dé équilibré. Si l'on se rend compte que, pour un dé donné, les prédictions du modèle ne se réalisent pas, on peut modifier les valeurs des coefficients p_k. Pour déterminer expérimentalement chaque coefficient, une méthode est de lancer un très grand nombre de fois le dé et de calculer des fréquences expérimentales. Des résultats théoriques permettent de donner une estimation du nombre de lancers à effectuer pour obtenir des valeurs suffisamment précises des p_k (voir chapitre 19).

Plus généralement, supposons que Ω soit un ensemble infini, mais *dénombrable* : choisissons une énumération⟨8⟩ de Ω :

$$\Omega = \{\omega_n\ ;\ n \in \mathbb{N}\}.$$

On considérera fréquemment les cas particuliers $\Omega = \mathbb{N}$ ou $\Omega = \mathbb{N}^*$.

Associer une probabilité à chacun des éléments ω de Ω revient donc à se donner une suite $(p_n)_{n\geqslant 0}$ de réels positifs, tels que la série $\sum p_n$ converge (ce qui revient à dire que la famille $(p_n)_{n\geqslant 0}$ est sommable) et

$$\sum_{n=0}^{\infty} p_n = 1.$$

Une telle suite est appelée **germe de probabilité**. On définit l'application $\mathbf{P} : \mathcal{P}(\Omega) \to \mathbb{R}^+$ ainsi : pour toute une partie A de Ω,

$$\mathbf{P}(\mathrm{A}) := \sum_{k\ ;\ \omega_k \in \mathrm{A}} p_k.$$

Dans cette dernière expression, la sommabilité de la famille p_n nous assure de la convergence de la somme restreinte à un sous-ensemble A de Ω ; il est alors élémentaire de vérifier que la fonction $\mathbf{P}$ ainsi définie répond aux exigences des trois axiomes d'une probabilité.

⟨8⟩. Voir le paragraphe A.1 page 548 de l'annexe A.

1.2.b Succession infinie de parties de pile ou face

Dès que l'on veut modéliser entièrement une expérience consistant en une succession infinie de tirages à *pile* ou *face*, l'univers obtenu est nécessairement infini et non dénombrable⟨9⟩ : *a minima*, on prendra $\Omega := \{\mathrm{P},\mathrm{F}\}^{\mathbb{N}^*}$, chaque expérience est une suite indexée par $\mathbb{N}^* = \{1,2,3,\ldots\}$ et à valeurs dans $\{\mathrm{P},\mathrm{F}\}$.

Pour toute réalisation ω du hasard (qu'on appellera *expérience* dans la suite, et qui comprend donc une infinité de tirages), on notera $u_n(\omega)$ son n-ième terme ; ainsi, $\omega = \big(u_n(\omega)\big)_{n\geqslant 1}$.

Il nous faut maintenant construire une tribu $\mathfrak{T}$ sur Ω, ainsi qu'une probabilité sur $(\Omega,\mathfrak{T})$.

- On désire évidemment pouvoir parler de l'ensemble des expériences de la forme « le k-ième lancer renvoie *pile* », que l'on notera

$$\mathrm{P}_k \qquad \text{ou} \qquad [*\ *\ *\ \cdots\ *\ \underset{\substack{\uparrow\\ k}}{\mathrm{P}}\ *\ \cdots\ *\ \cdots]$$

  le symbole « $*$ » jouant le rôle d'un joker. Ainsi, les événement P_k et leur complémentaire F_k, ainsi que les intersections d'un nombre fini ou infini (nécessairement dénombrable) de ces expériences devront être des événements, c'est-à-dire des éléments de la tribu $\mathfrak{T}$.
- On supposera que chaque tirage d'une pièce mène à *pile* avec une même probabilité p, et à *face* avec la probabilité $q := 1-p$. Ainsi, on exigera que $\mathbf{P}(\mathrm{P}_k) = p$ pour tout $k \in \mathbb{N}$.
- On supposera enfin une indépendance des tirages, c'est-à-dire que si $k_1,\ldots,k_n$ sont des indices distincts, S_{k_i} des symboles de $\{\mathrm{P},\mathrm{F}\}$ et s_{k_i} les probabilités correspondantes (p ou q), alors on définit la probabilité

$$\mathbf{P}_0(\mathrm{S}_{k_1}\cap\cdots\cap\mathrm{S}_{k_n}) = s_{k_1}\cdots s_{k_n}.$$

  Les parties de la forme $\mathrm{S}_{k_1}\cap\cdots\cap\mathrm{S}_{k_n}$ sont bien des événements, qu'on appelle des **cylindres**.

  Par convention, on acceptera également le cas $n=0$, le cylindre en question étant alors $[*\ *\ \cdots\ *\cdots]$, c'est-à-dire Ω tout entier.
- Notons $\mathfrak{C}$ l'ensemble constitué des cylindres : il contient Ω, est stable par passage au complémentaire et par réunion finie. C'est donc une algèbre.

  Il est possible de montrer que $\mathbf{P}_0$ est σ-additive sur $\mathfrak{C}$⟨10⟩, c'est-à-dire que si $(\mathrm{C}_n)_{n\geqslant 1}$ est une famille de cylindres disjoints dont la réunion est dans $\mathfrak{C}$, alors

$$\mathbf{P}_0\left(\bigsqcup_n \mathrm{C}_n\right) = \sum_{n=1}^{\infty}\mathbf{P}(\mathrm{C}_n).$$

  Le théorème de Carathéodory (B.13 page 575) montre que l'on peut étendre la fonction $\mathbf{P}_0$ en une probabilité sur la tribu $\mathfrak{T} = \sigma(\mathfrak{C})$ engendrée par les cylindres.

⟨9⟩. On rappelle en annexe l'argument de la diagonale de Cantor permettant de prouver que $\{0,1\}^{\mathbb{N}^*}$ n'est pas dénombrable, voir théorème A.4 page 548.

⟨10⟩. Cette démonstration est assez technique ; on la trouvera dans [2, Annexe B].

- C'est cette probabilité que l'on utilisera chaque fois que l'on aura affaire à une suite infinie de tirages à *pile* ou *face*. Bien que la tribu $\mathfrak{T}$ soit difficile à visualiser, on se contentera de retenir
  - ▷ qu'elle contient les cylindres (dont les probabilités se calculent aisément et conformément à l'intuition),
  - ▷ ainsi que le résultat élémentaire suivant : *si une infinité de symboles est spécifiée, alors la probabilité de l'événement est nulle.*

    En effet, si $(k_n)_{n\geqslant 0}$ est une suite strictement croissantes de symboles et si on note

$$\mathrm{A} := \bigcap_{i=1}^{\infty} \mathrm{S}_{k_i} = [* \cdots * \mathrm{S}_{k_1} * \cdots * \mathrm{S}_{k_2} * \cdots * \mathrm{S}_{k_i} * \cdots * * * \cdots],$$

    alors pour tout $n \geqslant 1$, on a par croissance et en notant $\alpha := \max(p,q) < 1$,

$$0 \leqslant \mathbf{P}(\mathrm{A}) \leqslant \mathbf{P}\left(\bigcap_{i=1}^{n} \mathrm{S}_{k_i}\right) = \prod_{i=1}^{n} s_{k_i} \leqslant \alpha^n \xrightarrow[n\to\infty]{} 0.$$

1.2.c Probabilités sur $\mathbb{R}$

Nous le verrons dès les prochains chapitres, une variable aléatoire X transporte une probabilité $\mathbf{P}$ (définie sur un espace probabilisable $(\Omega, \mathfrak{T})$) sur l'espace probabilisable $\big(\mathbb{R}, \mathfrak{B}(\mathbb{R})\big)$. Aussi l'étude de quelques probabilités classiques sur $\big(\mathbb{R}, \mathfrak{B}(\mathbb{R})\big)$ est-elle particulièrement importante.

Exemple 1.31 (Mesure de Lebesgue sur $[0\,;1[$) On munit l'espace $\Omega = [0\,;1]$ de la tribu des boréliens, c'est-à-dire la tribu engendrée par les intervalles (ou les ouverts, etc) de Ω. On note traditionnellement μ la mesure définie sur cet espace par la fonction de répartition $\mathrm{F}(x) = x$, et on l'appelle **mesure de Lebesgue sur $[0\,;1]$**. Elle vérifie donc

$$\mu\big([a\,;b]\big) = \mu\big([a\,;b[\big) = \mu\big(]a\,;b]\big) = \mu\big(]a\,;b[\big) = b - a.$$

Un résultat classique de la théorie de la mesure montre que c'est une probabilité sur $[0\,;1]$ muni de la tribu des boréliens. C'est la probabilité la plus naturelle qui se puisse imaginer sur $[0\,;1]$, celle qui « ne privilégie aucun point spécifique » ; on l'appelle également la **probabilité uniforme sur $[0\,;1]$**.

C'est également elle qui servira à la construction de variables aléatoires suivant des lois quelconques, aussi bien du point de vue théorique que lors de simulations informatiques (paragraphe § 8.4, page 233).

Exemple 1.32 (Probabilités discrètes sur $\mathbb{R}$) Soit $\{a_n\ ;\ n \in \mathbb{N}\}$ une famille dénombrable de réels distincts, et $\{p_n\ ;\ n \in \mathbb{N}\}$ une famille de réels positifs telle que $\sum_{n=0}^{\infty} p_n = 1$ — c'est-à-dire un **germe de probabilités**. On définit, sur l'espace probabilisable $\big(\mathbb{R}, \mathcal{P}(\mathbb{R})\big)$, la probabilité $\mathbf{P}$ en posant, pour toute partie A de $\mathbb{R}$:

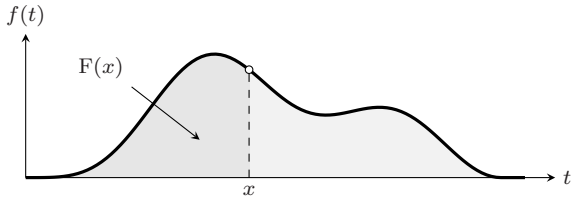
$$\mathbf{P}(\mathrm{A}) = \sum_{k\ ;\ a_k \in \mathrm{A}} p_k,$$

où la somme est définie sur l'ensemble des indices k tels que $a_k \in \mathrm{A}$ (cette somme est bien définie). Cela revient à poser

$$\mathbf{P} = \sum_{k=0}^{\infty} p_k\, \delta_{a_k}.$$

Exemple 1.33 (Probabilités à densité sur $\mathbb{R}$) À l'opposé d'une probabilité finie ou discrète, on trouve la famille des probabilités dites **absolument continues**, ou encore **probabilités à densité** : ce sont les probabilités $\mathbf{P}$ pour lesquelles il existe une fonction $f : \mathbb{R} \to \mathbb{R}^+$ mesurable telle que

$$\forall x \in \mathbb{R} \qquad \mathrm{F}(x) := \mathbf{P}\big(]-\infty\,;x]\big) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t.$$



La fonction f, appelée **densité** de la probabilité $\mathbf{P}$, est à valeurs positives et est égale, en presque tous les réels, à la dérivée de F. Elle vérifie de plus

$$\int_{-\infty}^{+\infty} f(t)\,\mathrm{d}t = \mathbf{P}(\mathbb{R}) = 1.$$

On démontre alors que $\mathbf{P}$ est effectivement une probabilité sur $\big(\mathbb{R}, \mathfrak{B}(\mathbb{R})\big)$, et que pour tout borélien B,

$$\mathbf{P}(\mathrm{B}) = \int_{\mathrm{B}} f = \int_{-\infty}^{+\infty} \mathbf{1}_{\mathrm{B}}(t)\, f(t)\,\mathrm{d}t.$$

1.3 Notion de limsup ou : « infiniment souvent »

Une question très courante est la suivante : quelle est la probabilité que, parmi les événements $\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_n, \ldots$, une infinité d'entre eux se réalise ?

Notons d'abord un résultat élémentaire, qui traduit le fait qu'une partie A de $\mathbb{N}$ est non bornée (ou encore : infinie) si et seulement si $\mathrm{A} \cap [\![n\,; +\infty [\![$ est non vide pour tout entier n :

■ **Lemme 1.34** — *Il y a disjonction de cas :*

(a) *soit une infinité d'événements de la suite* $(\mathrm{A}_n)_{n\geqslant 1}$ *se réalisent,*

(b) *soit il existe un entier* N *tel que, pour tout* $n \geqslant \mathrm{N}$*, l'événement* A_n^c *se réalise.*

On notera $\{\mathrm{A}_n \text{ ; i.s.}\}$ le premier événement, qu'on lira « A_n se réalise infiniment souvent », et $\{\mathrm{A}_n^c \text{ ; àpcr}\}$ le deuxième, qu'on lira « les A_n^c sont tous réalisés à partir d'un certain rang ».

