

Cours de Probabilités - MDI 104–114

P. Bianchi, T. Bonald, L. Decreusefond, G. Fort, J. Najim

2017–2018

Table des matières

Introduction	7
Guide de lecture	11
1. Evénements	13
1.1. Définitions	13
1.2. Probabilités sur un espace discret	14
1.3. Conditionnement et indépendance	18
1.4. Exercices	23
2. Variables aléatoires discrètes	27
2.1. Loi d'une variable discrète	27
2.2. Indépendance des v.a. discrètes	29
2.3. Espérance, moments	31
2.4. Fonction génératrice d'une v.a. à valeurs entières	38
2.5. Exercices	40
3. Chaînes de Markov	47
3.1. Préliminaires	47
3.2. Définition	48
3.3. Matrice de transition	49
3.4. Graphe de transition	51
3.5. Irréductibilité	51
3.6. Périodicité	54
3.7. Théorème ergodique	59
3.8. Exercices	64
4. Mesures et intégration	69
4.1. Introduction	69
4.2. Tribus	70
4.3. Fonctions mesurables	72
4.4. Mesures	76
4.5. Intégration	85
4.6. Exercices	102
5. Variables et vecteurs aléatoires réels	111
5.1. Généralités	111
5.2. Variables aléatoires réelles	112
5.3. Vecteurs aléatoires	120

5.4. Changement de variables	126
5.5. Exercices	132
6. Loi conditionnelle	149
6.1. Cas général	150
6.2. Calculs	152
6.3. Espérance conditionnelle	155
6.4. Exercices	157
7. Fonction caractéristique	161
7.1. Définition et propriétés	161
7.2. Fonctions caractéristiques de v.a. usuelles	164
7.3. Caractérisation de la loi	166
7.4. Caractérisation de l'indépendance	166
7.5. Calcul de moments	167
7.6. Exercices	171
8. Vecteurs gaussiens	173
8.1. Préliminaires	173
8.2. Définition, propriétés	173
8.3. Caractérisation de l'indépendance	175
8.4. Stabilité par transformation affine	176
8.5. Somme de vecteurs gaussiens indépendants	177
8.6. La loi d'un vecteur gaussien admet-elle une densité ?	178
8.7. Exercices	180
9. Convergences	183
9.1. Loi des grands nombres	183
9.2. Limite centrée	183
9.3. Exercices	185
10. Simulation	187
10.1. Générateurs de nombres aléatoires	187
10.2. Méthode d'inversion de la fonction de répartition	189
10.3. Méthode du rejet	191
10.4. Méthodes de Monte Carlo	193
10.5. Histogrammes	195
10.6. Exercices	196
Annexes	199
A. Ensembles	201
A.1. Opérations sur les ensembles	201
A.2. Espaces d'états dénombrables	203
B. Notions utiles d'analyse	205
B.1. Limite supérieure et limite inférieure	205
B.2. Séries	206

B.3. Convexité	207
--------------------------	-----

Notations

$\perp\!\!\!\perp$	Indépendance d'événement, indépendance de variables aléatoires.
$(\cdot)^T$	Transposée d'une matrice ou d'un vecteur.
\circ	Composition.
$ \Omega $	Cardinal d'un ensemble Ω .
$a_n \uparrow a$	La suite réelle $(a_n)_n$ est croissante et converge vers a .
$a_n \downarrow a$	La suite réelle $(a_n)_n$ est décroissante et converge vers a .
$A_n \uparrow A$	La suite d'ensembles $(A_n)_n$ est croissante et converge vers A <i>i.e.</i> , $\cup_n A_n = A$.
$A_n \downarrow A$	La suite d'ensembles $(A_n)_n$ est décroissante et converge vers A <i>i.e.</i> , $\cap_n A_n = A$.
\det	Déterminant.
I_n	Matrice identité de taille n .
$\mathcal{P}(\Omega)$ ou 2^Ω	Tribu des parties sur Ω .
Tr	Trace.
$x \vee y$	Maximum de x et de y .
$x \wedge y$	Minimum de x et de y .
$\arg\min_{x \in C} f(x)$	La valeur de la variable x (supposée atteinte et unique) en laquelle f est minimale sur C .

Introduction

Les premières formalisations des probabilités datent du XVIII^e siècle avec les travaux de Jacob Bernoulli (1713) et de Abraham de Moivre (1718). La probabilité d'un événement y était définie comme le rapport du nombre de cas favorables sur le nombre total de cas. Au début du XIX^e siècle, les « probabilités géométriques » firent leur apparition. Dans ce cadre, la probabilité d'un événement s'exprime comme un rapport de volumes ou d'aires. Ces approches permettaient de faire bon nombre de calculs mais butaient sur certains paradoxes.

Les probabilités sont au départ, une tentative de représentation mathématique de l'incertain. Elles doivent être tout à la fois suffisamment formalisées pour permettre des calculs justes et rigoureux et garder une connexion forte et immédiate avec les phénomènes « physiques » analysés. Cette tension a longtemps posé des problèmes. Notamment, à la fin du XIX^e, se posait le problème des événements « presque certains » ou « presque impossibles » : y-a-t'il un seuil en dessous un événement de probabilité inférieure à ce seuil ne peut se réaliser ?

Au début du XX^e, David Hilbert assigna aux mathématiciens, 23 problèmes, ou plutôt 23 défis, pour les années à venir. Parmi ceux-ci figurait l'axiomatisation de la « physique » par laquelle il fallait entendre l'axiomatisation des probabilités.

Le formalisme correct ne se fit jour qu'en 1930 dans les travaux d'Andréï Kolmogorov, qui réussit la synthèse des réflexions de Émile Borel, Jacques Hadamard, Maurice Fréchet et Paul Lévy entre autres.

Le concept de mesure permet d'avoir une vision unifiée des probabilités discrètes et des probabilités dites « continues ». Le vocabulaire de l'intégration permet de simplifier la présentation des différentes notions probabilistes. Par ailleurs, ainsi que l'illustre le deuxième paradoxe de Bertrand, la modélisation de certains phénomènes même simples impose de comprendre finement les liens entre théorie et interprétation physique. Enfin, la simulation, outil indispensable tellement est grande la complexité des systèmes, requiert de « construire » des variables et des processus aléatoires. Tout cela ne peut se faire sans une solide compréhension de la théorie sous-jacente.



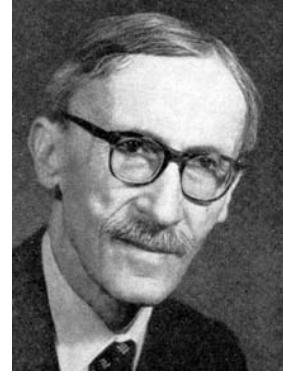
É. Borel (1871-1956),



M. Fréchet (1878-1973)



J. Hadamard (1865-1963),



P. Lévy (1886-1971)



A. Kolmogorov (1903-1987) (DR)

Ce document garde les traces de ses précédentes versions avec un chapitre sur les probabilités discrètes. Comme ces notions sont supposées connues, elles ont été mises à titre de rappel et pour fixer les notations.

Guide de lecture

Ce support de cours est destinée à plusieurs populations d'étudiants qui se distinguent par leur prérequis en mathématiques générales et en probabilités. La lecture n'est donc pas la même pour tous.

Pour ceux qui connaissent déjà les probabilités discrètes, le cheminement normal du cours (et donc de votre lecture) est le suivant :

- Chaînes de Markov, chapitre 3
- Mesure et intégration, chapitre 4
- Variables et vecteurs aléatoires, chapitre 5
- Loi conditionnelle, chapitre 6
- Fonctions caractéristiques, 7
- Vecteurs gaussiens, 8

Les deux chapitres 9 et 10 peuvent être abordés avant celui sur les fonctions caractéristiques ou en fin de programme après les vecteurs gaussiens.

Avertissement : Le lecteur pressé peut être tenté de sauter le chapitre « mesures et intégration ». Cela peut se faire en première lecture mais les renvois à ce chapitre tant du point de vue des notations que des concepts sont fréquents. Ne pas le lire risque d'obscurez votre compréhension tant du cours de probabilités que du cours d'analyse MDI 103 dont il constitue le fondement.

Pour ceux qui ne maîtrisent pas les manipulations de base sur les ensembles, il est recommandé de se référer à l'annexe A. Les résultats prérequis d'analyse générale notamment sur les séries sont accessibles dans l'annexe B. Les exercices des chapitres sur les probabilités discrètes et la théorie de la mesure sont tous corrigés sur la version en ligne de cet opus, accessible depuis le site pédagogique.

1. Événements

1.1. Définitions

Une *expérience aléatoire* est une expérience pouvant conduire à plusieurs résultats possibles. Formellement, une expérience aléatoire se décrit par la donnée de l'ensemble Ω des résultats possibles. L'ensemble Ω est appelé l'*univers* ou l'*espace des états*.

Traditionnellement, un résultat possible de l'expérience est noté ω . C'est un élément de l'univers Ω . Un tel élément $\omega \in \Omega$ est parfois appelé une *épreuve* ou une *issue*.

EXEMPLE 1.– *a)* Jet d'un dé : $\Omega = \{1, 2, 3, 4, 5, 6\}$.

b) Deux lancers consécutifs d'une pièce. L'univers est $\Omega = \{PP, PF, FP, FF\}$ où *P* et *F* signifient respectivement *pile* et *face*.

c) Durée de fonctionnement sans panne d'une machine : $\Omega = [0, +\infty[$.

d) Valeur d'un signal continu sur un intervalle de temps $[t_0, t_1]$: $\Omega = \mathcal{C}_b([t_0, t_1])$ est l'ensemble des fonctions continues de $[t_0, t_1]$ dans \mathbf{R} .

Un *événement aléatoire* est un événement dont la réalisation dépend du résultat de l'expérience. Formellement, un événement aléatoire se décrit comme un sous-ensemble de Ω .

EXEMPLE 2.– Considérons à nouveau les exemples précédents.

a) $\Omega = \{1, 2, \dots, 6\}$. L'événement *A* = « Le résultat est pair » s'identifie au sous-ensemble $A = \{2, 4, 6\}$.

b) $\Omega = \{PP, PF, FP, FF\}$. L'événement *A* = « on obtient deux faces identiques » s'identifie au sous-ensemble $A = \{PP, FF\}$.

c) $\Omega = [0, \infty[$. L'événement *A* = « La machine fonctionne au moins x unités de temps » s'identifie à $A = [x, +\infty[$.

d) $\Omega = \mathcal{C}_b([t_0, t_1])$. L'événement *A* = « L'amplitude du signal n'excède pas α » s'écrit $A = \{\omega \in \Omega : \sup_{t \in [t_0, t_1]} |\omega(t)| \leq \alpha\}$.

Pour une issue donnée $\omega \in \Omega$, on dit qu'un événement *A* est *réalisé* si $\omega \in A$.

L'espace d'état Ω est aussi appelé l'*événement certain* : il est réalisé quelle que soit l'issue. L'ensemble vide \emptyset est aussi appelé l'*événement impossible* : il n'est jamais réalisé.

La notation suivante sera d'un usage constant.

Définition 1.1. Soit Ω un espace d'état et $A \subset \Omega$ un ensemble. La fonction indicatrice de A est définie par :

$$\begin{aligned}\mathbf{1}_A : \Omega &\rightarrow \{0, 1\} \\ \omega &\mapsto 1 \text{ si } \omega \in A, 0 \text{ sinon.}\end{aligned}$$

L'exercice 1.1 fournit quelques propriétés importantes de l'indicatrice.

1.2. Probabilités sur un espace discret

Comme nous l'avons vu au début de ce chapitre, beaucoup d'expériences aléatoires peuvent être décrites par un univers Ω fini ou dénombrable. Citons comme exemples immédiats le tirage à pile ou face ($\Omega = \{P, F\}$), le lancer de dé ($\Omega = \{1, \dots, 6\}$), le nombre de requêtes reçues par un serveur en une unité de temps ($\Omega = \mathbb{N}$), etc.

