

Cours de probabilité et simulation

Licence de mathématiques

Version 2.0

Christian Léonard

DÉPARTEMENT DE MATHÉMATIQUES ET INFORMATIQUE, UNIVERSITÉ PARIS OUEST.
NANTERRE.

E-mail address: leonard@u-paris10.fr

Table des matières

Chapitre 1. Fondements de la théorie des probabilités	1
1.1. Événements	1
1.2. Probabilité	3
Chapitre 2. Variables aléatoires	7
2.1. Fonction de répartition	8
2.2. Variables aléatoires discrètes	11
2.3. Variables aléatoires continues	12
2.4. Quelques éléments de réflexion	14
Chapitre 3. Loi et espérance d'une variable aléatoire	17
3.1. Variables discrètes	17
3.2. Variables continues	21
3.3. Une notation commune	23
3.4. Fonction indicatrice d'ensemble	24
3.5. Variance et écart-type	24
3.6. Moments	25
3.7. Fonctions d'une variable aléatoire	26
3.8. Égalité en loi	28
3.9. Définition abstraite de la loi d'une variable aléatoire	29
Chapitre 4. Variables aléatoires usuelles	31
4.1. Exemples de variables aléatoires discrètes	31
4.2. Exemples de variables aléatoires continues	33
Chapitre 5. Fonctions génératrices et caractéristiques	39
5.1. Le cas des variables entières	39
5.2. Fonctions caractéristiques	41
Chapitre 6. Couples aléatoires	45
6.1. Lois jointe et marginales	45
6.2. Fonction de répartition	45
6.3. Indépendance	46
6.4. Couples discrets	49
6.5. Couples continus	52
6.6. Fonctions caractéristiques	56
6.7. Inégalité de Cauchy-Schwarz	57
Chapitre 7. Fonctions d'un couple aléatoire	59
7.1. Quelques exercices corrigés	59
7.2. Somme de deux variables aléatoires indépendantes	60

Chapitre 8. Conditionnement	63
8.1. Probabilité conditionnelle	63
8.2. Conditionnement dans le cas discret	64
8.3. Conditionnement dans le cas continu	65
Chapitre 9. Indépendance (revisitée)	69
9.1. Définition	70
9.2. Propriétés élémentaires	71
9.3. Échantillons	73
Chapitre 10. Construction d'une variable aléatoire réelle générale	77
10.1. Construction d'une variable aléatoire continue uniforme	77
10.2. Construction d'une variable aléatoire réelle générale	79
Chapitre 11. Simulation d'une variable aléatoire	81
11.1. Description rapide de certains générateurs	81
11.2. Simulation. Principe et applications	81
11.3. Histogrammes	85
Chapitre 12. Convergence des variables aléatoires	89
Chapitre 13. Inégalités de convexité	91
Annexe A. Dénombrabilité	93
Annexe B. Éléments de théorie de l'intégration	97
Annexe C. Espérance mathématique sans théorie de l'intégration	101
Annexe D. Convexité	105
Index	109

CHAPITRE 1

Fondements de la théorie des probabilités

1.1. Événements

Nous commençons par présenter les fondements axiomatiques de la théorie des probabilités.

DÉFINITION 1.1. L'ensemble des *réalisations* possibles d'une *expérience* est appelé *univers* de l'expérience. Il est généralement noté Ω .

EXEMPLE 1.2. On tire une fois à pile ou face. Il est naturel de considérer $\Omega = \{p, f\}$ où p et f sont les réalisations de l'expérience qui correspondent aux tirages respectifs de *pile* et de *face*. Voici quelques événements :

- (a) la réalisation est *face*
- (b) la réalisation est *face* ou *pile*
- (c) la réalisation est *face* et *pile* simultanément
- (d) la réalisation n'est pas *face*

Ces événements peuvent être décrits respectivement par les parties A de Ω suivantes :

- (a) $A = \{f\}$
- (b) $A = \{f\} \cup \{p\} = \{f, p\} = \Omega$
- (c) $A = \{f\} \cap \{p\} = \emptyset$
- (d) $A = \{f\}^c = \{p\}$

où A^c désigne le complémentaire de la partie A dans Ω .

EXEMPLE 1.3. On lance un dé une fois. Il est naturel de considérer $\Omega = \{1, 2, 3, 4, 5, 6\}$ dont les éléments correspondent aux différentes facettes du dé. Voici quelques événements :

- (a) la réalisation est 1
- (b) la réalisation est un nombre pair
- (c) la réalisation est un nombre pair inférieur à 3
- (d) la réalisation n'est pas un nombre pair

Ces événements peuvent être décrits respectivement par les parties A de Ω suivantes :

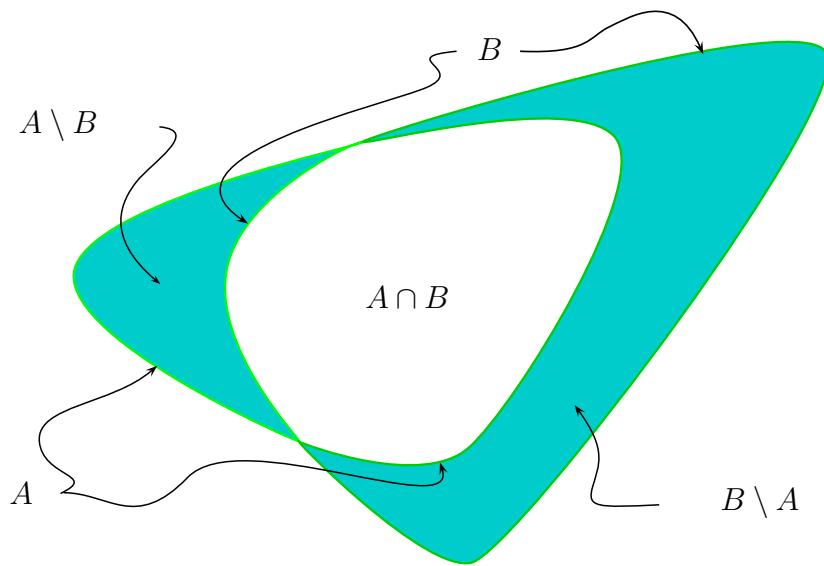
- (a) $A = \{1\}$
- (b) $A = \{2, 4, 6\}$
- (c) $A = \{2, 4, 6\} \cap \{1, 2, 3\} = \{2\}$
- (d) $A = \{2, 4, 6\}^c = \{1, 3, 5\}$

Si A et B sont des événements qui correspondent respectivement aux réalisations effectives a et b , on peut avoir besoin de considérer les événements composés :

a	\dashrightarrow	A
b	\dashrightarrow	B
non a	\dashrightarrow	A^c
a et b	\dashrightarrow	$A \cap B$
a mais pas b	\dashrightarrow	$A \setminus B$
a ou b	\dashrightarrow	$A \cup B$
a ou bien b	\dashrightarrow	$A \Delta B$

où

- $A \setminus B = A \cap B^c$ est la différence A moins B , c'est-à-dire l'ensemble des éléments qui se trouvent dans A mais pas dans B ;
- $A \Delta B = (A \cup B) \setminus (A \cap B)$ est la différence symétrique de A et B , c'est-à-dire l'ensemble des éléments qui se trouvent soit dans A , soit dans B , mais pas simultanément dans A et B .



La région colorée est $A \Delta B = (A \setminus B) \cup (B \setminus A)$. Remarquons la différence entre *ou bien* qui est exclusif et *ou* qui ne l'est pas et correspond à la réunion $A \cup B$.

Si $A \cap B = \emptyset$, on dit que les événements sont *incompatibles*, \emptyset est l'événement *impossible* et Ω est l'événement *certain*.

L'ensemble de tous les événements est noté \mathcal{A} , il est inclus dans l'ensemble de toutes les parties de Ω notée 2^Ω . Cette notation est justifiée par l'exercice suivant.

EXERCICE 1.4. En considérant l'ensemble des applications $\{\text{oui}, \text{non}\}^\Omega$ de Ω dans $\{\text{oui}, \text{non}\}$, montrer que lorsque le cardinal de Ω est n , celui de 2^Ω est 2^n .

Lorsque Ω n'est pas un ensemble dénombrable (voir la Définition A.1), pour des raisons subtiles (qui ne sont pas aisément compréhensibles au niveau de ce cours) on ne pourra pas en général prendre $\mathcal{A} = 2^\Omega$. Compte tenu de ce qui précède, \mathcal{A} doit au moins satisfaire :

$$(1) \quad A, B \in \mathcal{A} \implies A \cup B \in \mathcal{A} \text{ et } A \cap B \in \mathcal{A}$$

- (2) $A \in \mathcal{A} \implies A^c \in \mathcal{A}$
- (3) $\emptyset \in \mathcal{A}$.

EXEMPLE 1.5. On répète notre lancer de pile ou face jusqu'à ce qu'on obtienne *pile*. L'univers est alors $\Omega = \{\omega_1, \omega_2, \dots\}$ avec $\omega_1 = p$, $\omega_2 = fp$, $\omega_3 = ffp$, ... La réalisation ω_i est : "on observe *pile* pour la première fois au i -ème lancer". L'ensemble correspondant à l'événement : "l'instant de première apparition de *pile* est pair" est $A = \{\omega_2\} \cup \{\omega_4\} \cup \{\omega_6\} \cup \dots$, c'est une réunion infinie dénombrable. Cette remarque justifie la définition suivante.

DÉFINITION 1.6. Un ensemble \mathcal{A} de parties de Ω est appelée une *tribu* (ou une σ -algèbre) si

- (1) $A_1, A_2, \dots \in \mathcal{A} \implies \bigcup_{i=1}^{\infty} A_i := \{\omega \in \Omega; \exists i \geq 1, \omega \in A_i\} \in \mathcal{A}$
- (2) $A \in \mathcal{A} \implies A^c \in \mathcal{A}$
- (3) $\emptyset \in \mathcal{A}$

Les éléments de \mathcal{A} (ce sont des parties de Ω) sont appelés des *événements*.

EXEMPLE 1.7 (Exemples de tribus).

- (a) $\mathcal{A} = \{\emptyset, \Omega\}$ (c'est la plus petite tribu)
- (b) $\mathcal{A} = 2^{\Omega}$ (c'est la plus grande tribu)
- (c) Si $A \subset \Omega$, $\mathcal{A} = \{\emptyset, A, A^c, \Omega\}$.

À une expérience, on associe le couple (Ω, \mathcal{A}) où \mathcal{A} est une tribu de Ω . Dire que A est un événement, c'est dire : $A \in \mathcal{A}$.

REMARQUE 1.8.

Lorsque Ω est un ensemble dénombrable (en particulier fini), on prend toujours pour tribu $\mathcal{A} = 2^{\Omega}$: l'ensemble de toutes les parties de Ω .

1.2. Probabilité

Si on note $\mathbb{P}(A)$ la probabilité d'occurrence d'un événement $A \in \mathcal{A}$, on attend que :

- $0\% = 0 \leq \mathbb{P}(A) \leq 1 = 100\%$ (par convention)
- $\mathbb{P}(\Omega) = 1$ (condition de normalisation)
- pour tous $A, B \in \mathcal{A}$, si $A \cap B = \emptyset$ alors $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ (additivité)

Comme nous l'avons déjà remarqué, il peut être utile de considérer des événements constitués par une réunion dénombrable d'événements disjoints A_1, A_2, \dots . On note dans de cas leur réunion $\bigcup_{i=1}^{\infty} A_i = \bigsqcup_{i=1}^{\infty} A_i$ pour mettre l'emphase sur leur disjonction qui signifie : $\forall i, j, i \neq j \Rightarrow A_i \cap A_j = \emptyset$. D'où la définition suivante.

DÉFINITION 1.9. Une *mesure de probabilité* \mathbb{P} sur (Ω, \mathcal{A}) est une fonction $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ qui satisfait :

- (1) $\mathbb{P}(\Omega) = 1$
- (2) si A_1, A_2, \dots est une suite d'événements disjoints, alors :

$$\mathbb{P}\left(\bigsqcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ est appelé un *espace de probabilité*.

Il provient immédiatement de cette définition,

- en choisissant $A_1 = A_2 = \emptyset$, que $0 \leq \mathbb{P}(\emptyset) = \lim_{n \rightarrow \infty} n\mathbb{P}(\emptyset)$ et par conséquent $\mathbb{P}(\emptyset) = 0$;
- en choisissant $A_1 = A$, $A_2 = B$ et $A_3 = A_4 = \dots = \emptyset$, que pour tous $A, B \in \mathcal{A}$ disjoints, $\mathbb{P}(A \sqcup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
- Il en va de même pour toute réunion d'un nombre fini d'événements disjoints :

$$\mathbb{P}\left(\bigsqcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i).$$

EXEMPLES 1.10.

- (a) Pile ou face correspond à $\Omega = \{f, p\}$, avec $\mathcal{A} = \{\emptyset, \{f\}, \{p\}, \Omega\}$ et $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\{f\}) = \mathbb{P}(\{p\}) = 1/2$, $\mathbb{P}(\Omega) = 1$.
- (b) Un lancer de dé éventuellement pipé peut se modéliser comme suit : $\Omega = \{1, 2, \dots, 6\}$, $\mathcal{A} = 2^\Omega$ et $\mathbb{P}(\{i\}) = p_i \geq 0$, $1 \leq i \leq 6$ avec $p_1 + \dots + p_6 = 1$. Pour tout $A \subset \Omega$, nous obtenons $\mathbb{P}(A) = \sum_{i \in A} p_i$.
- (c) Si le dé est honnête, $p_1 = \dots = p_6 = 1/6$ et $\mathbb{P}(A) = \#(A)/6$ où $\#(A)$ désigne le cardinal de A .

Voici quelques conséquences immédiates de la définition de \mathbb{P} .

LEMME 1.11. *Pour tous $A, B \in \mathcal{A}$, nous avons*

- (1) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- (2) $A \subset B \implies \mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$
- (3) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

DÉMONSTRATION. Laissée en exercice. □

DÉFINITION 1.12 (Masse de Dirac). Soit $a \in \Omega$. On définit la fonction d'ensembles $\delta_a : \mathcal{A} \rightarrow \{0, 1\}$ par

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

On appelle δ_a la *masse de Dirac* au point a .

EXERCICE 1.13.

- (a) Vérifier que δ_a est une mesure de probabilité sur \mathcal{A} .
- (b) Si on prend trois éléments distincts a, b et c de Ω , alors $\mathbb{P} = \frac{1}{7}\delta_a + \frac{4}{7}\delta_b + \frac{2}{7}\delta_c$ est aussi une mesure de probabilité.
- (c) Montrer que $\mathbb{P}(\{a, b\}) = 5/7$ et calculer $\mathbb{P}(\{a, c\})$.

La mesure de probabilité $\mathbb{P} = \frac{1}{7}\delta_a + \frac{4}{7}\delta_b + \frac{2}{7}\delta_c$ de l'exercice précédent modélise l'expérience qui attribue les chances d'occurrence $1/7$, $4/7$ et $2/7$ aux réalisations élémentaires a, b et c .

EXEMPLE 1.14. On se donne une urne contenant 3 boules *rouges* appelées ω_1, ω_2 et ω_3 , 2 *bleues* appelées ω_4, ω_5 et 1 *verte* : ω_6 . On tire au hasard une boule et on note sa couleur.

On peut prendre $\Omega = \{\omega_1, \dots, \omega_6\}$ avec $\mathbb{P}(\omega_n) = 1/6$, $n = 1, \dots, 6$ puisque notre intuition nous suggère l'équiprobabilité. Bien sûr, on choisit $\mathcal{A} = 2^\Omega$ et on obtient pour tout $A \subset \Omega$, $\mathbb{P}(A) = \#(A)/6$. On constate que

$$\mathbb{P} = \sum_{n=1}^6 \frac{1}{6} \delta_{\omega_n}.$$

Notons les événements $R = \{\omega_1, \omega_2, \omega_3\}$, $B = \{\omega_4, \omega_5\}$, $V = \{\omega_6\}$ correspondant au tirage d'une boule *rouge*, *bleue* ou *verte*. On voit que $\mathbb{P}(B) = 1/6 \sum_{n=1}^6 \delta_{\omega_n}(B) = 1/6 \sum_{n=1}^6 \delta_{\omega_n}(\{\omega_4, \omega_5\}) = (0 + 0 + 0 + 1 + 1 + 0)/6 = 1/3$.

Si on n'est concerné que par la couleur de la boule, on peut prendre l'univers $\Omega' = \{r, b, v\}$ munit de la mesure de probabilité $\mathbb{P}' = \mathbb{P}(R)\delta_r + \mathbb{P}(B)\delta_b + \mathbb{P}(V)\delta_v = \frac{1}{2}\delta_r + \frac{1}{3}\delta_b + \frac{1}{6}\delta_v$.

Lorsque Ω est l'ensemble dénombrable $\Omega = \{\omega_n; n \geq 1\}$, toute mesure de probabilité sur $\mathcal{A} = 2^\Omega$ est de la forme

$$(1.15) \quad \mathbb{P} = \sum_{n \geq 1} p_n \delta_{\omega_n}$$

où $(p_n)_{n \geq 1}$ est tel que $p_n \geq 0$, $\forall n$ et $\sum_{n \geq 1} p_n = 1$. L'interprétation de cette formule est : $\mathbb{P}(\{\omega_n\}) = p_n$, $n \geq 1$.

Notre premier résultat concernant une quantité infiniment dénombrable d'opérations sur les événements est le suivant.

LEMME 1.16.

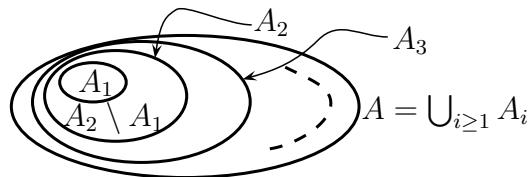
(1) Soient A_1, A_2, \dots une suite croissante (pour la relation d'inclusion) de \mathcal{A} : $A_1 \subset A_2 \subset \dots$ et $A = \bigcup_{n=1}^{\infty} A_n = \{\omega \in \Omega; \exists i \geq 1, \omega \in A_i\}$ sa limite. Alors

$$\mathbb{P}(A) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n).$$

(2) Soient B_1, B_2, \dots une suite décroissante (pour la relation d'inclusion) de \mathcal{A} : $B_1 \supset B_2 \supset \dots$ et $B = \bigcap_{n=1}^{\infty} B_n = \{\omega \in \Omega; \forall i \geq 1, \omega \in A_i\}$ sa limite. Alors

$$\mathbb{P}(B) = \lim_{n \rightarrow \infty} \mathbb{P}(B_n).$$

DÉMONSTRATION. Puisque $(A_n)_{n \geq 1}$ est une suite croissante,



$A = A_1 \sqcup (A_2 \setminus A_1) \sqcup (A_3 \setminus A_2) \sqcup \dots$ est la réunion disjointe d'une famille d'événements. Par conséquent,

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A_1) + \sum_{i=1}^{\infty} \mathbb{P}(A_{i+1} \setminus A_i) \\ &= \mathbb{P}(A_1) + \lim_{n \rightarrow \infty} \sum_{i=1}^{n-1} [\mathbb{P}(A_{i+1}) - \mathbb{P}(A_i)] \\ &= \lim_{n \rightarrow \infty} \mathbb{P}(A_n)\end{aligned}$$

Pour le résultat concernant la famille décroissante, passer aux complémentaires en utilisant la relation $(A \cup B)^c = A^c \cap B^c$. \square

EXEMPLE 1.17. On joue indéfiniment à pile ou face jusqu'à ce qu'on obtienne pour la première fois *pile*. Le premier instant d'obtention de *pile* est un entier qui peut être arbitrairement grand. On doit donc prendre un univers Ω de cardinal infini. Un bon choix est $\Omega = \{p, f\}^{\{1, 2, \dots\}}$: l'ensemble des suites $\omega = \omega_1 \omega_2 \dots \omega_n \dots$ constituées des lettres *p* et *f* avec l'interprétation que $\omega_n = p$ signifie qu'on a obtenu *pile* au n -ième lancer. Notons que nous choisissons un univers Ω différent de celui de l'Exemple 1.5, pour modéliser la même expérience.