Exemple 1.35 Considérons un espace probabilisé sur lequel existe une suite $(\mathrm{X}_n)_{n\geqslant 1}$ de fonctions à valeurs réelles. En mettant de côté toute la problématique liée à la tribu (les parties utiles de Ω sont supposées être des événements), on cherche à décrire l'ensemble B des ω tels que la suite $\big(\mathrm{X}_n(\omega)\big)$ ne tende pas vers 0.

Pour un ω fixé, il y a équivalence entre les énoncés

(a) la suite $\big(\mathrm{X}_n(\omega)\big)$ ne tend pas vers 0 ;

(b) $\neg\big[\forall \varepsilon > 0 \quad \exists \mathrm{N} \in \mathbb{N}^* \quad \forall n \geqslant \mathrm{N} \qquad \big|\mathrm{X}_n(\omega)\big| \leqslant \varepsilon\big]$;

(c) $\neg\Big[\forall \varepsilon \in \mathbb{Q}_+^* \quad \exists \mathrm{N} \in \mathbb{N}^* \quad \forall n \geqslant \mathrm{N} \qquad \big|\mathrm{X}_n(\omega)\big| \leqslant \varepsilon\Big]$;

(d) $\neg\Big[\forall \varepsilon \in \mathbb{Q}_+^* \quad \big\{\big|\mathrm{X}_n(\omega)\big| \leqslant \varepsilon \text{ ; àpcr}\big\}\Big]$;

(e) $\exists \varepsilon \in \mathbb{Q}_+^* \quad \big\{\big|\mathrm{X}_n(\omega)\big| > \varepsilon \text{ ; i.s.}\big\}$.

Autrement dit, l'événement « $\big(\mathrm{X}_n(\omega)\big)$ ne tend pas vers 0 » s'écrit

$$\bigcup_{\varepsilon \in \mathbb{Q}_+^*} \big\{\big|\mathrm{X}_n(\omega)\big| > \varepsilon \text{ ; i.s.}\big\}.$$

Le passage par $\mathbb{Q}_+^*$ n'a d'autre but que d'écrire notre événement sous la forme d'une réunion *dénombrable.*

L'événement $\{A_n^c\ ;\ \text{ÀPCR}\}$ s'écrit naturellement

$$\underset{\uparrow}{\bigcup_{N=1}^{\infty}}\ \underset{\uparrow}{\bigcap_{n=N}^{\infty}}\ A_n^c$$
$$\text{« } \exists N \text{ » « } \forall n \geqslant N \text{ »}$$

et donc, en passant au complémentaire, on a le théorème suivant :

Théorème 1.36 — *Soit $(A_n)_{n\geqslant 1}$ une suite d'événements. Alors*

$$\{A_n\ ;\ \text{I.S.}\} = \bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty} A_n.$$

Notons $S_n = \bigcup_{n=N}^{\infty} A_n = \sup_{n\geqslant N} A_n$. Alors la suite $(S_n)_{n\geqslant 1}$ est décroissante, donc

$$\{A_n\ ;\ \text{I.S.}\} = \lim_{N} \downarrow S_N = \lim_{N} \downarrow \sup_{n\geqslant N} A_n,$$

ce qui justifie la définition :

Définition 1.37 — L'ensemble $\{A_n\ ;\ \text{I.S.}\}$ est également appelé **limsup des A_n**, et noté

$$\{A_n\ ;\ \text{I.S.}\} = \limsup A_n.$$

Revenons maintenant à

$$\{A_n\ ;\ \text{ÀPCR}\} = \{A_n^c\ ;\ \text{I.S.}\}^c = \bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty} A_n.$$

Si l'on note $I_n = \bigcap_{n=N}^{\infty} A_n = \inf_{n\geqslant N} A_n$, la suite $(I_n)_{n\geqslant 1}$ est croissante et

$$\{A_n\ ;\ \text{ÀPCR}\} = \lim_{N} \uparrow I_N = \lim_{N} \uparrow \inf_{n\geqslant N} A_n,$$

ce qui justifie la définition suivante :

Définition 1.38 — L'ensemble $\{A_n\ ;\ \text{ÀPCR}\}$ est également appelé **liminf des A_n**, et noté

$$\{A_n\ ;\ \text{ÀPCR}\} = \liminf A_n.$$

Enfin, pour les applications probabilistes, il est indispensable de remarquer que $\{A_n\ ;\ \text{I.S.}\}$ et $\{A_n\ ;\ \text{ÀPCR}\}$ sont des événements :

Théorème 1.39 — *Soit $\mathfrak{T}$ une tribu. Alors $\mathfrak{T}$ est stable par passage à la limite supérieure et à la limite inférieure : si $(A_n)_{n\geqslant 0}$ est une suite d'éléments de $\mathfrak{T}$, alors $\{A_n\ ;\ \text{I.S.}\}$ et $\{A_n\ ;\ \text{ÀPCR}\}$ sont également des éléments de $\mathfrak{T}$.*

⟦ DÉMONSTRATION : Avec les notations précédentes, les éléments $S_N = \sup_{n\geqslant N} A_n$ sont dans la tribu (stabilité par union dénombrable), donc $\{A_n\ ;\ \text{I.S.}\}$ l'est également (stabilité par intersection dénombrable). Ce raisonnement appliqué à la suite $(A_n^c)_{n\geqslant 1}$ montre, par passage au complémentaire, que $\{A_n\ ;\ \text{ÀPCR}\}$ est encore dans $\mathfrak{T}$. ⟧

Thème 2 : Problème de la ruine du joueur

On considère deux joueurs, désignés par A et B, qui s'affrontent dans une partie de *pile* ou *face*, avec les règles suivantes :

- Le joueur A dispose au départ de n euros, le joueur B de $\mathrm{N}-n$ euros.
- Le jeu se déroule en une succession de manches opposant A et B ; à chaque manche, A gagne avec la probabilité $p \in \,]0\,;1[$, et perd avec la probabilité $q := 1-p$. Le perdant d'une manche donne un euro au vainqueur, et le jeu s'arrête avec la ruine d'un joueur.

On note a_n la probabilité que A gagne le jeu lorsque sa fortune de départ est de n euros (la fortune totale des deux joueurs étant fixée à N).

- Si le joueur A gagne la première manche (avec probabilité p), alors il se retrouve immédiatement dans la position d'un joueur de fortune $n+1$, qui gagnera le jeu avec probabilité a_{n+1}.
- Si au contraire il perd cette première manche (avec probabilité q), alors il se retrouve immédiatement dans la position d'un joueur de fortune $n-1$, qui gagnera le jeu avec probabilité a_{n-1}.

Cela nous fournit la relation

$$a_n = q\,a_{n-1} + p\,a_{n+1}.$$

Cette relation de récurrence a pour polynôme caractéristique

$$\mathrm{Q} = p\mathrm{X}^2 - \mathrm{X} + q,$$

dont on doit déterminer les racines.
Si l'on ne voit pas de racine évidente, on peut toujours utiliser la formule générale, et les racines sont

$$\frac{1 \pm \sqrt{1-4p(1-p)}}{2p} = \frac{1 \pm \sqrt{1-4p+4p^2}}{2p} = \frac{1 \pm \sqrt{(1-2p)^2}}{2p} = \frac{1 \pm |1-2p|}{2p}.$$

Le terme en valeur absolue peut être positif ou négatif, mais les racines seront dans tous les cas

$$\frac{1 \pm (1-2p)}{2p}$$

c'est-à-dire

$$1 \qquad \text{et} \qquad \lambda := \frac{1-p}{p} = \frac{q}{p}.$$

Remarque 1.40 Il paraît quand même judicieux, surtout dans un problème de probabilités, de vérifier si par hasard 1 n'est pas racine évidente du polynôme. Les relations coefficients-racines fournissent alors immédiatement la seconde racine.

On distingue ensuite deux cas.

- Si $p = \frac{1}{2}$, $\mathrm{Q} = \frac{1}{2}(\mathrm{X}-1)^2$ possède une racine double. On vérifie que $a_{n+1} - a_n$ est une quantité constante, ou bien on cherche la solution sous forme $u_n = (\alpha n + \beta)\cdot 1^n$ et on trouve, avec les conditions $a_0 = 0$ et $a_{\mathrm{N}} = 1$:

  $$\forall n \in [\![0\,;\mathrm{N}]\!] \qquad u_n = \frac{n}{\mathrm{N}}.$$

  Ainsi, un joueur qui aurait les trois quarts de la fortune initiale pourrait ruiner l'autre avec une probabilité 0,75.

► Si $p \neq \frac{1}{2}$, on recherche les solutions sous la forme $a_n = \alpha + \beta\,\lambda^n$ ce qui, avec les conditions $a_0 = 0$ et $a_{\mathrm{N}} = 1$ donne

$$a_n = \frac{1-\lambda^n}{1-\lambda^{\mathrm{N}}}.$$

On remarque alors que, lorsque $p > 1/2$, alors $0 \leqslant \lambda < 1$ et la suite est concave. Au contraire, lorsque $p < 1/2$, $\lambda > 1$ et la suite est convexe (voir figure 1.3).
(On peut également vérifier que, si $p \to 1/2$, alors $\lambda \to 1$ et un développement limité redonne le résultat précédent.)

Remarque 1.41 (La partie finit presque sûrement) Le même raisonnement donne, en notant b_n la probabilité que B gagne sachant que sa fortune de départ est n, la même expression en changeant p en $1-p$ ou, ce qui revient au même, λ en λ^{-1} :

$$b_n = \frac{1-\lambda^{-n}}{1-\lambda^{-\mathrm{N}}}$$

et alors, en réduisant au même dénominateur : $a_n + b_{\mathrm{N}-n} = 1$.
Ainsi, avec probabilité 1, le joueur A gagne ou le joueur B gagne. En d'autres termes, comme dans le cas d'une infinité de tirages *pile* lors de lancers de pièce, le cas d'une partie infinie est possible, mais de probabilité nulle.

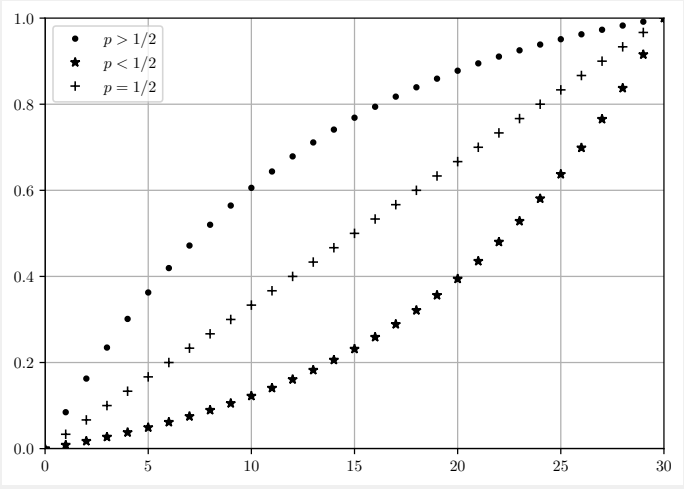


Figure 1.3 — *Ruine du joueur. On affiche la probabilité que* A *gagne la partie (c'est-à-dire que* B *soit ruiné le premier) en fonction de sa fortune n initiale, la fortune totale de* A *et de* B *étant constante égale à* $\mathrm{N} = 30$. *Les valeurs prises par p sur le graphe sont* $p = 0{,}48$, $p = 0,5$ *et* $p = 0{,}52$.

EXERCICES

♦ ***Exercice 1.1 (Problème du facteur et fonctions croissantes)*** (⋆⋆)

i) Un facteur distribue q prospectus indiscernables dans p boîtes aux lettres, sans contrainte. Quelle est le nombre de possibilités de distribution ?

ii) Combien y a-t-il de fonctions $f : [\![0\,;p]\!] \to [\![0\,;q]\!]$, croissantes, et vérifiant $f(0) = 0$ et $f(p) = q$?

iii) Combien y a-t-il de telles fonctions *strictement* croissantes ?