1.2.1. Définition

Définition 1.2. Une mesure μ , sur un ensemble E au plus dénombrable, est une application de $\mathcal{P}(E)$, l'ensemble des parties de E , dans \mathbb{R} qui satisfait les deux propriétés suivantes :

- $\mu(\emptyset) = 0$,
- pour toute famille $(A_j, j \in \mathbb{N}^*)$ de parties deux à deux disjointes de E ,

$$\mu\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{+\infty} \mu(A_j). \quad (1.1)$$

Les parties de E s'appellent plus souvent des « événements ».

Définition 1.3. Une mesure μ est dite mesure de probabilité (ou probabilité) lorsque $\mu(E) = 1$. Dans ce cas, on la note usuellement \mathbf{P} et non μ .

1.2.2. Propriétés générales

Proposition 1.1. Soient $A, B, (A_n)_{n \in \mathbb{N}^*}$ des ensembles.

- a) $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$.
- b) $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.
- c) Si $A \subset B$, alors $\mathbf{P}(A) \leq \mathbf{P}(B)$.

d) Si $(A_n)_{n \in \mathbb{N}^*}$ est une partition de Ω , alors

$$\mathbf{P}(B) = \sum_{n=1}^{\infty} \mathbf{P}(A_n \cap B) .$$

- e) Si $A_n \uparrow A$, alors $\mathbf{P}(A) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n)$.
Si $A_n \downarrow A$, alors $\mathbf{P}(A) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n)$.
- f) Si $\mathbf{P}(A_n) = 1$ pour tout $n \in \mathbb{N}^*$, alors $\mathbf{P}(\bigcap_{n=1}^{\infty} A_n) = 1$.
- g) Pour une famille quelconque $(A_n)_{n \in \mathbb{N}^*}$ dans \mathcal{F} , on a la *borne de l'union* :

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

Démonstration. a) On applique l'axiome de σ -additivité (1.1) en posant $A_1 = A$, $A_2 = A^c$ et $A_n = \emptyset$ pour tout $n \geq 3$. Il en résulte que $1 = \mathbf{P}(\Omega) = \mathbf{P}(A \cup A^c \cup \emptyset \cup \emptyset \cup \dots) = \mathbf{P}(A) + \mathbf{P}(A^c) + 0 + 0 + \dots$ et finalement $1 = \mathbf{P}(A) + \mathbf{P}(A^c)$.

b) On écrit que $A \cup B$ s'écrit comme l'union disjointe $(A \setminus B) \cup (B \setminus A) \cup (A \cap B)$. La règle de σ -additivité conduit à :

$$\mathbf{P}(A \cup B) = \mathbf{P}(A \setminus B) + \mathbf{P}(B \setminus A) + \mathbf{P}(A \cap B) . \quad (1.2)$$

Par ailleurs, A s'écrit comme l'union disjointe $(A \setminus B) \cup (A \cap B)$ et donc $\mathbf{P}(A) = \mathbf{P}(A \setminus B) + \mathbf{P}(A \cap B)$. De même, $\mathbf{P}(B) = \mathbf{P}(B \setminus A) + \mathbf{P}(A \cap B)$. On a donc : $\mathbf{P}(A \setminus B) + \mathbf{P}(B \setminus A) = \mathbf{P}(A) + \mathbf{P}(B) - 2\mathbf{P}(A \cap B)$. En faisant la substitution dans (1.2), nous obtenons le résultat.

c) Si $A \subset B$, on a en particulier : $B = A \cup (B \setminus A)$ et comme l'union est disjointe, $\mathbf{P}(B) = \mathbf{P}(A) + \mathbf{P}(B \setminus A) \geq \mathbf{P}(A)$.

d) Comme les (A_n) sont deux à deux disjoints, il en est de même pour les événements $(A_n \cap B)$. Par σ -additivité, on a $\sum_n \mathbf{P}(A_n \cap B) = \mathbf{P}(\bigcup_n (A_n \cap B)) = \mathbf{P}((\bigcup_n A_n) \cap B) = \mathbf{P}(B)$, où la dernière égalité provient du fait que $\bigcup_n A_n = \Omega$.

e) Soit $A_n \uparrow A$. On introduit la suite (B_n) définie par récurrence de la façon suivante : $B_1 = A_1$ et $B_{n+1} = A_{n+1} \setminus B_n$. On vérifie sans peine que les (B_n) sont deux à deux disjoints, ce qui implique :

$$\mathbf{P}\left(\bigcup_{k=1}^{\infty} B_k\right) = \sum_{k=1}^{\infty} \mathbf{P}(B_k) . \quad (1.3)$$

On vérifie également que pour tout n , $A_n = \bigcup_{k=1}^n B_k$, et donc, par passage à la limite, $A = \bigcup_{k=1}^{\infty} B_k$. Ainsi, le membre de gauche de (1.3) n'est autre que $\mathbf{P}(A)$. Le membre de droite se réécrit comme la limite quand $n \rightarrow \infty$ de la suite $\sum_{k=1}^n \mathbf{P}(B_k)$. Mais comme les (B_k) sont deux à deux disjoints, $\sum_{k=1}^n \mathbf{P}(B_k) = \mathbf{P}(\bigcup_{k=1}^n B_k) = \mathbf{P}(A_n)$. On a donc bien montré que $\mathbf{P}(A) = \lim_n \mathbf{P}(A_n)$.

Soit maintenant $A_n \downarrow A$. Dans ce cas, $A_n^c \uparrow A^c$. En appliquant le résultat précédent, $\mathbf{P}(A^c) = \lim_n \mathbf{P}(A_n^c)$. Par la propriété a), cette égalité se réécrit $1 - \mathbf{P}(A) = \lim_n (1 - \mathbf{P}(A_n))$, d'où on déduit $\mathbf{P}(A) = \lim_n \mathbf{P}(A_n)$.

f) La suite $(\cap_{k=1}^n A_k)$ est décroissante et converge vers $\cap_{k=1}^\infty A_k$. Puisque $\mathbf{P}(A_n) = 1$ pour tout n , il s'en suit que $1 = \lim_n \mathbf{P}(A_n) = \mathbf{P}(\cap_{k=1}^\infty A_k)$.

g) On montre d'abord la borne pour un nombre fini d'éléments :

$$\forall n, \mathbf{P}\left(\bigcup_{k=1}^n A_k\right) \leq \sum_{k=1}^n \mathbf{P}(A_k). \quad (1.4)$$

L'inégalité est vraie au rang $n = 1$. Supposons qu'elle soit vraie au rang n , $\mathbf{P}\left(\bigcup_{k=1}^{n+1} A_k\right) = \mathbf{P}(A_{n+1} \cup (\bigcup_{k=1}^n A_k)) \leq P(A_{n+1}) + \mathbf{P}(\bigcup_{k=1}^n A_k)$ d'après la propriété b). En injectant l'hypothèse de récurrence dans le membre de droite, la proposition est démontrée au rang $n + 1$.

L'inégalité (1.4) implique que $\mathbf{P}(\bigcup_{k=1}^n A_k) \leq \sum_{k=1}^\infty \mathbf{P}(A_k)$ pour tout n . Or la suite $(\bigcup_{k=1}^n A_k)$ est croissante, de limite $\bigcup_{k=1}^\infty A_k$. Par passage à la limite dans la dernière inégalité, on obtient le résultat voulu en invoquant la propriété e).

□

1.2.3. Représentation des probabilités sur un espace discret

La propriété suivante établit qu'une probabilité \mathbf{P} sur un espace discret est *entièrement caractérisée par la valeur qu'elle prend sur les singletons*.

Proposition 1.2. Soit Ω un espace discret et \mathbf{P} une mesure de probabilité définie sur la tribu des parties $\mathcal{P}(\Omega)$. Alors pour tout événement A ,

$$\mathbf{P}(A) = \sum_{\omega \in A} \mathbf{P}(\{\omega\}).$$

Démonstration. Comme Ω est au plus dénombrable, on peut indexer ses éléments sous la forme $\Omega = \{\omega_1, \omega_2, \dots\}$. Tout sous-ensemble A de Ω est donc de la forme $A = \{\omega_{i_1}, \omega_{i_2}, \dots\}$ où i_1, i_2, \dots sont des entiers. Par conséquent, A est l'union dénombrable des singletons $\{\omega_{i_1}\}, \{\omega_{i_2}\}$, etc. Par σ -additivité de \mathbf{P} , on a donc $\mathbf{P}(A) = \mathbf{P}(\{\omega_{i_1}\}) + \mathbf{P}(\{\omega_{i_2}\}) + \dots$, ce qui prouve le résultat. □

Ainsi, il suffit de connaître la probabilité des événements élémentaires pour connaître la probabilité de n'importe quel événement. Cette affirmation est caractéristique des probabilités sur un espace discret, elle est clairement fausse dans le cas général des probabilités sur un espace non dénombrable.

La propriété suivante va un peu plus loin : elle établit que, pour qu'une famille de nombres positifs définissent une probabilité, il faut et il suffit que leur somme soit égale à un.

Proposition 1.3. Soit Ω un ensemble discret. Soit $(p_\omega)_{\omega \in \Omega}$ une suite de nombres positifs satisfaisant $\sum_{\omega \in \Omega} p_\omega = 1$. Alors il existe une (unique) mesure de probabilité \mathbf{P} sur $\mathcal{P}(\Omega)$ telle que pour tout $\omega \in \Omega$, $\mathbf{P}(\{\omega\}) = p_\omega$.

Démonstration. L'unicité est une conséquence de la propriété précédente. Afin de montrer l'existence, il suffit de poser $\mathbf{P}(A) = \sum_{\omega \in A} p_\omega$. On montre sans peine que cette application satisfait les axiomes d'une mesure de probabilité. \square

Ainsi, concrètement, une probabilité sur un espace discret se ramène à une famille de nombres positifs sommant à un : se donner l'un revient à se donner l'autre.

1.2.4. Exemples de probabilités sur un espace discret

Cas où Ω est fini

Loi uniforme Soit Ω un ensemble fini quelconque. La *probabilité uniforme sur Ω* est définie par :

$$\mathbf{P}(A) = \frac{|A|}{|\Omega|}$$

où $|A|$ représente ici le cardinal de l'ensemble A . Autrement dit, $\mathbf{P}(A)$ est le ratio entre le nombre d'issues pour lesquelles A est réalisé, et le nombre total d'issues. D'après la propriété précédente, on aurait pu définir la probabilité uniforme de façon équivalente comme l'unique probabilité pour laquelle toutes les issues sont *équiprobables*, c'est-à-dire pour tout ω ,

$$\mathbf{P}(\{\omega\}) = \frac{1}{|\Omega|} .$$

Loi de Bernoulli Soit $p \in [0, 1]$. La *probabilité de Bernoulli de paramètre p* , notée $B(p)$, est la probabilité définie sur $\Omega = \{0, 1\}$ par :

$$\mathbf{P}(\{1\}) = p , \quad \mathbf{P}(\{0\}) = 1 - p .$$

La probabilité de Bernoulli permet de décrire la probabilité de succès ou d'échec d'une expérience. Par exemple, elle permet de décrire la probabilité qu'une pièce tombe sur pile : si la pièce est parfaitement équilibrée, on choisira $p = 1/2$ et la probabilité de Bernoulli se ramène à la loi uniforme sur $\{0, 1\}$; dans le cas d'une pièce non équilibrée ou d'un jeu truqué, le paramètre p est possiblement différent de $1/2$.

Loi binomiale Soit $n \in \mathbb{N}^*$ et $p \in [0, 1]$. La *probabilité binomiale de paramètres n, p* , notée $B(n, p)$, est la probabilité définie sur $\Omega = \{0, 1, \dots, n\}$ par :

$$\mathbf{P}(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}$$

pour tout $k = 0, \dots, n$ où l'on rappelle que $\binom{n}{k} = n!/(k!(n-k)!)$. La probabilité binomiale est utilisée pour décrire le nombre de succès obtenus lorsqu'on réitère n fois une expérience ayant même probabilité de succès p (voir l'exercice 2.1).