L'événement qui correspond à l'obtention pour la première fois de *pile* au n -ième lancer est $P_n = \{\omega \in \Omega; \omega_1 = \dots = \omega_{n-1} = f, \omega_n = p\}$. C'est un ensemble infini qui a le même cardinal que Ω puisque seul le début des suites ω est spécifié (Exercice : le prouver). Il est naturel de demander lors de notre modélisation de cette expérience que $\mathbb{P}(P_n) = 2^{-n}$ puisqu'il y a 2^n mots de longueur n constitués des lettres *p* et *f* et que chacun de ces mots qui code la réalisation de n lancers de pile ou face doit avoir la même probabilité (situation d'équiprobabilité).

Soit $B_n = \{\omega \in \Omega; \omega_1 = \dots = \omega_n = f\} = \bigcup_{i \geq n+1} P_i$ l'événement "il n'y a pas eu *pile* pendant les n premiers lancers". L'additivité des probabilités d'événements disjoints s'écrit $\mathbb{P}(B_n) = \sum_{i=n+1}^{\infty} \mathbb{P}(P_i)$ c'est-à-dire $2^{-n} = \sum_{i=n+1}^{\infty} 2^i$. On vient de retrouver une formule bien connue.

La suite $(B_n)_{n \geq 1}$ est décroissante avec $\bigcap_{n \geq 1} B_n = P_{\infty} = \{\tilde{\omega}\}$ où $\tilde{\omega} = f f f f \dots$ est la suite constituée de *f* uniquement : l'événement "*pile* n'apparaît jamais". Le lemme précédent nous assure de $\mathbb{P}(P_{\infty}) = \lim_{n \rightarrow \infty} 2^{-n} = 0$. C'est-à-dire que $\mathbb{P}(\tilde{\omega}) = 0$. En d'autres termes, avec cette modélisation de l'expérience, on conclut que l'événement complémentaire "*pile* finit par apparaître" est de probabilité $1 - 0 = 1$; il est certain.

Un paradoxe. Compte tenu de la symétrie de notre modélisation, tous les ω sont équiprobables : $\forall \omega \in \Omega, \mathbb{P}(\omega) = \mathbb{P}(\tilde{\omega}) = 0$. Or la "somme" des probabilités de tous les événements élémentaires doit être égale à 1 : " $\sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1$ ". Ce qui nous mène à " $\sum_{\omega \in \Omega} 0 = 1$ ". Une somme de zéros égale à un ! Cette somme ne peut donc pas être la somme d'une série car $\sum_{n \in \mathbb{N}} 0 = 0$. C'est la raison pour laquelle on a mis " " \sum " entre guillemets. On lève le paradoxe en se rappelant que Ω est un ensemble non-dénombrable (voir le Lemme A.7-2), c'est-à-dire qu'il ne peut pas être mis en injection dans \mathbb{N} , il est beaucoup plus gros. De ce fait " $\sum_{\omega \in \Omega}$ " est une opération indéfinie ; en particulier elle n'est pas une série.

CHAPITRE 2

Variables aléatoires

Pour définir une variable aléatoire, seul (Ω, \mathcal{A}) suffit. On laisse \mathbb{P} de côté pour le moment. On se donne (Ω, \mathcal{A}) .

Essentiellement, une variable aléatoire est une fonction numérique sur l'univers Ω souvent notée $X : \Omega \rightarrow \mathbb{R}$.

EXEMPLE 2.1. On joue deux fois de suite à pile ou face. Notre univers est $\Omega = \{pp, pf, fp, ff\}$ (l'ordre des lancers est pris en compte). Le nombre d'apparitions de *pile* est la variable aléatoire suivante

$$X(\omega) = \begin{cases} 2 & \text{si } \omega = pp \\ 1 & \text{si } \omega \in \{pf, fp\} \\ 0 & \text{si } \omega = ff \end{cases}$$

EXEMPLE 2.2. On jette une flèche par terre et on note l'angle de sa direction avec le nord magnétique. Une telle expérience peut être décrite à l'aide de $\Omega = [0, 2\pi[$. Quant à la tribu \mathcal{A} , contentons-nous de dire qu'elle contient entre autres toutes les réunions dénombrables d'intervalles. L'application

$$X(\omega) = \omega, \quad \omega \in [0, 2\pi[$$

est la variable aléatoire qui correspond à l'angle de la flèche. Si l'on considère le cosinus de cet angle : $Y = \cos X$, on obtient à nouveau une variable aléatoire sur (Ω, \mathcal{A}) . Nous reviendrons sur la question du choix de \mathbb{P} à l'Exemple 2.7.

Il est très pratique d'introduire la notation suivante

$$\{\omega \in \Omega; X(\omega) \in C\} := \{X \in C\}, \quad C \subset \mathbb{R}.$$

En particulier, nous noterons $\{\omega \in \Omega; X(\omega) \leq x\} = \{X \leq x\}$.

DÉFINITION 2.3. Une application $X : \Omega \rightarrow \mathbb{R}$ est une *variable aléatoire réelle* si pour tout $x \in \mathbb{R}$, l'ensemble $\{X \leq x\}$ appartient à \mathcal{A} .

Lorsque Ω est dénombrable on prend $\mathcal{A} = 2^\Omega$ et bien sûr toute fonction numérique X sur Ω est une variable aléatoire. Mais lorsque Ω n'est pas dénombrable, comme c'est le cas dans l'Exemple 2.2, pour des raisons techniques délicates d'une difficulté dépassant le niveau de ce cours, on ne peut pas considérer toutes les fonctions numériques X sur Ω mais seulement celles qui sont spécifiées dans la définition précédente.

REMARQUES 2.4.

- (1) Notons que X est une fonction. Elle n'est donc ni variable, ni aléatoire ! Le vocable *variable aléatoire* date du début de la théorie des probabilités avec Pierre de Fermat (?-1665) et Blaise Pascal (1623-1662), bien avant que les mathématiques soient formalisées. Il faut donc prendre l'expression *variabla aléatoire* sans lui accorder une portée sémantique – n'hésitez pas à ouvrir votre dictionnaire.

- (2) Les premières formalisations rigoureuses de la théorie des probabilités datent du début du vingtième siècle. Nous pratiquons celle de Kolmogorov, mathématicien, physicien, génial et soviétique.

2.1. Fonction de répartition

Dès lors que l'on réintroduit la mesure de probabilité \mathbb{P} , le comportement aléatoire de X peut être quantifié. L'objet fondamental de cette description est la fonction de répartition.

DÉFINITION 2.5. On se donne $(\Omega, \mathcal{A}, \mathbb{P})$ et une variable aléatoire X sur (Ω, \mathcal{A}) . La *fonction de répartition* de X est définie par

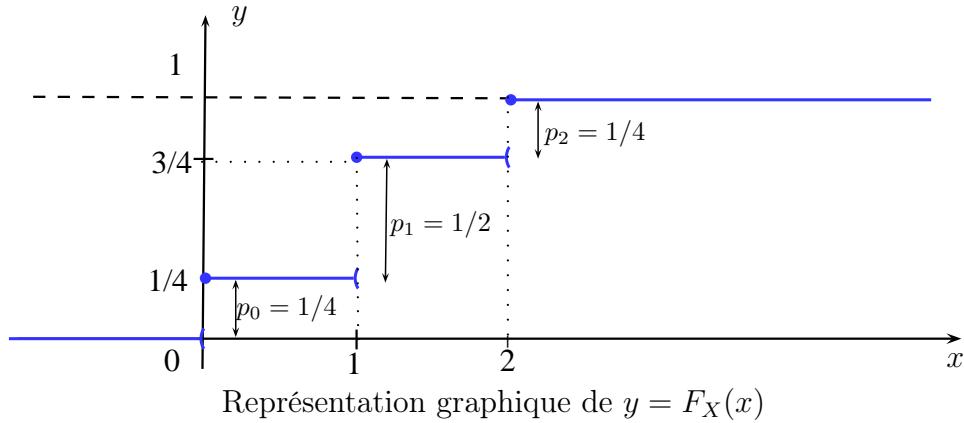
$$F_X(x) = \mathbb{P}(X \leq x), \quad x \in \mathbb{R}.$$

Notons que pour pouvoir écrire $\mathbb{P}(X \leq x)$, il faut que X soit une variable aléatoire au sens de la Définition 2.3.

EXEMPLE 2.6. On reprend la variable aléatoire X de l'Exemple 2.1. Notre espace probabilisé est $(\Omega, \mathcal{A}, \mathbb{P})$ avec $\Omega = \{pp, pf, fp, ff\}$, $\mathcal{A} = 2^\Omega$ et $\mathbb{P}(pp) = \mathbb{P}(pf) = \mathbb{P}(fp) = \mathbb{P}(ff) = 1/4$. Nous avons bien sûr, $\mathbb{P}(X = 0) = \mathbb{P}(X = 2) = 1/4$ et $\mathbb{P}(X = 1) = 1/2$. La fonction de répartition de X est

$$F_X(x) = \begin{cases} 0 & \text{si } x \in]-\infty, 0[\\ 1/4 & \text{si } x \in [0, 1[\\ 3/4 & \text{si } x \in [1, 2[\\ 1 & \text{si } x \in [2, +\infty[\end{cases}$$

et son graphe est



On constate que F_X ne croît que pour les valeurs effectivement fréquentées par X : 0, 1 et 2. La hauteur de chacune des marches est respectivement $p_0 = \mathbb{P}(X = 0)$, $p_1 = \mathbb{P}(X = 1)$ et $p_2 = \mathbb{P}(X = 2)$.

EXEMPLE 2.7 (suite de l'Exemple 2.2). Compte tenu de la symétrie de l'expérience, il semble raisonnable d'en modéliser le hasard à l'aide de la mesure de probabilité qui satisfait $\mathbb{P}([a, b]) = (b - a)/(2\pi)$, $0 \leq a < b < 2\pi$. Soient $X(\omega) = \omega$ et $Y(\omega) = \cos \omega$. Les

fonctions de répartition de X et Y sont

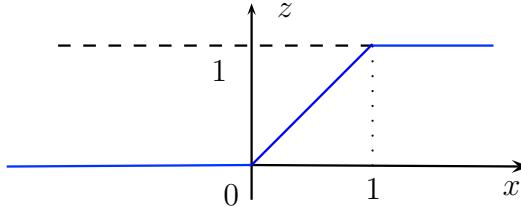
$$F_X(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x/(2\pi) & \text{si } 0 \leq x < 2\pi \\ 1 & \text{si } x \geq 2\pi \end{cases}$$

et

$$F_Y(y) = \begin{cases} 0 & \text{si } y < -1 \\ 1 - (\arccos y)/\pi & \text{si } -1 \leq y < 1 \\ 1 & \text{si } y \geq 1 \end{cases}$$

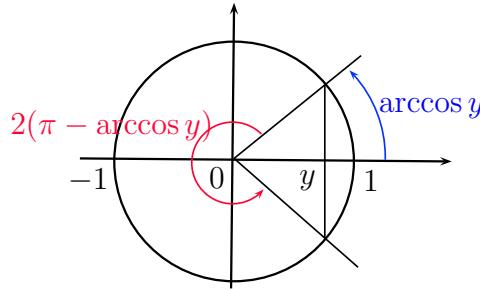
En effet, pour $0 \leq x < 2\pi$

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) \\ &= \mathbb{P}(\{\omega \in \Omega; 0 \leq \omega \leq x\}) = \mathbb{P}([0, x]) = x/(2\pi) \end{aligned}$$

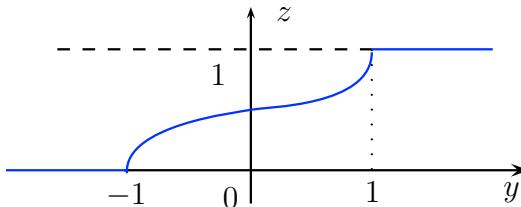


Représentation graphique de $z = F_X(x)$

et pour $-1 \leq y < 1$



$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) \\ &= \mathbb{P}(\{\omega \in \Omega; \cos \omega \leq y\}) = \mathbb{P}(X \in [-(\pi - \arccos y), \pi - \arccos y]) \\ &= 2(\pi - \arccos y)/(2\pi) = 1 - (\arccos y)/\pi \end{aligned}$$



Représentation graphique de $z = F_Y(y)$

Les fonctions de répartition jouissent d'un certain nombre de propriétés.

PROPOSITION 2.8. *Une fonction de répartition F possèdent les propriétés suivantes :*

(1) $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow \infty} F(x) = 1$,

(2) F est croissante

(3) pour tous $a < b$, $\mathbb{P}(a < X \leq b) = F(b) - F(a)$

(4) F est continue à droite

DÉMONSTRATION. • Preuve de (1). Soit $B_n = \{X \leq -n\}$. Alors B_1, B_2, \dots est une suite décroissante d'événements de limite vide. Par conséquent, grâce au Lemme 1.16, $\lim_{n \rightarrow \infty} \mathbb{P}(B_n) = \mathbb{P}(\emptyset) = 0$. Pour l'autre limite, considérer $A_n = \{X \leq n\}$.

• Preuve de (2) et (3). Soient $a < b$ et $A(a) = \{X \leq a\}$, $A(a, b) = \{a < X \leq b\}$. Alors, $A(b) = A(a) \sqcup A(a, b)$ est une union disjointe, de sorte que

$$\mathbb{P}(A(b)) = \mathbb{P}(A(a)) + \mathbb{P}(A(a, b))$$

d'où il vient que

$$F(b) = F(a) + \mathbb{P}(a < X \leq b) \geq F(a)$$

qui est (3) et prouve (2).

• Preuve de (4). Avec la notation précédente, pour tout $a \in \mathbb{R}$, $A(a, a+h)$ décroît vers le vide lorsque $h > 0$ décroît vers zéro. Par conséquent, grâce à (3), $\lim_{h \downarrow 0} F(a+h) - F(a) = \lim_{n \rightarrow \infty} F(a+1/n) - F(a) = \lim_{n \rightarrow \infty} \mathbb{P}(X \in]a, a+1/n]) \stackrel{(*)}{=} \mathbb{P}(X \in \lim_{n \rightarrow \infty}]a, a+1/n]) = \mathbb{P}(X \in \emptyset) = 0$, où l'égalité (*) est une conséquence du Lemme 1.16 et l'existence de la limite $\lim_{h \downarrow 0} F(a+h)$ est garantie par la croissance de F démontrée au point (2). \square

Le résultat suivant montre que la fonction de répartition permet d'évaluer la probabilité $\mathbb{P}(X \in I)$ pour n'importe quel intervalle I .

PROPOSITION 2.9. *Soient $-\infty \leq a \leq b \leq +\infty$. Alors,*

- (1) $\mathbb{P}(X \in]a, b]) = F_X(b) - F_X(a)$;
- (2) $\mathbb{P}(X \in [a, b]) = F_X(b) - F_X(a^-)$;
- (3) $\mathbb{P}(X \in]a, b[) = F_X(b^-) - F_X(a)$;
- (4) $\mathbb{P}(X \in [a, b[) = F_X(b^-) - F_X(a^-)$

où $F_X(c^-) := \lim_{x \uparrow c} F_X(x)$ est la limite à gauche de F_X en c et par convention $F_X(-\infty) := \lim_{x \rightarrow -\infty} F_X(x) = 0$ et $F_X(+\infty) := \lim_{x \rightarrow +\infty} F_X(x) = 1$, d'après la Proposition 2.8-(1).

On notera que la limite à gauche $F_X(c^-)$ existe puisque F_X est une fonction croissante de sorte que $\lim_{x \uparrow c} F_X(x) = \sup_{x < c} F_X(x)$.

DÉMONSTRATION. • Preuve de (1). Dans ce cas, $b < \infty$. Lorsque $a = -\infty$, c'est évident et lorsque a est fini, ce résultat a été obtenu à la Proposition 2.8.

• Preuve de (2). Dans ce cas, a et b sont finis. Puisque, $[a, b] = \bigcap_{n \geq 1}]a - 1/n, b]$ on a $\{X \in [a, b]\} = \bigcap_{n \geq 1} \{X \in]a - 1/n, b]\}$ et on obtient à l'aide de (1) et du Lemme 1.16, $\mathbb{P}(X \in [a, b]) = \lim_{n \rightarrow \infty} \mathbb{P}(X \in]a - 1/n, b]) = \lim_{n \rightarrow \infty} F_X(b) - F_X(a - 1/n) = F_X(b) - F_X(a^-)$.

• Preuve de (3). Prenons $a = -\infty$. Si $b = \infty$, le résultat est évident et si $b < \infty$, $\mathbb{P}(X \in]-\infty, b]) = \mathbb{P}(X \in \bigcup_{n \geq 1}]-\infty, b - 1/n]) = \lim_{n \rightarrow \infty} \mathbb{P}(X \in]-\infty, b - 1/n]) = \lim_{n \rightarrow \infty} F_X(b - 1/n) = F_X(b^-)$. Lorsque a est fini, $\mathbb{P}(X \in]a, b[) = \mathbb{P}(X \in]-\infty, b[) - \mathbb{P}(X \in]-\infty, a]) = F_X(b^-) - F_X(a)$.

• Preuve de (4). Dans ce cas a est fini et en tenant compte de (3), $\mathbb{P}(X \in [a, b[) = \lim_{n \rightarrow \infty} \mathbb{P}(X \in]a - 1/n, b[) = \lim_{n \rightarrow \infty} F_X(b^-) - F_X(a - 1/n) = F_X(b^-) - F_X(a^-)$. \square

2.2. Variables aléatoires discrètes

Commençons par rappeler la définition d'une variable aléatoire discrète.

DÉFINITION 2.10. La variable aléatoire X est dite *discrète* si elle prend ses valeurs dans une partie dénombrable $\{x_n; n \in N\}$ de \mathbb{R} où N est un ensemble d'indices.

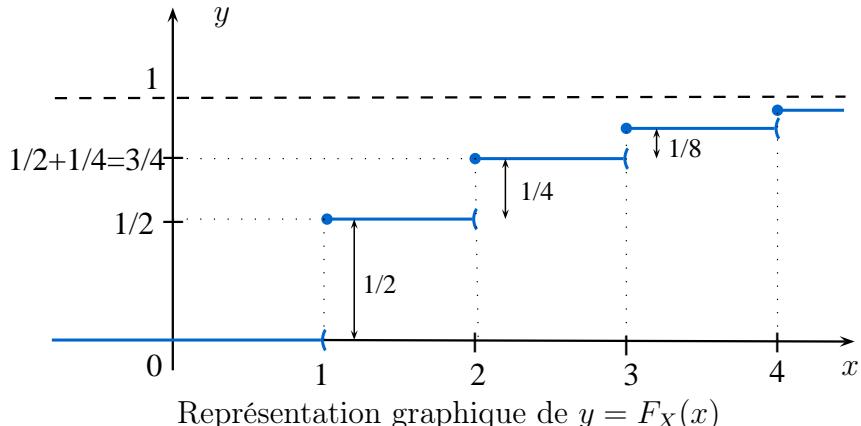
Rappelons que certains des résultats les plus simples au sujet de la dénombrabilité sont présentés en Annexe A.