♦ ***Exercice 1.2*** (⋆) Au loto d'un village, n personnes participent. Les lots sont nombreux : on note r leur nombre et on suppose que $r \geqslant n$. Pour la distribution, on choisit r fois au hasard (avec probabilité uniforme et de manière indépendante pour chaque tirage) un des participants.

i) Donner en fonction de r et n la probabilité qu'une personne P_1 ne reçoive aucun lot, puis celle que k $(1 \leqslant k \leqslant n-1)$ personnes $\mathrm{P}_1, \mathrm{P}_2, \ldots, \mathrm{P}_k$ ne reçoivent aucun lot.

ii) En utilisant la formule du crible, en déduire la probabilité que chacune des n personnes reçoive au moins un lot.

On remarquera que la probabilité ainsi calculée est également *la probabilité qu'une application de* $[\![1\,;r]\!]$ *dans* $[\![1\,;n]\!]$ *soit surjective.*

♦ ***Exercice 1.3 (Continuité monotone en $\varnothing$)*** (⋆⋆) Soit $(\Omega, \mathfrak{T})$ un espace probabilisable. Soit $\mathbf{P} : \mathfrak{T} \to \mathbb{R}^+$ une application additive et telle que $\mathbf{P}(\Omega) = 1$.

Montrer qu'il y a équivalence entre

(a) $\mathbf{P}$ est σ-additive

(b) pour toute suite décroissante $(\mathrm{A}_n)_{n\geqslant 0}$ d'événements de limite $\varnothing$, c'est-à-dire telle que $\bigcap_n \mathrm{A}_n = \varnothing$, on a $\lim \mathbf{P}(\mathrm{A}_n) = 0$.

♦ ***Exercice 1.4 (Tribus dénombrables ?)*** (☕) Soit $\mathfrak{T}$ une tribu au plus dénombrable sur un ensemble E. Le but de l'exercice est de montrer que $\mathfrak{T}$ est nécessairement finie. Ainsi, toute tribu est finie ou indénombrable.

i) **Un résultat préliminaire :** Soit $\{\mathrm{F}_n\ ;\ n \in \mathbb{N}\}$ une partition dénombrable d'un ensemble A. On note $\mathfrak{B}$ la plus petite tribu contenant tous les F_n. Montrer que $\mathfrak{B}$ n'est pas dénombrable.

ii) Soit $a \in \mathrm{E}$. Montrer que

$$\mathrm{C}(a) := \bigcap_{\mathrm{A}\in\mathfrak{T}\ ;\ a\in\mathrm{A}} \mathrm{A}$$

est élément de $\mathfrak{T}$.

iii) On définit sur E la relation $\sim$ par

$$a \sim b \iff b \in \mathrm{C}(a).$$

Montrer que c'est une relation d'équivalence ; en décrire les classes d'équivalence.

iv) Montrer que la plus petite tribu contenant $\mathfrak{P} = \big\{\mathrm{C}(a)\ ;\ a \in \mathrm{E}\big\}$ est $\mathfrak{T}$; en déduire que $\mathfrak{T}$ est finie.

♦ ***Exercice 1.5*** (○) Soit $(\mathrm{A}_n)_{n\geqslant 0}$ une suite d'événements. On note $\mathbf{1}_{\mathrm{A}_n}$ la fonction indicatrice de A_n, c'est-à-dire la fonction définie par $\mathbf{1}_{\mathrm{A}_n}(\omega) = 1$ si $\omega \in \mathrm{A}_n$ et $\mathbf{1}_{\mathrm{A}_n}(\omega) = 0$ si $\omega \notin \mathrm{A}_n$. Montrer que :

i) $\liminf \mathrm{A}_n \subset \limsup \mathrm{A}_n$.

ii) Si $(\mathrm{A}_n)_{n\geqslant 0}$ est croissante, alors $\liminf\limits_n \mathrm{A}_n = \limsup\limits_n \mathrm{A}_n = \lim\limits_n \uparrow \mathrm{A}_n$.

iii) Si $(\mathrm{A}_n)_{n\geqslant 0}$ est décroissante, alors $\liminf\limits_n \mathrm{A}_n = \limsup\limits_n \mathrm{A}_n = \lim\limits_n \downarrow \mathrm{A}_n$.

iv) $\mathbf{1}_{\liminf \mathrm{A}_n} = \liminf \mathbf{1}_{\mathrm{A}_n}$ et $\mathbf{1}_{\limsup \mathrm{A}_n} = \limsup \mathbf{1}_{\mathrm{A}_n}$.

Cette dernière propriété relie les dénominations de « limite supérieure » et « limite inférieure » vues ici pour des familles d'ensembles, à celles concernant les suites réelles.

♦ ***Exercice 1.6 (Inégalités de Fatou)*** (⋆) Soit $(\mathrm{A}_n)_{n\geqslant 0}$ une suite d'événements de $(\Omega, \mathfrak{T}, \mathbf{P})$. Montrer que

$$\mathbf{P}(\liminf \mathrm{A}_n) \leqslant \liminf \mathbf{P}(\mathrm{A}_n) \leqslant \limsup \mathbf{P}(\mathrm{A}_n) \leqslant \mathbf{P}(\limsup \mathrm{A}_n).$$

En déduire que, si la suite $(\mathrm{A}_n)_{n\geqslant 0}$ converge (au sens où $\liminf \mathrm{A}_n = \limsup \mathrm{A}_n$) alors la suite de terme général $\mathbf{P}(\mathrm{A}_n)$ converge et $\lim\limits_{n\to\infty} \mathbf{P}(\mathrm{A}_n) = \mathbf{P}(\lim\limits_{n\to\infty} \mathrm{A}_n)$.

♦ ***Exercice 1.7*** (⋆⋆) On note $\mathcal{A}$ l'ensemble des parties de A de $\mathbb{N}$ admettant une densité, c'est-à-dire telles que la quantité

$$\delta(\mathrm{A}) := \lim_{n\to+\infty} \frac{\big|\mathrm{A}\cap[\![1\,;n]\!]\,\big|}{n}$$

existe.

Montrer que $\mathcal{A}$ n'est pas égale à $\mathcal{P}(\mathbb{N})$.

L'ensemble $\mathcal{A}$ est-il une tribu ?

Indications

◊ *Indication 1.3 :* On se rappellera que la σ-additivité est équivalente en réalité à la continuité monotone croissante et à la décroissante.

◊ *Indication 1.4 :*

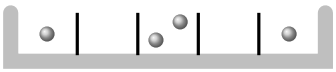
i) $\mathfrak{B}$ est exactement l'ensemble des parties de la forme $\bigcup_{n\in\mathrm{I}} \mathrm{F}_n$ pour I parcourant $\mathcal{P}(\mathbb{N})$.

ii) Il suffit de prouver que l'intersection est dénombrable.

SOLUTIONS

♦ **Solution 1.1**

i) Dessinons les p boîtes aux lettres au moyen d'un cadre (en gris) et de $p-1$ cloisons (en noir). On cherche le nombre de façon de distribuer q prospectus indiscernables, représentés sous la forme de boules, tous identique. Voici une configuration possible, pour $p=5$ et $q=4$.



Convenons maintenant de coder une telle configuration en alignant successivement $p+q-1$ symboles, qui peuvent être une cloison « | » ou une boule « ● ». Par exemple, la configuration précédente s'écrit

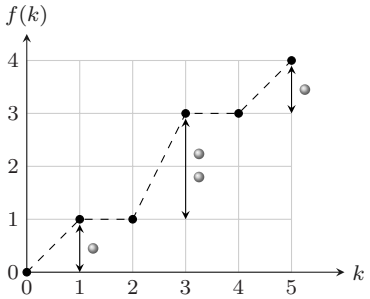
Il est clair que le passage d'une représentation à l'autre est bien inversible. Déterminer le nombre de configurations des prospectus reviens donc à déterminer la manière de distribuer q symboles « ● » parmi les $p+q-1$ symboles.

Le nombre de possibilités de distribution est donc $\binom{p+q-1}{q}$.

(On trouvera une application de ce résultat à l'exercice 7.6 page 206.)

ii) Pour déterminer une telle fonction, on doit décider, pour chacune des valeurs $1,\dots,p$, la valeur du saut $f(k)-f(k-1)$, qui est un nombre positif ou nul ; la somme de tous les sauts doit être égale à $f(p)-f(0)=q$.

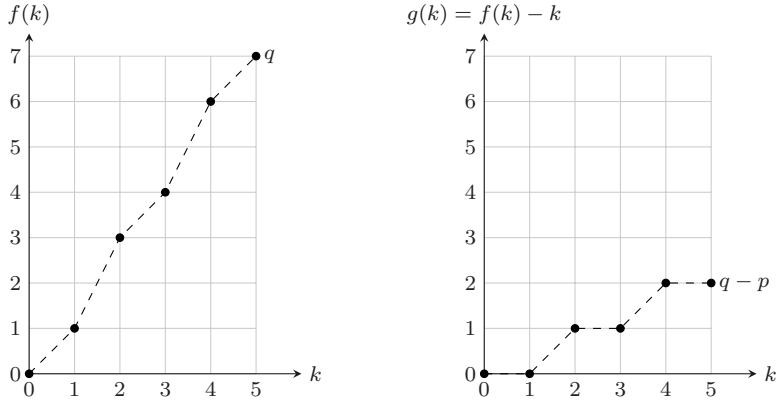
Ainsi, l'exemple précédent code la fonction suivante :



On est donc ramenés au problème de distribuer q « sauts unités » sur p emplacements (chaque emplacement pouvant recevoir plusieurs sauts).

Il y a donc $\binom{p+q-1}{q}$ fonctions possibles.

iii) Une fonction $f : [\![0\,;p]\!] \to [\![0\,;q]\!]$ est strictement croissante si et seulement si $g := f - \mathrm{Id}$ est croissante (on doit alors évidemment avoir $q \geqslant p$) ; la fonction g vérifie $g(0) = 0$ et $g(p) = q-p$. Plus précisément, l'application $f \mapsto f - \mathrm{Id}$ établit une bijection de l'ensemble des fonctions strictement croissantes vérifiant $f(0) = 0$ et $f(p) = q$, et l'ensemble des fonctions croissantes vérifiant $g(0) = 0$ et $g(p) = q - p$.



Le résultat précédent montre donc qu'il existe $\binom{q-1}{q-p} = \binom{q-1}{p-1}$ fonctions strictement croissantes.

♦ **Solution 1.2** Puisque l'on ne travaille ici qu'avec un nombre fini de possibilités, on peut modéliser l'expérience avec un univers Ω fini, muni de la tribu triviale $\mathfrak{T} = \mathcal{P}(\Omega)$.

i) La personne P_1 ne reçoit aucun lot lorsqu'elle est successivement perdante à chacun des r tirages, donc avec la probabilité $(1 - 1/n)^r$. De même, $\mathrm{P}_1, \ldots, \mathrm{P}_k$ ne reçoivent aucun lot au premier tirage avec la probabilité $1 - k/n$, et à aucun des r tirages successifs avec le probabilité $(1 - k/n)^r$.

ii) Pour $i = 1, \ldots, n$, notons B_i l'événement

$$\mathrm{B}_i : \text{« la personne } \mathrm{P}_i \text{ n'a pas eu de lot »}.$$

Alors, l'événement A : « chacune des n personnes reçoit un lot » s'écrit

$$\overline{\mathrm{A}} = \mathrm{B}_1 \cup \mathrm{B}_2 \cup \cdots \cup \mathrm{B}_n.$$

Les événements B_i n'étant pas incompatibles, on invoque la formule du crible :

$$\mathbf{P}(\overline{\mathrm{A}}) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} \mathbf{P}\big(\mathrm{B}_{i_1} \cap \cdots \cap \mathrm{B}_{i_k}\big).$$

Pour une valeur de k fixée et pour des indices $i_1, \ldots, i_k$ fixés également, puisque tous les participants jouent des rôles symétriques,

$$\mathbf{P}\big(\mathrm{B}_{i_1} \cap \cdots \cap \mathrm{B}_{i_k}\big) = \mathbf{P}\big(\mathrm{B}_1 \cap \cdots \cap \mathrm{B}_k\big) = \left(\frac{n-k}{n}\right)^r.$$

De plus, toujours pour k fixé, il y a $\binom{n}{k}$ manières de choisir les indices ordonnés (en effet, il y a autant de manières de choisir k indices ordonnés dans le sens croissant que de manières de choisir un ensemble de k indices distincts). On en déduit que

$$\mathbf{P}(\overline{\mathrm{A}}) = \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \left(\frac{n-k}{n}\right)^r.$$

Par passage au complémentaire,

$$\mathbf{P}(\mathrm{A}) = 1 - \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \left(\frac{n-k}{n}\right)^r = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \left(\frac{n-k}{n}\right)^r.$$

♦ **Solution 1.3** Supposons la continuité monotone en $\varnothing$. Soit $(\mathrm{B}_n)_{n\geqslant 0}$ une suite croissante d'éléments de $\mathfrak{T}$, et posons

$$\mathrm{B} = \lim_{n\to\infty} \uparrow \mathrm{B}_n = \bigcup_{n=0}^{\infty} \mathrm{B}_n.$$

On peut alors écrire pour tout $n \in \mathbb{N}$:

$$\mathrm{B} = (\mathrm{B} - \mathrm{B}_n) \sqcup \mathrm{B}_n.$$

La suite de terme général $(\mathrm{B} - \mathrm{B}_n)$ est décroissante et de limite $\varnothing$ et la propriété d'additivité permet d'écrire que $\mathbf{P}(\mathrm{B}) = \mathbf{P}(\mathrm{B} - \mathrm{B}_n) + \mathbf{P}(\mathrm{B}_n)$, donc

$$\mathbf{P}(\mathrm{B}_n) = \mathbf{P}(\mathrm{B}) - \mathbf{P}(\mathrm{B} - \mathrm{B}_n) \xrightarrow[n\to\infty]{} \mathbf{P}(\mathrm{B}).$$

Réciproquement, la σ-additivité implique la continuité monotone décroissante, qui implique évidemment la propriété voulue.