Cas où Ω est dénombrable

Loi géométrique Soit $p \in]0, 1]$. La *probabilité géométrique de paramètre p sur \mathbf{N}^** , notée $\mathcal{G}(p)$, est la probabilité définie sur $\Omega = \mathbf{N}^*$ par :

$$\mathbf{P}(\{k\}) = p(1-p)^{k-1} \quad (1.5)$$

pour tout $k \in \mathbf{N}^*$. Imaginons que l'on réitère autant de fois que nécessaire une certaine expérience ayant un probabilité de succès p . Alors la probabilité géométrique est utilisée pour décrire le nombre d'expériences qui ont été nécessaires pour obtenir un succès (voir l'exercice 2.1).

REMARQUE.— On peut aussi définir la probabilité géométrique sur \mathbf{N} (et non \mathbf{N}^*) par $\mathbf{P}(\{k\}) = p(1-p)^k$ pour tout $k = 0, 1, 2, \dots$. Dans ce dernier cas, on cherche à décrire non pas l'instant du premier succès, mais le nombre d'échecs qui ont précédé le premier succès.

Loi de Poisson Soit $\alpha > 0$. La *probabilité de Poisson de paramètre α* , notée $\mathcal{P}(\alpha)$, est la probabilité définie sur $\Omega = \mathbf{N}$ par :

$$\mathbf{P}(\{k\}) = \frac{\alpha^k}{k!} e^{-\alpha} .$$

La probabilité de Poisson est souvent utilisée pour modéliser des quantités telles que le nombre de requêtes reçues par un serveur par unité de temps, ou le nombre de clients qui se présentent à un guichet pendant une unité de temps.

1.3. Conditionnement et indépendance

1.3.1. Probabilité conditionnelle : définition

De façon informelle, la probabilité d'un événement vise à quantifier l'occurrence de cet événement. La probabilité conditionnelle d'un événement A sachant un événement B vise à quantifier l'occurrence de A lorsque l'on sait que B s'est produit. D'un point de vue plus formel, on a la définition suivante.

Soit $(\Omega, \mathcal{F}, \mathbf{P})$ un espace de probabilité.

Définition 1.4. Pour tous événements $A, B \in \mathcal{F}$ tels que $\mathbf{P}(B) \neq 0$, on définit la probabilité conditionnelle de A sachant B , et on note $\mathbf{P}(A | B)$, la quantité :

$$\mathbf{P}(A | B) := \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} .$$

Si on associe probabilité et « poids », la probabilité d'un ensemble étant son poids relatif par rapport à celui de l'ensemble total, la probabilité conditionnelle de A sachant B est le poids de la trace de A sur B relativement au poids total de B .

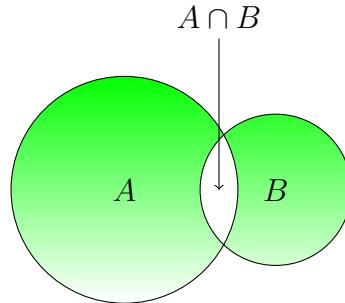


Figure 1.1. – Interprétation graphique du conditionnement.

Considérons le cas où \mathbf{P} est la probabilité uniforme sur un ensemble Ω fini, c'est-à-dire $\mathbf{P}(A) = |A|/|\Omega|$. On a alors $\mathbf{P}(A | B) = |A \cap B|/|B|$. Cette expression justifie la remarque suivante : $\mathbf{P}(A | B)$ peut être interprétée comme la probabilité de l'événement $A \cap B$ dans ce nouvel univers qu'est B .

APPLICATION 1.– Considérons le lancer d'un dé : \mathbf{P} est la probabilité uniforme sur $\Omega = \{1, 2, \dots, 6\}$. Calculer la probabilité d'obtenir « 6 » sachant que le résultat est pair.

Proposition 1.4. Soit $B \in \mathcal{F}$ tel que $\mathbf{P}(B) \neq 0$. L'application définie sur \mathcal{F} par $A \mapsto \mathbf{P}(A | B)$ est une mesure de probabilité. On la nomme la probabilité conditionnelle à B .

Démonstration. i) $\mathbf{P}(\Omega | B) = \mathbf{P}(\Omega \cap B)/\mathbf{P}(B) = 1$ et $\mathbf{P}(\emptyset | B) = \mathbf{P}(\emptyset \cap B)/\mathbf{P}(B) = 0$.
ii) Soit (A_n) une famille d'événement deux à deux disjoints. $\mathbf{P}(\bigcup_n A_n | B) = \mathbf{P}(\bigcup_n A_n \cap B)/\mathbf{P}(B) = \mathbf{P}(\bigcup_n (A_n \cap B))/\mathbf{P}(B) = \sum_n \mathbf{P}(A_n \cap B)/\mathbf{P}(B) = \sum_n \mathbf{P}(A_n | B)$. \square

1.3.2. Propriétés

La première propriété est connue sous le nom de *formule des probabilités totales*.

Proposition 1.5. a) Soient $A, B \in \mathcal{F}$ tels que $0 < \mathbf{P}(B) < 1$. Alors,

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

b) Soit $(B_n)_{n \in \mathbb{N}^*}$ une partition de Ω telle que pour tout n , $\mathbf{P}(B_n) \neq 0$. Alors,

$$\mathbf{P}(A) = \sum_{n=1}^{\infty} \mathbf{P}(A | B_n) \mathbf{P}(B_n) .$$

Démonstration. A s'écrit comme l'union disjointe $A = (A \cap B) \cup (A \cap B^c)$ donc $\mathbf{P}(A) = \mathbf{P}(A \cap B) + \mathbf{P}(A \cap B^c)$. Le résultat provient du fait que $\mathbf{P}(A \cap B) = \mathbf{P}(A | B) \mathbf{P}(B)$ et $\mathbf{P}(A \cap B^c) = \mathbf{P}(A | B^c) \mathbf{P}(B^c)$. La preuve de b) est fondée sur le même principe. \square

EXEMPLE 3. On dispose de trois pièces de monnaie : l'une est bien équilibrée, l'une comporte deux côtés pile, l'autre deux côtés face. On choisit une pièce au hasard. Evaluons la probabilité de tomber sur pile.

Désignons par E , $2P$ et $2F$ les événements « la pièce bien équilibrée est choisie », « la pièce comportant deux côtés pile est choisie », etc. D'après la propriété ci-dessus,

$$\begin{aligned} \mathbf{P}(\text{pile}) &= \mathbf{P}(\text{pile} | E) \mathbf{P}(E) + \mathbf{P}(\text{pile} | 2P) \mathbf{P}(2P) + \mathbf{P}(\text{pile} | 2F) \mathbf{P}(2F) \\ &= \frac{1}{2} \times \frac{1}{3} + 1 \times \frac{1}{3} + 0 \times \frac{1}{3} = \frac{1}{2} . \end{aligned}$$

La seconde propriété est connue sous le nom de *formule de Bayes*. La preuve est immédiate.

Proposition 1.6. Soient $A, B \in \mathcal{F}$ deux événements tels que $\mathbf{P}(A) \neq 0$ et $\mathbf{P}(B) \neq 0$. Alors,

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

La formule de Bayes permet typiquement d'évaluer des probabilités du type :

$$\mathbf{P}(\text{une action 'a' a été effectuée} | \text{le résultat 'r' a été observé})$$

lorsqu'on connaît le modèle $\mathbf{P}(\text{le résultat 'r' est observé} | \text{l'action 'a' est effectuée})$.

EXEMPLE 4. Reprenons l'exemple précédent des trois pièces. Sachant qu'on obtient le résultat pile, quelle est la probabilité que la pièce à deux côtés pile ait été choisie ? La réponse est donnée par la formule de Bayes :

$$\mathbf{P}(2P | \text{pile}) = \frac{\mathbf{P}(\text{pile} | 2P) \mathbf{P}(2P)}{\mathbf{P}(\text{pile})} = \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3} .$$

1.3.3. Événements indépendants

Dans l'exemple précédent, l'événement $B = \langle\text{pile est le résultat}\rangle$ apporte une information sur la probabilité que l'événement $A = \langle\text{la pièce à deux côtés pile a été choisie}\rangle$. Avant l'expérience qui a vu B se réaliser, la probabilité de A valait $\frac{1}{2}$. Après l'expérience, elle vaut $\frac{2}{3}$. Le fait que B soit réalisé ne dit pas si A est ou non réalisé, mais par contre, il change

notre croyance en A . A l'inverse, il existe des événements A, B tels que la réalisation de B n'apporte aucune information sur A . De tels événements sont dits *indépendants*. Voici une définition plus formelle.

Définition 1.5. Deux événements $A, B \in \mathcal{F}$ sont dits *indépendants*, noté $A \perp\!\!\!\perp B$, si

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

Si $\mathbf{P}(B) \neq 0$, la définition revient bien à $\mathbf{P}(A | B) = \mathbf{P}(A)$: la réalisation de B ne modifie pas la croyance en A .

REMARQUE.– a) Les propriétés suivantes sont équivalentes : $A \perp\!\!\!\perp B$, $B \perp\!\!\!\perp A$, $A \perp\!\!\!\perp B^c$, $A^c \perp\!\!\!\perp B$, $A^c \perp\!\!\!\perp B^c$.
b) Si $\mathbf{P}(B) = 0$ ou $\mathbf{P}(B) = 1$, alors A et B sont indépendants quel que soit A .

Définition 1.6. Soit I un ensemble quelconque. Une famille $(A_i)_{i \in I}$ d'événements est dite indépendante si pour tout sous-ensemble **fini** $J \subset I$, on a :

$$\mathbf{P} \left(\bigcap_{j \in J} A_j \right) = \prod_{j \in J} \mathbf{P}(A_j) .$$

Illustrons la formule ci-dessus lorsque la famille contient trois événements A, B, C . Les événements A, B, C sont indépendants si

$$\begin{aligned} \mathbf{P}(A \cap B) &= \mathbf{P}(A)\mathbf{P}(B), \quad \mathbf{P}(A \cap C) = \mathbf{P}(A)\mathbf{P}(C), \quad \mathbf{P}(B \cap C) = \mathbf{P}(B)\mathbf{P}(C), \\ \text{et } \mathbf{P}(A \cap B \cap C) &= \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C) . \end{aligned}$$

Il est important de souligner que la première ligne d'équations ci-dessus n'implique pas la deuxième : ce n'est pas parce que des événements sont deux à deux indépendants qu'ils forment une famille indépendante. L'exercice 1.10 fournit un contre-exemple.

Définition 1.7. Soit $C \in \mathcal{F}$ tel que $\mathbf{P}(C) \neq 0$. On dit que A et B sont *indépendants conditionnellement à C* , noté $A \perp\!\!\!\perp B | C$ si $\mathbf{P}(A \cap B | C) = \mathbf{P}(A | C)\mathbf{P}(B | C)$.

La notion de famille indépendante conditionnellement à C se définit selon le même principe.

REMARQUE.– Attention : des propositions $A \perp\!\!\!\perp B$ et $A \perp\!\!\!\perp B | C$, aucune n'implique l'autre. Là encore, l'exercice 1.10 fournit un contre-exemple.

Ω	Probabilité	Expression de \mathbf{P}	Notation
Ω fini	Probabilité uniforme	$\mathbf{P}(x) = \frac{1}{ \Omega }$	
$\{0, 1\}$	Bernoulli de paramètre $p \in [0, 1]$	$\mathbf{P}(\{1\}) = p$ $\mathbf{P}(\{0\}) = 1 - p$	$\mathcal{B}(p)$
$\{1, \dots, n\}$	Binomiale de paramètres $n, p \in [0, 1]$	$\mathbf{P}(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}$	$\mathcal{B}(n, p)$
\mathbb{N}^*	Géométrique de paramètre $p \in]0, 1]$	$\mathbf{P}(\{k\}) = p(1-p)^{k-1}$	$\mathcal{G}(p)$
\mathbb{N}	Géométrique de paramètre $p \in]0, 1]$	$\mathbf{P}(\{k\}) = p(1-p)^k$	$\mathcal{G}(p)$
\mathbb{N}	Poisson de paramètre $\alpha > 0$	$\mathbf{P}(\{k\}) = \frac{\alpha^k}{k!} e^{-\alpha}$	$\mathcal{P}(\alpha)$

Table 1.1. – Quelques exemples de probabilités sur un espace discret.