REMARQUES 2.11.

- (1) Bien sûr, on peut sans restriction supposer que les x_n sont tous distincts.
- (2) Puisque N est dénombrable, on peut choisir $N = \{1, \dots, K\}$ si X prend $K = \#(X(\Omega)) < \infty$ valeurs ou bien $N = \{1, 2, \dots\}$ si X prend une infinité de valeurs.

EXEMPLES 2.12.

- (1) La variable aléatoire de l'Exemple 2.1 est discrète.
- (2) On note X le premier instant d'obtention de *pile* dans l'Exemple 1.17. C'est une variable aléatoire à valeurs dans $\{1, 2, \dots\} \cup \{\infty\}$ où $X = \infty$ signifie que *pile* n'apparaît jamais. On a vu que $\mathbb{P}(X = \infty) = 0$ de sorte que X est effectivement à valeurs dans \mathbb{R} et qu'on peut considérer sa fonction de répartition. On a déjà vu que pour tout $n \geq 1$, $\mathbb{P}(X = n) = \mathbb{P}(B_n) = 2^{-n}$. La représentation graphique de F_X est



Comme nous allons le voir, de telles fonctions de répartition sont typiques des variables discrètes.

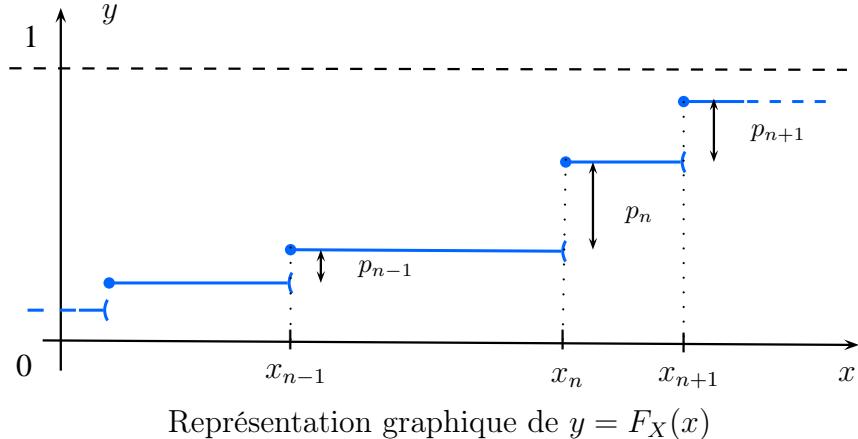
Le comportement d'une variable discrète X est décrit par la donnée de $(x_n, p_n)_{n \in N}$ où les x_n sont supposés distincts et $p_n := \mathbb{P}(X = x_n) \geq 0$. Du fait que $1 = \mathbb{P}(X \in \mathbb{R})$, nous obtenons la condition de normalisation

$$(2.13) \quad \sum_{n \in N} p_n = 1.$$

On peut toujours choisir pour N une partie de \mathbb{Z} constituée de nombres consécutifs de sorte que les valeurs de X soient rangées par ordre croissant : $\dots < x_{n-1} < x_n < x_{n+1} < \dots$. À l'aide de la Proposition 2.8-(3), on voit que $\mathbb{P}(X = x_n) = \mathbb{P}(x_{n-1} < X \leq x_n) = F_X(x_n) - F_X(x_{n-1})$, soit

$$(2.14) \quad p_n = F_X(x_n) - F_X(x_{n-1}), \quad n \in N$$

avec les conventions $x_{(\inf(N)-1)} = -\infty$ et $F_X(-\infty) = 0$. De plus, pour tous $x_{n-1} \leq x \leq y < x_n$, nous avons $0 \leq F_X(y) - F_X(x) = \mathbb{P}(x < X \leq y) \leq \mathbb{P}(x_{n-1} < X < x_n) = 0$. Par conséquent $F_X(x) = F_X(y)$, ce qui signifie que F_X est constante sur les intervalles semi-ouverts $[x_{n-1}, x_n]$. La forme générale de F_X est donc



Une telle fonction de répartition est dite *atomique* : c'est-à-dire qu'elle est constante entre ses discontinuités qui sont des sauts positifs.

2.3. Variables aléatoires continues

La situation précédente est radicalement différente de celle des variables aléatoires continues.

DÉFINITIONS 2.15.

- (1) Une fonction numérique est dite *continue par morceaux* si tous ses points de discontinuité sont isolés. Ceci signifie que pour tout point de discontinuité il existe un intervalle ouvert qui le contient et ne contient pas d'autre point de discontinuité.
- (2) La *variable aléatoire* X est dite *continue* si sa fonction de répartition peut s'écrire sous la forme

$$(2.16) \quad F_X(x) = \int_{-\infty}^x f_X(u) du, \quad x \in \mathbb{R}$$

pour une certaine fonction $f_X : \mathbb{R} \rightarrow [0, \infty[$ continue par morceaux et intégrable.

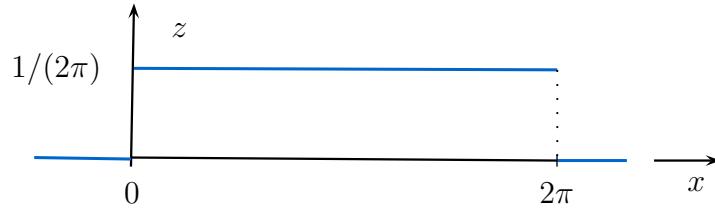
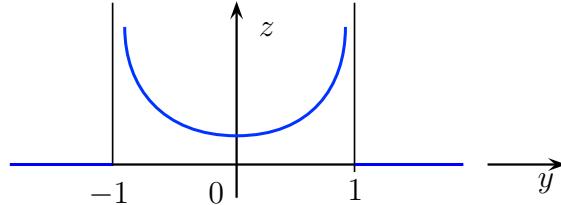
- (3) Dans ce cas, la fonction f_X est appelée *fonction de densité* de la variable aléatoire X .

EXEMPLE 2.17 (suite de l'Exemple 2.7). On constate que X et Y sont continues puisque

$$F_X(x) = \int_{-\infty}^x f_X(u) du, \quad F_Y(y) = \int_{-\infty}^y f_Y(u) du$$

avec les fonctions de densité

$$f_X(x) = \begin{cases} 1/(2\pi) & \text{si } x \in [0, 2\pi] \\ 0 & \text{sinon} \end{cases}, \quad f_Y(y) = \begin{cases} 1/(\pi\sqrt{1-y^2}) & \text{si } y \in [-1, 1] \\ 0 & \text{sinon} \end{cases}$$

Représentation graphique de $z = f_X(x)$ Représentation graphique de $z = f_Y(y)$

Par souci de lisibilité, ces deux représentations ne sont pas à la même échelle. Notons l'explosion en -1 et 1 de la densité de Y .

REMARQUES 2.18.

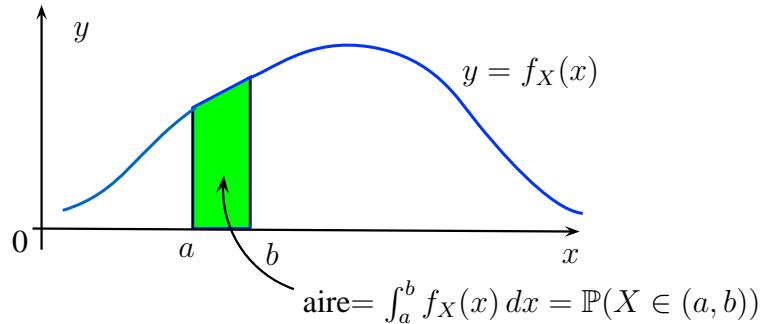
- (1) Il est clair que la fonction de répartition F_X d'une variable continue est continue. En fait, elle est un peu plus régulière : des fonctions F_X qui admettent une représentation (2.16) sont dites *absolument continues*.
- (2) Si f_X est elle-même continue, F_X est dérivable (de classe C^1) et $F'_X = f_X$.
- (3) Remarquons que F_X n'est pas dérivable aux points de discontinuité de f_X .

Si X est une variable aléatoire continue, F_X est une fonction continue et toutes les expressions des membres de droite des égalités de la Proposition 2.9 sont égales. On en déduit immédiatement le

COROLLAIRE 2.19. *Si X est une variable aléatoire continue de densité f_X , pour tous $a \leq b$ nous avons*

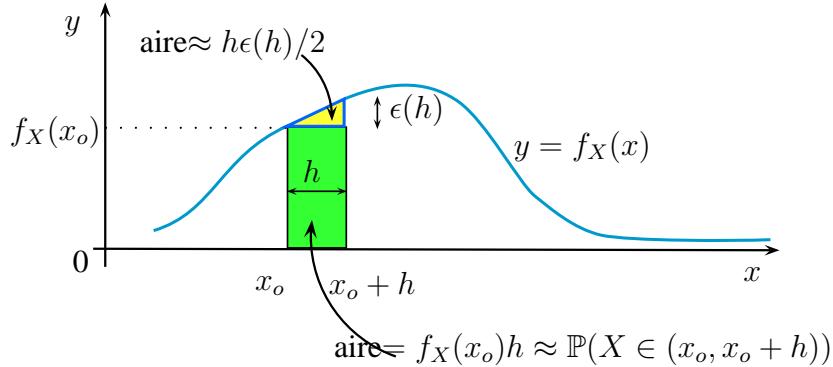
$$\begin{aligned} \mathbb{P}(X \in]a, b]) &= \mathbb{P}(X \in [a, b]) = \mathbb{P}(X \in]a, b[) \\ &= \mathbb{P}(X \in [a, b[) = \int_a^b f_X(x) dx. \end{aligned}$$

Lorsque X est continue, on notera parfois $\mathbb{P}(X \in (a, b))$ chacune des quantités égales $\mathbb{P}(X \in]a, b]) = \mathbb{P}(X \in [a, b]) = \mathbb{P}(X \in]a, b[) = \mathbb{P}(X \in [a, b[)$.



En se souvenant de la définition de l'intégrale de Riemann comme limite de sommes de Darboux, on obtient en tout point x de continuité de la densité f_X que lorsque $h \geq 0$ tend vers zéro, $\mathbb{P}(X \in (x, x + h)) = \int_x^{x+h} f_X(t) dt = f_X(x)h + \eta(h)h$ où $\lim_{h \rightarrow 0} \eta(h) = 0$. De façon informelle, on traduit ceci par

$$(2.20) \quad \mathbb{P}(X \in (x, x + h)) \underset{h \rightarrow 0}{\approx} f_X(x)h.$$



On constate donc que la variable aléatoire X a plus de chance de prendre des valeurs dans les régions où f_X est grande. En particulier, X ne prend pas de valeur dans l'ensemble $\{f_X = 0\} := \{x \in \mathbb{R}; f_X(x) = 0\}$.

Bien évidemment, puisque $1 = \mathbb{P}(\Omega) = \mathbb{P}(X \in \mathbb{R})$, nous avons toujours la condition de normalisation

$$(2.21) \quad \int_{\mathbb{R}} f_X(x) dx = 1.$$

qui est l'analogue de (2.13).

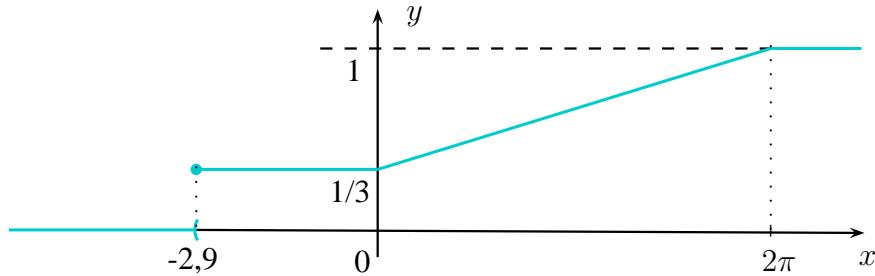
2.4. Quelques éléments de réflexion

Nous concluons ce chapitre en donnant un exemple de variable aléatoire qui n'est ni continue, ni discrète ; ainsi qu'une remarque au sujet de la tribu \mathcal{A} lorsque X prend un nombre non-dénombrable de valeurs.

EXEMPLE 2.22 (Une variable aléatoire ni continue, ni discrète). On tire une boule d'une urne qui contient 1 boule rouge et 2 boules vertes. Si la boule obtenue est verte, alors on lance notre flèche par terre et on mesure son angle. L'univers de l'expérience est $\Omega = \{r\} \cup \{(v, x); 0 \leq x < 2\pi\}$. Soit $X : \Omega \rightarrow \mathbb{R}$ donnée par

$$X(r) = -2, 9, \quad X((v, x)) = x.$$

X prend ses valeurs dans $\{-2, 9\} \cup [0, 2\pi[$ et sa fonction de répartition admet la représentation graphique suivante.



Représentation graphique de $y = F_X(x)$

EXEMPLE 2.23 (L'escalier du diable).

REMARQUE 2.24.

Clairement, si X prend un nombre non-dénombrable de valeurs, il est nécessaire que Ω ne soit pas dénombrable. C'est le cas pour les variables continues. En revenant à la Remarque 1.8, on peut se demander pourquoi dans cette situation on ne pourrait pas prendre la tribu 2^Ω de toutes les parties. C'est à l'évidence une tribu et on peut donc considérer une probabilité \mathbb{P} construite sur elle. Le problème que l'on rencontre est le suivant. On peut montrer qu'il n'existe pas de mesures de probabilités sur 2^Ω autres que celles de la forme (1.15) : $\sum_{n \geq 1} p_n \delta_{\omega_n}$ car 2^Ω est un ensemble trop gros.

CHAPITRE 3

Loi et espérance d'une variable aléatoire

Nous commençons par présenter les notions de loi et d'espérance dans la situation la plus simple qui est celle des variables discrètes. Puis, nous étendons par analogie ces notions au cas des variables continues. Finalement, nous montrons qu'il existe un cadre mathématique général qui permet de comprendre et définir ces notions pour toutes les variables aléatoires.

3.1. Variables discrètes

Soit X une variable aléatoire qui prend les valeurs $\{x_n; n \in N\}$ où les x_n sont distincts et N est un ensemble d'indices inclus dans l'ensemble $\{1, 2, \dots\}$ des entiers positifs non nuls, voir les Remarques 2.11. On décrit le comportement aléatoire de X par la donnée de $(x_n, p_n)_{n \in N}$ avec $p_n := \mathbb{P}(X = x_n)$, $n \in N$. Cette donnée est moins informative a priori que celle de (X, \mathbb{P}) qui décrit le phénomène ω par ω , mais elle est suffisante pour obtenir toutes les quantités moyennes que nous désirons.

DÉFINITION 3.1. La loi de la variable aléatoire discrète X est

$$(3.2) \quad P_X = \sum_{n \in N} p_n \delta_{x_n}$$

Une loi de cette forme est dite *atomique*. Ses atomes sont les x_n tels que $p_n > 0$.

On rappelle que δ_x est la masse de Dirac au point x , c'est-à-dire que pour toute partie $B \subset \mathbb{R}$, $\delta_x(B) = \begin{cases} 1 & \text{si } x \in B \\ 0 & \text{sinon} \end{cases}$, voir la Définition 1.12. La loi P_X est une mesure de probabilité sur \mathbb{R} .

EXEMPLES 3.3.

- (1) La variable aléatoire X de l'Exemple 2.12-(1) a pour loi $P_X = \frac{1}{4}\delta_0 + \frac{1}{2}\delta_1 + \frac{1}{4}\delta_2$.
- (2) La loi de celle de l'Exemple 2.12-(2) est $P_X = \sum_{n \geq 1} 2^{-n} \delta_n$.

Soit B une partie de \mathbb{R} , nous constatons que

$$(3.4) \quad \mathbb{P}(X \in B) = P_X(B), \quad B \subset \mathbb{R}$$

puisque

$$\begin{aligned} P_X(B) &= \sum_{n \in N} p_n \delta_{x_n}(B) = \sum_{n \in N: x_n \in B} p_n \\ &= \sum_{n \in N: x_n \in B} \mathbb{P}(X = x_n) = \mathbb{P}(X \in B). \end{aligned}$$

On voit clairement à l'aide de (2.14) que la donnée de $(x_n, p_n)_{n \in N}$ est équivalente à celle de la fonction de répartition F_X , de même qu'elle est équivalente à celle de la loi

P_X . En résumé, le comportement aléatoire de X est décrit de manière équivalente par la donnée de

- $(x_n, p_n)_{n \in N}$ ou
- la fonction de répartition F_X ou
- la loi P_X .

La valeur moyenne de X pondérée par les probabilités de réalisation des événements est appelée son espérance mathématique.

DÉFINITION 3.5. Soit X une variable discrète de loi $P_X = \sum_{n \in N} p_n \delta_{x_n}$. L'*espérance mathématique* de X est

$$\mathbb{E}X := \sum_{n \in N} p_n x_n.$$

Pour que cette quantité soit définie correctement, il est nécessaire de supposer que

$$\mathbb{E}|X| := \sum_{n \in N} p_n |x_n| < \infty$$

c'est-à-dire que $\sum_{n \in N} p_n x_n$ est une série absolument convergente.

EXEMPLES 3.6.

- (1) La variable X de l'Exemple 3.3-(1) a pour loi $P_X = \frac{1}{4}\delta_0 + \frac{1}{2}\delta_1 + \frac{1}{4}\delta_2$. Son espérance est $\mathbb{E}X = \frac{1}{4} \times 0 + \frac{1}{2} \times 1 + \frac{1}{4} \times 2 = 1$.
- (2) La variable X de l'Exemple 3.3-(2) a pour loi $P_X = \sum_{n \geq 1} 2^{-n} \delta_n$. Son espérance est $\mathbb{E}X = \sum_{n \geq 1} 2^{-n} n$.

REMARQUES 3.7.

- (1) Lorsque X est une variable aléatoire positive, son espérance $\mathbb{E}X = \sum_{n \in N} p_n x_n$ est une série à termes positifs. Elle est donc toujours définie à condition de lui donner la valeur $+\infty$ lorsqu'elle est divergente. En particulier, pour toute variable aléatoire, on a $\mathbb{E}|X| = \sum_{n \in N} p_n |x_n|$ et l'on peut écrire $\mathbb{E}|X|$ sans précaution en tant que nombre dans $[0, +\infty] = [0, +\infty] \cup \{+\infty\}$. De plus, $\mathbb{E}|X| < \infty$ signifie que la série $\sum_{n \in N} p_n x_n$ est absolument convergente et donc que $\mathbb{E}X$ est bien défini.
- (2) On définit la loi d'une variable aléatoire discrète X à valeurs dans un ensemble quelconque \mathcal{X} exactement comme lorsque $\mathcal{X} \subset \mathbb{R}$, par la donnée de $(x_n, p_n)_{n \in N}$ où les x_n sont dans \mathcal{X} . La loi de X est donnée par la Définition 3.1 : $P_X = \sum_{n \in N} p_n \delta_{x_n}$. C'est une mesure de probabilité sur \mathcal{X} muni de la tribu $2^{\mathcal{X}}$ de ses parties.
- (3) En revanche, pour considérer $\mathbb{E}X$, il faut pouvoir additionner les x et les multiplier par des poids $0 \leq p \leq 1$. La notion d'espérance de X n'a donc de sens que si \mathcal{X} est un espace vectoriel. L'espérance de X est donnée par la Définition 3.5 : $\mathbb{E}X = \sum_{n \in N} p_n x_n \in \mathcal{X}$ sous réserve que cette série soit absolument convergente, c'est-à-dire que la série à termes positifs $\mathbb{E}|X| = \sum_{n \in N} p_n \|x_n\| < \infty$ soit convergente, où $\|\cdot\|$ est une norme sur l'espace vectoriel \mathcal{X} . Un cas très important est celui de $\mathcal{X} = \mathbb{R}^d$ muni de la norme euclidienne ou de n'importe quelle autre norme équivalente.