♦ **Solution 1.4** .

i) $\mathfrak{B}$ est exactement l'ensemble des parties de la forme $\bigcup_{n\in \mathrm{I}} \mathrm{F}_n$ pour I parcourant $\mathcal{P}(\mathbb{N})$. Ainsi, $\mathfrak{B}$ est en bijection avec $\mathcal{P}(\mathbb{N})$ qui est non dénombrable.

ii) L'élément $\mathrm{A} = \mathrm{E}$ est dans $\mathfrak{T}$ et $a \in \mathrm{A}$, donc $\mathrm{C}(a)$ est non vide.
Ensuite, $\mathfrak{T}$ étant dénombrable, $\mathrm{C}(a)$ est une intersection dénombrable d'éléments de $\mathfrak{T}$, donc est élément de $\mathfrak{T}$.
$\mathrm{C}(a)$ est donc « le plus petit événement contenant a ».

iii) On vérifie rapidement que $a \sim b$ si et seulement si $\mathrm{C}(a) = \mathrm{C}(b)$. La classe de a est exactement $\mathrm{C}(a)$.

iv) L'ensemble $\mathfrak{P}$ est une partition de E, dénombrable en tant que partie de $\mathfrak{T}$.
La tribu engendrée $\sigma(\mathfrak{P})$ est évidemment incluse dans $\mathfrak{T}$.
Réciproquement, soit $\mathrm{A} \in \mathfrak{T}$, alors $\mathrm{A} = \bigcup_{a\in \mathrm{A}} \mathrm{C}(a) \in \sigma(\mathfrak{P})$.
Ainsi, $\sigma(\mathfrak{P}) = \mathfrak{T}$.
Or si $\mathfrak{T}$ était dénombrable mais infinie, alors $\mathfrak{P}$ serait également de cardinal infini, donc d'après la première question, $\mathfrak{T} = \sigma(\mathfrak{P})$ serait non dénombrable : contradiction.

♦ **Solution 1.6** Notons $\mathrm{B}_n := \bigcap_{k=n}^{\infty} \mathrm{A}_k$ et $\mathrm{C}_n := \bigcup_{k=n}^{\infty} \mathrm{A}_k$, et montrons chacune des inégalités :

- Tout d'abord, la suite de terme général $\bigcup_{n=0}^{\mathrm{N}} \mathrm{B}_n$ est croissante, et sa réunion est, par définition, $\liminf \mathrm{A}_n$:

$$\bigcup_{n=0}^{\mathrm{N}} \mathrm{B}_n \uparrow \mathrm{B} := \liminf_n \mathrm{A}_n.$$

Par continuité monotone,

$$\mathbf{P}(\mathrm{B}_n) = \mathbf{P}\left(\bigcap_{k=n}^{\infty} \mathrm{A}_k\right) \xrightarrow[n\to\infty]{} \mathbf{P}(\mathrm{B}),$$

or $\mathrm{B}_n \subset \mathrm{A}_k$ pour $k \geqslant n$, donc $\mathbf{P}(\mathrm{B}_n) \leqslant \mathbf{P}(\mathrm{A}_k)$ pour tout $k \geqslant n$, c'est-à-dire que

$$\mathbf{P}(\mathrm{B}_n) \leqslant \inf_{k\geqslant n} \mathbf{P}(\mathrm{A}_k).$$

Par continuité monotone croissante, on peut passer à la limite pour conclure que

$$\mathbf{P}(\mathrm{B}) \leqslant \liminf \mathbf{P}(\mathrm{A}_n).$$

- L'inégalité $\liminf \mathbf{P}(\mathrm{A}_n) \leqslant \limsup \mathbf{P}(\mathrm{A}_n)$ est immédiate : si l'on note $s_n = \sup_{k\geqslant n} \mathbf{P}(\mathrm{A}_k)$ et $i_n = \inf_{k\geqslant n} \mathbf{P}(\mathrm{A}_k)$, alors $i_n \leqslant s_n$ et

$$\liminf \mathbf{P}(\mathrm{A}_n) := \lim_{n\to\infty} i_n \leqslant \lim_{n\to\infty} s_n =: \limsup \mathbf{P}(\mathrm{A}_n).$$

- De même, $\bigcap_{n=0}^{\mathrm{N}} \mathrm{C}_n \downarrow \limsup_n \mathrm{A}_n = \mathrm{C}$, donc

$$\mathbf{P}(\mathrm{C}_n) = \mathbf{P}\left(\bigcup_{k=n}^{\infty} \mathrm{A}_k\right) \xrightarrow[n\to\infty]{} \mathbf{P}(\mathrm{C}),$$

or $\mathrm{C}_n \supset \mathrm{A}_k$ pour $k \geqslant n$, donc $\mathbf{P}(\mathrm{C}_n) \geqslant \mathbf{P}(\mathrm{C}_k)$ et finalement, $\mathbf{P}(\mathrm{C}_n) \geqslant \sup_{k\geqslant n} \mathbf{P}(\mathrm{A}_k)$, puis on passe à la limite.

Ainsi, on a bien prouvé

$$\mathbf{P}(\liminf \mathrm{A}_n) \leqslant \liminf \liminf \mathbf{P}(\mathrm{A}_n) \leqslant \limsup \mathbf{P}(\mathrm{A}_n) \leqslant \mathbf{P}(\limsup \mathrm{A}_n).$$

Enfin si $(\mathrm{A}_n)_{n\geqslant 0}$ converge, alors le premier et le dernier membre sont égaux, donc les membres intermédiaires aussi, ce qui montre que

$$\lim_{n\to\infty} \mathbf{P}(\mathrm{A}_n) = \mathbf{P}\left(\lim_{n\to\infty} \mathrm{A}_n\right).$$

♦ **Solution 1.7** On fabrique une partie $\mathcal{A}$ en réunissant une collection d'intervalles entiers de taille rapidement croissante, les morceaux étant trois fois plus grands chaque fois, alternativement vides et pleins :

$$\mathrm{A} = \{1\} \sqcup \bigsqcup_{n=1}^{\infty} \rrbracket 4^{2n-1}\,;4^{2n} \rrbracket .$$

Alors

$$\frac{\left|\mathrm{A} \cap \llbracket 1\,;4^{2n} \rrbracket\right|}{n} \geqslant \frac{3}{4} \qquad \text{et} \qquad \frac{\left|\mathrm{A} \cap \llbracket 1\,;4^{2n-1} \rrbracket\right|}{n} \leqslant \frac{1}{4}$$

ce qui montre que $\mathrm{A} \notin \mathcal{A}$.

Comme de plus $\mathrm{A}_n := \rrbracket 4^{2n-1}\,;4^{2n} \rrbracket \in \mathcal{A}$ pour tout n, on en déduit que $\mathcal{A}$ n'est pas stable par réunion dénombrable, donc n'est pas tribu.

Chapitre 2

Conditionnement et indépendance d'événements

Le Calcul des Probabilités est l'art de calculer des probabilités inconnues en fonction de probabilités connues.

Dans ce chapitre

2.1 Probabilités conditionnelles

2.1.a Intuition de définition

Soit $(\Omega, \mathfrak{T}, \mathbf{P})$ un espace probabilisé, et soit A un événement fixé, de probabilité strictement positive.

Si l'on cherche à modéliser intuitivement le calcul de la probabilité d'un événement B, *sachant que* A *est réalisé*[1], on peut faire l'expérience de pensée suivante :

- on répète un « grand nombre » n de fois une même expérience ;
- on oublie toutes celles pour lesquelles A n'est pas réalisé, et on note a le nombre d'expériences retenues ;
- on compte alors, parmi ces a expériences, le nombre b de fois que B est réalisé.

On accède intuitivement à la *probabilité de* B *sachant que* A *est réalisé* par le rapport b/a (nombre de succès sur nombre total d'expériences retenues). En écrivant

$$\frac{b}{a} = \frac{b/n}{a/n},$$

on conçoit que la manière « naturelle » de définir la probabilité de B sachant A est de décréter qu'elle est égale $\mathbf{P}(\mathrm{A} \cap \mathrm{B})/\mathbf{P}(\mathrm{A})$.

Définition 2.1 — Soit A un événement de probabilité non nulle : $\mathbf{P}(\mathrm{A}) > 0$. Pour tout événement B, on définit la **probabilité conditionnelle de B sachant A** comme le réel, noté $\mathbf{P}_{\mathrm{A}}(\mathrm{B})$ ou $\mathbf{P}(\mathrm{B} \,|\, \mathrm{A})$, valant

$$\mathbf{P}_{\mathrm{A}}(\mathrm{B}) = \mathbf{P}\left(\mathrm{B} \,|\, \mathrm{A}\right) := \frac{\mathbf{P}\left(\mathrm{A} \cap \mathrm{B}\right)}{\mathbf{P}(\mathrm{A})}.$$

Théorème 2.2 — *L'application*

$$\begin{aligned} \mathbf{P}_{\mathrm{A}} : \ \mathfrak{T} &\longrightarrow [0\,;1]\,, \\ \mathrm{B} &\longmapsto \mathbf{P}(\mathrm{B} \,|\, \mathrm{A}) \end{aligned}$$

est une probabilité sur $(\Omega, \mathfrak{T})$ *; on l'appelle* **probabilité conditionnelle sachant A**, *ou encore* **probabilité conditionnée par A**.

⟦ Démonstration : Vérifions explicitement que les trois axiomes définissant une probabilité sont bien respectés.

- $\mathbf{P}_{\mathrm{A}}$ est bien à valeurs positives.
- De plus
$$\mathbf{P}_{\mathrm{A}}(\Omega) = \frac{\mathbf{P}(\Omega \cap \mathrm{A})}{\mathbf{P}(\mathrm{A})} = \frac{\mathbf{P}(\mathrm{A})}{\mathbf{P}(\mathrm{A})} = 1.$$
- Enfin, si $(\mathrm{B}_n)_{n \geqslant 0}$ est une suite d'éléments deux à deux disjoints de $\mathfrak{T}$, alors les $\mathrm{B}_n \cap \mathrm{A}$

[1]. Prenons l'exemple d'un tirage de dé, on peut considérer l'événement B « le 6 sort », dont on cherche la probabilité sachant que l'événement A, définie par « un nombre pair sort », est réalisé.

sont également deux à deux disjoints, donc

$$\mathbf{P}_{\mathrm{A}}\left(\bigsqcup_{n=0}^{\infty}\mathrm{B}_n\right)=\frac{\mathbf{P}\left[\left(\bigsqcup_{n=0}^{\infty}\mathrm{B}_n\right)\cap\mathrm{A}\right]}{\mathbf{P}(\mathrm{A})}=\frac{\mathbf{P}\left(\bigsqcup_{n=0}^{\infty}\mathrm{B}_n\cap\mathrm{A}\right)}{\mathbf{P}(\mathrm{A})}$$
$$=\sum_{n=0}^{\infty}\frac{\mathbf{P}(\mathrm{B}_n\cap\mathrm{A})}{\mathbf{P}(\mathrm{A})}=\sum_{n=0}^{\infty}\mathbf{P}_{\mathrm{A}}(\mathrm{B}_n).$$ ⟧

2.1.b Formule des probabilités composées

Soit A est un événement tel que $\mathbf{P}(\mathrm{A})>0$; pour tout événement B, la formule

$$\mathbf{P}(\mathrm{A}\cap\mathrm{B})=\mathbf{P}(\mathrm{A})\cdot\mathbf{P}(\mathrm{B}\,|\,\mathrm{A}) \tag{2.1}$$

reçoit couramment une interprétation en terme de *connaissance croissante des événements* : pour calculer la probabilité que les événements A et B aient lieu tous deux,

- on calcule d'abord la probabilité de réalisation de A,
- puis la probabilité de réalisation de B sachant que A est réalisé.