1.4. Exercices

Pour apprendre

EXERCICE 1.1.-

On rappelle que

$$\mathbf{1}_A(x) = 1 \text{ si } x \in A, 0 \text{ sinon.}$$

- 1.1.a) Montrer que $\mathbf{1}_{A^c} = 1 - \mathbf{1}_A$, $\mathbf{1}_{A \cap B} = \mathbf{1}_A \mathbf{1}_B$, $\mathbf{1}_{A \cup B} = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_A \mathbf{1}_B$.
- 1.1.b) Montrer que pour toute famille dénombrable $(A_i)_{i \in I}$ d'ensembles deux à deux disjoints,

$$\mathbf{1}_{\bigcup_{i \in I} A_i} = \sum_{i \in I} \mathbf{1}_{A_i}.$$

EXERCICE 1.2.-

On lance simultanément trois dés à 6 faces non pipés.

- 1.2.a) Quel est l'espace des événements ?
- 1.2.b) Quelle est la probabilité d'avoir au moins 1 as ?
- 1.2.c) Montrer que les événements « la somme des faces est paire » et « la somme des faces est impaire » ont même probabilité.
- 1.2.d) Quelle est la probabilité que la somme des faces soit paire ?

EXERCICE 1.3 (Loto).-

Un joueur coche 6 numéros sur une grille de 49 numéros. On tire 6 boules parmi 49 boules numérotées. Quelle est la probabilité pour que le joueur ait exactement n bons numéros ($n = 1, \dots, 6$) ?

Pour s'entraîner

EXERCICE 1.4.-

Dans un lot de 20 articles, 12 sont parfaits, 6 comportent un défaut mineur et 2 un défaut majeur. Deux articles sont choisis au hasard, calculer les probabilités suivantes :

- 1.4.a) Les deux sont parfaits,
 1.4.b) Les deux ont un défaut majeur,
 1.4.c) Au moins l'un d'entre eux est parfait,
 1.4.d) Exactement un est parfait
 1.4.e) Au plus l'un d'entre eux est parfait.

Un lot de 20 articles est accepté lorsque 3 éléments choisis au hasard n'ont pas de défaut majeur.

1.4.f) Quelle est la probabilité que le lot décrit ci-dessus soit accepté ?

EXERCICE 1.5.-

Une boîte contient 4 piles usagées et 6 piles neuves. On tire deux piles au hasard. L'une d'entre elles seulement est testée.

1.5.a) Quelle est la probabilité que l'autre soit bonne si la pile testée est bonne ?

1.5.b) Même question si la pile testée est usagée.

On teste l'ensemble de la boîte par la méthode suivante : les piles sont tirées les unes après les autres au hasard sans remise. À chaque tirage, on teste la pile courante, le protocole s'arrête lorsque l'on a sorti les 4 piles usagées.

1.5.c) Quelle est la probabilité que le test s'arrête au cinquième tirage (au dixième tirage) ?

EXERCICE 1.6 (Formule de Bayes).-

On suppose que l'on dispose d'un test déterminant d'une maladie donnée. Malheureusement, comme tout test, celui-ci est faillible : 1% des individus que l'on sait sains sont positifs au test et 2% des individus que l'on sait malades sont négatifs au test. On suppose que la maladie atteint 1% de la population testée. Quelle est la probabilité qu'un individu réagissant positivement au test soit effectivement malade ?

Pour aller plus loin

EXERCICE 1.7.-

Lors d'un bal, n couples dansent. Les cavaliers ont choisi leur cavalière aléatoirement. Quelle est la probabilité qu'aucun des couples d'origine ne soit réuni ?

EXERCICE 1.8 (Arnaque ou pas?).-

Dans le jeu « Vegas », il est vendu 500 000 tickets à 3€ chaque. Ces tickets sont distribués aux buralistes sous forme de bandes de 50 tickets attachés les uns aux autres. La répartition des gains est la suivante :

Nb de lots	Gains
1	40 000€
1	20 000€
2	10 000€
5	1 000€
18	500€
800	200€
850	100€
2 020	50€
4 000	20€
9 000	10€
28 000	6€
25 000	4€
47 500	3€

- 1.8.a) Quel est le montant moyen des gains ?
- 1.8.b) Quelle est la probabilité d'avoir un lot supérieur à 20 € ?
- 1.8.c) Sur 50 tickets, quelle est la probabilité (exacte et approchée) d'avoir 0 ou 1 lot supérieur à 20 € ?
- 1.8.d) M. R. a acheté 100 bandes de 50 tickets et il a constaté qu'aucune d'entre elles ne comportait plus d'un lot supérieur à 20 €. Quelle est la probabilité (approchée, en supposant que 5 000 est négligeable devant 500 000) d'un tel événement ?

EXERCICE 1.9.–

Soient \mathbf{P} et \mathbf{Q} deux mesures de probabilité sur \mathbf{N} . On note $p_i = \mathbf{P}(\{i\})$ et $q_i = \mathbf{Q}(\{i\})$. On définit la distance en *variation totale* entre \mathbf{P} et \mathbf{Q} par

$$d_{TV}(\mathbf{P}, \mathbf{Q}) = \sup_{A \in \mathbf{N}} |\mathbf{P}(A) - \mathbf{Q}(A)|.$$

- 1.9.a) Montrer que

$$\sum_{i=0}^{+\infty} (p_i - q_i)^+ = \frac{1}{2} \sum_{i=0}^{+\infty} |p_i - q_i|.$$

On pourra utiliser le fait $\sum_i p_i = \sum_i q_i = 1$.

- 1.9.b) Montrer que pour toute partie A de \mathbf{N} ,

$$|\mathbf{P}(A) - \mathbf{Q}(A)| \leq \sum_{i=0}^{+\infty} (p_i - q_i)^+.$$

- 1.9.c) En choisissant convenablement l'ensemble A , montrer que

$$d_{TV}(\mathbf{P}, \mathbf{Q}) = \frac{1}{2} \sum_{i=0}^{+\infty} |p_i - q_i|.$$

On suppose maintenant que \mathbf{P} est donnée par $\mathbf{P}(\{0\}) = p = 1 - \mathbf{P}(\{1\})$ et que \mathbf{Q} est une mesure de Poisson de paramètre $\lambda = -\ln(p)$, c'est-à-dire que

$$q_i = e^{-\lambda} \frac{\lambda^i}{i!}.$$

- 1.9.d) Calculer $d_{TV}(\mathbf{P}, \mathbf{Q})$.

EXERCICE 1.10.–

On lance deux dés. Soient les événements : $A = \text{« le premier dé affiche un résultat pair »}$, $B = \text{« le deuxième dé affiche un résultat pair »}$, $C = \text{« la somme des deux dés est paire »}$. Montrer que A , B , C sont deux à deux indépendants, mais ne forment pas une famille indépendante (on montrera que $\mathbf{P}(A \cap B \cap C) \neq \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C)$).

2. Variables aléatoires discrètes

2.1. Loi d'une variable discrète

2.1.1. Définitions

On se donne un univers Ω au plus dénombrable, équipé d'une probabilité \mathbf{P} . De manière informelle, une variable aléatoire discrète est une grandeur à valeur dans un ensemble discret E qui dépend du résultat de l'expérience. C'est donc une *fonction* de l'issue ω (en ce sens, la terminologie de *variable* est assez malencontreuse). On la notera souvent :

$$\begin{aligned} X : \Omega &\rightarrow E \\ \omega &\mapsto X(\omega), \end{aligned}$$

où Ω est l'univers, supposé muni d'une probabilité \mathbf{P} et où E est un ensemble **au plus dénombrable**.

EXEMPLE 5.— Considérons un lancer de n dés. Une issue ω est un n -uplet sur $\Omega = \{1, \dots, 6\}^n$. On peut par exemple définir la variable aléatoire $X(\omega)$ qui est égale au nombre de « 6 » obtenus : c'est bien une fonction de ω .

Nous nous intéressons à la probabilité la forme « la variable X vaut x » ou, plus généralement,

« la variable X appartient à l'ensemble H » = $\{\omega \in \Omega : X(\omega) \in H\} = X^{-1}(H)$.

Nous utiliserons souvent les notations $[X \in H]$ ou $\{X \in H\}$ pour désigner l'événement $\{\omega \in \Omega : X(\omega) \in H\}$.

Définition 2.1. On appelle *loi de la v.a.* X la fonction définie pour tout $H \subset E$ par :

$$\mathbf{P}_X(H) = \mathbf{P}(X^{-1}(H)).$$

Avec une écriture plus compacte : $\mathbf{P}_X := \mathbb{P} \circ X^{-1}$ où \circ représente la composition.
Nous pouvons écrire de manière équivalente mais sans doute plus « parlante » :

$$\mathbf{P}_X(H) := \mathbb{P}[X \in H].$$

Autrement dit, $\mathbf{P}_X(H)$ est la probabilité pour que X appartienne à H .

Proposition 2.1. \mathbf{P}_X est une probabilité sur E .

Démonstration. On vérifie les axiomes i) et ii) que doit satisfaire une mesure de probabilité.

i) $\mathbf{P}_X(E) = \mathbb{P}(X^{-1}(E)) = \mathbb{P}(\Omega) = 1$ et $\mathbf{P}_X(\emptyset) = \mathbb{P}(X^{-1}(\emptyset)) = \mathbb{P}(\emptyset) = 0$.

ii) Soit $(H_n)_{n \in \mathbb{N}^*}$ une famille d'éléments de E deux à deux disjoints. On a $X^{-1}(\bigcup_n H_n) = \bigcup_n X^{-1}(H_n)$ et on montre aisément que les événements $(X^{-1}(H_n))$ sont deux à deux disjoints. Ainsi en appliquant \mathbb{P} aux deux membres de l'égalité précédente, on obtient $\mathbf{P}_X(\bigcup_n H_n) = \mathbb{P}(\bigcup_n X^{-1}(H_n)) = \sum_n \mathbb{P}(X^{-1}(H_n))$. \square

On sait grâce au paragraphe 1.2.3 que \mathbf{P}_X est entièrement caractérisée par la valeur qu'elle prend sur les singletons. Afin d'alléger les notations, nous écrirons $\mathbf{P}_X(x)$ au lieu de $\mathbf{P}_X(\{x\})$, soit :

$$\mathbf{P}_X(x) = \mathbf{P}[X = x].$$

Proposition 2.2. Pour toute partie H de l'ensemble d'arrivée E , on a :

$$\mathbf{P}[X \in H] = \sum_{x \in H} \mathbf{P}_X(x).$$

EXEMPLE 6.- Soit X le nombre de « 6 » obtenus lorsqu'on lance n dés. La probabilité d'obtenir au plus deux « 6 » s'écrit : $\mathbf{P}[X \in \{0, 1, 2\}] = \mathbf{P}_X(0) + \mathbf{P}_X(1) + \mathbf{P}_X(2)$.

2.1.2. Loi jointe, lois marginales

Soient X, Y deux v.a.d. de Ω dans E de lois respectives \mathbf{P}_X et \mathbf{P}_Y . L'application :

$$\begin{aligned} (X, Y) &: \Omega \rightarrow E \times E \\ \omega &\mapsto (X(\omega), Y(\omega)) \end{aligned}$$

définit une v.a.d. sur $E \times E$.

Définition 2.2. La loi du couple (X, Y) est appelée la *loi jointe* de X et Y , notée $\mathbf{P}_{X,Y}$. Les lois \mathbf{P}_X et \mathbf{P}_Y sont appelées les *lois marginales* de X et Y respectivement.

D'après ce qui précède, la loi jointe est définie pour tout $(x, y) \in E \times E$ par $\mathbf{P}_{X,Y}(x, y) = \mathbf{P}[(X, Y) = (x, y)]$ soit :

$$\mathbf{P}_{X,Y}(x, y) = \mathbf{P}[X = x, Y = y].$$

Proposition 2.3. La loi marginale \mathbf{P}_X est liée à la loi jointe par :

$$\forall x \in E, \mathbf{P}_X(x) = \sum_{y \in E} \mathbf{P}_{X,Y}(x, y).$$

Démonstration. La famille d'événements de la forme $[Y = y]$ où y décrit E , forme une partition de l'univers Ω . D'après la formule des probabilités totales, $\mathbf{P}[X = x] = \sum_y \mathbf{P}[X = x, Y = y]$. \square

Généralisation au cas d'une famille finie de v.a.d.