Considérons la variable aléatoire $Y = \varphi(X)$, image de X par la fonction numérique $\varphi : \mathbb{R} \rightarrow \mathbb{R}$. Sa loi est $P_Y = \sum_{m \in M} q_m \delta_{y_m}$ où $\{y_m; m \in M\} = \{\varphi(x_n); n \in N\}$ les y_m étant tous distincts et

$$\begin{aligned}
 q_m &:= \mathbb{P}(Y = y_m) \\
 &= \mathbb{P}(\varphi(X) = y_m) \\
 &= \sum_{x \in X(\Omega): \varphi(x) = y_m} \mathbb{P}(X = x) \\
 (3.8) \quad &= \sum_{n \in N(m)} p_n
 \end{aligned}$$

où $N(m) = \{n \in N : \varphi(x_n) = y_m\}$ est l'ensemble des indices des x_n dont l'image par φ est y_m .

Notons que $(N(m))_{m \in M}$ constitue une partition de N . C'est-à-dire que les parties $N(m)$ sont disjointes : $m \neq m' \Rightarrow N(m) \cap N(m') = \emptyset$ (puisque les y_m sont tous distincts), et

$$(3.9) \quad N = \bigsqcup_{m \in M} N(m).$$

THÉORÈME 3.10. *On suppose que $\sum_{n \in N} p_n |\varphi(x_n)| < \infty$. Alors,*

$$(3.11) \quad \mathbb{E}[\varphi(X)] = \sum_{n \in N} p_n \varphi(x_n).$$

DÉMONSTRATION. En notant $Y = \varphi(X)$ comme précédemment, nous avons

$$\begin{aligned}
 \mathbb{E}[\varphi(X)] &= \mathbb{E}Y \\
 &\stackrel{(a)}{=} \sum_{m \in M} q_m y_m \\
 &\stackrel{(b)}{=} \sum_{m \in M} \sum_{n \in N(m)} p_n y_m \\
 &\stackrel{(c)}{=} \sum_{m \in M} \sum_{n \in N(m)} p_n \varphi(x_n) \\
 &\stackrel{(d)}{=} \sum_{n \in N} p_n \varphi(x_n)
 \end{aligned}$$

où (a) est la définition de l'espérance, (b) provient de (3.8), (c) est une conséquence de $y_m = \varphi(x_n)$, $\forall n \in N(m)$ et (d) vient de (3.9).

Bien évidemment, il faut s'assurer que toutes ces séries sont absolument convergentes. Or, en reprenant le précédent calcul en remplaçant Y par $|Y|$ et donc φ par $|\varphi|$, on voit que c'est le cas sous notre hypothèse : $\sum_{n \in N} p_n |\varphi(x_n)| < \infty$. \square

THÉORÈME 3.12. *La loi de $\varphi(X)$ est $P_{\varphi(X)} = \sum_{n \in N} p_n \delta_{\varphi(x_n)}$.*

DÉMONSTRATION. On reprend en la transposant la preuve du Théorème 3.10. Ce qui donne :

$$\begin{aligned} P_{\varphi(X)} &= P_Y = \sum_{m \in M} q_m \delta_{y_m} = \sum_{m \in M} \sum_{n \in N(m)} p_n \delta_{y_m} \\ &= \sum_{m \in M} \sum_{n \in N(m)} p_n \delta_{\varphi(x_n)} = \sum_{n \in N} p_n \delta_{\varphi(x_n)} \end{aligned}$$

qui est le résultat désiré. \square

Reprendons l'Exemple 3.3-(1), c'est-à-dire $P_X = \frac{1}{4}\delta_0 + \frac{1}{2}\delta_1 + \frac{1}{4}\delta_2$ et considérons $\varphi(x) = (x-1)^2$. On obtient alors $P_{\varphi(X)} = \frac{1}{4}\delta_{\varphi(0)} + \frac{1}{2}\delta_{\varphi(1)} + \frac{1}{4}\delta_{\varphi(2)} = \frac{1}{4}\delta_1 + \frac{1}{2}\delta_0 + \frac{1}{4}\delta_1 = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$.

En prenant $N = \{1, 2, 3\}$, $x_1 = 0$, $x_2 = 1$ et $x_3 = 2$, ainsi que $M = \{1, 2\}$ avec $y_1 = 0 = \varphi(1)$ et $y_2 = 1 = \varphi(0) = \varphi(2)$, nous obtenons $N(1) = \{2\}$ et $N(2) = \{1, 3\}$. La formule (3.8) s'écrit $q_1 = \sum_{n \in N(1)} p_n = p_2$ et $q_2 = \sum_{n \in N(2)} p_n = p_1 + p_3$, ce qui donne $\mathbb{P}(\varphi(X) = 0) = 1/2$ et $\mathbb{P}(\varphi(X) = 1) = 1/4 + 1/4 = 1/2$.

LEMME 3.13 (Positivité de l'espérance).

- (1) Soit X une variable positive : $X \geq 0$, c'est-à-dire $X(\omega) \geq 0, \forall \omega \in \Omega$. Alors, $0 \leq \mathbb{E}X \leq \infty$.
- (2) Soient φ et ψ deux fonctions positives telles que $0 \leq \varphi \leq \psi$. Alors, $0 \leq \mathbb{E}[\varphi(X)] \leq \mathbb{E}[\psi(X)] \leq \infty$.

DÉMONSTRATION. • Preuve de (1). Nous avons $x_n \geq 0$ et $p_n \geq 0$ pour tout $n \in N$. Donc $\mathbb{E}X = \sum_{n \in N} p_n x_n \geq 0$.

• Preuve de (2). Pour tout $n \in N$, $0 \leq p_n \varphi(x_n) \leq p_n \psi(x_n)$. Donc les séries à termes positifs correspondantes sont ordonnées de façon similaire : $0 \leq \mathbb{E}[\varphi(X)] = \sum_{n \in N} p_n \varphi(x_n) \leq \sum_{n \in N} p_n \psi(x_n) \leq \mathbb{E}[\psi(X)] \leq \infty$. \square

THÉORÈME 3.14 (Linéarité de l'espérance). Soient $\varphi, \psi : \mathbb{R} \rightarrow \mathbb{R}$ deux fonctions numériques telles que $\mathbb{E}|\varphi(X)| < \infty$ et $\mathbb{E}|\psi(X)| < \infty$. Pour tous réels a, b , nous avons

$$\mathbb{E}[a\varphi(X) + b\psi(X)] = a\mathbb{E}[\varphi(X)] + b\mathbb{E}[\psi(X)]$$

où toutes les espérances sont bien définies.

DÉMONSTRATION. Puisque $|a\varphi(X) + b\psi(X)| \leq |a||\varphi(X)| + |b||\psi(X)|$, grâce au Lemme 3.13-(2), nous avons $\mathbb{E}|a\varphi(X) + b\psi(X)| \leq |a|\mathbb{E}|\varphi(X)| + |b|\mathbb{E}|\psi(X)| < \infty$ de sorte que toutes les espérances sont bien définies. Grâce au Théorème 3.10,

$$\begin{aligned} \mathbb{E}[a\varphi(X) + b\psi(X)] &= \sum_{n \in N} p_n [a\varphi(x_n) + b\psi(x_n)] \\ &= a \sum_{n \in N} p_n \varphi(x_n) + b \sum_{n \in N} p_n \psi(x_n) \\ &= a\mathbb{E}[\varphi(X)] + b\mathbb{E}[\psi(X)] \end{aligned}$$

ce qui achève la preuve. \square

THÉORÈME 3.15 (Croissance de l'espérance). Soient φ et ψ deux fonctions numériques telles que $\mathbb{E}|\varphi(X)| < \infty$, $\mathbb{E}|\psi(X)| < \infty$ et $\varphi \leq \psi$. Alors, $\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\psi(X)]$.

DÉMONSTRATION. $\psi(X) - \varphi(X) \geq 0$, donc par linéarité et positivité de l'espérance $\mathbb{E}[\psi(X)] - \mathbb{E}[\varphi(X)] = \mathbb{E}[\psi(X) - \varphi(X)] \geq 0$. \square

REMARQUE 3.16.

En reprenant la Remarque 3.7-(2), on peut étendre les Théorèmes 3.14 et 3.15 au cas des variables aléatoires discrètes à valeurs dans un ensemble \mathcal{X} quelconque, en prenant des fonctions $\varphi, \psi : \mathcal{X} \rightarrow \mathbb{R}$, puisque $\varphi(X)$ et $\psi(X)$ sont des variables aléatoires *réelles*.

3.2. Variables continues

Nous allons procéder par analogie avec les variables discrètes. Nous gardons les notations introduites à la Définition 2.15, en particulier la densité f_X de la loi de la variable aléatoire continue X est supposée continue par morceaux.

DÉFINITION 3.17.

- (1) On note \mathcal{C}_X l'ensemble des fonctions de $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ qui sont continues par morceaux et telles que l'intégrale généralisée $\int_{\mathbb{R}} |\varphi(x)| f_X(x) dx$ soit convergente, c'est-à-dire $\int_{\mathbb{R}} |\varphi(x)| f_X(x) dx < \infty$.
- (2) Soit $\varphi \in \mathcal{C}_X$. L'espérance mathématique de la variable aléatoire $\varphi(X)$ est définie par

$$(3.18) \quad \mathbb{E}\varphi(X) := \int_{\mathbb{R}} \varphi(x) f_X(x) dx.$$

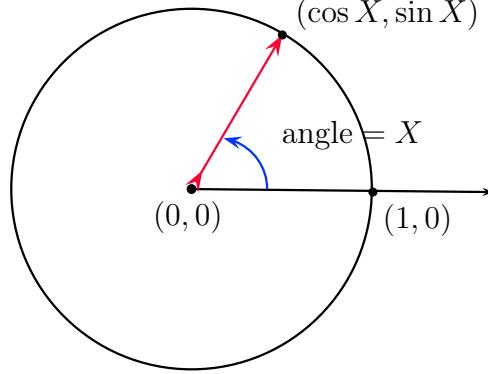
- Une justification rigoureuse de cette définition peut être obtenue en montrant qu'elle est l'extension naturelle de la Définition 3.5 de l'espérance d'une variable discrète.
- En tenant compte de (2.20), lorsqu'on se souvient de la construction de l'intégrale de Riemann comme limite de sommes de Darboux, on voit que cette définition est analogue au résultat obtenu en (3.11) pour les variables discrètes.
- Du fait que f_X et φ sont continues par morceaux, il en est de même pour leur produit φf_X qui, par conséquent, est localement intégrable au sens de Riemann.

REMARQUES 3.19.

- (1) Si $\varphi \geq 0$ est une fonction continue par morceaux et positive, on peut définir l'espérance (3.18) en posant $\mathbb{E}\varphi(X) = +\infty$ lorsque l'intégrale généralisée positive $\int_{\mathbb{R}} \varphi(x) f_X(x) dx$ est divergente.
En particulier, pour toute fonction φ continue par morceaux, on note $\mathbb{E}|\varphi(X)| = \int_{\mathbb{R}} |\varphi(x)| f_X(x) dx \in [0, \infty]$.
- (2) L'hypothèse d'intégrabilité $\mathbb{E}|\varphi(X)| = \int_{\mathbb{R}} |\varphi(x)| f_X(x) dx < \infty$ exprime que l'intégrale généralisée $\int_{\mathbb{R}} \varphi(x) f_X(x) dx$ est absolument convergente.

EXEMPLE 3.20. Si X est l'angle de la flèche de l'Exemple 2.17 : $f_X(x) = \mathbf{1}_{[0,2\pi]}(x)/(2\pi)$ de sorte que $\mathbb{E}(X) = \int_0^{2\pi} \frac{x}{2\pi} dx = \pi$.

REMARQUE 3.21. On peut se demander ce que signifie la valeur moyenne de l'angle $\mathbb{E}X = \pi$. En effet, si l'on avait choisi de coder l'angle dans $[-\pi, \pi[$, on aurait obtenu $\mathbb{E}X = 0$ pour la même expérience. En revanche, les coordonnées cartésiennes $(\cos X, \sin X)$ sur le cercle trigonométrique sont indépendantes du choix de l'origine des angles.



On définit $\mathbb{E}(\cos X, \sin X) = (\mathbb{E}[\cos X], \mathbb{E}[\sin X])$ et on obtient la direction moyenne $\mathbb{E}(\cos X, \sin X) = (0, 0)$ puisque $\mathbb{E}[\cos X] = \frac{1}{2\pi} \int_0^{2\pi} \cos x dx = 0$ et $\mathbb{E}[\sin X] = \frac{1}{2\pi} \int_0^{2\pi} \sin x dx = 0$. Ce qui signifie bien qu'aucune direction n'est privilégiée.

THÉORÈME 3.22 (Linéarité de l'espérance). *L'ensemble \mathcal{C}_X est un sous-espace vectoriel de l'espace des fonctions numériques.*

Pour tous $\varphi, \psi \in \mathcal{C}_X$ et tous réels a, b , nous avons

$$\mathbb{E}[a\varphi(X) + b\psi(X)] = a\mathbb{E}[\varphi(X)] + b\mathbb{E}[\psi(X)].$$

DÉMONSTRATION. Soient φ et ψ deux fonctions continues par morceaux. L'ensemble des points de discontinuité de $\varphi + \psi$ est inclus dans la réunion des ensembles de points de discontinuité de φ et ψ et une réunion finie de points isolés reste un ensemble de points isolés. Donc $\varphi + \psi$ est continue par morceaux. Il en est de même pour $a\varphi$ pour tout $a \in \mathbb{R}$.

D'autre part, $\int_{\mathbb{R}} |a\varphi(x)| f_X(x) dx = |a| \int_{\mathbb{R}} |\varphi(x)| f_X(x) dx < \infty$. Ce qui prouve que \mathcal{C}_X est un espace vectoriel.

La linéarité de l'intégrale nous assure de

$$\begin{aligned} \mathbb{E}[a\varphi(X) + b\psi(X)] &= \int_{\mathbb{R}} [a\varphi(x) + b\psi(x)] f_X(x) dx \\ &= a \int_{\mathbb{R}} \varphi(x) f_X(x) dx + b \int_{\mathbb{R}} \psi(x) f_X(x) dx \\ &= a\mathbb{E}[\varphi(X)] + b\mathbb{E}[\psi(X)], \end{aligned}$$

qui est le résultat annoncé. □

THÉORÈME 3.23 (Croissance de l'espérance).

- (1) *Soient $\varphi, \psi \geq 0$ deux fonctions positives continues par morceaux telles que $0 \leq \varphi \leq \psi$. Alors la Remarque 3.19-(1) nous assure du sens des quantités $\mathbb{E}[\varphi(X)]$ et $\mathbb{E}[\psi(X)]$ et nous avons $0 \leq \mathbb{E}[\varphi(X)] \leq \mathbb{E}[\psi(X)] \leq \infty$.*
- (2) *Soient $\varphi, \psi \in \mathcal{C}_X$ telles que $\varphi \leq \psi$, alors $\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\psi(X)]$.*

DÉMONSTRATION. Ces résultats sont des conséquences immédiates des propriétés de croissance des intégrales généralisées. □

Par analogie avec la relation (3.4), nous introduisons la

DÉFINITION 3.24. La loi de X est la mesure de probabilité sur \mathbb{R}

$$P_X(dx) := f_X(x) dx$$

qui est définie par

$$P_X(B) := \mathbb{P}(X \in B) = \int_a^b f_X(x) dx \stackrel{\text{notation}}{=} \int_B f_X(x) dx$$

pour tout intervalle $B = (a, b) \subset \mathbb{R}$.

3.3. Une notation commune

Nous venons de voir que les résultats de croissance (Théorèmes 3.15 et 3.23) et de linéarité (Théorèmes 3.14 et 3.22) s'expriment de façon analogue pour les variables aléatoires discrètes et continues. C'est l'indice qu'il existe une théorie générale qui englobe ces deux situations. Il s'agit de la théorie de l'intégration de Lebesgue que nous n'aborde-rons pas dans ce cours. En revanche, nous allons introduire des notations issues de cette théorie qui permettront de traiter simultanément ces deux types de variables aléatoires. Les principaux résultats de cette théorie sont collectés à l'Annexe B.

On note

$$\int_{\mathbb{R}} \varphi(x) P_X(dx) = \int_{\mathbb{R}} \varphi dP_X = \mathbb{E}\varphi(X)$$

(1) la quantité

$$\int_{\mathbb{R}} \varphi dP_X = \sum_{n \in N} \varphi(x_n) p_n$$

lorsque X est discrète de loi $P_X = \sum_{n \in N} p_n \delta_{x_n}$ ou bien

(2) la quantité

$$\int_{\mathbb{R}} \varphi dP_X = \int_{\mathbb{R}} \varphi(x) f_X(x) dx$$

lorsque X est continue de loi $P_X(dx) = f_X(x) dx$.

Nous avons montré aux Théorèmes 3.15, 3.23, 3.14 et 3.22 que, pour φ et ψ dans une bonne classe de fonctions, les propriétés suivantes sont satisfaites.

– *Linéarité*. Pour tous $a, b \in \mathbb{R}$,

$$(3.25) \quad \mathbb{E}[a\varphi(X) + b\psi(X)] = a\mathbb{E}\varphi(X) + b\mathbb{E}\psi(X)$$

ou avec notre nouvelle notation :

$$\int_{\mathbb{R}} [a\varphi + b\psi] dP_X = a \int_{\mathbb{R}} \varphi dP_X + b \int_{\mathbb{R}} \psi dP_X$$

– *Croissance*. Si $\varphi \leq \psi$, alors

$$(3.26) \quad \mathbb{E}\varphi(X) \leq \mathbb{E}\psi(X)$$

ou avec notre nouvelle notation :

$$\int_{\mathbb{R}} \varphi dP_X \leq \int_{\mathbb{R}} \psi dP_X.$$

– *Normalisation*. On note **1** la fonction constante égale à 1.

$$(3.27) \quad \mathbb{E}(\mathbf{1}) = \int_{\mathbb{R}} dP_X = P_X(\mathbb{R}) = \mathbb{P}(\Omega) = 1.$$

3.4. Fonction indicatrice d'ensemble

On introduit maintenant une fonction très pratique en calcul des probabilités.

DÉFINITION 3.28 (Fonction indicatrice). Soit V un ensemble quelconque et $W \subset V$ une partie de V . La fonction indicatrice de W est

$$\mathbf{1}_W(v) := \begin{cases} 1 & \text{si } v \in W \\ 0 & \text{sinon} \end{cases}, \quad v \in V.$$

REMARQUES 3.29.