La formule (2.1) se généralise à une famille finie d'événements :

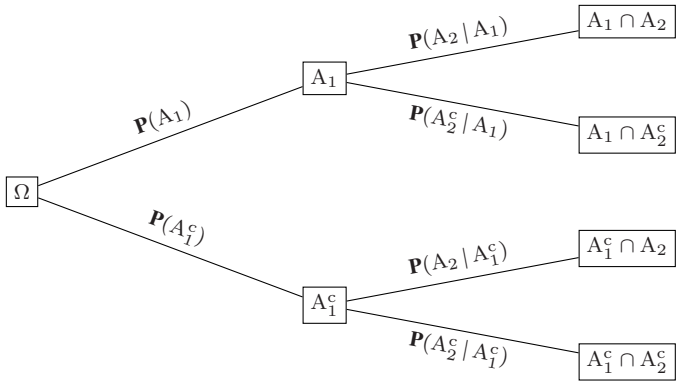
Théorème 2.3 (Formule des probabilités composées) — *Soient $n\geqslant 2$ un entier et soient $\mathrm{A}_1,\mathrm{A}_2,\dots,\mathrm{A}_n$ des événements ; on suppose que $\mathrm{A}=\mathrm{A}_1\cap\mathrm{A}_2\cap\cdots\cap\mathrm{A}_{n-1}$ est de probabilité non nulle. Alors*

$$\mathbf{P}(\mathrm{A}_1\cap\mathrm{A}_2\cap\cdots\cap\mathrm{A}_n)=\mathbf{P}(\mathrm{A}_1)\cdot\mathbf{P}_{\mathrm{A}_1}(\mathrm{A}_2)\cdot\mathbf{P}_{\mathrm{A}_1\cap\mathrm{A}_2}(\mathrm{A}_3)\cdots\mathbf{P}_{\mathrm{A}_1\cap\cdots\cap\mathrm{A}_{n-1}}(\mathrm{A}_n).$$

⟦ Démonstration : On démontre cette propriété par récurrence sur n.

- **Le cas $n=2$** correspond à l'équation (2.1).
- **Soit $n\geqslant 2$ et supposons la propriété démontrée au rang n.** Soient $\mathrm{A}_1,\dots,\mathrm{A}_{n+1}$ des événements tels que $\mathbf{P}(\mathrm{A}_1\cap\cdots\cap\mathrm{A}_n)>0$. Notons $\mathrm{A}'=\mathrm{A}_1\cap\cdots\cap\mathrm{A}_n$; puisque $\mathrm{A}_1\cap\cdots\cap\mathrm{A}_{n-1}$ est de probabilité non nulle (car au moins égale à $\mathbf{P}(\mathrm{A}')>0$), la propriété de récurrence permet d'écrire
$$\mathbf{P}(\mathrm{A}')=\mathbf{P}(\mathrm{A}_1)\cdot\mathbf{P}_{\mathrm{A}_1}(\mathrm{A}_2)\cdot\mathbf{P}_{\mathrm{A}_1\cap\mathrm{A}_2}(\mathrm{A}_3)\cdots\mathbf{P}_{\mathrm{A}_1\cap\cdots\cap\mathrm{A}_{n-1}}(\mathrm{A}_n).$$
Enfin, la relation à l'ordre 2 s'écrit
$$\mathbf{P}(\mathrm{A}'\cap\mathrm{A}_{n+1})=\mathbf{P}(\mathrm{A}')\cdot\mathbf{P}_{\mathrm{A}'}(\mathrm{A}_{n+1}),$$
ce qui est la formule voulue à l'ordre $n+1$. ⟧

Le calcul par probabilités composées peut agréablement se résumer sous la forme d'un arbre de probabilités.



2.1.c Formule des probabilités totales

Théorème 2.4 (Formule des probabilités totales) — *Soit* $(\mathrm{A}_i)_{i\in\mathrm{I}}$ *une famille, finie ou dénombrable, formant un système complet d'événements. Soit* B *un événement. On suppose que* $\mathbf{P}(\mathrm{A}_i) > 0$ *pour tout* $i \in \mathrm{I}$. *Alors*

$$\mathbf{P}(\mathrm{B}) = \sum_{i\in\mathrm{I}} \mathbf{P}(\mathrm{A}_i)\,\mathbf{P}(\mathrm{B}\,|\,\mathrm{A}_i).$$

⟦ Démonstration : Puisque $(\mathrm{A}_i)_{i\in\mathrm{I}}$ est une partition de Ω, on peut écrire

$$\mathbf{P}(\mathrm{B}) = \mathbf{P}(\mathrm{B}\cap\Omega) = \mathbf{P}\Big(\mathrm{B}\cap\bigsqcup_{i\in\mathrm{I}}\mathrm{A}_i\Big) = \mathbf{P}\Big(\bigsqcup_{i\in\mathrm{I}}\mathrm{B}\cap\mathrm{A}_i\Big).$$

puis, par σ-additivité,

$$\mathbf{P}(\mathrm{B}) = \sum_{i\in\mathrm{I}}\mathbf{P}(\mathrm{B}\cap\mathrm{A}_i) = \sum_{i\in\mathrm{I}}\mathbf{P}(\mathrm{A}_i)\,\mathbf{P}(\mathrm{B}\,|\,\mathrm{A}_i). \qquad ⟧$$

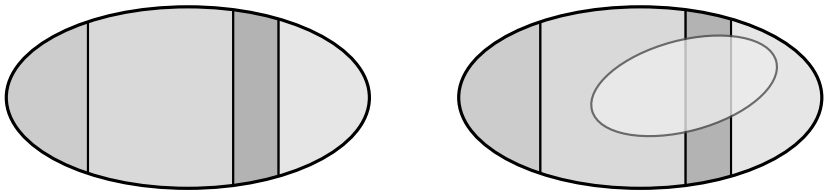


Figure 2.1 — À gauche : *un système complet d'événements.* À droite : *l'événement* B *se découpe selon le système précédent.*

Remarque 2.5 Il est courant, pour les étudiants et étudiantes qui commencent l'étude des probabilités, d'hésiter lorsque l'on doit interpréter les probabilités apparaissant dans un énoncé « en français » : est-ce $\mathbf{P}(\mathrm{A}\cap\mathrm{B})$ ou $\mathbf{P}(\mathrm{B}\,|\,\mathrm{A})$ que l'on connaît ? Le plus souvent, si la question se pose, on retiendra que c'est la probabilité conditionnelle $\mathbf{P}(\mathrm{B}\,|\,\mathrm{A})$ qui est la donnée « naturelle ».

On considère l'énoncé suivant.

> *La probabilité qu'il pleuve un matin de semaine est* $p = 1/5$. *Lorsqu'il pleut, la probabilité que je prenne le bus est* $9/10$, *sinon je vais faire cours à pieds. Lorsqu'il ne pleut pas, la probabilité que je prenne le bus est* $1/10$. *Quelle est la probabilité que je prenne le bus un matin donné ?*

La première étape dans la résolution d'un tel problème est de formaliser les données. On note A l'événement « il pleut », et A^{c} l'événement « il ne pleut pas » ; l'univers Ω est bien sûr égal à l'union disjointe $\mathrm{A}\cup\mathrm{A}^{\mathrm{c}}$. Notons B l'événement « je prends le bus ». La phrase « Lorsqu'il pleut, la probabilité que je prenne le bus est 9/10 » est donc àinterpréter comme une *probabilité conditionnelle*, c'est-à-dire comme $\mathbf{P}(\mathrm{B}\,|\,\mathrm{A})$ (« sachant qu'il pleut »), et non comme la probabilité d'une conjonction $\mathbf{P}(\mathrm{B}\cap\mathrm{A})$ (« il pleut *et* je prends le bus »). On a ainsi

$$\mathbf{P}(\mathrm{B}\,|\,\mathrm{A}) = \frac{9}{10} \qquad \text{et} \qquad \mathbf{P}(\mathrm{B}\,|\,\mathrm{A}^{\mathrm{c}}) = \frac{1}{10}.$$

La formule des probabilités totales donne alors

$$\mathbf{P}(\mathrm{B}) = \mathbf{P}(\mathrm{A})\cdot\mathbf{P}(\mathrm{B}\,|\,\mathrm{A}) + \mathbf{P}(\mathrm{A}^{\mathrm{c}})\cdot\mathbf{P}(\mathrm{B}\,|\,\mathrm{A}^{\mathrm{c}}) = \frac{1}{5}\cdot\frac{9}{10} + \frac{4}{5}\cdot\frac{1}{10} = \frac{13}{50}.$$

2.1.d Théorème de Bayes

La relation $\mathbf{P}(\mathrm{A}\cap\mathrm{B}) = \mathbf{P}(\mathrm{A})\cdot\mathbf{P}(\mathrm{B}\,|\,\mathrm{A})$ introduit une dissymétrie dans la manière d'envisager les événements A et B. Dans le membre de droite, A est considéré comme une « cause » de l'événement B, ou du moins comme un événement « antérieur » à B.

La probabilité de voir se réaliser les événements A et B s'obtient en évaluant tout d'abord la probabilité d'obtenir l'événement A puis, sachant que A est réalisé, la probabilité d'obtenir l'événement B.

Rétablissons la symétrie, en écrivant la double égalité

$$\mathbf{P}(\mathrm{A} \cap \mathrm{B}) = \mathbf{P}(\mathrm{A}) \cdot \mathbf{P}(\mathrm{B} \,|\, \mathrm{A}) = \mathbf{P}(\mathrm{B}) \cdot \mathbf{P}(\mathrm{A} \,|\, \mathrm{B}),$$

valable dès que A et B sont des événements de probabilité non nulle. Cela permet d'en déduire le résultat élémentaire suivant, connu sous le nom de **théorème de Bayes**⟨2⟩, que l'on considère parfois comme une *formule de probabilité des causes (connaissant les conséquences)*, ou de *renversement de chronologie*, puisqu'il permet de renverser le sens de la causalité ou d'antériorité⟨3⟩ entre les événements.

Théorème 2.6 (Théorème de Bayes) — *Soient* A *et* B *deux événements de probabilité non nulle, alors*

$$\mathbf{P}(\mathrm{A} \,|\, \mathrm{B}) = \frac{\mathbf{P}(\mathrm{A}) \cdot \mathbf{P}(\mathrm{B} \,|\, \mathrm{A})}{\mathbf{P}(\mathrm{B})}.$$

Remarque 2.7 Le théorème de Bayes se généralise sous la forme suivante. On se donne un système quasi-complet d'événements $(\mathrm{A}_i)_{i\in\mathrm{I}}$, c'est-à-dire qu'il existe un événement négligeable N tel que

$$\Omega = \mathrm{N} \sqcup \left(\bigsqcup_{i\in\mathrm{I}} \mathrm{A}_i \right)$$

et on peut donc supposer naturellement que $\mathbf{P}(\mathrm{A}_i) > 0$ pour tout $i \in \mathrm{I}$.

Supposons de plus que $\mathbf{P}(\mathrm{B}) > 0$.