Soient X_1, \dots, X_n des v.a. de $\Omega \rightarrow E$. Le n -uplet (X_1, \dots, X_n) définit une v.a.d. sur E^n . Sa loi est appelée la *loi jointe* de X_1, \dots, X_n , notée $\mathbf{P}_{X_1, \dots, X_n}$. Pour tout k , la loi \mathbf{P}_{X_k} est appelée la loi marginale de X_k .

Proposition 2.4. Pour tout $k = 1, \dots, n$ et tout $x_k \in E$,

$$\mathbf{P}_{X_k}(x_k) = \sum_{x_1 \dots x_{k-1}, x_{k+1} \dots x_n} \mathbf{P}_{X_1, \dots, X_n}(x_1, \dots, x_n),$$

où la somme s'étend sur l'ensemble des $(n - 1)$ -uplets $(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n)$ sur E .

Ainsi, à partir de la loi jointe, on peut déduire les lois marginales en éliminant les variables non-souhaitées par sommation sur toutes les valeurs possibles prises par celles-ci.

Définition 2.3. Une famille $(X_i)_{i \in I}$ de variables aléatoires sur le même espace E est dite identiquement distribuée si toutes les variables ont la même loi : $\forall i \in I$, $\mathbf{P}_{X_i} = \mathbf{P}_{X_1}$.

REMARQUE. Il est évident que deux v.a. X et Y différentes peuvent avoir la même loi ($\mathbf{P}_X = \mathbf{P}_Y$). Par exemple, si $X \in \{0, 1\}$ suit une loi de Bernoulli de paramètre $1/2$, alors la v.a. $Y = 1 - X$ est différente de X et suit néanmoins la même loi :

$$\mathbf{P}[Y = 0] = \mathbf{P}[1 - X = 0] = \mathbf{P}[X = 1] = \frac{1}{2} = \mathbf{P}[X = 0].$$

2.2. Indépendance des v.a. discrètes

Soient X et Y deux v.a.d. à valeurs dans E .

Définition 2.4. X et Y sont dites *indépendantes* si pour tout $G, H \subset E$, les événements $[X \in G]$ et $[Y \in H]$ sont indépendants, autrement dit si :

$$\mathbf{P}[X \in G, Y \in H] = \mathbf{P}[X \in G]\mathbf{P}[Y \in H],$$

où $[X \in G, Y \in H]$ désigne l'ensemble $[X \in G] \cap [Y \in H]$.

Proposition 2.5. Deux v.a.d X et Y sont indépendantes si et seulement si pour tout $(x, y) \in E^2$,

$$\mathbf{P}_{X,Y}(x, y) = \mathbf{P}_X(x) \mathbf{P}_Y(y) \quad (2.1)$$

REMARQUE.– Par définition, X et Y sont indépendantes lorsque les événements $[X \in H]$ et $[Y \in G]$ sont indépendants quel que soit le choix de H et G . Le résultat ci-dessus montre qu'il suffit de vérifier cette propriété sur les singletons $H = \{x\}$ et $G = \{y\}$.

Démonstration. Le sens \Rightarrow est immédiat. On montre la réciproque. On a pour tout $H, G \subset E$, $\mathbf{P}[X \in H, Y \in G] = \mathbf{P}_{X,Y}(H \times G)$. Comme $\mathbf{P}_{X,Y}$ est une mesure de probabilité sur un espace discret, $\mathbf{P}_{X,Y}(H \times G) = \sum_{(x,y) \in H \times G} \mathbf{P}_{X,Y}(x, y)$. En appliquant l'hypothèse, $\mathbf{P}_{X,Y}(H \times G) = \sum_{(x,y) \in H \times G} \mathbf{P}_X(x) \mathbf{P}_Y(y) = \sum_{x \in H} \mathbf{P}_X(x) \sum_{y \in G} \mathbf{P}_Y(y) = \mathbf{P}_X(H) \mathbf{P}_Y(G)$, ce qui prouve le résultat. \square

Soit E' un autre espace discret et soient $f, g : E \rightarrow E'$. On désigne par $f(X)$ la v.a.d. $\omega \mapsto f(X(\omega))$, c'est à dire $f(X) = f \circ X$.

Proposition 2.6. Si X et Y sont indépendantes, alors $f(X)$ et $g(Y)$ sont des v.a. indépendantes.

Démonstration. Soient H, G deux parties de E' . Les ensembles $[f(X) \in H]$ et $[g(Y) \in G]$ s'écrivent respectivement $[X \in f^{-1}(H)]$ et $[Y \in g^{-1}(G)]$ et sont donc indépendants. \square

Généralisation au cas d'une famille finie de v.a.d.

Soient X_1, \dots, X_n des v.a.d. sur E . Elles sont dites indépendantes si pour toute suite d'ensembles (H_1, \dots, H_n) , les événements $([X_k \in H_k])_{k=1, \dots, n}$ sont indépendants. Autrement dit,

$$\mathbb{P}[X_1 \in H_1, \dots, X_n \in H_n] = \mathbb{P}[X_1 \in H_1] \times \dots \times \mathbb{P}[X_n \in H_n],$$

où l'on utilise la notation $[X_1 \in H_1, \dots, X_n \in H_n] = \bigcap_k [X_k \in H_k]$.

Théorème 2.7. X_1, \dots, X_n sont indépendantes si et seulement si $\forall (x_1, \dots, x_n) \in E^n$,

$$\mathbf{P}_{X_1 \dots X_n}(x_1, \dots, x_n) = \prod_{k=1}^n \mathbf{P}_X(x_k).$$

La preuve suit le même principe que dans le cas $n = 2$ traité plus haut.

Définition 2.5. Une famille de variables aléatoires est dite indépendante si toute sous-famille *finie* est indépendante.

n.b. : on utilise souvent l'abréviation *i.i.d.* pour désigner une famille *indépendante* et *identiquement distribuée* de variables aléatoires.

Proposition 2.8. Soit $(X_i)_{i \in I}$ une famille indépendante de v.a., chacune étant à valeur dans un espace E_i . On se donne pour tout i une application mesurable f_i sur E_i . Alors la famille de v.a. $(f_i(X_i))_{i \in I}$ est indépendante.

2.3. Espérance, moments

2.3.1. Introduction

Un joueur de « pile ou face » gagne 10 euros lorsque la pièce tombe sur `pile` et perd 5 euros lorsqu'elle tombe sur `face`. Soit X le gain réalisé après l'expérience. X peut prendre deux valeurs : $a = 10$ ou $b = -5$. On définit l'*espérance* du gain par

$$\begin{aligned}\mathbb{E}(X) &= a \mathbf{P}[X = a] + b \mathbf{P}[X = b] \\ &= 10 \cdot \frac{1}{2} + (-5) \cdot \frac{1}{2} = 2,5 \text{ euros.}\end{aligned}$$

L'espérance est donc une *moyenne pondérée* des gains. D'un point de vue physique, c'est le *centre de gravité* des points a et b auxquel on a affecté les masses $\mathbf{P}[X = a]$ et $\mathbf{P}[X = b]$ respectivement.

Imaginons que le joueur précédent effectue n lancers de pièce : on note X_1, \dots, X_n les gains respectifs réalisés à chaque expérience. La *moyenne empirique* des gains est définie par

$$S_n = \frac{1}{n} \sum_{k=1}^n X_k .$$

Nous verrons à la fin de ce chapitre un résultat important appelé la *loi des grands nombres* que nous énonçons pour l'instant de manière informelle : la moyenne empirique S_n converge vers l'espérance lorsque $n \rightarrow \infty$. Naturellement, il conviendrait de préciser de quelle « convergence » il est question (n'oublions pas que l'on parle ici de variables aléatoires et non d'une simple suite de nombres). Mais cette remarque donne une première illustration de l'importance de l'espérance en probabilité.

2.3.2. Définition

On suppose dorénavant que E est une partie au plus dénombrable de \mathbb{R} .

Définition 2.6. On définit l'espérance $\mathbb{E}(X)$ d'une v.a.d. X par

$$\begin{aligned}\mathbb{E}(X) &= \sum_{x \in E} x \mathbf{P}[X = x] \\ &= \sum_{x \in E} x \mathbf{P}_X(x) .\end{aligned}\tag{2.2}$$

Pour que cette somme ait un sens, il suffit que l'une de ces deux conditions soit

vérifiée :

1. ses termes sont tous positifs : $\mathbf{P}_X(x) = 0$ pour tout $x < 0$ (auquel cas $\mathbb{E}(X)$ peut éventuellement être égal à $+\infty$) ;
2. ses termes sont absolument sommables, c'est à dire :

$$\sum_{x \in E} |x| \mathbf{P}[X = x] < \infty . \quad (2.3)$$

Lorsque la première condition est vraie *i.e.*, $\mathbf{P}_X(x) = 0$ pour tout $x < 0$, nous dirons que la v.a.d. X est *positive presque partout* et nous noterons $X \geq 0$ p.p.

Soulignons que l'espérance $\mathbb{E}(X)$ est une constante, elle ne dépend pas de l'issue ω . Elle ne dépend de X qu'au travers de sa loi \mathbf{P}_X . En particulier, deux v.a. identiquement distribuées ont même espérance.

Une variable d'espérance nulle est dite *centrée*.

2.3.3. Propriétés

Soient E, F deux espaces discrets avec $F \subset \mathbb{R}$. Soit $f : E \rightarrow F$ une fonction. La composée $f(X)$ définit une nouvelle variable aléatoire $\omega \mapsto f(X(\omega))$. Nous nous intéressons à son espérance.

Proposition 2.9. Si f est positive,

$$\mathbb{E}(f(X)) = \sum_{x \in E} f(x) \mathbf{P}[X = x] . \quad (2.4)$$

La formule reste vraie pour f quelconque pourvu que $\sum_{x \in E} |f(x)| \mathbf{P}[X = x] < \infty$.

Démonstration. Donnons d'abord la preuve pour f positive :

$$\begin{aligned} \mathbb{E}(f(X)) &= \sum_{y \in F} y \mathbf{P}[f(X) = y] = \sum_{y \in F} y \mathbf{P}[X \in f^{-1}(\{y\})] \\ &= \sum_{y \in F} \sum_{x \in f^{-1}(\{y\})} y \mathbf{P}[X = x] \\ &= \sum_{y \in F} \sum_{x \in f^{-1}(\{y\})} f(x) \mathbf{P}[X = x] \\ &= \sum_{x \in E} f(x) \mathbf{P}[X = x] , \end{aligned} \quad (2.5)$$

où on a utilisé le fait que les ensembles de la forme $f^{-1}(\{y\})$ sont une partition de E . Dans le cas où f n'est pas positive, on doit d'abord vérifier que $\mathbb{E}(f(X))$ est bien définie. En appliquant le résultat déjà démontré à la fonction « valeur absolue », nous avons :

$$\sum_{y \in F} |y| \mathbf{P}[f(X) = y] = \mathbb{E}(|f(X)|)$$

et en appliquant le résultat à la fonction $|f|$, $\mathbb{E}(|f(X)|) = \sum_{x \in E} |f(x)| \mathbf{P}[X = x]$ qui est fini par hypothèse. L'équivalent pour la v.a. $f(X)$ de la condition (2.3) est satisfaite : $\mathbb{E}(f(X))$ est bien définie. La preuve de (2.4) est obtenue par le même calcul qu'en (2.5). \square

REMARQUE.— Si l'on devait évaluer $\mathbb{E}(f(X))$ en utilisant la définition (2.2) de l'espérance, on devrait au préalable calculer la loi $\mathbf{P}_{f(X)}$ de la v.a.d. $f(X)$. L'équation (2.4) montre que l'espérance $\mathbb{E}(f(X))$ s'exprime directement en fonction de la loi de la variable X .