(1) Notons que $\mathbf{1}_W(v) = \delta_v(W)$.

(2) Pour tout $B \subset \mathbb{R}$, $\mathbf{1}_{\{X \in B\}}(\omega) = \mathbf{1}_B(X(\omega)) = \begin{cases} 1 & \text{si } X(\omega) \in B \\ 0 & \text{sinon} \end{cases}$.

PROPOSITION 3.30.

(1) Pour $B \subset \mathbb{R}$, $\mathbb{E}[\mathbf{1}_{\{X \in B\}}] = \mathbb{E}[\mathbf{1}_B(X)] = \mathbb{P}(X \in B) = P_X(B)$.

(2) Pour tout réel c , $\mathbb{E}(c\mathbf{1}_\Omega) = c$.

On notera souvent la variable aléatoire égale à la constante c : $c\mathbf{1}_\Omega = c$; donc $\mathbb{E}(c) = c$. Une telle variable aléatoire est dite *déterministe*.

DÉMONSTRATION. • Preuve de (1). Commençons par le cas où X est discrète. Grâce au Théorème 3.10, $\mathbb{E}[\mathbf{1}_{\{X \in B\}}] = \mathbb{E}[\mathbf{1}_B(X)] =$
 $= \sum_{n \in N} p_n \mathbf{1}_B(x_n) = \sum_{n \in N; x_n \in B} p_n = \mathbb{P}(X \in B) = P_X(B)$.

Lorsque X est continue, $\mathbb{E}[\mathbf{1}_{\{X \in B\}}] = \mathbb{E}[\mathbf{1}_B(X)] = \int_{\mathbb{R}} \mathbf{1}_B(x) f_X(x) dx = \int_B f_X(x) dx = P_X(B)$.

• Preuve de (2). Avec (3.27) : $\mathbb{E}(c) = c\mathbb{E}(\mathbf{1}) = c \times 1$. □

3.5. Variance et écart-type

Pour mesurer la moyenne des fluctuations de X autour de sa moyenne $\mu := \mathbb{E}X$, on peut prendre la moyenne de l'écart à la moyenne : $X - \mu$. C'est-à-dire $\mathbb{E}(X - \mu)$. Mais on voit que $\mathbb{E}(X - \mu) = \mathbb{E}X - \mathbb{E}\mu = \mu - \mu = 0$. En moyenne, les écarts par défaut compensent exactement les écarts par excès. Une idée naturelle est donc de considérer la moyenne de l'écart absolu à la moyenne : $\mathbb{E}|X - \mu|$. Mais personne n'aime beaucoup travailler avec les valeurs absolues qui demandent des découpages fastidieux. C'est la raison pour laquelle on préfère considérer la moyenne du carré de l'écart à la moyenne : $\mathbb{E}[(X - \mu)^2]$. Si on change d'échelle de mesure, par exemple si X est une longueur exprimée en mètres et X' la même longueur exprimée en millimètres, on a $X' = 1000X$ d'où $\mathbb{E}[(X' - \mathbb{E}(X'))^2] = \mathbb{E}[(1000X - 1000\mathbb{E}(X))^2] = 1000^2 \mathbb{E}[(X - \mathbb{E}X)^2]$. Ces quantités diffèrent du facteur 1000^2 et s'expriment comme des longueurs au carré. Il est donc pertinent de considérer la quantité $\sqrt{\mathbb{E}[(X - \mu)^2]}$ qui conserve les bonnes unités et les facteurs d'échelle.

DÉFINITION 3.31. On suppose que $\mathbb{E}|X| < \infty$ de sorte que $\mathbb{E}X$ est bien défini. La *variance* de X est

$$\text{Var}(X) := \mathbb{E}[(X - \mathbb{E}X)^2] \in [0, +\infty]$$

Son *écart-type* est

$$\sigma(X) := \sqrt{\text{Var}(X)} \in [0, +\infty].$$

On remarque qu'en tant qu'espérance de la variable positive $(X - \mu)^2$, $\text{Var}(X)$ est un nombre positif.

Il est pratique lors de certains calculs d'utiliser les formules suivantes.

PROPOSITION 3.32. *Soit X tel que $\mathbb{E}|X| < \infty$. Nous avons*

$$(1) \text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2.$$

$$(2) \text{Var}(aX) = a^2\text{Var}(X) \text{ et } \sigma(aX) = |a|\sigma(X), \text{ pour tout réel } a \neq 0, \text{ avec la convention } a^2 \times \infty = |a| \times \infty = \infty$$

Bien sûr, si $a = 0$, $\text{Var}(0) = \sigma(0) = 0$.

$$(3) \text{Var}(X + c) = \text{Var}(X) \text{ pour tout réel } c.$$

$$(4) \text{Var}(c) = 0 \text{ pour tout réel } c.$$

DÉMONSTRATION. • Preuve de (1). Grâce à la linéarité de l'espérance (3.25) et à la Proposition 3.30-(2), en posant $\mu = \mathbb{E}X$, $\text{Var}(X) = \mathbb{E}[(X - \mu)^2] = \mathbb{E}[X^2 - 2\mu X + \mu^2] = \mathbb{E}(X^2) - 2\mu\mathbb{E}X + \mathbb{E}(\mu^2) = \mathbb{E}(X^2) - 2\mu^2 + \mu^2 = \mathbb{E}(X^2) - \mu^2$.

• Preuve de (2). A nouveau, par la linéarité de l'espérance, $\text{Var}(aX) = \mathbb{E}[(aX - a\mu)^2] = \mathbb{E}[a^2(X - \mu)^2] = a^2\mathbb{E}[(X - \mu)^2] = a^2\text{Var}(X)$.

• Preuve de (3). $\text{Var}(X + c) = \mathbb{E}[\{(X + c) - \mathbb{E}(X + c)\}^2] = \mathbb{E}[\{X + c - (\mathbb{E}X + c)\}^2] = \mathbb{E}[\{X - \mathbb{E}X\}^2] = \text{Var}(X)$.

• Preuve de (4). $\text{Var}(c) = \text{Var}(c - c) = \text{Var}(0) = 0$. □

3.6. Moments

Commençons par la définition des moments d'une variable aléatoire.

DÉFINITION 3.33. Soit X une variable aléatoire réelle.

- Si $X \geq 0$ est une variable aléatoire positive, pour tout réel $p > 0$, on appelle *moment d'ordre p* de X la quantité $\mathbb{E}[X^p] \in [0, \infty]$.
- Dans le cas général où X est une variable aléatoire réelle, pour tout *entier* $p \geq 1$ tel que $\mathbb{E}[|X|^p] < \infty$, on appelle *moment d'ordre p* de la variable aléatoire réelle X la quantité $\mathbb{E}(X^p)$.

On rappelle que les puissances non-entières ne sont définies que pour les nombres positifs par $x^p := \exp(p \ln(x))$, $x > 0, p \in \mathbb{R}$ et $0^p = 0$ si $p > 0$.

PROPOSITION 3.34 (Comparaison des moments). *On se donne deux réels $0 < p \leq q$. Soit $X \geq 0$ une variable aléatoire positive : $\mathbb{E}[X^q] < \infty \Rightarrow \mathbb{E}[X^p] < \infty$.*

Pour toute variable aléatoire réelle X : $\mathbb{E}[|X|^q] < \infty \Rightarrow \mathbb{E}[|X|^p] < \infty$.

DÉMONSTRATION. Soit $X \geq 0$. On utilise les fonctions indicatrices $\mathbf{1}_W$, voir la Définition 3.28, en remarquant que $\mathbf{1} = \mathbf{1}_W + \mathbf{1}_{W^c}$:

$$\begin{aligned} \mathbb{E}[X^p] &= \mathbb{E}[(\mathbf{1}_{\{X < 1\}} + \mathbf{1}_{\{X \geq 1\}})X^p] \\ &\stackrel{(a)}{=} \mathbb{E}[\mathbf{1}_{\{X < 1\}}X^p] + \mathbb{E}[\mathbf{1}_{\{X \geq 1\}}X^p] \\ &\stackrel{(b)}{\leq} 1 + \mathbb{E}[\mathbf{1}_{\{X \geq 1\}}X^q] \\ &\stackrel{(c)}{\leq} 1 + \mathbb{E}[X^q] < \infty. \end{aligned}$$

L'égalité (a) est une application de la linéarité de l'espérance. L'inégalité (b) vient de $\mathbf{1}_{\{0 \leq x < 1\}} x^p \leq 1$ et $x^p \leq x^q$ lorsque $x \geq 1$ et $0 < p \leq q$. On obtient l'inégalité (c) en remarquant que $\mathbf{1}_{\{x \geq 1\}} x^q \leq x^q$ lorsque $x \geq 0$. On a invoqué (3.26) pour des fonctions positives pour ces deux inégalités.

La dernière assertion de la proposition s'en déduit immédiatement. \square

COROLLAIRE 3.35. *Si $\mathbb{E}(X^2) < \infty$, alors $\mathbb{E}|X| < \infty$.
De plus, $\mathbb{V}\text{ar}(X) < \infty$ si et seulement si $\mathbb{E}(X^2) < \infty$.*

DÉMONSTRATION. La première assertion est un cas particulier de la Proposition 3.34 et la seconde s'en déduit à l'aide de la Proposition 3.32-(1). \square

3.7. Fonctions d'une variable aléatoire

Si φ est une fonction numérique suffisamment régulière et X est une variable aléatoire, alors $Y = \varphi(X)$ est aussi une variable aléatoire. Pour tout intervalle $B \subset \mathbb{R}$, notons $\varphi^{-1}(B) := \{x \in \mathbb{R}; \varphi(x) \in B\}$.

EXERCICE 3.36. Montrer que si φ est continue par morceaux, $\varphi^{-1}(B)$ est une réunion dénombrable d'intervalles.

Grâce à l'exercice précédent et à l'identité (3.46) plus bas, on peut considérer $P_X(\varphi^{-1}(B))$ et écrire

$$\begin{aligned} P_Y(B) &= \mathbb{P}(Y \in B) \\ &= \mathbb{P}(\varphi(X) \in B) \\ &= \mathbb{P}(X \in \varphi^{-1}(B)) \\ &= P_X(\varphi^{-1}(B)) \end{aligned}$$

ce qui spécifie la loi de Y . Avec $B = \bigsqcup_{n \geq 1} I_n$ où les I_n sont des intervalles disjoints, nous avons

$$(3.37) \quad \mathbb{P}(X \in B) = \sum_{n \geq 1} \mathbb{P}(X \in I_n).$$

(Notons que si B est la réunion finie de N intervalles, on peut toujours prendre $I_n = \emptyset$ pour $n > N$). Or cette quantité est entièrement déterminée par la fonction de répartition F_X de X comme le montre la Proposition 2.9.

Par exemple, lorsque φ est une application strictement monotone son application réciproque φ^{-1} est bien définie et en prenant $B =]-\infty, y]$ nous obtenons lorsque φ est strictement croissante

$$\begin{aligned} F_Y(y) &= \mathbb{P}(\varphi(X) \leq y) \\ &= \mathbb{P}(X \leq \varphi^{-1}(y)) \\ &= F_X(\varphi^{-1}(y)) \end{aligned}$$

et lorsque φ est strictement décroissante

$$\begin{aligned} F_Y(y) &= \mathbb{P}(\varphi(X) \leq y) \\ &= \mathbb{P}(X \geq \varphi^{-1}(y)) \\ &= 1 - F_X((\varphi^{-1}(y))^-) \end{aligned}$$

Donnons quelques exemples d'application de cette méthode.

- (a) Soit X une variable continue de densité f_X continue par morceaux. On cherche la loi de $Y = aX + b$ avec a et b réels.

Remarquons avant tout que lorsque $a = 0$, Y vaut b quoiqu'il arrive, sa loi est donc $P_Y = \delta_b$. On note en passant que ceci nous donne un exemple de $\varphi(X)$ discrète alors que X est continue.

Prenons maintenant $a \neq 0$ et calculons la fonction de répartition de $Y = aX + b$.

- Si $a > 0$, $F_Y(y) = \mathbb{P}(aX + b \leq y) = \mathbb{P}(X \leq (y - b)/a) = F_X((y - b)/a)$. Ce qui donne $f_Y(y) = F'_Y(y) = f_X((y - b)/a)/a$.
- Si $a < 0$, $F_Y(y) = \mathbb{P}(aX + b \leq y) = \mathbb{P}(X \geq (y - b)/a) = 1 - F_X((y - b)/a)$. Ce qui donne $f_Y(y) = F'_Y(y) = -f_X((y - b)/a)/a$.

Finalement, nous obtenons dans les deux cas

$$(3.38) \quad f_Y(y) = \frac{f_X((y - b)/a)}{|a|}, \quad y \in \mathbb{R}$$

- (b) Soit X une variable aléatoire quelconque, la fonction de répartition F_Y de $Y = X^2$ s'exprime en fonction de F_X de la manière suivante. Pour tout $y \geq 0$,

$$\begin{aligned} F_Y(y) &= \mathbb{P}(X^2 \leq y) \\ &= \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X((-\sqrt{y})^-) \end{aligned}$$

alors que pour tout $y < 0$, $F_Y(y) = 0$.

En particulier, si X admet une densité f_X continue par morceaux, F_X est dérivable partout sauf en un nombre fini de points et $F'_X = f_X$. Par conséquent Y admet la densité (définie partout sauf en un nombre fini de points)

$$(3.39) \quad f_Y(y) = F'_Y(y) = \mathbf{1}_{(y>0)} \frac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}}.$$

EXEMPLE 3.40. Si X est l'angle de la flèche de l'Exemple 2.17 et $Y = X^2$, $f_X(x) = \mathbf{1}_{[0,2\pi]}(x)/(2\pi)$ et avec (3.39) : $f_Y(y) = \mathbf{1}_{[0,4\pi^2]}/(4\pi\sqrt{y})$ de sorte que

$$\begin{aligned} \mathbb{E}(X^2) &= \int_0^{2\pi} \frac{x^2}{2\pi} dx = \frac{4}{3}\pi^2 \\ \mathbb{E}(Y) &= \int_0^{4\pi^2} \frac{\sqrt{y}}{4\pi} dy = \frac{4}{3}\pi^2 \end{aligned}$$

On constate bien évidemment que $\mathbb{E}(Y) = \mathbb{E}(X^2)$.

- (c) Les choses sont plus simples si l'on considère $Z = X^3$. En effet, pour tout $z \in \mathbb{R}$, nous avons

$$F_Z(z) = \mathbb{P}(X^3 \leq z) = \mathbb{P}(X \leq z^{1/3}) = F_X(z^{1/3}).$$

La simplicité de ce calcul vient du fait que z^3 est injective, alors que la non-injectivité de z^2 créait quelques difficultés dans l'exemple précédent. Si X admet une fonction de densité continue par morceaux, $Z = X^3$ admet la fonction de densité

$$f_Z(z) = \frac{f_X(z^{1/3})}{3z^{2/3}}.$$

Notons que cette fonction n'est pas définie en $z = 0$, mais ça n'est pas un problème puisque des fonctions de densité égales sauf sur un ensemble de longueur nulle

(Lebesgue-presque partout) correspondent à la même loi, voir la Proposition 3.43 plus bas.

3.8. Egalité en loi

Cette notion est spécifique à la théorie des probabilités.

DÉFINITION 3.41 (Egalité en loi). Deux variables aléatoires X_1 et X_2 construites respectivement sur (Ω_1, \mathbb{P}_1) et (Ω_2, \mathbb{P}_2) sont égales en loi si et seulement si elles ont la même loi : $P_{X_1} = P_{X_2}$. On note dans ce cas : $X_1 \stackrel{\mathcal{L}}{=} X_2$.

Cela ne signifie pas que

- (1) $X_1 = X_2$ ni même que
- (2) $\mathbb{P}(X_1 = X_2) = 1$, même lorsque $(\Omega_1, \mathbb{P}_1) = (\Omega_2, \mathbb{P}_2)$.

Bien sûr, (1) implique (2) qui implique l'égalité en loi.

L'égalité en loi est la notion la plus faible permettant d'identifier deux phénomènes aléatoires.

EXEMPLES 3.42.

- (1) On joue deux fois de suite à pile ou face de sorte que $\Omega_1 = \{pp, pf, fp, ff\}$ et $\mathbb{P}_1 = \frac{1}{4}(\delta_{pp} + \delta_{pf} + \delta_{fp} + \delta_{ff})$. On considère X_1 défini par : $X_1(pp) = X_1(pf) = -3$ et $X_1(fp) = X_1(ff) = \sqrt{5}$.

On lance un dé de sorte que $\Omega_2 = \{a, b, c, d, e, f\}$ avec $\mathbb{P}_2 = \frac{1}{6}(\delta_a + \delta_b + \delta_c + \delta_d + \delta_e + \delta_f)$. On considère X_2 défini par $X_2(a) = X_2(b) = X_2(c) = -3$ et $X_2(d) = X_2(e) = X_2(f) = \sqrt{5}$.

On voit que $P_{X_1} = P_{X_2} = \frac{1}{2}(\delta_{-3} + \delta_{\sqrt{5}})$, c'est-à-dire $X_1 \stackrel{\mathcal{L}}{=} X_2$.

- (2) Soit X la variable de l'Exemple 2.6 dont la loi est $\frac{1}{4}\delta_0 + \frac{1}{2}\delta_1 + \frac{1}{4}\delta_2$. Montrer que $X \stackrel{\mathcal{L}}{=} 2 - X$.

- (3) Soit X une variable aléatoire continue dont la densité est une fonction paire ; $f_X(-x) = f_X(x), \forall x$. Alors nous avons $X \stackrel{\mathcal{L}}{=} -X$. En effet, pour tout réel y nous avons

$$\begin{aligned} F_{-X}(y) &= \mathbb{P}(X \geq -y) \\ &= \int_{-y}^{+\infty} f_X(x) dx \\ &\stackrel{(a)}{=} \int_{-\infty}^y f_X(-z) dz \\ &\stackrel{(b)}{=} \int_{-\infty}^y f_X(z) dz \\ &= F_X(y) \end{aligned}$$

où l'égalité (a) s'obtient avec le changement de variable $z = -x$ et (b) est une conséquence de la parité de f_X .

Nous avons déjà remarqué que les données de F_X et P_X sont équivalentes. On en déduit le résultat suivant.

PROPOSITION 3.43. *Deux variables aléatoires X_1 et X_2 construites respectivement sur (Ω_1, \mathbb{P}_1) et (Ω_2, \mathbb{P}_2) sont égales en loi si et seulement si elles ont la même fonction de répartition :*

$$F_{X_1} = F_{X_2}.$$

Si elles sont discrètes, cela signifie qu'il existe une suite (éventuellement finie) $(x_n)_{n \in \mathbb{N}}$ de réels distincts telle que $\sum_{n \in \mathbb{N}} \mathbb{P}_1(X_1 = x_n) = 1$ et

$$\mathbb{P}_1(X_1 = x_n) = \mathbb{P}_2(X_2 = x_n), \quad \forall n \in \mathbb{N}$$

Si elles sont continues, cela signifie que leurs densités ont le même ensemble de points de discontinuité (Cf. les Définitions 2.15 et 3.17) et qu'elles sont égales partout sauf éventuellement sur cet ensemble de "longueur nulle". On dit alors qu'elles sont égales Lebesgue-presque partout et on note

$$f_{X_1} = f_{X_2}, \quad \text{Lebesgue-p.p.}$$

3.9. Définition abstraite de la loi d'une variable aléatoire

Spécifier complètement le comportement d'une variable aléatoire X devrait permettre en principe d'évaluer les quantités $\mathbb{P}(X \in B)$ pour toute partie B de \mathbb{R} . Mais cela n'est possible que si l'ensemble $\{X \in B\}$ est un événement, c'est-à-dire un élément de la tribu \mathcal{A} .