Choisissons un indice p particulier ; le théorème précédent permet de mettre en relation les quantités $\mathbf{P}(\mathrm{A}_p \,|\, \mathrm{B})$ et $\mathbf{P}(\mathrm{B} \,|\, \mathrm{A}_p)$:

$$\mathbf{P}(\mathrm{A}_p \,|\, \mathrm{B}) = \frac{\mathbf{P}(\mathrm{A}_p) \cdot \mathbf{P}(\mathrm{B} \,|\, \mathrm{A}_p)}{\mathbf{P}(\mathrm{B})}.$$

En écrivant B sous la forme $\mathrm{B} = (\mathrm{B} \cap \mathrm{N}) \sqcup \left(\bigsqcup_i \mathrm{B} \cap \mathrm{A}_i \right)$, on obtient :

$$\forall p \in \mathrm{I} \qquad \mathbf{P}(\mathrm{A}_p \,|\, \mathrm{B}) = \frac{\mathbf{P}(\mathrm{A}_p) \cdot \mathbf{P}(\mathrm{B} \,|\, \mathrm{A}_p)}{\sum\limits_{i\in\mathrm{I}} \mathbf{P}(\mathrm{A}_i) \cdot \mathbf{P}(\mathrm{B} \,|\, \mathrm{A}_i)}.$$

2.2 Notions d'indépendance

2.2.a Indépendance de deux événements

D'un point de vue intuitif, dire qu'un événement B est indépendant d'un événement « antérieur » A, c'est dire que le calcul (ou la mesure expérimentale) de la probabilité de B, conditionné par la connaissance de la réalisation (ou non) de l'événement A, conduit au même résultat que le calcul absolu de cette même probabilité, c'est-à-dire

$$\mathbf{P}(\mathrm{B} \,|\, \mathrm{A}) = \mathbf{P}(\mathrm{B} \,|\, \mathrm{A}^{\mathrm{c}}) = \mathbf{P}(\mathrm{B}).$$

⟨2⟩. Thomas Bayes (1702–1761), prêtre et théologien protestant anglais, étudia les mathématiques durant son temps libre, notamment la théorie des probabilités [3]. Cependant, il fut peu publié de son vivant, et ne fut pas reconnu de ses pairs. Son *Essay towards solving a problem in the doctrine of chances* fut publié, de manière posthume, en 1764, dans les comptes-rendus de la *Royal Society*. Il défendit également les bases théoriques de la dérivation (la « *theorie des fluxions* ») d'Isaac Newton.

⟨3⟩. Ce qui évite de parler de causalité, le terme étant sémantiquement très chargé aussi bien dans les domaines scientifiques que dans la vie quotidienne.

Multipliant cette égalité par la probabilité de A, on obtient la formule, plus symétrique,

$$\mathbf{P}(\mathrm{B}\cap\mathrm{A}) = \mathbf{P}(\mathrm{B})\,\mathbf{P}(\mathrm{A}).$$

qui a deux avantages :

- elle met sur le même plan les événements A et B ;
- elle ne fait pas appel à une probabilité conditionnelle — le cas $\mathbf{P}(\mathrm{A}) = 0$ étant problématique.

Définition 2.8 — Deux événements A et B d'un espace probabilisé $(\Omega, \mathfrak{T}, \mathbf{P})$ sont **indépendants** (ou : **stochastiquement indépendants**) s'ils vérifient

$$\mathbf{P}(\mathrm{A}\cap\mathrm{B}) = \mathbf{P}(\mathrm{A})\cdot\mathbf{P}(\mathrm{B}).$$

Remarque 2.9 L'indépendance de deux événements est une propriété d'espace probabilisé (elle dépend de la probabilité **P**), tandis que l'incompatibilité est une propriété ensembliste (elle ne dépend que de la définition de A et de B dans Ω).

■ **Proposition 2.10** — *Soient* A *et* B *deux événements de* $(\Omega, \mathfrak{T}, \mathbf{P})$.

i) Si $\mathbf{P}(\mathrm{A}) > 0$, *alors* A *et* B *sont indépendants si et seulement si* $\mathbf{P}(\mathrm{B}\,|\,\mathrm{A}) = \mathbf{P}(\mathrm{B})$.

ii) Les événements Ω *et* ∅ *sont indépendants de tout autre événement.*

iii) A *est indépendant de lui-même si et seulement si il est de probabilité* 0 *ou* 1.

iv) Si A *et* B *sont indépendants, alors* A *et* B^c *le sont, ainsi bien sûr que* A^c *et* B, *et que* A^c *et* B^c.

v) Si A *est négligeable ou presque certain, alors* A *est indépendant de* B.

⟦ Démonstration : Les deux premières propriétés sont immédiates. Pour la troisième, A est indépendant de lui-même si et seulement si $\mathbf{P}(\mathrm{A}\cap\mathrm{A}) = \mathbf{P}(\mathrm{A})^2$, c'est-à-dire $\mathbf{P}(\mathrm{A}) = \mathbf{P}(\mathrm{A})^2$, ce qui équivaut à $\mathbf{P}(\mathrm{A}) \in \{0, 1\}$. Pour la quatrième propriété, supposons A et B indépendants, alors

$$\begin{aligned}\mathbf{P}(\mathrm{A}\cap\mathrm{B}^{\mathrm{c}}) = \mathbf{P}\big(\mathrm{A}\cap(\Omega-\mathrm{B})\big) &= \mathbf{P}\big(\mathrm{A}-(\mathrm{A}\cap\mathrm{B})\big) = \mathbf{P}(\mathrm{A}) - \mathbf{P}(\mathrm{A}\cap\mathrm{B})\\ &= \mathbf{P}(\mathrm{A}) - \mathbf{P}(\mathrm{A})\,\mathbf{P}(\mathrm{B}) = \mathbf{P}(\mathrm{A})\cdot\big[1-\mathbf{P}(\mathrm{B})\big] = \mathbf{P}(\mathrm{A})\cdot\mathbf{P}(\mathrm{B}^{\mathrm{c}}).\end{aligned}$$

Enfin, pour la dernière propriété, il suffit de remarquer que, si $\mathbf{P}(\mathrm{A}) = 0$, alors $\mathbf{P}(\mathrm{A}\cap\mathrm{B}) \leqslant \mathbf{P}(\mathrm{A})$ donc $\mathbf{P}(\mathrm{A}\cap\mathrm{B}) = 0 = \mathbf{P}(\mathrm{A})\,\mathbf{P}(\mathrm{B})$. Si A est presque certain, alors A^{c} est négligeable, donc indépendant de B, et on conclut avec *iv*). ⟧

2.2.b Indépendance (mutuelle) d'une famille d'événements

Définition 2.11 — Des événements $\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_n$ sont dit **(mutuellement) indépendants** si, pour tout entier $p \leqslant n$ et pour tout p-uplet d'indices $(i_1, \ldots, i_p)$ deux à deux distincts, on a

$$\mathbf{P}(\mathrm{A}_{i_1}\cap\mathrm{A}_{i_2}\cap\cdots\cap\mathrm{A}_{i_p}) = \mathbf{P}(\mathrm{A}_{i_1})\,\mathbf{P}(\mathrm{A}_{i_2})\cdots\mathbf{P}(\mathrm{A}_{i_p}). \tag{2.2}$$

Plus généralement, soit I un ensemble, fini ou infini, d'indices, et soit $(\mathrm{A}_i)_{i\in\mathrm{I}}$ une famille d'événements. On dit que ces événements sont **mutuellement indépendants**

si, pour toute partie *finie* $\mathrm{J} \subset \mathrm{I}$ non vide, on a

$$\mathbf{P}\Big(\bigcap_{j\in\mathrm{J}} \mathrm{A}_j\Big) = \prod_{j\in\mathrm{J}} \mathbf{P}(\mathrm{A}_j).$$

Remarque 2.12 On peut également adopter le point de vue suivant : on définit des événements $\mathrm{B}_1, \mathrm{B}_2, \ldots, \mathrm{B}_n$ tels que pour chaque indice k, on ait $\mathrm{B}_k = \mathrm{A}_k$ ou $\mathrm{B}_k = \Omega$. Les événements $\mathrm{A}_1, \ldots, \mathrm{A}_n$ sont alors indépendants si et seulement si, pour tout choix des B_k,

$$\mathbf{P}(\mathrm{B}_1 \cap \mathrm{B}_2 \cap \cdots \cap \mathrm{B}_n) = \mathbf{P}(\mathrm{B}_1)\,\mathbf{P}(\mathrm{B}_2)\cdots\mathbf{P}(\mathrm{B}_n).$$

L'indépendance mutuelle implique l'indépendance deux à deux (il suffit de prendre une partie J à deux éléments). La réciproque n'est pas vraie, et l'exemple canonique suivant doit être connu :

Exemple 2.13 Considérons un lancer de deux dés et les événements

- A : « la somme des deux dés est paire » ;
- B : « le premier dé est pair » ;
- C : « le deuxième dé est pair ».

Ces événements sont bien deux à deux indépendants, puisque

$$\mathbf{P}(\mathrm{A}) = \mathbf{P}(\mathrm{B}) = \mathbf{P}(\mathrm{C}) = \tfrac{1}{2} \qquad \text{et} \qquad \mathbf{P}(\mathrm{A}\cap\mathrm{B}) = \mathbf{P}(\mathrm{B}\cap\mathrm{C}) = \mathbf{P}(\mathrm{C}\cap\mathrm{A}) = \tfrac{1}{4},$$

mais

$$\mathbf{P}(\mathrm{A}\cap\mathrm{B}\cap\mathrm{C}) = \mathbf{P}(\mathrm{B}\cap\mathrm{C}) = \tfrac{1}{4} \qquad \text{et} \qquad \mathbf{P}(\mathrm{A})\,\mathbf{P}(\mathrm{B})\,\mathbf{P}(\mathrm{C}) = \tfrac{1}{8},$$

ce qui prouve qu'ils ne sont pas mutuellement indépendants.

La propriété suivante montre que l'on préserve l'indépendance d'une famille en passant au complémentaire pour un nombre arbitraire d'éléments de la famille :

■ **Proposition 2.14** — *Soit $(\mathrm{A}_i)_{i\in\mathrm{I}}$ une famille quelconque d'événements mutuellement indépendants. On considère une application $\varepsilon : \mathrm{I} \to \{-1, +1\}$, et on pose*

$$\mathrm{B}_i = \begin{cases} \mathrm{A}_i & \text{si } \varepsilon_i = +1 \\ \mathrm{A}_i^c & \text{si } \varepsilon_i = -1. \end{cases}$$

Alors $(\mathrm{B}_i)_{i\in\mathrm{I}}$ est une famille d'événements mutuellement indépendants.

⟦ Démonstration : Par définition, il suffit de le prouver pour une famille finie (la définition ne fait intervenir *que* des sous-familles finies). Il suffit alors simplement de remarquer que si $\mathrm{A}_1, \ldots, \mathrm{A}_n$ sont indépendants, et si i est un indice particulier, alors

$$\begin{aligned}
\mathbf{P}(\mathrm{A}_1 \cap \cdots \cap \mathrm{A}_i^c \cap \cdots \cap \mathrm{A}_n) &= \mathbf{P}(\mathrm{A}_1 \cap \cdots \cap (\Omega - \mathrm{A}_i) \cap \cdots \cap \mathrm{A}_n) \\
&= \mathbf{P}(\mathrm{A}_1 \cap \cdots \cap \Omega \cap \cdots \cap \mathrm{A}_n) - \mathbf{P}(\mathrm{A}_1 \cap \cdots \cap \mathrm{A}_i \cap \cdots \cap \mathrm{A}_n) \\
&= \prod_{k\neq i} \mathbf{P}(\mathrm{A}_k) - \prod_{k} \mathbf{P}(\mathrm{A}_k) \\
&= \big[1 - \mathbf{P}(\mathrm{A}_i)\big] \prod_{k\neq i} \mathbf{P}(\mathrm{A}_k) = \mathbf{P}(\mathrm{A}_1)\cdots\mathbf{P}(\mathrm{A}_k^c)\cdots\mathbf{P}(\mathrm{A}_n).
\end{aligned}$$

De proche en proche, on peut donc passer au complémentaire pour autant d'éléments que l'on veut. ⟧

Exemple 2.15 (Indépendance et fonction indicatrice d'Euler) ✐ *Arithmétique*

Soit $n \geqslant 1$ un entier. On note $p_1, \ldots, p_q$ ses diviseurs premiers distincts. On munit l'univers $\Omega = \{1, 2, \ldots, n\}$ de la probabilité uniforme.

Pour tout entier $k \in [\![1\,;q]\!]$, on note

$$\mathrm{B}_k = \{\omega \in \Omega\ ;\ p_k \mid \omega\}\,.$$

Puisque p_k un diviseur de n, on en déduit que B_k est un ensemble des multiples de p_k, au nombre de n/p_k précisément :

$$\mathrm{B}_k = \left\{p_k, 2p_k, \dots, \frac{n}{p_k}\,p_k\right\}.$$

Notamment,

$$\mathbf{P}(\mathrm{B}_k) = \frac{1}{n}\frac{n}{p_k} = \frac{1}{p_k}.$$

On peut maintenant montrer que les événements $\mathrm{B}_1, \dots, \mathrm{B}_q$ sont mutuellement indépendants.