Lorsqu'on choisit pour f l'indicatrice d'un ensemble H , on a le corollaire suivant :

$$\mathbb{E}(\mathbf{1}_H(X)) = \mathbf{P}_X(H) . \quad (2.6)$$

La propriété (2.4) permet d'écrire la condition de sommabilité (2.3) de manière plus compacte : on écrira simplement $\mathbb{E}|X| < \infty$.

Soient deux X et Y deux v.a.d. sur $E \subset \mathbb{R}$. Pour tous coefficients réels α, β , la somme $\alpha X + \beta Y$ est bien une v.a.d. en tant que fonction du couple (X, Y) . Dans la suite, on utilisera la notation « p.p. » pour « presque partout », c'est-à-dire avec probabilité 1. Par exemple, on écrira que « $X = a$ p.p. » pour signifier que $\mathbf{P}(X = a) = 1$.

Proposition 2.10. Soient X et Y deux variables aléatoires dans un ensemble $E \subset \mathbb{R}$ discret. Supposons que $\mathbb{E}|X| < \infty$ et $\mathbb{E}|Y| < \infty$. Soient $(\alpha, \beta) \in \mathbb{R}^2$ et $a \in E$. Alors :

- a) $\mathbb{E}(\alpha X + \beta Y)$ est bien définie et $\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}(X) + \beta \mathbb{E}(Y)$.
- b) Si $X \geq 0$ p.p., alors $\mathbb{E}(X) \geq 0$.
- c) Si $X \geq 0$ p.p. et si $\mathbb{E}(X) = 0$, alors $X = 0$ p.p.
- d) $|\mathbb{E}(X)| \leq \mathbb{E}|X|$.
- e) Si $X \leq Y$ p.p., alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$.
- f) Si $X = a$ p.p., alors $\mathbb{E}(X) = a$.

Démonstration. Montrons que $\mathbb{E}(\alpha X + \beta Y)$ est bien définie. D'après la propriété précédente,

$$\begin{aligned} \mathbb{E}|\alpha X + \beta Y| &= \sum_{(x,y) \in E^2} |\alpha x + \beta y| \mathbf{P}_{X,Y}(x,y) \\ &\leq |\alpha| \sum_{(x,y)} |x| \mathbf{P}_{X,Y}(x,y) + |\beta| \sum_{(x,y)} |y| \mathbf{P}_{X,Y}(x,y) \\ &= |\alpha| \sum_x |x| \sum_y \mathbf{P}_{X,Y}(x,y) + |\beta| \sum_y |y| \sum_x \mathbf{P}_{X,Y}(x,y) \\ &= |\alpha| \sum_x |x| \mathbf{P}_X(x) + |\beta| \sum_y |y| \mathbf{P}_Y(y) . \end{aligned}$$

Ainsi, $\mathbb{E}|\alpha X + \beta Y| \leq |\alpha| \mathbb{E}|X| + |\beta| \mathbb{E}|Y| < \infty$ par hypothèse. On évalue l'espérance :

$$\begin{aligned} \mathbb{E}(\alpha X + \beta Y) &= \sum_{(x,y)} (\alpha x + \beta y) \mathbf{P}_{X,Y}(x,y) \\ &= \alpha \sum_{(x,y)} x \mathbf{P}_{X,Y}(x,y) + \beta \sum_{(x,y)} y \mathbf{P}_{X,Y}(x,y) , \end{aligned}$$

où la dernière égalité se justifie par le fait que les deux dernières sommes convergent absolument (nous l'avons prouvé plus haut). Par le même calcul que ci-dessus, ces deux sommes sont égales à $\mathbb{E}(X)$ et $\mathbb{E}(Y)$ respectivement, ce qui démontre a). Les preuves des autres propositions sont laissées au lecteur. \square

2.3.4. Inégalités

Proposition 2.11. (Inégalité de Markov) Pour tout $\epsilon > 0$, $p \geq 1$,

$$\mathbf{P} [|X| > \epsilon] \leq \frac{\mathbb{E}(|X|^p)}{\epsilon^p} .$$

Démonstration. On donne d'abord la preuve pour $\epsilon = p = 1$ et $X \geq 0$. D'après (2.6), $\mathbf{P}[X > 1] = \mathbb{E}(\mathbf{1}_{[1,+\infty]}(X)) \leq \mathbf{E}(X)$ car $\mathbf{1}_{[1,+\infty]}(X) \leq X$. Dans le cas général, on utilise le fait que $\mathbf{P}[|X| > \epsilon] = \mathbf{P}[|X|^p/\epsilon^p > 1]$ et on applique le résultat précédent. \square

Lorsque $p = 2$, l'inégalité de Markov est aussi connue sous le nom d'inégalité de Bienaymé-Tchebychev.

Proposition 2.12. (Inégalité de Cauchy-Schwarz)

$$\mathbf{E}(|XY|) \leq \sqrt{\mathbf{E}(X^2) \mathbf{E}(Y^2)} .$$

Démonstration. Si $\mathbf{E}(X^2) = 0$, la v.a. X^2 est nulle p.p. donc $XY = 0$ p.p. ce qui implique que le membre de gauche est nul. L'inégalité est triviale dans ce cas. Le seul cas non-trivial est celui pour lequel $\mathbf{E}(X^2) \neq 0$ et $\mathbf{E}(Y^2) \neq 0$.

On utilise l'inégalité $U^2 + V^2 \geq 2UV$ en posant $U = |X|/\sqrt{\mathbf{E}(X^2)}$ et $V = |Y|/\sqrt{\mathbf{E}(Y^2)}$. Comme $\mathbf{E}(U^2) = \mathbb{E}(V^2) = 1$, on obtient en prenant l'espérance de chaque membre de l'inégalité : $1 + 1 \geq 2\mathbf{E}(|XY|/\sqrt{\mathbf{E}(X^2) \mathbf{E}(Y^2)})$ ce qui démontre le résultat. \square

2.3.5. Moments, variance, covariance

Définition 2.7. Soit $p \geq 0$. Soit une v.a.d. réelle X telle que $\mathbb{E}(|X|^p) < \infty$. La quantité $\mathbb{E}(X^p)$ est appelée le *moment d'ordre p* de X .

On dit d'une telle variable qu'elle est d'*ordre p*, ou qu'elle *possède un moment d'ordre p*.

REMARQUE.– Le moment d'ordre 1 coïncide avec l'espérance. Une variable bornée possède tous ses moments.

Proposition 2.13. Une variable d'ordre p possède tous ses moments d'ordre inférieur.

Démonstration. Soit $0 \leq q \leq p$. De l'inégalité $|x|^q \leq 1 + |x|^p$, on déduit $\mathbb{E}|X|^q \leq 1 + \mathbb{E}|X|^p < \infty$. \square

Définition 2.8. La *variance* d'une v.a.d. X d'ordre 2 est définie par

$$\text{Var}(X) := \mathbb{E}((X - \mathbb{E}(X))^2).$$

Son *écart-type* est la racine carrée de la variance, noté $\sigma_X := \sqrt{\text{Var}(X)}$.

EXEMPLE 7. Un joueur lance une pièce, gagne un euro si le résultat est pile, perd un euro sinon. L'espérance du gain X est $\mathbb{E}(X) = 0$. La variance est $\text{Var}(X) = 1 \times \frac{1}{2} + 1 \times \frac{1}{2} = 1$ et l'écart-type est 1. Si le joueur gagne ou perd 10 euros à chaque partie, l'espérance de gain est toujours nulle. En revanche, la variance vaut 100 et l'écart type vaut 10. La variance donne donc une information sur l'amplitude des fluctuations de la X autour de son espérance.

EXEMPLE 8. La variance d'une loi de Bernoulli $B(p)$ vaut $p(1-p)$.

Définition 2.9. Soient X et Y deux v.a.d. d'ordre 2. Leur *covariance* est définie par :

$$\text{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))].$$

L'inégalité de Cauchy-Schwarz garantit que la quantité ci-dessus est bien définie. En statistique et en traitement du signal, on utilise souvent une version renormalisée de la covariance, le *coefficient de corrélation* qui est défini par :

$$\rho_{X,Y} := \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}.$$

Lorsque $\text{Cov}(X, Y) = 0$, on dit que X et Y sont *décorrélées*.

Proposition 2.14. Soient X et Y deux v.a.d. d'ordre 2 et $(\alpha, \beta) \in \mathbb{R}^2$. On a :

- a) $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$;
- b) $\text{Cov}(X, X) = \text{Var}(X)$;
- c) $\text{Cov}(Y, X) = \text{Cov}(X, Y)$;
- d) $\text{Var}(\alpha X + \beta) = \alpha^2 \text{Var}(X)$;
- e) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$.

La preuve est laissée à titre d'exercice.

2.3.6. Cas des variables indépendantes

Proposition 2.15. Soient X et Y des v.a. indépendantes telles que $\mathbf{E}|X|, \mathbf{E}|Y| < \infty$. Alors $\mathbf{E}|XY| < \infty$, et on a l'égalité :

$$\mathbb{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y) .$$

Démonstration. $\mathbb{E}|XY| = \sum_{(x,y)} |xy|\mathbf{P}_{X,Y}(x,y)$ et comme X et Y sont indépendantes, $\mathbf{P}_{X,Y}(x,y) = \mathbf{P}_X(x)\mathbf{P}_Y(y)$. Ainsi, $\mathbb{E}|XY| = \sum_x |x|\mathbf{P}_X(x) \sum_y |y|\mathbf{P}_Y(y) = \mathbf{E}|X|\mathbf{E}|Y| < \infty$. Le même calcul, sans les valeurs absolues, montre que $\mathbb{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$. \square

Cette propriété admet une généralisation immédiate. Si X et Y sont indépendantes, on sait que pour des fonctions f et g arbitraires, les v.a.d. $f(X)$ et $g(Y)$ restent indépendantes. Par conséquent,

$$\mathbb{E}(f(X)g(Y)) = \mathbf{E}(f(X))\mathbf{E}(g(Y)) , \quad (2.7)$$

dès lors que les deux sommes du membre de droite sont absolument convergentes. On a même une réciproque à ce résultat.

Théorème 2.16. Deux variables aléatoires $X : \Omega \rightarrow (E, \mathbf{E})$ et $Y : \Omega \rightarrow (F, \mathcal{F})$ sont indépendantes si et seulement pour toutes les fonctions mesurables bornées $f : E \rightarrow \mathbf{R}$ et $g : F \rightarrow \mathbf{R}$,

$$\mathbf{E}[f(X)g(Y)] = \mathbf{E}[f(X)]\mathbf{E}[g(Y)]. \quad (2.8)$$

Démonstration. On vient de voir que l'indépendance implique (2.8).

Réciproquement, si l'équation (2.8) est vérifiée, alors quels que soient $x \in E$ et $y \in F$, on obtient (2.1) en spécialisant (2.8) pour $f = \mathbf{1}_{\{x\}}$ et $g = \mathbf{1}_{\{y\}}$. \square

Un cas particulier intéressant est obtenu en posant $f(x) = x - \mathbb{E}(X)$ et $g(y) = y - \mathbb{E}(Y)$. Dans ce cas, le membre de gauche de (2.7) n'est autre que la covariance $\text{Cov}(X, Y)$ et les deux facteurs du membre de droite sont nuls. On en déduit la propriété suivante :

Proposition 2.17. Si X et Y sont indépendantes et d'ordre 2, alors

$$\text{Cov}(X, Y) = 0 .$$

Cette propriété implique en particulier que pour des v.a. indépendantes :

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) . \quad (2.9)$$

Notons que deux variables décorrélées ne sont pas nécessairement indépendantes. L'exercice 2.4 permet de s'en convaincre.

Généralisation au cas d'une famille finie de v.a.d.