Lorsque X est une variable discrète, on peut prendre Ω dénombrable et $\mathcal{A} = 2^\Omega$ de sorte que pour tout $B \subset \mathbb{R}$, $\{X \in B\}$ est un événement.

Lorsque X est une variable aléatoire continue, comme nous l'avons déjà évoqué à la Remarque 2.24, les choses se compliquent du point de vue mathématique : on ne peut pas prendre n'importe quelle partie B . Les "bonnes" parties B de \mathbb{R} sont celles de la tribu de Borel.

DÉFINITION 3.44. La tribu de Borel de \mathbb{R} est la plus petite tribu contenant l'ensemble \mathcal{I} de tous les intervalles de \mathbb{R} . On la notera \mathcal{B} .

EXERCICE 3.45. Montrer que si $(\mathcal{A}_\gamma, \gamma \in \Gamma)$ est une collection quelconque de tribus sur le même ensemble Ω , alors l'ensemble $\bigcap_{\gamma \in \Gamma} \mathcal{A}_\gamma$ constitué des parties de Ω qui se trouvent dans toutes les tribus \mathcal{A}_γ lorsque γ parcourt l'ensemble d'indices Γ , est aussi une tribu.

La plus petite tribu contenant l'ensemble \mathcal{I} de tous les intervalles de \mathbb{R} est par définition l'intersection de toutes les tribus contenant \mathcal{I} . Cette intersection existe puisque $2^\mathbb{R}$ est une tribu qui contient \mathcal{I} , de plus en tant qu'intersection de tribus, c'est une tribu d'après l'exercice précédent. Ceci justifie la définition de la tribu de Borel \mathcal{B} .

On peut montrer, mais ça n'est pas simple, qu'il existe des parties de \mathbb{R} qui ne sont pas dans \mathcal{B} .

On retiendra que la tribu de Borel contient toutes les réunions dénombrables d'intervalles.

Avec $B = \bigsqcup_{n \geq 1} I_n$ où les I_n sont des intervalles disjoints, nous avons

$$(3.46) \quad \mathbb{P}(X \in B) = \sum_{n \geq 1} \mathbb{P}(X \in I_n).$$

(Notons que si B est la réunion finie de N intervalles, on peut toujours prendre $I_n = \emptyset$ pour $n > N$). Or cette quantité est entièrement déterminée par la fonction de répartition F_X de X comme le montre la Proposition 2.9.

DÉFINITION 3.47. La *loi* de la variable aléatoire (quelconque) X est la mesure de probabilité P_X sur $(\mathbb{R}, \mathcal{B})$ définie par

$$P_X(B) = \mathbb{P}(X \in B), \quad B \in \mathcal{B}.$$

La connaissance de P_X sur tous les intervalles de la forme $]a, b]$ permet de retrouver $F_X(x) = \mathbb{P}(X \in]-\infty, x]) = \lim_{n \rightarrow \infty} P_X(]-n, x])$, $x \in \mathbb{R}$.

Réciprocement, si on se donne F_X , grâce à la Proposition 2.9, P_X est connue sur tous les intervalles et par suite, grâce à (3.46), sur toutes les réunions dénombrables d'intervalles. On peut montrer, mais c'est assez délicat et dépasse le niveau de ce cours, qu'en fait F_X spécifie P_X complètement sur \mathcal{B} .

En résumé, F_X et P_X encodent la même information sur le comportement aléatoire de X .

De plus, P_X n'est autre que l'image sur $(\mathbb{R}, \mathcal{B})$ de la mesure de probabilité \mathbb{P} sur (Ω, \mathcal{A}) par l'application X :

$$P_X = X \# \mathbb{P}.$$

La notion de *mesure image* est présentée à l'Annexe ??.

CHAPITRE 4

Variables aléatoires usuelles

Nous présentons ici les lois des variables aléatoires les plus usitées. Certaines, comme la loi normale, sont extrêmement importantes tant sur le plan théorique que pratique (utilisation très fréquente en statistique).

4.1. Exemples de variables aléatoires discrètes

Nous présentons dans cette section les lois de Bernoulli, binomiales, de Poisson et géométriques.

Loi de Bernoulli. Il s'agit d'une des lois les plus simples. La variable aléatoire X suit la loi de Bernoulli $\mathcal{B}(p)$ de paramètre $0 \leq p \leq 1$ si sa loi est

$$P_X = q\delta_0 + p\delta_1.$$

Ceci signifie que X peut prendre les valeurs 0 et 1 avec les probabilités respectives $q = 1 - p$ et p . On obtient immédiatement que $\mathbb{E}X = q0 + p1 = p$ et que puisque $X^2 = X$ sous cette loi, $\mathbb{E}(X^2) = p$. Par conséquent, $\text{Var}X = p - p^2 = pq$.

Une variante immédiate de cette loi est $P_Y = q\delta_a + p\delta_b$ avec a, b réels. On a immédiatement $\mathbb{E}Y = qa + pb$ et du fait que $Y = a + (b - a)X$ avec $X \sim \mathcal{B}(p)$, $\text{Var}Y = (b - a)^2\text{Var}X = (b - a)^2pq$, grâce à la Proposition 3.32.

Loi binomiale. La variable aléatoire X suit la loi binomiale $\mathcal{B}(n, p)$ de paramètres $n \geq 1$ et $0 \leq p \leq 1$ si sa loi est

$$P_X = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} \delta_k$$

où comme précédemment on pose $q = 1 - p$. Ceci signifie que X peut prendre les valeurs $0, 1, \dots, n$ avec $\mathbb{P}(X = k) = \binom{n}{k} p^k q^{n-k}$ pour $0 \leq k \leq n$. On constate qu'avec $n = 1$, on retrouve $\mathcal{B}(1, p) = \mathcal{B}(p)$.

EXERCICE 4.1.

- (a) Vérifier que P_X est une mesure de probabilité.
- (b) Montrer que $\mathbb{E}X = np$ et $\text{Var}X = npq$.

SOLUTION. Nous donnons seulement la solution de $\mathbb{E}X = np$. Nous avons

$$\begin{aligned}
 \mathbb{E}X &= \sum_{k=0}^n k \frac{n!}{k!(n-k)!} p^k q^{n-k} \\
 &= np \sum_{k=0}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} q^{n-k} \\
 &\stackrel{(a)}{=} np \sum_{l=0}^n \binom{n-1}{l} p^l q^{n-1-l} \\
 &\stackrel{(b)}{=} np(p+q)^{n-1} \\
 &= np
 \end{aligned}$$

où l'on a effectué le changement de variable $l = k-1$ en (a) (on notera que $n-k = n-1-l$) et utilisé la formule du binôme de Newton en (b).

Une indication pour calculer $\text{Var}X$: commencer par calculer $\mathbb{E}[X(X-1)]$ en procédant dans le même esprit que ce que nous venons de faire. \square

Loi géométrique. La variable aléatoire X suit la loi géométrique $\mathcal{G}(p)$ de paramètre $0 < p \leq 1$ si sa loi est

$$P_X = \sum_{k=1}^{\infty} q^{k-1} p \delta_k$$

où comme précédemment on pose $q = 1-p$. Ceci signifie que X peut prendre les valeurs $1, 2, \dots$ avec $\mathbb{P}(X = k) = q^{k-1}p$ pour $k \geq 1$.

EXERCICE 4.2.

- (a) Vérifier que P_X est une mesure de probabilité.
- (b) Montrer que $\mathbb{E}X = 1/p$.

SOLUTION. On pose $\varphi(q) = \sum_{k=0}^{\infty} q^k$, $0 \leq q < 1$. On sait que $\varphi(q) = \lim_{n \rightarrow \infty} \sum_{k=0}^n q^k = \lim_{n \rightarrow \infty} (1 - q^{n+1})/(1 - q) = 1/(1 - q)$. De ce fait, $P_X(\mathbb{N}) = p \sum_{k=1}^{\infty} q^{k-1} = p \sum_{k=0}^{\infty} q^k = p/(1 - q) = 1$, ce qui montre (a). Grâce au Théorème de dérivation sous le signe somme B.3, en dérivant terme à terme la série $\sum_{k=0}^{\infty} q^k$ on obtient $\sum_{k=1}^{\infty} kq^{k-1} = \varphi'(q)$ et puisque $\varphi'(q) = \frac{d}{dq}(1/(1-q)) = 1/(1-q)^2$, on voit que $\mathbb{E}X = \sum_{k=1}^{\infty} kq^{k-1}p = p/(1 - q)^2 = 1/p$. \square

Loi de Poisson. La variable aléatoire X suit la loi de Poisson $\mathcal{P}(\lambda)$ de paramètre $\lambda > 0$ si sa loi est

$$P_X = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \delta_k.$$

Ceci signifie que X peut prendre les valeurs $0, 1, 2, \dots$ avec $\mathbb{P}(X = k) = e^{-\lambda} \lambda^k / k!$ pour $k \geq 0$ avec la conventions habituelles $\lambda^0 = 1$ et $0! = 1$ de sorte que $\mathbb{P}(X = 0) = e^{-\lambda}$.

EXERCICE 4.3.

- (a) Vérifier que P_X est une mesure de probabilité.
- (b) Montrer que $\mathbb{E}X = \text{Var}X = \lambda$.

SOLUTION. Commençons par rappeler que pour tout réel x

$$(4.4) \quad e^x = \sum_{l \geq 0} \frac{x^l}{l!}$$

On en déduit immédiatement que $P_X(\mathbb{N}) = e^{-\lambda} \sum_{k \geq 0} \lambda^k / k! = e^{-\lambda} e^\lambda = 1$. Montrons que $\mathbb{E}X = \lambda$. Nous avons

$$\begin{aligned} \mathbb{E}X &= \sum_{k \geq 0} k e^{-\lambda} \frac{\lambda^k}{k!} = \sum_{k \geq 1} k e^{-\lambda} \frac{\lambda^k}{k!} \\ &= \lambda e^{-\lambda} \sum_{k \geq 1} \frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda e^{-\lambda} \sum_{l \geq 0} \frac{\lambda^l}{l!} = \lambda e^{-\lambda} e^\lambda = \lambda \end{aligned}$$

où l'on a effectué le changement de variable $l = k - 1$ et utilisé la formule (4.4).

Calculons de façon similaire

$$\begin{aligned} \mathbb{E}[X(X-1)] &= \sum_{k \geq 0} k(k-1) e^{-\lambda} \frac{\lambda^k}{k!} = \sum_{k \geq 2} k(k-1) e^{-\lambda} \frac{\lambda^k}{k!} \\ &= \lambda^2 e^{-\lambda} \sum_{k \geq 2} \frac{\lambda^{k-2}}{(k-2)!} \\ &= \lambda^2 e^{-\lambda} \sum_{l \geq 0} \frac{\lambda^l}{l!} = \lambda^2 e^{-\lambda} e^\lambda = \lambda^2 \end{aligned}$$

On en déduit que $\text{Var}X = \mathbb{E}[X(X-1)] + \mathbb{E}X - (\mathbb{E}X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$. □

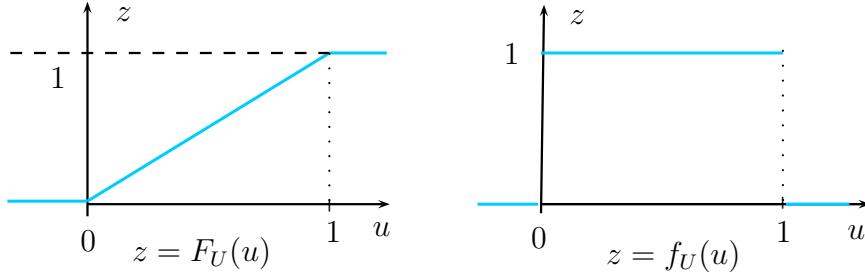
EXERCICE 4.5. En vous inspirant de la solution précédente, montrer que pour tout entier $k \geq 1$, $\mathbb{E}[X(X-1) \cdots (X-k+1)] = \lambda^k$.

4.2. Exemples de variables aléatoires continues

Nous présentons dans cette section les lois uniformes, exponentielles, normales, Gamma et de Cauchy.

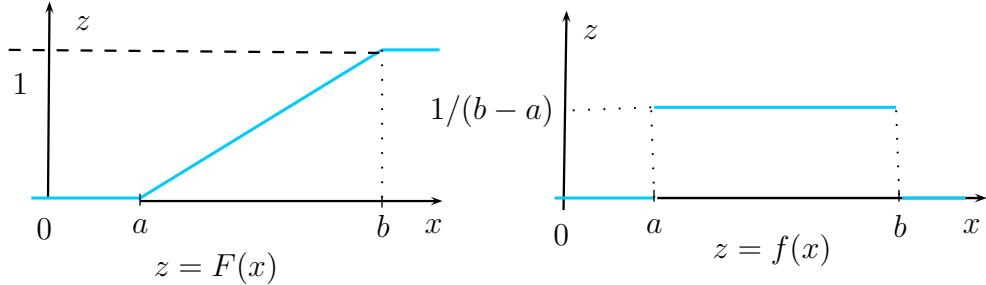
Loi uniforme. Nous avons déjà rencontré la variable U de loi uniforme sur $[0, 1]$. Ses fonctions de répartition et de densité sont

$$F_U(u) = \begin{cases} 0 & \text{si } u \leq 0 \\ u & \text{si } 0 \leq u \leq 1 \\ 1 & \text{si } u \geq 1 \end{cases} \quad \text{et} \quad f_U(u) = \mathbf{1}_{(0 \leq u \leq 1)}, \quad u \in \mathbb{R}.$$



Une variable aléatoire X suit une loi uniforme sur $[a, b]$ si elle a la même loi (c'est-à-dire la même fonction de répartition) que $a + (b - a)U$. Ses fonctions de répartition et de densité (voir (3.38)) sont

$$F(x) = \begin{cases} 0 & \text{si } x \leq a \\ (x - a)/(b - a) & \text{si } a \leq x \leq b \\ 1 & \text{si } x \geq b \end{cases} \quad \text{et} \quad f(x) = \frac{1_{(a \leq x \leq b)}}{b - a}, \quad x \in \mathbb{R}.$$



On note $\mathcal{U}(a, b)$ la loi uniforme sur $[a, b]$. Nous avons donc

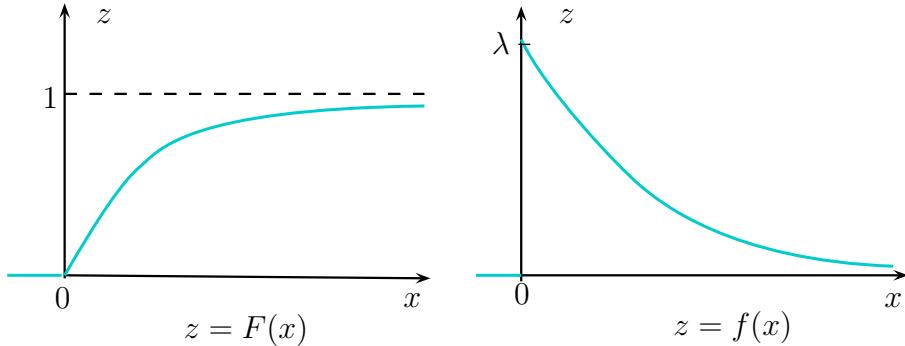
$$(4.6) \quad a + (b - a)U \sim \mathcal{U}(a, b)$$

lorsque $U \sim \mathcal{U}(0, 1)$.

EXERCICE 4.7. Vérifier que $\mathbb{E}(X) = (a + b)/2$ et que $\text{Var}(X) = (b - a)^2/12$.

Loi exponentielle. Une variable aléatoire X suit la loi exponentielle de paramètre λ , notée $\mathcal{E}(\lambda)$, si ses fonction de répartition et fonction de densité sont

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 - e^{-\lambda x} & \text{si } x \geq 0 \end{cases} \quad \text{et} \quad f(x) = \mathbf{1}_{(x \geq 0)} \lambda e^{-\lambda x}, \quad x \in \mathbb{R}.$$



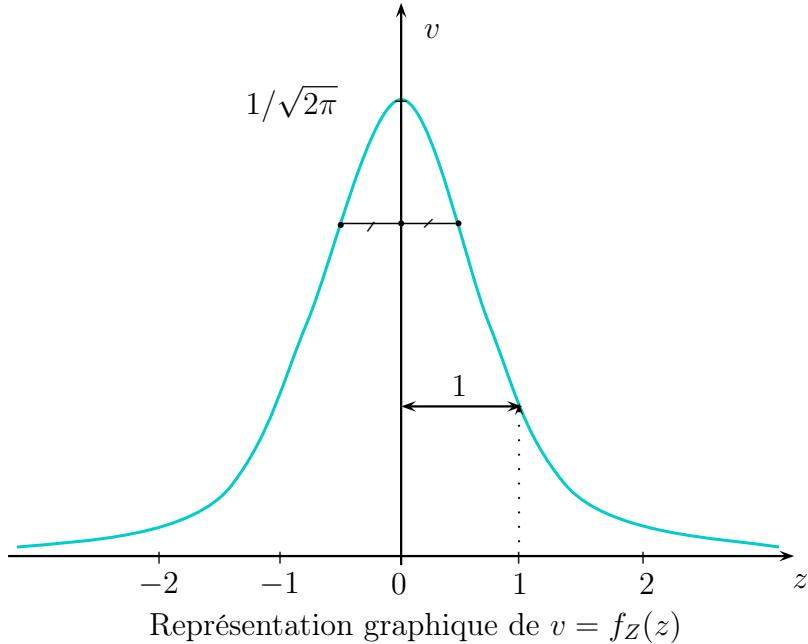
EXERCICE 4.8. Vérifier que $\mathbb{E}(X) = 1/\lambda$ et que $\text{Var}(X) = 1/\lambda^2$.

Cette variable aléatoire sert souvent à modéliser des temps d'attente. Elle intervient de façon fondamentale dans la construction des processus de Markov à temps continu que l'on rencontre lors de la modélisation de système de files d'attente (réseaux informatiques, guichets, etc...).

Loi normale. C'est probablement la loi continue la plus importante. On l'appelle aussi *loi de Gauss* ou loi gaussienne. On dit qu'une variable aléatoire Z suit une *loi normale centrée réduite* si sa fonction de densité est

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right), \quad z \in \mathbb{R}$$

Cette loi est notée $\mathcal{N}(0, 1)$.