Soit $r \in [\![1\,;q]\!]$ et soient $k_1, \dots, k_r$ des indices distincts de $[\![1\,;q]\!]$. L'événement $\mathrm{B}_{k_1} \cap \dots \cap \mathrm{B}_{k_r}$ est l'ensemble des entiers $\omega \in [\![1\,;n]\!]$ tels que ω est divisible à la fois par $p_{k_1}, \dots, p_{k_r}$, ce qui équivaut à « ω est multiple de $p_{k_1} \dots p_{k_r}$ ». De la même manière, Le nombre de multiples de $p_{k_1} \dots p_{k_r}$ dans $[\![1\,;n]\!]$ est exactement $n/p_{k_1} \dots p_{k_r}$, donc

$$\mathbf{P}\left(\mathrm{B}_{k_1} \cap \dots \cap \mathrm{B}_{k_r}\right) = \frac{1}{p_{k_1} \dots p_{k_r}} = \prod_{i=1}^{r} \mathbf{P}(\mathrm{B}_{k_i}).$$

Ainsi, $\mathrm{B}_1, \dots, \mathrm{B}_q$ sont mutuellement indépendants.

En voici une application classique. L'**indicatrice d'Euler** est la fonction $\varphi : \mathbb{N}^* \to \mathbb{N}$ définie par

$$\varphi(n) = \mathrm{Card}\left\{\omega \in [\![0\,;n-1]\!]\ ;\ \omega \wedge n = 1\right\}$$

(voir § A.11, page 566).

Notons

$$\mathrm{B} := \{\omega \in [\![1\,;n]\!]\ ;\ \omega \wedge n = 1\}\,.$$

Un entier $1 \leqslant \omega \leqslant n$ est dans D si et seulement si il n'est multiple ni de p_1, ni de p_2,..., ni de p_q. Ainsi, $\mathrm{D} = \mathrm{B}_1^{\mathrm{c}} \cap \dots \cap \mathrm{B}_q^{\mathrm{c}}$ et, par indépendance mutuelle des $\mathrm{B}_k^{\mathrm{c}}$,

$$\mathbf{P}(\mathrm{D}) = \prod_{i=1}^{q} \left(1 - \mathbf{P}(\mathrm{B}_{k_i})\right) = \prod_{i=1}^{q} \left(1 - \frac{1}{p_i}\right).$$

Puisque $\mathbf{P}(\mathrm{D}) = \mathrm{Card}(\mathrm{D})/n$, on en déduit que

$$\varphi(n) = n\,\mathbf{P}(\mathrm{D}) = n \prod_{i=1}^{q} \left(1 - \frac{1}{p_i}\right).$$

L'indépendance d'une famille d'événements n'utilise que des sous-familles finies. Cependant, on utilisera très souvent une formule pour une famille dénombrable, ou une suite, d'événements :

Théorème 2.16 — *Si $(\mathrm{A}_n)_{n\geqslant 1}$ est une suite d'événements indépendants,*

$$\mathbf{P}\left(\bigcap_{n=1}^{\infty} \mathrm{A}_n\right) = \prod_{n=1}^{\infty} \mathbf{P}(\mathrm{A}_n).$$

⟦ Démonstration : Par continuité décroissante monotone

$$\mathbf{P}\left(\bigcap_{n=1}^{\infty} \mathrm{A}_n\right) = \lim_{n\to\infty} \mathbf{P}\left(\bigcap_{k=1}^{n} \mathrm{A}_k\right) = \lim_{n\to\infty} \prod_{k=1}^{n} \mathbf{P}\left(\mathrm{A}_k\right) =: \prod_{k=1}^{\infty} \mathbf{P}(\mathrm{A}_k). \qquad ⟧$$

On remarquera que le second membre est toujours correctement défini, puisque la suite des produits partiels est positive et décroissante. Ce produit infini peut, en revanche, être nul⟨4⟩.

⟨4⟩. La théorie générale des produits infinis traite ce cas comme un cas de divergence, mais cela est sans incidence pour nous.

2.2.c Indépendance de tribus (∗)

Généralisons encore la notion d'indépendance.

Définition 2.17 — Deux tribus $\mathfrak{T}_1$ et $\mathfrak{T}_2$ sont dites **indépendantes** si, quels que soient les événements $\mathrm{A}_1 \in \mathfrak{T}_1$ et $\mathrm{A}_2 \in \mathfrak{T}_2$, ces événements sont indépendants :

$$\mathbf{P}(\mathrm{A}_1 \cap \mathrm{A}_2) = \mathbf{P}(\mathrm{A}_1)\,\mathbf{P}(\mathrm{A}_2).$$

Soient $\mathfrak{C}_1, \mathfrak{C}_2, \dots, \mathfrak{C}_n$ des sous-tribus de $\mathfrak{T}$[5]. On dit que ces tribus sont **indépendantes** lorsque, pour tout choix d'éléments C_k de $\mathfrak{C}_k$ $(k = 1, \dots, n)$, les événements $\mathrm{C}_1, \mathrm{C}_2, \dots, \mathrm{C}_n$ sont indépendants.
Une famille $(\mathfrak{C}_i)_{i\in\mathrm{I}}$ de tribus est indépendante lorsque, quels que soient l'entier $n \geqslant 1$ et les indices $i_1, i_2, \dots, i_n$, les tribus $\mathfrak{C}_{i_1}, \dots, \mathfrak{C}_{i_n}$ le sont.

En pratique, lorsque des tribus sont engendrées par des « π-systèmes », c'est-à-dire d'un ensemble de parties de Ω, stable par intersections finies, on peut se limiter à vérifier l'indépendance des événements de ces π-systèmes :

■ **Lemme 2.18 (Indépendance et π-systèmes)** — *Soient $n \geqslant 1$ un entier et soient $\mathfrak{A}_1, \mathfrak{A}_2, \dots, \mathfrak{A}_n$ des sous-tribus de $\mathfrak{T}$. On suppose que chacune des tribus $\mathfrak{A}_k$ est engendrée par un π-système $\mathfrak{C}_k$: $\mathfrak{A}_k = \sigma(\mathfrak{C}_k)$. Si $\mathfrak{C}_1, \mathfrak{C}_2, \dots, \mathfrak{C}_n$ sont indépendantes, c'est-à-dire si, pour tout choix de $(\mathrm{C}_1, \dots, \mathrm{C}_n) \in \mathfrak{C}_1 \times \cdots \times \mathfrak{C}_n$, on a*

$$\mathbf{P}(\mathrm{C}_1 \cap \mathrm{C}_2 \cap \cdots \cap \mathrm{C}_n) = \mathbf{P}(\mathrm{C}_1)\,\mathbf{P}(\mathrm{C}_2)\,\dots\,\mathbf{P}(\mathrm{C}_n),$$

alors les tribus $\mathfrak{A}_1, \mathfrak{A}_2, \dots, \mathfrak{A}_n$ sont indépendantes.

Pour mesurer l'importance de ce théorème, rappelons que la tribu des boréliens de $\mathbb{R}$ (qui est immense) est engendrée par le (relativement petit) π-système

$$\mathfrak{C} = \Big\{]-\infty\,;a] \ ;\ a \in \mathbb{R} \Big\}.$$

Exemple 2.19 On considère l'univers $\Omega = \{\mathrm{P},\mathrm{F}\}^2$, muni de la tribu $\mathfrak{T} = \mathcal{P}(\Omega)$ et de la probabilité uniforme. Cet univers permet de modeliser deux tirages successifs à *pile* ou *face*.

On note

$$\mathfrak{T}_1 = \Big\{ \varnothing, \big\{(\mathrm{P},\mathrm{P}),(\mathrm{P},\mathrm{F})\big\}, \big\{(\mathrm{F},\mathrm{P}),(\mathrm{F},\mathrm{F})\big\}, \Omega \Big\}$$

et

$$\mathfrak{T}_2 = \Big\{ \varnothing, \big\{(\mathrm{P},\mathrm{P}),(\mathrm{F},\mathrm{P})\big\}, \big\{(\mathrm{P},\mathrm{F}),(\mathrm{F},\mathrm{F})\big\}, \Omega \Big\}.$$

On peut vérifier explicitement que ce sont des tribus, et qu'elles sont indépendantes : par exemple

$$\mathbf{P}\big(\big\{(\mathrm{P},\mathrm{P}),(\mathrm{P},\mathrm{F})\big\}\big) = \frac{1}{2} \qquad \mathbf{P}\big(\big\{(\mathrm{P},\mathrm{F}),(\mathrm{F},\mathrm{F})\big\}\big) = \frac{1}{2}$$

et

$$\mathbf{P}\big(\big\{(\mathrm{P},\mathrm{P}),(\mathrm{P},\mathrm{F})\big\} \cap \big\{(\mathrm{P},\mathrm{F}),(\mathrm{F},\mathrm{F})\big\}\big) = \mathbf{P}\big(\big\{(\mathrm{P},\mathrm{F})\big\}\big) = \frac{1}{4}$$

et les autres calculs sont du même type.

Si l'on note (en anticipant sur les chapitres concernant les variables aléatoires) X_1 la variable aléatoire donnant le résultat du premier tirage, et définie par $\mathrm{X}_1\big((a,b)\big) = a$, et X_2 la variable aléatoire donnant le résultat du second tirage, définie par $\mathrm{X}_2\big((a,b)\big) = b$, alors les tribus $\mathfrak{T}_1$ et $\mathfrak{T}_2$ sont les tribus engendrées par les variables aléatoires X_1 et X_2, et leur indépendance reflète l'indépendance des variables.

⟨5⟩. Ou, plus généralement, des « classes », c'est-à-dire des ensembles, d'événements. Ce sont donc des parties de $\mathfrak{T}$.

2.3 Les lemmes de Borel-Cantelli

✍ Rapport du jury ✍

« Illustration de la notion d'indépendance en probabilités. » *Le programme fournit plusieurs utilisations élémentaires de l'indépendance : [...] lemmes de Borel-Cantelli [...].*

Les lemmes de Borel⟨6⟩-Cantelli⟨7⟩ montrent que, pour une famille $(A_n)_{n\geqslant 0}$ d'événements indépendants, la réalisation d'une infinité d'entre eux est, soit presque sûre, soit négligeable ! (Dans le second cas, l'hypothèse d'indépendance n'est même pas nécessaire.)

2.3.a Le premier lemme de Borel-Cantelli

Théorème 2.20 (1er lemme de Borel-Cantelli) — *Soit $(A_n)_{n\geqslant 0}$ une suite d'événements telle que la série $\sum \mathbf{P}(A_n)$ converge, alors $\mathbf{P}\{A_n \text{ ; i.s.}\} = 0$, ou encore : presque sûrement, seul un nombre* fini *d'événements A_n se réalisent.*

⟦ Démonstration : Notons $S_n = \sup_{k\geqslant n} A_k = \bigcup_{k=n}^{\infty} A_k$, de telle sorte que

$$\{A_i \text{ ; i.s.}\} = \lim_n \downarrow S_n = \bigcap_{n=1}^{\infty} S_n.$$

et, notamment, $\{A_i \text{ ; i.s.}\} \subset S_n$ pour tout n. Par σ-sous-additivité,

$$\mathbf{P}\{A_i \text{ ; i.s.}\} \leqslant \mathbf{P}(S_n) = \mathbf{P}\Big(\bigcup_{k=n}^{\infty} A_k\Big) \leqslant \sum_{k=n}^{\infty} \mathbf{P}(A_k).$$

Le membre de droite étant le reste d'une série convergente, il tend vers 0 quand $n \to \infty$, et donc $\mathbf{P}\{A_i \text{ ; i.s.}\} = 0$. ⟧

Remarque 2.21 (Démonstration alternative du premier lemme) En anticipant sur la notion générale d'espérance d'une variable aléatoire, on peut donner une démonstration alternative de ce premier lemme. Notons $X_k = \mathbf{1}_{A_k}$ la variable aléatoire valant 1 sur A_k et 0 en dehors ; on suppose que $\sum \mathbf{P}(A_n)$ converge, et l'on définit $S = \sum_{k=1}^{\infty} X_k$, variable aléatoire à valeurs dans $\mathbb{N} \cup \{+\infty\}$.

Le théorème de convergence monotone montre que

$$\mathbf{E}(S) = \mathbf{E}\Big(\lim_{n\to\infty} S_n\Big) = \lim_{n\to\infty} \mathbf{E}(S_n) = \sum_{n=1}^{\infty} \mathbf{E}(X_n) = \sum_{n=1}^{\infty} \mathbf{P}(A_n) < +\infty.$$

Notamment, la variable aléatoire positive S est presque sûrement finie, c'est-à-dire $\mathbf{P}\{S = +\infty\} = 0$, ou encore $\mathbf{P}\{A_n \text{ ; i.s.}\} = 0$.