Proposition 2.18. Pour tout $k = 1, \dots, n$, soit X_k une v.a. sur un espace discret E_k et $f_k : E_k \rightarrow E'_k$ une fonction sur $E'_k \subset \mathbb{R}$ telle que $\mathbb{E}|f_k(X_k)| < \infty$. On suppose X_1, \dots, X_n indépendantes. Alors,

$$\mathbf{E} \left(\prod_{k=1}^n f_k(X_k) \right) = \prod_{k=1}^n \mathbf{E}(f_k(X_k))$$

Proposition 2.19. Si X_1, \dots, X_n sont des v.a.d. indépendantes d'ordre 2, alors

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n).$$

Démonstration. La propriété est vraie au rang $n = 1$. Supposons la vraie au rang $n - 1$. Posons $Z_n = X_1 + \dots + X_{n-1}$. Les v.a. X_n et Z_n sont indépendantes. Par l'égalité (2.9), $\text{Var}(X_n + Z_n) = \text{Var}(X_n) + \text{Var}(Z_n)$ or $\text{Var}(Z_n) = \text{Var}(X_1) + \dots + \text{Var}(X_{n-1})$ par l'hypothèse de récurrence. La propriété est donc démontrée. \square

2.3.7. Application : Loi faible des grands nombres *

Soit $(X_k)_{k \in \mathbb{N}^*}$ une famille indépendante et identiquement distribuée de v.a. sur un ensemble $E \subset \mathbb{R}$ au plus dénombrable. On s'intéresse au comportement de la moyenne empirique des n premières variables :

$$S_n = \frac{1}{n} \sum_{k=1}^n X_k.$$

Théorème 2.20. Soit $(X_k)_{k \in \mathbb{N}^*}$ une famille indépendante, identiquement distribuée de v.a.d.. On suppose que $\mathbf{E}(X_1^2) < \infty$. Alors,

$$\forall \epsilon > 0, \quad \lim_{n \rightarrow \infty} \mathbf{P}[|S_n - \mathbf{E}(X_1)| > \epsilon] = 0.$$

On dit de la variable aléatoire S_n qu'elle *converge en probabilité* vers $\mathbf{E}(X_1)$.

Démonstration. En utilisant le fait que $\mathbf{E}(X_1) = \mathbf{E}(X_k)$, on a :

$$\begin{aligned} \mathbf{P}[|S_n - \mathbf{E}(X_1)| > \epsilon] &= \mathbf{P}\left[\left|\sum_k (X_k - \mathbf{E}(X_k))\right| > n\epsilon\right] \\ &\leq \frac{\mathbf{E}((\sum_k (X_k - \mathbf{E}(X_k)))^2)}{n^2 \epsilon^2}, \end{aligned}$$

en utilisant l'inégalité de Bienaymé-Tchebichev. La somme dans l'espérance est une v.a. centrée, donc son moment d'ordre 2 et sa variance coïncident. Par indépendance de X_k , sa variance satisfait : $\text{Var}(\sum_k X_k) = \sum_k \text{Var}(X_k) = n \text{Var}(X_1)$, où la seconde égalité

provient du fait que toutes les variances sont égales, les X_k étant identiquement distribués. Finalement,

$$\mathbf{P} [|S_n - \mathbf{E}(X_1)| > \epsilon] \leq \frac{n \operatorname{Var}(X_1)}{n^2 \epsilon^2},$$

et le membre de gauche converge bien vers zéro quand n tend vers l'infini. \square

Le théorème précédent se nomme « loi faible des grands nombres ». Donnons-en une illustration. Un joueur lance une pièce, gagne un euro si le résultat est pile, perd un euro sinon. Il réitère l'expérience n fois. X_k représente son gain à l'instant k et S_n la moyenne des gains. L'espérance du gain X_k est $\mathbf{E}(X_k) = 0$. La loi faible des grands nombres implique que $\mathbf{P}[|S_n| > \epsilon]$ tend vers zéro. Quel que soit ϵ aussi petit qu'on veut, le gain moyen est plus petit que ϵ avec forte probabilité lorsque n est grand.

REMARQUE.— Plus loin dans ce cours, nous étendrons la loi faible des grands nombres à des v.a. quelconques, pas nécessairement discrètes. Nous montrerons également un résultat plus puissant appelé « loi forte des grands nombres ». La loi forte établit que *quelle que soit l'issue* ω , hormis peut-être pour ω dans un ensemble de probabilité nulle, nous avons $\lim_n S_n(\omega) = \mathbf{E}(X_1)$.

2.4. Fonction génératrice d'une v.a. à valeurs entières

Dans ce paragraphe, on se limite au cas où la v.a. X est à valeurs dans \mathbb{N} (ou bien dans un sous-ensemble $E \subset \mathbb{N}$: dans ce dernier cas, on étend X à une fonction dans \mathbb{N} en imposant que $\mathbf{P}[X = k] = 0$ pour $k \notin E$).

Définition 2.10. La *fonction génératrice* de X , notée Φ_X , est définie pour tout s dans l'intervalle $[-1, +1]$ par :

$$\begin{aligned}\Phi_X(s) &= \mathbf{E}(s^X) \\ &= \sum_{k=0}^{\infty} \mathbf{P}[X = k] s^k.\end{aligned}$$

La fonction génératrice est donc la série entière de terme général $\mathbf{P}[X = k]$. Le rayon de convergence de cette série est supérieur ou égal à un.

Proposition 2.21. Pour toute v.a.d. à valeurs entières, sa fonction génératrice Φ_X satisfait les propriétés suivantes.

- a) Φ_X est continue sur $[-1, +1]$ et de classe C^∞ sur $] -1, +1 [$.
- b) Pour tout n ,

$$\mathbf{P}[X = n] = \frac{\Phi_X^{(n)}(0)}{n!},$$

où $\Phi_X^{(n)}$ est la dérivée n ème de Φ_X .

Démonstration. Il suffit d'appliquer des résultats connus sur les séries entières. On sait (voir l'annexe et [Rud95, Théorème 8.1]) qu'une série entière de terme général a_k et de rayon de convergence R est de classe C^∞ sur $] -r, r[$, et sa dérivée n ème vaut $\sum_{k \geq n} n(n-1) \cdots (n-k+1) a_k s^{n-k}$. L'application de ce résultat démontre b). Il ne reste qu'à montrer la continuité de Φ_X en ± 1 , ce qui peut être fait par un argument de convergence dominée. \square

Du deuxième résultat, on en déduit le corollaire suivant.

Corollaire 2.22. Si X et Y sont deux v.a.d. de même fonction génératrice alors $\mathbf{P}_X = \mathbf{P}_Y$, i.e. X et Y ont la même loi.

Si la fonction génératrice caractérise la loi, elle caractérise *a fortiori* les moments. La propriété suivante permet de déduire les moments de la fonction caractéristique.

Notation : Pour toute fonction f ayant un limite à gauche (resp. à droite) en b , on note $f(b^-)$ cette limite (resp. $f(b^+)$).

Théorème 2.23. Une v.a. $X : \Omega \rightarrow \mathbb{N}$ admet un moment d'ordre p si et seulement si $\Phi_X^{(p)}$ admet une limite à gauche en 1. Alors,

$$\Phi_X^{(p)}(1^-) = \mathbf{E}(X(X-1) \cdots (X-p+1)) .$$

Démonstration. On traite le cas $p = 1$, le cas général suit le même principe. Rappelons que $\Phi'_X(s) = \sum_{k \geq 1} kp_X(k) s^{k-1}$. Supposons que $\mathbf{E}(X) < \infty$. Comme $\mathbf{E}(X) = \sum_{k \geq 1} kp_X(k)$, les termes $kp_X(k) s^{k-1}$ de la série $\Phi'_X(s)$ sont dominés par une suite sommable $kp_X(k)$. Par convergence dominée, $\lim_{s \uparrow 1} \Phi'_X(s) = \sum_{k \geq 1} \lim_{s \uparrow 1} kp_X(k) s^{k-1} = \mathbf{E}(X)$.

Réciproquement, supposons que $\Phi'_X(1^-)$ existe. Comme Φ'_X est croissante sur $[0, 1[$, on a pour tout $s < 1$, $\sum_{k \geq 1} kp_X(k) s^{k-1} \leq \Phi'_X(1^-)$ et comme tous les termes sont positifs, $\sum_{k=1}^n kp_X(k) s^{k-1} \leq \Phi'_X(1^-)$ quel que soit n . En faisant $s \uparrow 1$ dans la dernière inégalité, on en déduit que la suite $(\sum_{k=1}^n kp_X(k))_n$ est bornée. C'est une suite croissante, elle est donc convergente. On a bien $\sum_{k=1}^n kp_X(k) < \infty$, autrement dit $\mathbf{E}(X) < \infty$. \square

L'exercice 2.2 fournit des exemples d'applications.

2.5. Exercices

Pour apprendre

EXERCICE 2.1.-

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite i.i.d. de loi de Bernoulli de paramètre p sur $\{0, 1\}$.

- 2.1.a) Pour tout n , caractériser la loi de $S_n = \sum_{k=1}^n X_k$.
- 2.1.b) En déduire l'espérance et la variance d'une variable binomiale de paramètres (n, p) .
- 2.1.c) On pose $Y = \min\{n : X_n = 1\}$ lorsque cet ensemble est non-vide, $Y = +\infty$ sinon.
Caractériser la loi de Y .

EXERCICE 2.2.-

On rappelle que Φ_X est la fonction génératrice de la variable aléatoire X (cf. définition 2.10).

- 2.2.a) Calculer Φ_X , $\mathbf{E}(X)$ et $\text{Var}(X)$ pour une v.a. de Bernoulli $B(p)$, une v.a. de loi géométrique $G(p)$, une v.a. de Poisson $\mathcal{P}(\lambda)$.
- 2.2.b) Soient X_1, \dots, X_n des v.a. indépendantes. X_k suit une loi de Poisson de paramètre λ_k . Caractériser la loi de $\sum_{k=1}^n X_k$.

EXERCICE 2.3.-

Soient (X_1, X_2, X_3) des variables aléatoires indépendantes de même loi à valeurs dans \mathbb{N} .

On note $p_i = \mathbf{P}(X_l = i)$, $l = 1, 2, 3$. On introduit Z de loi uniforme sur $\{1, 2\}$.

- 2.3.a) Quelle est la loi de $Y = (X_Z, X_{3-Z})$?
- 2.3.b) Soit W le vecteur aléatoire défini par :

$$W = (X_1, X_3) \text{ si } Z = 2 \text{ et } W = (X_3, X_2) \text{ si } Z = 1.$$

Quelle est la loi de W ?

EXERCICE 2.4.-

Soit X de loi uniforme sur $\{0, 1\}$ et Z de loi uniforme sur $\{-1, +1\}$ indépendante de X .

Soit $Y = ZX$. Montrer que X et Y sont décorrélées mais ne sont pas indépendantes.

EXERCICE 2.5.-

On veut calculer les moments d'une v.a. de loi hypergéométrique. On se donne donc une urne contenant r boules rouges et b boules blanches de sorte que $N = r + b$. Muni d'une épuisette à boules, on tire m boules parmi les N présentes. On range ces boules dans des cases numérotées de 1 à m . On note X le nombre de boules rouges ressorties et

$$X_i = \begin{cases} 1 & \text{si la case } i \text{ contient une boule rouge,} \\ 0 & \text{sinon.} \end{cases}$$

On a donc $X = \sum_{i=1}^m X_i$.

- 2.5.a) Soit σ une permutation de $\{1, \dots, n\}$. Pourquoi les vecteurs aléatoires (X_1, \dots, X_m) et $(X_{\sigma(1)}, \dots, X_{\sigma(n)})$ ont-ils la même loi?
- 2.5.b) Calculer $\mathbf{P}(X_i = 1)$ et $\mathbf{P}(X_i X_j = 1)$ pour $i \neq j$.
- 2.5.c) En déduire $\mathbf{E}[X]$ et $\text{Var}(X)$.

Pour s'entraîner

EXERCICE 2.6.–

Soient $1 \leq n \leq N$ deux entiers. Soit M une v.a. de loi binomiale (N, θ) et X une v.a. dont la loi est donnée par

$$\mathbf{P}(X = k \mid M = m) = \frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}} \text{ pour tout } k \in \{0, \dots, n\}.$$

- 2.6.a) Calculer $\mathbf{P}(X = k)$.
- 2.6.b) Calculer la loi de M sachant $X = k$, dite loi *a posteriori* de M .
- 2.6.c) Pour $k = 0$, identifier cette loi.