Il n'existe pas d'expression analytique de la fonction de répartition de Z . On la note traditionnellement

$$(4.9) \quad \Phi(y) = \mathbb{P}(Z \leq y) = \int_{-\infty}^y \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz.$$

Toutefois, on peut vérifier que $\lim_{y \rightarrow +\infty} \Phi(y) = \int_{\mathbb{R}} f_Z(z) dz = 1$. Pour cela posons $I = \int_{\mathbb{R}} f_Z(z) dz$. Nous avons par un simple jeu d'écriture sur les variables d'intégration

$$\begin{aligned} I^2 &= \int_{\mathbb{R}} f_Z(x) dx \int_{\mathbb{R}} f_Z(y) dy = \iint_{\mathbb{R}^2} f_Z(x) f_Z(y) dx dy \\ &= \frac{1}{2\pi} \iint_{\mathbb{R}^2} e^{-x^2/2} e^{-y^2/2} dx dy = \frac{1}{2\pi} \iint_{\mathbb{R}^2} e^{-(x^2+y^2)/2} dx dy \\ &\stackrel{(a)}{=} \frac{1}{2\pi} \int_0^{2\pi} \int_0^\infty e^{-r^2/2} r dr d\theta = \frac{1}{2\pi} \left(\int_0^{2\pi} d\theta \right) \left(\int_0^\infty e^{-r^2/2} r dr \right) \\ &\stackrel{(b)}{=} \int_0^\infty e^{-u} du \\ &= 1 \end{aligned}$$

où nous avons effectué

- en (a) : le changement de variables en coordonnées polaires : $x = r \cos \theta$, $y = r \sin \theta$ avec $r \geq 0$ et $0 \leq \theta < 2\pi$ de sorte que $r^2 = x^2 + y^2$ et $dxdy$ est remplacé par $rdrd\theta$;
- en (b) : le changement de variable $u = r^2/2$.

Puisque $I > 0$ et $I^2 = 1$, nous venons de montrer que

$$(4.10) \quad \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-z^2/2} dz = \int_{\mathbb{R}} f_Z(z) dz = 1.$$

EXERCICE 4.11. Vérifier que $\mathbb{E}(Z) = 0$ et que $\text{Var}(Z) = 1$.

SOLUTION. L'intégrale $\mathbb{E}Z = \int_{\mathbb{R}} z f_Z(z) dz$ est nulle car la fonction $z \mapsto z f_Z(z)$ est impaire et intégrable. Donc $\mathbb{E}Z = 0$ et $\text{Var}Z = \mathbb{E}Z^2 = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} z^2 e^{-z^2/2} dz$. On effectue une intégration par parties $\int uv' = [uv] - \int u'v$ avec $u'(z) = ze^{-z^2/2}$ et $v(z) = z$. Nous avons $u(z) = -e^{-z^2/2}$ et $v'(z) = 1$, de sorte que $\int_{\mathbb{R}} z^2 e^{-z^2/2} dz = [-ze^{-z^2/2}]_{-\infty}^{+\infty} + \int_{\mathbb{R}} e^{-z^2/2} dz = 0 + \sqrt{2\pi} \int_{\mathbb{R}} f_Z(z) dz$. On en déduit avec (4.10) que $\mathbb{E}Z^2 = 1$. \square

EXERCICE 4.12. Montrer que $-Z \stackrel{\mathcal{L}}{=} Z$.

SOLUTION. Pour tout réel y , $F_{-Z}(y) = \mathbb{P}(-Z \leq y) = \mathbb{P}(Z \geq -y) = \int_{-y}^{\infty} f_Z(z) dz = \int_{-y}^{\infty} f_Z(-z) dz = - \int_{y}^{-\infty} f_Z(x) dx = \int_{-\infty}^y f_Z(x) dx = F_Z(y)$ où nous avons utilisé successivement la parité de f_Z : $f_Z(z) = f_Z(-z)$ et le changement de variable $x = -z$. Par conséquent Z et $-Z$ ont la même fonction de répartition. \square

DÉFINITION 4.13. De manière générale, une variable aléatoire X est dite *centrée* si $\mathbb{E}(X) = 0$ et *réduite* si $\text{Var}(X) = 1$.

Une variable aléatoire X suit une loi normale de paramètres μ et σ^2 ($\mu \in \mathbb{R}$, $\sigma > 0$) notée $\mathcal{N}(\mu, \sigma^2)$, si elle peut s'écrire sous la forme

$$(4.14) \quad X = \mu + \sigma Z$$

où Z suit une loi $\mathcal{N}(0, 1)$. Cette loi est notée $\mathcal{N}(\mu, \sigma^2)$.

EXERCICE 4.15. Vérifier que $\mathbb{E}(X) = \mu$ et que $\text{Var}(X) = \sigma^2$.

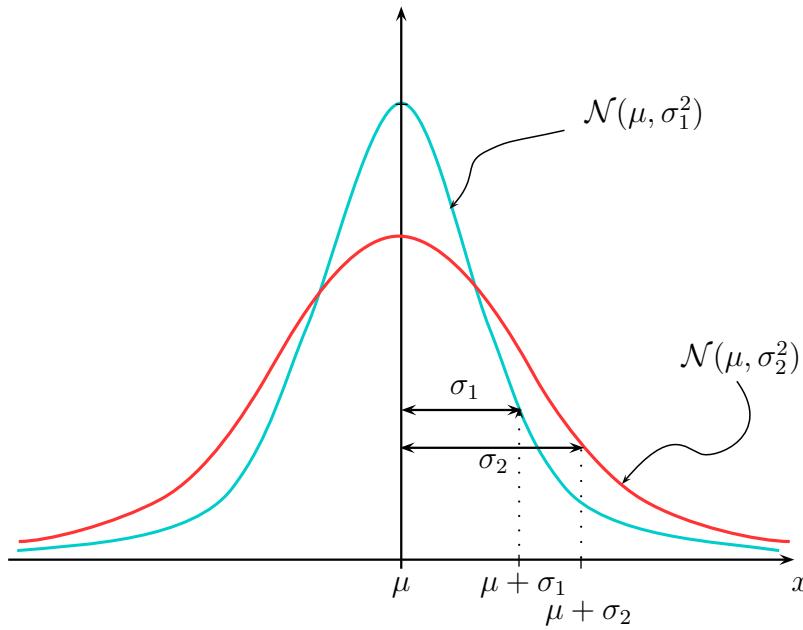
La fonction de répartition de X est

$$\begin{aligned} F(x) &= \mathbb{P}(X \leq x) = \mathbb{P}(\mu + \sigma Z \leq x) = \mathbb{P}(Z \leq (x - \mu)/\sigma) \\ &= \Phi((x - \mu)/\sigma), \end{aligned}$$

de sorte qu'avec $f(x) = F'(x)$, nous obtenons l'expression de la fonction de densité de X suivante :

$$(4.16) \quad f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.$$

La figure suivante donne la représentation graphique des densités de probabilité des lois $\mathcal{N}(\mu, \sigma_1^2)$ et $\mathcal{N}(\mu, \sigma_2^2)$ avec $0 < \sigma_1 < \sigma_2$. On constate que ces densités sont symétriques par rapport à la moyenne μ et que les aires situées entre les courbes et l'axe des x sont les mêmes pour les deux densités. De plus, la densité de $\mathcal{N}(\mu, \sigma_1^2)$ est plus concentrée autour de la moyenne μ que celle de $\mathcal{N}(\mu, \sigma_2^2)$.



L'exercice suivant permet de donner une approximation de la fonction de répartition Φ définie en (4.9) bien qu'on n'en connaisse pas d'expression analytique exacte.

EXERCICE 4.17. Pour tout $y > 0$, nous avons

$$(a) \mathbb{P}(Z \geq y) = 1 - \Phi(y) \leq \frac{e^{-y^2/2}}{y\sqrt{2\pi}} \text{ et}$$

$$(b) \mathbb{P}(|Z| \geq y) \leq \frac{2e^{-y^2/2}}{y\sqrt{2\pi}}.$$

SOLUTION. En remarquant que $z/y \geq 1$ pour tout $z \geq y$, nous avons

$$\begin{aligned} \mathbb{P}(Z \leq y) &= \int_y^\infty \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \\ &\leq \int_y^\infty \frac{1}{\sqrt{2\pi}} \frac{z}{y} e^{-z^2/2} dz = \frac{1}{y\sqrt{2\pi}} \int_y^\infty z e^{-z^2/2} dz \\ &= \frac{1}{y\sqrt{2\pi}} [-e^{-z^2/2}]_y^\infty = \frac{e^{-y^2/2}}{y\sqrt{2\pi}} \end{aligned}$$

ce qui prouve (a). On en déduit (b) en remarquant que $\mathbb{P}(|Z| \geq y) = \mathbb{P}(Z \leq -y) + \mathbb{P}(Z \geq y) = \mathbb{P}(-Z \geq y) + \mathbb{P}(Z \geq y) = 2\mathbb{P}(Z \geq y)$ puisque $-Z$ a la même loi que Z , voir l'Exercice 4.12. \square

Notons que les majorations de l'exercice précédent sont très mauvaises pour y proche de 0, puisqu'elles sont en $1/y$ au voisinage zéro. En revanche ces estimées s'améliorent beaucoup pour des grandes valeurs de y . On trouve $\mathbb{P}(|Z| \geq 3) \leq 0,0533$ ainsi que $\mathbb{P}(|Z| \geq 4) \leq 0,0021$, $\mathbb{P}(|Z| \geq 5) \leq 3 \cdot 10^{-5}$ et $\mathbb{P}(|Z| \geq 6) \leq 2 \cdot 10^{-7}$. En pratique, c'est-à-dire plus de 997 fois sur 1000, Z prend ses valeurs entre -4 et 4.

CHAPITRE 5

Fonctions génératrices et caractéristiques

Nous allons présenter des méthodes efficaces pour calculer les moments de certaines lois, ainsi que les lois de sommes de variables indépendantes. Nous commençons par étudier les variables aléatoires à valeurs entières, puis les variables générales.

Rappelons que le moment d'ordre k de la variable aléatoire X est $\mathbb{E}(X^k)$, voir la Définition 3.33. Les principaux résultats abstraits concernant les moments sont présentés en Chapitre 13.

Dans ce qui suit on notera $f^{(k)}$ la dérivée d'ordre k de la fonction f .

5.1. Le cas des variables entières

On dit qu'une variable aléatoire X est entière si elle prend ses valeurs dans l'ensemble \mathbb{N} des nombres entiers. Sa loi est donc de la forme $P_X = \sum_{n \geq 0} p_n \delta_n$. C'est le cas des variables binomiales, géométriques et de Poisson.

DÉFINITION 5.1. Soit X une variable entière. Sa *fonction génératrice* est définie pour tous $0 \leq t \leq 1$ par $G_X(t) = \mathbb{E}(t^X)$.

On remarque que puisque $0 \leq t \leq 1$ et X est entier, nous avons $0 \leq t^X \leq 1$ de sorte que $0 \leq \mathbb{E}(t^X) \leq 1$ est bien défini. En notant $p_n = \mathbb{P}(X = n)$, $n \in \mathbb{N}$, nous obtenons bien sûr

$$(5.2) \quad G_X(t) = \sum_{n \geq 0} p_n t^n = p_0 + \sum_{n \geq 1} p_n t^n, \quad 0 \leq t \leq 1$$

avec $G_X(1) = \mathbb{E}(1) = 1$ et $G_X(0) = p_0$. Cette dernière égalité est une convention puisque $G_X(0) = p_0 0^0$: nous avons choisi de prendre $0^0 = 1$. Cette convention est justifiée du fait qu'elle garantit la continuité de $G_X(t)$ en $t = 0$. En effet, grâce au Théorème B.2, puisque $0 \leq t^X \leq 1$ est borné, $\lim_{t \downarrow 0} G_X(t) = p_0 + \lim_{t \downarrow 0} \sum_{n \geq 1} p_n t^n = p_0 + \sum_{n \geq 1} 0 = p_0$.

PROPOSITION 5.3. *Pour tout entier $k \geq 1$ tel que $\mathbb{E}(X^k) < \infty$, nous avons*

$$\mathbb{E}[X(X-1) \cdots (X-k+1)] = G_X^{(k)}(1)$$

où $G_X^{(k)}(1)$ est la dérivée à gauche d'ordre k de G_X en 1.

On remarque que puisque X ne prend que des valeurs entières, $X(X-1) \cdots (X-k+1) = 0$ si $X \in \{0, \dots, k-1\}$ de sorte que $\mathbb{E}[X(X-1) \cdots (X-k+1)] \geq 0$.

On appelle $\mathbb{E}[X(X-1) \cdots (X-k+1)]$ le k -ième moment factoriel de X .

DÉMONSTRATION. Du fait que $\mathbb{E}(X^k) < \infty$, nous avons aussi grâce à la Proposition 3.34 : $\mathbb{E}(X^l) < \infty$ pour tous $1 \leq l \leq k$. Ce qui implique clairement que $\mathbb{E}[X(X-1) \cdots (X-l+1)] < \infty$ pour tous $1 \leq l \leq k$.

Commençons par le cas $k = 1$ sous l'hypothèse $\mathbb{E}X < \infty$. On peut donc appliquer le théorème de dérivation sous le signe somme énoncé au Théorème B.3 pour obtenir

$G'_X(1) = \sum_{n \geq 1} p_n n t^{n-1} \Big|_{t=1} = \sum_{n \geq 1} p_n n$ puisque $\mathbb{E}X = \sum_{n \geq 1} p_n n < \infty$. En recommandant, on montre de même que $G''_X(1) = \sum_{n \geq 2} p_n n(n-1) t^{n-2} \Big|_{t=1} = \sum_{n \geq 2} p_n n(n-1)$ sous l'hypothèse $\sum_{n \geq 2} p_n n(n-1) = \mathbb{E}[X(X-1)] < \infty$. En dérivant k fois, nous obtenons

$$G_X^{(k)}(1) = \sum_{n \geq k} p_n n(n-1) \cdots (n-k+1) = \mathbb{E}[X(X-1) \cdots (X-l+1)]$$

sous l'hypothèse $\mathbb{E}[X(X-1) \cdots (X-l+1)] < \infty$. \square

EXEMPLES 5.4.

- (a) La loi de Bernoulli $\mathcal{B}(p)$ de paramètre $0 \leq p \leq 1$ est $P_X = q\delta_0 + p\delta_1$ où $q = 1 - p$. Par conséquent, pour tout $0 \leq t \leq 1$, $G_X(t) = qt^0 + pt^1 = q + pt$. On a bien sûr, $G_X(0) = q$, $G_X(1) = q + p = 1$ et $\mathbb{E}X = G'_X(1) = p$.
- (b) La loi binomiale $\mathcal{B}(n, p)$ de paramètres $n \geq 1$ et $0 \leq p \leq 1$ est $\sum_{k=0}^n \binom{n}{k} p^k q^{n-k} \delta_k$ de sorte que $G_X(t) = \sum_{k=0}^n p^k q^{n-k} t^k = \sum_{k=0}^n (pt)^k q^{n-k} = (q + pt)^n$ en utilisant la formule du binôme de Newton. Avec $n = 1$, on retrouve la formule précédente pour $\mathcal{B}(p)$. On obtient $\mathbb{E}X = G'_X(1) = np(q + pt)^{n-1} \Big|_{t=1} = np(q + p) = np$ ainsi que $\mathbb{E}[X(X-1)] = G''_X(1) = n(n-1)p^2(q + pt)^{n-2} \Big|_{t=1} = n(n-1)p^2$. On en déduit que $\text{Var}(X) = \mathbb{E}[X(X-1)] + \mathbb{E}X - (\mathbb{E}X)^2 = n(n-1)p^2 + np - (np)^2 = npq$.
- (c) La loi de Poisson $\mathcal{P}(\lambda)$ de paramètre $\lambda > 0$ est $\sum_{n \geq 0} e^{-\lambda} \lambda^n / n! \delta_n$ de sorte que $G_X(t) = e^{-\lambda} \sum_{n \geq 0} \lambda^n / n! t^n = e^{-\lambda} \sum_{n \geq 0} (\lambda t)^n / n! = e^{-\lambda} e^{\lambda t} = e^{\lambda(t-1)}$. On a $\mathbb{E}X = G'_X(1) = \lambda e^{\lambda(t-1)} \Big|_{t=1} = \lambda$, ainsi que $\mathbb{E}[X(X-1)] = G''_X(1) = \lambda^2 e^{\lambda(t-1)} \Big|_{t=1} = \lambda^2$. On en déduit que $\text{Var}(X) = \mathbb{E}[X(X-1)] + \mathbb{E}X - (\mathbb{E}X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.
- (d) La loi géométrique $\mathcal{G}(p)$ est $\sum_{n \geq 1} q^{n-1} p \delta_n$. Par conséquent $G_X(t) = \sum_{n \geq 1} q^{n-1} p t^n = pt \sum_{n \geq 1} (qt)^{n-1} = pt \sum_{n \geq 0} (qt)^n = pt / (1 - qt)$. On obtient donc $\mathbb{E}X = G'_X(1) = [p(1 - qt) + pqt] / (1 - qt)^2 \Big|_{t=1} = 1/p$.

Comme le montre le résultat suivant, la fonction génératrice permet de retrouver la loi de X .

PROPOSITION 5.5. *Soit X une variable aléatoire entière de fonction génératrice G_X . Nous avons*

$$p_n = G_X^{(n)}(0) / n!, \quad n \geq 0$$

où $G_X^{(n)}(0)$ est la dérivée n -ième à droite de G_X en 0.

DÉMONSTRATION. La preuve est analogue à celle de la Proposition 5.3. En dérivant n fois terme à terme la série (5.2), on obtient $G_X^{(n)}(t) = \sum_{k=n}^{\infty} p_k k(k-1) \cdots (k-n+1) t^{k-n} = p_n n! + \sum_{k=n+1}^{\infty} p_k k(k-1) \cdots (k-n+1) t^{k-n}$ et en $t = 0$: $G_X^{(n)}(0) = p_n n! + 0$. \square

De ce fait G_X caractérise la loi de la variable entière X .

Un développement illimité formel en $t = 0$ de G_X donne $G_X(t) = \sum_{n \geq 0} G_X^{(n)}(0) / n! t^n$ (un tel développement s'appelle un développement en série entière). La proposition précédente exprime que l'on peut identifier terme à terme cette série formelle avec la série (5.2) : $G_X(t) = \sum_{n \geq 0} p_n t^n$.

5.2. Fonctions caractéristiques

On considère maintenant une variable X générale. On cherche une fonction analogue à G_X qui permette de calculer aisément à l'aide de dérivations successives les moments de X . La généralisation naturelle de la fonction $X \mapsto t^X$ lorsque X peut prendre des valeurs non-entières s'obtient en posant $t = e^s$ ce qui nous donne $X \mapsto e^{sX}$. De sorte que la généralisation de $G_X(t) = \mathbb{E}t^X$ est $L_X(s) = \mathbb{E}e^{sX}$.

DÉFINITIONS 5.6.

(1) La *transformée de Laplace* de la loi de X est définie par

$$s \in \mathbb{R} \mapsto L_X(s) = \mathbb{E}e^{sX} \in [0, \infty]$$

(2) La *transformée de Fourier* de la loi de X est définie par

$$s \in \mathbb{R} \mapsto \phi_X(s) = \mathbb{E}e^{isX} \in \mathbb{C}$$

où i est le nombre imaginaire tel que $i^2 = -1$. On appelle aussi ϕ_X la *fonction caractéristique* de la loi de X .

REMARQUES 5.7.

- (1) Puisque $e^{sX} \geq 0$, son espérance $L_X(s) = \mathbb{E}e^{sX}$ est toujours définie dans $[0, \infty]$ (en incluant la valeur $+\infty$).
- (2) De même, $e^{isX} = \cos(sX) + i \sin(sX)$ est une variable bornée et son espérance $\phi_X(s) = \mathbb{E}e^{isX} = \mathbb{E}[\cos(sX)] + i\mathbb{E}[\sin(sX)]$ est un nombre complexe bien défini puisque ses parties réelle et imaginaire sont intégrables puisque bornées.
- (3) En particulier, la fonction caractéristique $\phi_X(s)$ est définie pour tout réel s alors qu'on peut avoir $L_X(s) = +\infty$ pour tout s non nul comme par exemple lorsque X suit une loi de Cauchy, voir (??).
- (4) Lorsque X est une variable entière, nous avons $L_X(s) = G_X(e^s)$ et $\phi_X(s) = G_X(e^{is})$, $s \in \mathbb{R}$.