⟨6⟩. Émile Borel (1871–1956) a étudié la théorie de la mesure, défini ce qu'on appelle aujourd'hui les boréliens, et s'est également intéressé à la physique mathématique. Il eut pour élève Henri Lebesgue.

⟨7⟩. Francesco Paolo Cantelli (1875–1966), mathématicien italien. On lui doit une version de la loi forte des grands nombres (§ 16.2.b, page 415), le théorème de Glivenko-Cantelli (§ 16.3, page 426) et une clarification des lemmes de Borel-Cantelli.

2.3.b Le second lemme (loi du 0–1 de Borel)

Le second lemme requiert une hypothèse d'indépendance supplémentaire, mais sa conclusion est intéressante : l'événement $\{A_n \text{ ; i.s.}\}$ ne peut être que de probabilité 0 ou 1, à l'exclusion de tout autre résultat.

C'est le premier cas d'une série de résultats du même style, qui disent en substance que *de l'aléatoire on peut tirer des certitudes.*

Théorème 2.22 (2e lemme, ou loi du 0–1 de Borel) — *Soit* $(A_n)_{n\geqslant 0}$ *une suite d'événements* (mutuellement) indépendants. *Alors l'événement* $\{A_n \text{ ; i.s.}\}$ *ne peut être que de probabilité* 0 *ou* 1 *:*

- *si* $\sum \mathbf{P}(A_n)$ *converge, alors* $\mathbf{P}\{A_n \text{ ; i.s.}\} = 0$ *;*
- *si* $\sum \mathbf{P}(A_n)$ *diverge, alors* $\mathbf{P}\{A_n \text{ ; i.s.}\} = 1$.

⟦ Démonstration : Le cas convergent est déjà démontré dans le premier lemme. On suppose donc que $\sum \mathbf{P}(A_n)$ diverge. Les A_k étant indépendants, leurs complémentaires le sont également. Ainsi

$$\begin{aligned}\mathbf{P}\left(\bigcap_{k=n}^{N} A_k^c\right) &= \prod_{k=n}^{N} \mathbf{P}(A_k^c)\\ &= \prod_{k=n}^{N} \big(1-\mathbf{P}(A_k)\big)\\ &\leqslant \prod_{k=n}^{N} e^{-\mathbf{P}(A_k)} = \exp\left(-\sum_{k=n}^{N} \mathbf{P}(A_k)\right)\end{aligned}$$

par convexité de la fonction exponentielle.

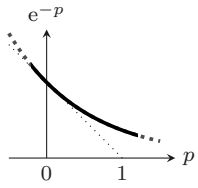


Figure 2.2 — $1-p \leqslant e^{-p}$.

Or la série $\sum \mathbf{P}(A_n)$ diverge, donc le membre de droite tend vers 0 quand N tend vers $+\infty$, et par continuité monotone décroissante,

$$\mathbf{P}\left(\bigcap_{k=n}^{\infty} A_k^c\right) = \lim_{N\to\infty} \mathbf{P}\left(\bigcap_{k=n}^{N} A_k^c\right) = 0.$$

Une réunion dénombrable d'ensembles négligeables est négligeable, donc

$$\mathbf{P}\left(\bigcup_{n=0}^{\infty}\bigcap_{k=n}^{\infty} A_k^c\right) = 0.$$

Enfin, par passage au complémentaire,

$$\mathbf{P}\{A_n \text{ ; i.s.}\} = \mathbf{P}\left(\bigcap_{n=0}^{\infty}\bigcup_{k=n}^{\infty} A_k\right) = 1. \qquad ⟧$$

Remarque 2.23 Sans l'hypothèse d'indépendance, la divergence de la série $\sum \mathbf{P}(A_n)$ peut ne rien entraîner. Voici un exemple purement artificiel : lors d'une succession infinie de tirages à *pile* ou *face*, on note A_n l'événement « le *premier* tirage retourne *pile* ». (Oui oui, tous les A_n sont alors égaux). On a alors $\mathbf{P}(A_n) = \frac{1}{2}$ pour tout $n \geqslant 1$, la série $\sum \mathbf{P}(A_n)$ diverge mais

$$\{A_n \text{ ; i.s.}\} = A_1 \qquad \text{donc} \qquad \mathbf{P}\{A_n \text{ ; i.s.}\} = \tfrac{1}{2}.$$

2.4 Exemples

Thème 3 : La marche aléatoire sur $\mathbb{Z}$ est presque sûrement non bornée

Un marcheur aléatoire, partant de la position $S_0 = 0$, parcourt l'ensemble $\mathbb{Z}$ en temps discret, en effectuant, indépendamment du passé, un pas vers la droite avec une probabilité p ou vers la gauche avec la probabilité $1-p$. Sa position à l'instant n est notée S_n, de sorte que si à l'instant n le marcheur est en position S_n,

- il se retrouve en $S_{n+1} = S_n + 1$ avec la probabilité p ..... événement D_n
- en $S_{n+1} = S_n - 1$ avec la probabilité $q = 1-p$ ..... événement $G_n = D_n^c$.

Par hypothèse, les événements $(D_n)_{n\geqslant 0}$ sont mutuellement indépendants.
Notons B l'événement « la suite $(S_n)_{n\geqslant 0}$ est bornée » et, pour tout entier $m \geqslant 1$,

$$B_m : \quad \text{la suite } (S_n)_{n\geqslant 0} \text{ est à valeurs dans }]-m+1\,;m-1[\ ».$$

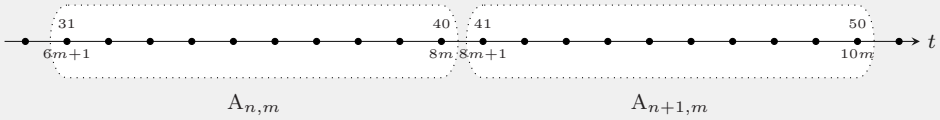
On remarque d'abord que

$$B = \bigcup_{m=1}^{\infty} B_m.$$

Fixons momentanément un entier $m \geqslant 1$.
Notons, pour tout n entier, $A_{n,m}$ l'événement : « le marcheur a effectué une succession de $2m$ pas successifs vers la droite, le premier étant à l'instant $2nm+1$ », c'est-à-dire

$$A_{n,m} = \bigcap_{k=2nm+1}^{2nm+2m} D_k.$$

Par exemple, $A_{0,m}$ représente « le marcheur n'a marché que vers la droite entre les instants n et $2m$ ».
Par indépendance des D_k, $\mathbf{P}(A_{n,m}) = p^{2m}$.
Or, les événements $A_{n+1,m}$ et $A_{n,m}$ couvrent des intervalles de temps disjoints : voici par exemple pour $m = 5$ les instants concernés pour les valeurs $n = 3$ et $n = 4$



Ainsi, par indépendance, la probabilité qu'*aucun* événement $A_{n,m}$ se réalise est

$$\prod_{n=1}^{\infty} \mathbf{P}(A_{m,n}^c) = \prod_{n=1}^{\infty} \left(1 - p^{2m}\right) = 0.$$

La probabilité qu'*au moins* un événement $A_{n,m}$ se réalise est donc 1.
Or dès qu'un des $A_{m,n}$ est réalisé, $2m$ pas sont effectués successivement sur la droite, et il est impossible que le mouvement reste confiné dans l'intervalle entier ouvert $\llbracket -m\,;m \rrbracket$:

$$A_{n,m} \subset B_m^c$$

et donc

$$\bigcup_{n=1}^{\infty} A_{n,m} \subset B_m^c.$$

Par croissance, $\mathbf{P}(B_m^c) = 1$ et, par passage au complémentaire, $\mathbf{P}(B_m) = 0$.

Enfin, par σ-sous-additivité (ou en rappelant que la réunion dénombrable d'événements négligeable est négligeable),

$$\mathbf{P}(\mathrm{B}) \leqslant \sum_{m=1}^{\infty} \mathbf{P}(\mathrm{B}_m) = 0.$$

La valeur $\mathbf{P}(\mathrm{B}) = 0$ montre bien que la marche aléatoire symétrique sur $\mathbb{Z}$ est presque sûrement non bornée.

Remarque 2.24 Dans le chapitre 18 on établira de nouveau ce résultat par d'autres méthodes, et on montrera un peu plus : la marche aléatoire symétrique sur $\mathbb{Z}$ n'est ni majorée ni minorée. On établira également que la marche dans le plan $\mathbb{Z}^2$ est non bornée.

Thème 4 : Inexistence de probabilité raisonnable sur $\mathbb{N}$

On munit l'ensemble N de la tribu $\mathfrak{T} = \mathcal{P}(\mathbb{N})$ et on cherche à définir une probabilité « raisonnable » sur $\mathbb{N}$, au sens où par exemple la probabilité qu'un entier soit pair devrait être $1/2$, la probabilité que ce soit un multiple de 10 devrait être $1/10$ et, plus généralement, pour tout $k \in \mathbb{N}^*$,

$$\mathbf{P}(k\mathbb{N}) = \frac{1}{k}.$$

Nous allons montrer par l'absurde que c'est impossible. On suppose donc l'existence d'une telle probabilité $\mathbf{P}$.

- Notons $(p_n)_{n\geqslant 1}$ la suite croissante des nombres premiers et $\mathcal{A}_k = p_k\mathbb{N}$.
- **Les événements $\boldsymbol{\mathcal{A}_k = p_k\mathbb{N}}$ sont mutuellement indépendants.** En effet, pour tout entier n et tout choix d'indices $i_1, \dots, i_n$ *distincts*, un entier est multiple de p_{i_1}, de p_{i_2} etc. et de p_{i_n}, si et seulement si il est multiple de $p_{i_1} \cdots p_{i_n}$, donc

$$\begin{aligned}\mathbf{P}\left(\bigcap_{k=1}^{n} \mathcal{A}_{i_k}\right) &= \mathbf{P}\big(\text{multiples de } p_{i_1} \cdots p_{i_n}\big) \\ &= \frac{1}{p_{i_1} \cdots p_{i_n}} = \prod_{k=1}^{n} \frac{1}{p_{i_k}} = \prod_{k=1}^{n} \mathbf{P}(\mathcal{A}_{i_k}).\end{aligned}$$

- Or $\sum \mathbf{P}(\mathcal{A}_k) = \sum 1/p_k$ est une série classiquement divergente ; du deuxième lemme de Borel-Cantelli (qui nécessite l'indépendance des A_k !) on déduit

$$\mathbf{P}\{\mathrm{A}_n \text{ ; i.s.}\} = 1$$

ou encore, pour presque tout entier, il existe une infinité de diviseurs premiers... ce qui est absurde, puisque seul 0 est dans $\{\mathrm{A}_n \text{ ; i.s.}\}$.

Mathématiques pour l'agrégation externe
Probabilités

La préparation des candidats aux concours de recrutement de l'Éducation nationale réclame des outils et des méthodes qu'il leur est souvent bien difficile de se procurer, faute d'une littérature adaptée aux exigences de la situation.

Taillé sur mesure pour ceux de l'agrégation externe de mathématiques, ce cours intègre principalement les notions du programme de probabilités spécifiques à ce concours : espaces probabilisés, variables aléatoires et lois, théorèmes de convergence et quelques éléments de statistiques. Toutes les notions y sont abordées dans le détail et leur assimilation est facilitée par près de 120 exercices corrigés dont beaucoup peuvent être utilisés par les candidats pour leurs leçons à l'épreuve orale. Ce manuel sera également très utile aux étudiants en M2 de mathématiques.

Il vient en complément de *Mathématiques pour l'agrégation externe. Analyse*, rédigé par Jean-Étienne Rombaldi et publié dans la même collection.

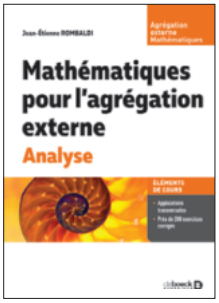
SOMMAIRE

LES PLUS

- **Les principaux théorèmes sont suivis d'applications**
- **Nombreuses remarques issues des rapports du jury**
- **30 thèmes qui pourront être intégrés à des leçons**

Walter Appel est normalien, agrégé de mathématiques et docteur ès sciences physiques. Il a enseigné les mathématiques à l'ENS Lyon et la physique en prépa agrég. Il enseigne actuellement les mathématiques en classes préparatoires MP* au lycée du Parc à Lyon.

Chez le même éditeur



37,90 €

ISBN : 978-2-8073-6166-9

9 782807 361669

deboeck SUPÉRIEUR

www.deboecksuperieur.com