EXERCICE 2.7.–

En codage correcteur d'erreurs, les erreurs interviennent au hasard sur l'un quelconque des bits. Si on transmet des mots de n bits, on pose $\Omega = \{0, 1\}^n$, que l'on munit de la loi uniforme. On introduit $X_i(\omega) = \omega_i$ pour $i = 1, \dots, n$. La distance de Hamming entre mots de code $x = (x_1, \dots, x_n)$ et $y = (y_1, \dots, y_n)$, est définie par :

$$d(x, y) = \sum_{i=1}^n \mathbf{1}_{\{x_i \neq y_i\}}.$$

On appelle longueur d'un mot x , sa distance au mot nul $0 = (0, \dots, 0)$.

- 2.7.a) Montrer que sous la loi uniforme sur Ω , les variables $(X_i, i \in \{1, \dots, n\})$ sont indépendantes et identiquement distribuées de loi de Bernoulli de paramètre $1/2$.
- 2.7.b) Quelle est la longueur moyenne d'un mot ?
- 2.7.c) Quelle est la variance de la longueur d'un mot ?
- 2.7.d) On choisit deux mots au hasard indépendamment l'un de l'autre, soit X et Y les variables aléatoires correspondantes. Calculer

$$\mathbf{E} [d(X, Y)^2].$$

Pour aller plus loin

EXERCICE 2.8.–

Un étang contient un nombre de poissons N inconnu. Pour estimer N , on prélève un échantillon de r poissons que l'on marque et que l'on remet dans l'étang. Une semaine plus tard, un autre échantillon de $s < r$ individus est prélevé. On appelle X le nombre de poissons marqués lors du premier prélèvement qui sont aussi dans le deuxième échantillon.

- 2.8.a) Calculer la loi de X (dite loi hypergéométrique).

On note pour la suite de cet exercice

$$p_k = \frac{\binom{r}{k} \binom{N-r}{s-k}}{\binom{N}{s}},$$

pour $k \leq \min(r, s)$ et $k \geq \max(s + r - N, 0)$.

2.8.b) Montrer que $p_k^2 \geq p_{k-1}p_{k+1}$.

2.8.c) En déduire qu'il existe une unique valeur de k telle que $p_k = \max_j p_j$.

2.8.d) Soit k_0 tel cette valeur. Par définition, $p_{k_0+1} < p_{k_0}$ et $p_{k_0-1} < p_{k_0}$. En déduire que

$$k_0 = \left\lceil \frac{(r+1)(s+1)}{N+2} \right\rceil.$$

On pourra poser pour simplifier les calculs, $r' = r + 1$, $s' = s + 1$, $N' = N + 2$.

2.8.e) En déduire une estimation de N .

EXERCICE 2.9 (***, Erdős et Renyi (1960)).—

On fabrique un graphe sur n sommets en choisissant ses arêtes au hasard. Plus précisément, on considère le graphe $G_{n,p}$ obtenu en choisissant chacune des $\binom{n}{2}$ arêtes potentielles indépendamment avec probabilité p . Le but de ce problème est d'étudier la probabilité que $G_{n,p}$ soit connexe. On s'intéressera au cas où p est de la forme

$$p = p(n) = \frac{\ln n}{n} + \frac{c}{n}$$

où c est une constante fixée.

2.9.a) Soit $(X_i, 1 \leq i \leq n)$ un n -uple de variables aléatoires à valeurs dans $\{0, 1\}$ et soit $X = \sum_{i=1}^n X_i$. Montrer que pour tout r tel que $r \geq 1$ et $2r + 1 \leq n$ on a :

$$\sum_{k=0}^{2r+1} (-1)^k F^{(k)} \leq \mathbf{P}(X = 0) \leq \sum_{k=0}^{2r} (-1)^k F^{(k)}$$

où l'on a posé $F^{(0)} = 1$ et pour $k \geq 1$

$$F^{(k)} = \sum_{j_1 < j_2 < \dots < j_k} \mathbf{E}[X_{j_1} X_{j_2} \dots X_{j_k}].$$

Suggestion. On pourra montrer que

$$\mathbf{P}(X = 0) = \mathbf{E} \left[\prod_{i=1}^n (1 - X_i) \right]$$

et appliquer une formule de Taylor à la fonction $\prod_{i=1}^n (1 - x_i)$.

On dira qu'un sommet est isolé s'il n'est l'extrémité d'aucune arête. Dans un premier temps, on étudie le nombre X de sommets isolés. On peut écrire $X = \sum_{i=1}^n X_i$ où X_i est la variable aléatoire qui vaut 1 si le sommet i est isolé, 0 sinon.

2.9.b) Que valent $\mathbf{E}[X_i]$ et $\mathbf{E}[X]$?

2.9.c) On suppose dorénavant c fixé. Montrer que la quantité $F^{(k)}$, pour la variable X , converge, lorsque n tend vers l'infini, vers $e^{-ck}/k!$.

2.9.d) Montrer que $\lim_{n \rightarrow \infty} \mathbf{P}(X = 0) = e^{-e^{-c}}$.

2.9.e) Calculer l'espérance du nombre de composantes connexes à 2 sommets, et constater que celle-ci tend vers zéro quand n tend vers l'infini.

- 2.9.f) Plus généralement, soit C_t le nombre de composantes connexes à t sommets. Montrer que pour $2 \leq t \leq n/2$,

$$\mathbf{E}[C_t] \leq \frac{1}{t!} \sum_{t-1 \leq k \leq \binom{t}{2}} \binom{\binom{t}{2}}{k} \left(\frac{p}{1-p}\right)^k.$$

En déduire que la probabilité que $G_{n,p}$ soit connexe tend, quand $n \rightarrow \infty$, vers $e^{-e^{-c}}$.
On admettra que $\sum_{2 \leq t \leq n/2} \mathbf{E}[C_t] \rightarrow 0$ quand $n \rightarrow \infty$.

- 2.9.g) Que peut-on dire de la probabilité que $G_{n,p}$ soit connexe ?

EXERCICE 2.10.-

On rappelle qu'une suite de variables aléatoires $(X_n, n \in \mathbf{N})$ converge en probabilité vers la variable aléatoire X si et seulement si pour tout $\epsilon > 0$,

$$\lim_{n \rightarrow +\infty} \mathbf{P}(|X_n - X| \geq \epsilon) = 0.$$

Soit $(X_n, n \in \mathbf{N})$ une suite de v.a. de moyenne μ_n et de variance σ_n^2 . Soit $(b_n, n \in \mathbf{N})$ une suite de réels positifs tels que σ_n^2/b_n^2 tende vers 0. Montrer que

$$\frac{X_n - \mu_n}{b_n} \text{ tend vers 0 en probabilité.}$$

EXERCICE 2.11 (Borne de Chernoff).-

Soit X une v.a. de loi de Poisson de paramètre λ .

- 2.11.a) Montrer que $\mathbf{P}(X \geq \eta) = \mathbf{P}(\exp(\theta X) \geq \exp(\theta \eta))$ pour tout $\theta > 0$.
2.11.b) Montrer que, pour tout $\theta \geq 0$,

$$\mathbf{P}(X \geq \eta) \leq e^{-\eta\theta} \mathbf{E}[\exp(\theta X)]. \quad (2.10)$$

- 2.11.c) Calculer $\mathbf{E}[\exp(\theta X)]$.

- 2.11.d) Trouver θ qui minimise le terme de droite de (2.10).

- 2.11.e) Trouver K tel que $\mathbf{P}(X \geq K\lambda) \leq 0,001$.

EXERCICE 2.12.-

On veut collectionner N images dont une et une seule apparaît dans chaque tablette de chocolat achetée. Les images sont mises au hasard dans les tablettes. On appelle T_i le nombre de tablettes nécessaires avant d'avoir i images distinctes. On pose $T_0 = 0$.

- 2.12.a) Montrer que $T_{i+1} - T_i$ suit une loi géométrique de paramètre $1 - i/N$.
2.12.b) Montrer que les variables aléatoires $T_0, T_1 - T_0, \dots, T_N - T_{N-1}$ sont indépendantes dans leur ensemble.
2.12.c) Calculer l'espérance et la variance de T_N . Trouver un équivalent de l'espérance et montrer que la variance est un $O(N^2)$ quand N tend vers $+\infty$.
2.12.d) En utilisant l'exercice 2.10, montrer que $T_N/(N \log N)$ tend vers 1 en probabilité.

EXERCICE 2.13.–

Les règles du jeu du **not-seven** sont les suivantes : on part d'un score $X_0 = 0$. À chaque coup, on lance deux dés non pipés, si la somme des faces égale 7, le score retourne à 0 et la partie est terminée. Sinon, le score augmente de la somme des faces et on a le droit de rejouer ou pas. Si l'on ne rejoue pas, le score est acquis et la partie est terminée. Si l'on rejoue, on relance les deux dés avec la même règle.

- 2.13.a) Calculer la loi de la somme S des deux faces. Calculer son espérance.

On considère une suite $(S_n, n \in \mathbf{N})$ de variables aléatoires indépendantes de même loi que S .

- 2.13.b) Soit $\tau = \inf\{n \geq 1, S_n = 7\}$, trouver la loi de τ . Quelle est la moyenne de τ ?
 2.13.c) Quelle est la stratégie d'un Initié (celui qui sait le résultat du prochain lancer de dés) ?
 2.13.d) Calculer son gain moyen.
 2.13.e) On appelle X_n le score au n -ième coup en l'absence de stratégie d'arrêt. Montrer que

$$\mathbf{E}[X_{n+1} | X_n = i] = \frac{5}{6}i + \frac{35}{6},$$

où l'espérance conditionnelle par rapport à un événement B est définie comme l'espérance associée à la loi de probabilité $A \mapsto \mathbf{P}(A | B)$.

- 2.13.f) En déduire que la stratégie optimale consiste à jouer tant que l'on n'a pas atteint 35 et à s'arrêter immédiatement après avoir franchi ce seuil.
 2.13.g) Calculer numériquement le gain moyen avec cette stratégie.

EXERCICE 2.14.–

Peut-on piper deux dés de sorte que la loi de leur somme soit la loi uniforme sur $\{2, \dots, 12\}$?

EXERCICE 2.15 (, Processus de branchement).–**

Soit X_0 une v.a. à valeurs entières. Soit $(X_{n,j}, n \geq 1, 1 \leq j \leq n)$ une famille dénombrable de variables aléatoires indépendantes, de loi \mathbf{P}_{X_0} . On note Φ la fonction génératrice de \mathbf{P}_{X_0} . On considère un individu « racine » qui a un nombre X_0 de descendants. Chacun de ses descendants a un nombre aléatoire de descendant, ce nombre est indépendant de celui des autres descendants et de loi \mathbf{P}_{X_0} . On pose Z_n le nombre total d'individus au rang n .

1. Calculer la fonction génératrice de Z_n en fonction de celle de Z_{n-1} .
2. Soit $u_n = \mathbf{P}(Z_n = 0)$. Montrer que $u_n = \Phi(u_{n-1})$.
3. Trouver des conditions nécessaires et suffisantes sur \mathbf{P}_{X_0} qui garantissent que Φ est strictement convexe.
4. Montrer que u converge vers une limite non nulle si et seulement si $\mathbf{E}[X_0] < 1$.

Ce processus représente tout aussi bien l'évolution de la contamination par un virus (X_0 est le nombre d'individus contaminés par le malade initial), que la transmission d'un nom de famille (X_0 étant alors le nombre d'enfants portant le nom de leur père) et bien d'autres situations.

EXERCICE 2.16 (**).—

Dans le tri rapide (quicksort), on note M_n le nombre de comparaisons nécessaires pour ordonner un tableau de n nombres. Montrer que $\mathbf{E}[M_n]$ vérifie la relation

$$\mathbf{E}[M_n] = n - 1 + \frac{2}{n} \sum_{k=1}^{n-1} \mathbf{E}[M_k].$$

En déduire que

$$\mathbf{E}[M_n] = 2(n+1) \sum_{i=1}^{n-1} \frac{i}{(i+1)(i+2)}$$

et trouver un équivalent asymptotique de M_n quand n tend vers $+\infty$.