THÉORÈME 5.8.

(1) *On suppose qu'il existe $s_o > 0$ tel que $\mathbb{E}e^{s_o|X|} < \infty$. Alors, pour tout $k \geq 1$, $\mathbb{E}|X|^k < \infty$ et*

$$\mathbb{E}(X^k) = L_X^{(k)}(0).$$

(2) *Sous les mêmes hypothèses qu'en (1), nous avons*

$$(\ln L_X)'(0) = \mathbb{E}X \quad \text{et} \quad (\ln L_X)''(0) = \mathbb{V}\text{ar}X.$$

(3) *Si $\mathbb{E}|X|^k < \infty$ alors ϕ_X est k fois différentiable et*

$$\mathbb{E}X^k = (-i)^k \phi_X^{(k)}(0).$$

La première assertion du théorème montre que l'hypothèse $\mathbb{E}e^{s_o|X|} < \infty$ faite en (1) et (2) est bien plus restrictive que celle faite en (3). Ceci justifie l'usage de la fonction caractéristique plutôt que celui de la transformée de Laplace dans certaines situations. Notons que les calculs sont essentiellement les mêmes avec L_X et ϕ_X du fait que formellement $\phi_X(s) = L_X(is)$.

DÉMONSTRATION. C'est une application directe du Théorème B.3 de dérivation sous le signe somme.

- Preuve de (1). Pour tout k , il existe $c > 0$ tel que $|x|^k \leq c + e^{s_0|x|}, \forall x \in \mathbb{R}$. Par conséquent, $\mathbb{E}|X|^k \leq c + \mathbb{E}e^{s_0|X|} < \infty$.

La dérivée k -ième de $s \mapsto e^{sX}$ est $X^k e^{sX}$. Or nous avons $|X^k e^{sX}| = |X|^k e^{sX} \leq c + e^{s_0|X|}$ dès que $|s| \leq s_1$ avec $0 < s_1 < s_0$ pour une certaine constante c . Sous notre hypothèse, nous avons $\mathbb{E}|X^k e^{sX}| \leq c + \mathbb{E}e^{s_0|X|} < \infty$ pour tout s tel que $|s| \leq s_1$, ce qui permet d'appliquer le Théorème B.3 de dérivation en $s = 0$ (avec $Y = c + e^{s_0|X|}$). Ceci nous donne $L_X^{(k)}(0) = \mathbb{E}(X^k e^{0X}) = \mathbb{E}X^k$ qui est le résultat annoncé.

- Preuve de (2). Nous avons $(\ln L_X)' = L'_X/L_X$ et $(\ln L_X)'' = L''_X/L_X - L'^2_X/L_X^2$. En particulier en 0, nous obtenons grâce à (1), $(\ln L_X)'(0) = L'_X(0)/L_X(0) = \mathbb{E}X$ puisque $L_X(0) = 1$ et $(\ln L_X)''(0) = L''_X(0)/L_X(0) - L'^2_X(0)/L_X^2(0) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \text{Var}X$.

- Preuve de (3). Elle est analogue à celle de la seconde partie de (1). La dérivée k -ième de $s \mapsto e^{isX}$ est $i^k X^k e^{isX}$. Or nous avons $|i^k X^k e^{isX}| = |X|^k$ pour tout s et nous faisons l'hypothèse que $\mathbb{E}|X|^k < \infty$. À l'aide du Théorème B.3 de dérivation en $s = 0$ nous obtenons $\phi_X^{(k)}(0) = \mathbb{E}(i^k X^k e^{0X}) = i^k \mathbb{E}X^k$ qui est le résultat annoncé. \square

REMARQUE 5.9. Le développement formel en série entière de $L_X : L_X(s) = \sum_{k \geq 0} L_X^{(k)}(0) s^k / k!$, peut nous permettre d'identifier rapidement les dérivées $L_X^{(k)}(0)$ lorsqu'on en connaît l'expression $L_X(s) = \sum_{k \geq 0} a_k s^k$. Nous avons alors $L_X^{(k)}(0) = k! a_k$, $k \geq 0$.

Un raisonnement analogue fonctionne lorsqu'on ne connaît qu'un développement limité en 0 à l'ordre K : $L_X(s) = \sum_{k=0}^K a_k s^k + s^k \epsilon(s)$, pour identifier les K premières dérivées en 0 de L_X .

EXEMPLES 5.10.

- Loi de Poisson $\mathcal{P}(\lambda)$. En reprenant l'Exemple 5.4-(c), avec la Remarque 5.7-(4) nous obtenons $L_X(s) = \exp(\lambda(e^s - 1))$ donc $\ln L_X(s) = \lambda(e^s - 1)$ de sorte que $(\ln L_X)'(s) = (\ln L_X)''(s) = \lambda e^s$. Avec le Théorème 5.8-(2) on retrouve $\mathbb{E}X = \text{Var}X = \lambda$.
- Loi géométrique $\mathcal{G}(p)$. En reprenant l'Exemple 5.4-(d), avec la Remarque 5.7-(4) nous obtenons $L_X(s) = pe^s / (1 - qe^s)$ donc $\ln L_X(s) = \ln p + s - \ln(1 - qe^s)$ de sorte que $(\ln L_X)'(s) = 1 + qe^s / (1 - qe^s)$ et $(\ln L_X)''(s) = \frac{qe^s(1-qe^s) + q^2e^{2s}}{(1-qe^s)^2}$. Avec le Théorème 5.8-(2) on retrouve $\mathbb{E}X = 1/p$ et on obtient $\text{Var}X = (qp + q^2) / p^2 = (1 - p) / p^2$.
- Loi exponentielle $\mathcal{E}(\lambda)$. Puisque $f_X(x) = \mathbf{1}_{\{x \geq 0\}} \lambda e^{-\lambda x}$, nous avons $L_X(s) = \lambda \int_0^\infty e^{sx} e^{-\lambda x} dx = \lambda \int_0^\infty e^{-(\lambda-s)x} dx$. Cette intégrale est convergente si et seulement si $s < \lambda$ et dans ce cas $L_X(s) = \lambda / (\lambda - s)$. Nous sommes bien dans les conditions d'application du Théorème 5.8-(1). Lorsque $|s| / \lambda < 1$, nous avons $L_X(s) = 1 / (1 - s/\lambda) = \sum_{k \geq 0} (s/\lambda)^k = \sum_{k \geq 0} \frac{s^k}{k!} \frac{k!}{\lambda^k}$. En tenant compte de la Remarque 5.9, nous obtenons $L_X^{(k)}(0) = k! / \lambda^k$, donc $\mathbb{E}X^k = k! / \lambda^k$.

Compte tenu de l'importance des variables aléatoires normales nous isolons le calcul de leurs transformées de Laplace et fonctions caractéristiques.

PROPOSITION 5.11.

- Soit Z une variable aléatoire normale standard : $Z \sim \mathcal{N}(0, 1)$. Nous avons pour tout réel s , $L_Z(s) = e^{s^2/2}$ et $\phi_Z(s) = e^{-s^2/2}$.

(2) Soit X une variable aléatoire normale de loi $\mathcal{N}(\mu, \sigma^2)$. Nous avons pour tout réel s , $L_X(s) = e^{\mu s + \sigma^2 s^2/2}$ et $\phi_X(s) = e^{i\mu s - \sigma^2 s^2/2}$.

DÉMONSTRATION. • Preuve de (1). Nous ne donnons que la preuve concernant L_Z en admettant que le lien formel $\phi_Z(s) = L_Z(is)$ est rigoureux dans ce cas. Cette identité nécessite la notion de prolongement analytique (prolongement de \mathbb{R} à \mathbb{C}) qui n'est pas du niveau de ce cours.

Pour tout réel s ,

$$\begin{aligned} L_Z(s) &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{sz} e^{-z^2/2} dz \\ &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{sz - z^2/2} dz \\ &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z^2 - 2sz + s^2)} e^{s^2/2} dz \\ &= e^{s^2/2} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-s)^2} dz \\ &= e^{s^2/2} \end{aligned}$$

où la dernière égalité provient de $\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-s)^2} dz = 1$, la condition de normalisation de la densité $\mathcal{N}(s, 1)$, voir (4.16).

En admettant $\phi_Z(s) = L_Z(is)$, on voit que $\phi_Z(s) = e^{-s^2/2}$.

• Preuve de (2). Grâce à (4.14) nous avons $X = \mu + \sigma Z$ de sorte que $L_X(s) = \mathbb{E}e^{s(\mu + \sigma Z)} = e^{s\mu} L_Z(\sigma s)$ et $\phi_X(s) = \mathbb{E}e^{is(\mu + \sigma Z)} = e^{is\mu} \phi_Z(\sigma s)$. \square

CHAPITRE 6

Couples aléatoires

Beaucoup d'énoncés probabilistes intéressants s'expriment à l'aide d'une paire de variables aléatoires X, Y . Nous allons étudier le problème de leur variation conjointe sur le même domaine Ω . Dans tout ce qui va suivre, les variables aléatoires sont définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

6.1. Lois jointe et marginales

La *loi du couple* (X, Y) est la mesure de probabilité $P_{X,Y}$ sur \mathbb{R}^2 qui est spécifiée par

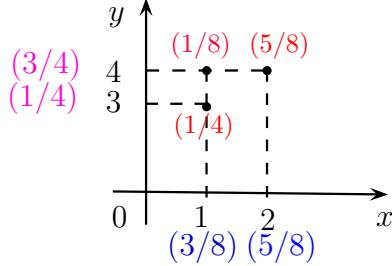
$$P_{X,Y}(A \times B) = \mathbb{P}(X \in A \text{ et } Y \in B)$$

pour tous intervalles A et B . On appelle *lois marginales* du couple (X, Y) les lois P_X et P_Y de X et de Y . Nous avons pour tous intervalles A et B ,

$$\begin{aligned} P_X(A) &= P_{X,Y}(A \times \mathbb{R}) \\ P_Y(B) &= P_{X,Y}(\mathbb{R} \times B) \end{aligned}$$

Pour distinguer la loi $P_{X,Y}$ des lois marginales, on l'appelle parfois la *loi jointe* de (X, Y) .

EXEMPLE 6.1. Soit un couple aléatoire (X, Y) qui prend les valeurs $(1, 3)$, $(1, 4)$ et $(2, 4)$ avec les probabilités respectives $1/4$, $1/8$ et $5/8$.



Sa loi est $P_{X,Y} = \frac{1}{4}\delta_{(1,3)} + \frac{1}{8}\delta_{(1,4)} + \frac{5}{8}\delta_{(2,4)}$. Ses lois marginales sont $P_X = \frac{3}{8}\delta_1 + \frac{5}{8}\delta_2$ et $P_Y = \frac{1}{4}\delta_3 + \frac{3}{4}\delta_4$.

6.2. Fonction de répartition

Nous introduisons une notion de fonction de répartition d'un couple de variables aléatoires analogue à celle des variables réelles.

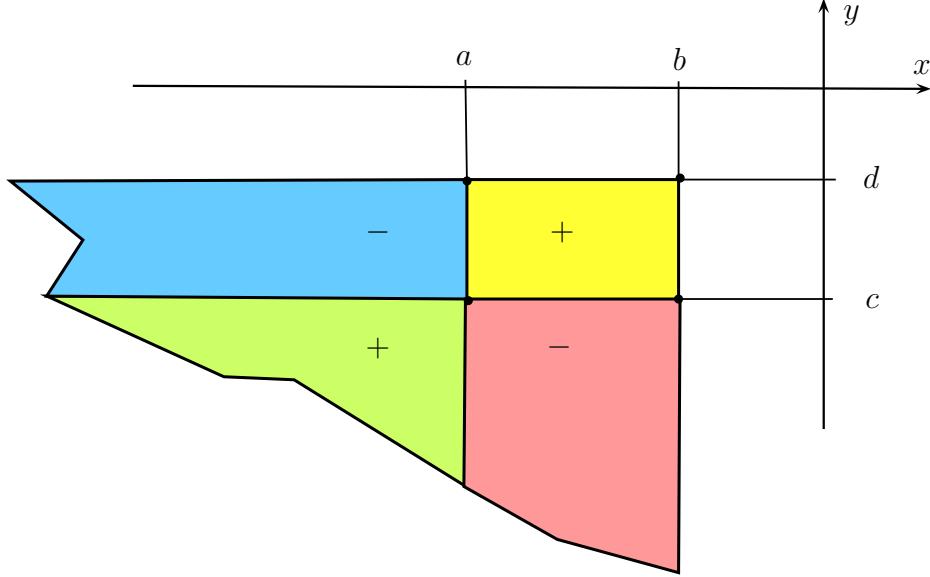
DÉFINITIONS 6.2. Une application $(X, Y) : \Omega \rightarrow \mathbb{R}^2$ est un *couple aléatoire* si pour tout $x, y \in \mathbb{R}$, l'ensemble $\{\omega \in \Omega; X(\omega) \leq x \text{ et } Y(\omega) \leq y\}$ appartient à \mathcal{A} .

La *fonction de répartition jointe* de (X, Y) est la fonction $F_{X,Y} : \mathbb{R}^2 \rightarrow [0, 1]$ donnée par

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y).$$

On montre aisément que pour tous $a \leq b, c \leq d \in \mathbb{R}$

$$\begin{aligned} & \mathbb{P}(a < X \leq b, c < Y \leq d) \\ &= F_{X,Y}(b, d) - F_{X,Y}(a, d) - F_{X,Y}(b, c) + F_{X,Y}(a, c). \end{aligned}$$



En d'autres termes, nous pouvons évaluer la probabilité que le point aléatoire (X, Y) "tombe" dans la région rectangulaire $[a, b] \times [c, d]$ du plan \mathbb{R}^2 . En travaillant de façon analogue à la Proposition 2.9, on récupère les probabilités de tomber dans des régions rectangulaires quelconques, puis leurs réunions dénombrables, etc... De fil en aiguille, il est possible de montrer, grâce aux propriétés des mesures de probabilité, l'assertion suivante :

PROPOSITION 6.3. *$F_{X,Y}$ spécifie de manière unique $\mathbb{P}((X, Y) \in C)$ pour toutes les parties ouvertes C de \mathbb{R}^2 . En d'autres termes, $F_{X,Y}$ spécifie entièrement la loi jointe $P_{X,Y}$.*

Les *fonctions de répartition marginales* de X et de Y sont

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) = \lim_{n \rightarrow \infty} \mathbb{P}(X \leq x \text{ et } Y \leq n) \\ &= F_{X,Y}(x, \infty) := \lim_{y \rightarrow \infty} F_{X,Y}(x, y), \\ F_Y(y) &= \mathbb{P}(Y \leq y) = \lim_{n \rightarrow \infty} \mathbb{P}(X \leq n \text{ et } Y \leq y) \\ &= F_{X,Y}(\infty, y) = \lim_{x \rightarrow \infty} F_{X,Y}(x, y), \end{aligned}$$

On constate que, même sur l'Exemple 6.1 qui est très simple, la fonction de répartition $F_{X,Y}$ est pénible à expliciter. En effet, elle nécessite de découper le plan en 5 zones rectangulaires. Nous n'emploierons donc que très peu souvent les fonctions de répartition dans les calculs explicites.

6.3. Indépendance

Deux variables aléatoires discrètes X et Y sont dites indépendantes si pour tous $x, y \in \mathbb{R}$, $\mathbb{P}(X = x \text{ et } Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$. Nous revisiterons plus en détail cette notion importante au Chapitre 9.

Il est clair que cette définition de l'indépendance ne peut pas être conservée si l'une au moins des variables (par exemple X) est continue, puisque dans ce cas $\mathbb{P}(X = x) = 0$, pour tout $x \in \mathbb{R}$. Nous adopterons la définition générale suivante.

DÉFINITION 6.4. Les variables aléatoires X et Y sont dites *indépendantes* si

$$\mathbb{P}(X \leq x \text{ et } Y \leq y) = \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y), \quad \forall x, y \in \mathbb{R}.$$

On vérifie que pour des variables aléatoires discrètes, cette définition de l'indépendance est équivalente à celle rappelée plus haut.

Une formulation équivalente est : X et Y sont indépendantes si et seulement si

$$F_{X,Y}(x, y) = F_X(x)F_Y(y), \quad \forall x, y \in \mathbb{R}.$$

PROPOSITION 6.5. Soient X et Y deux variables aléatoires indépendantes. Alors pour toute réunion dénombrable d'intervalles A et B , nous avons

$$\mathbb{P}(X \in A \text{ et } Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

et pour toutes fonctions numériques continues par morceaux φ et ψ , les variables aléatoires $\varphi(X)$ et $\psi(Y)$ sont indépendantes.

Notons que lorsque X et Y sont des variables discrètes dont toutes les valeurs sont isolées, toutes les fonctions φ et ψ sont continues (en restriction à $X(\Omega)$ et $Y(\Omega)$).

IDÉE DE LA PREUVE. Nous n'avons pas les outils suffisants pour donner une preuve complète (donc une preuve) de ce résultat. Notons toutefois qu'il est possible de montrer, de façon similaire à la preuve de la Proposition 6.3, que X et Y sont indépendantes si et seulement si pour toutes réunions dénombrables de parties ouvertes A et B de \mathbb{R} , $\mathbb{P}(X \in A \text{ et } Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$.

Maintenant, nous pouvons écrire pour toute paire d'ouverts A, B :

$$\begin{aligned} \mathbb{P}(\varphi(X) \in A \text{ et } \psi(Y) \in B) &= \mathbb{P}(X \in \varphi^{-1}(A) \text{ et } Y \in \psi^{-1}(B)) \\ &= \mathbb{P}(X \in \varphi^{-1}(A))\mathbb{P}(Y \in \psi^{-1}(B)) \\ &= \mathbb{P}(\varphi(X) \in A)\mathbb{P}(\psi(Y) \in B) \end{aligned}$$

où l'avant-dernière égalité est une conséquence de l'indépendance de X et Y et du fait que φ et ψ sont continues par morceaux, les ensembles $\varphi^{-1}(A)$ et $\psi^{-1}(B)$ sont des réunions dénombrables d'ouverts. \square

Cette notion mathématique de l'indépendance est cohérente avec la notion intuitive que nous en avons. Pour étayer cette affirmation, donnons-en une illustration simple.

EXEMPLE 6.6. Nous avons deux urnes contenant des boules de couleur numérotées.

- La première urne contient 5 boules numérotées : 1,2,3,4 et 5. Les boules 1,2,3 sont jaunes et les boules 4,5 sont rouges.
- La deuxième urne contient 3 boules numérotées : a,b,c. Les boules a,b sont vertes et la boule c est bleue.

On note X et Y les numéros aléatoires des boules tirées au hasard dans la première et la seconde urne. On suppose que ces tirages sont uniformes sur $\{1, 2, 3, 4, 5\}$ et $\{a, b, c\}$. De même, on note U et V les couleurs aléatoires des boules tirées au hasard dans la première et la seconde urne : $U = \varphi(X)$ et $V = \psi(Y)$ avec $\varphi(1) = \varphi(2) = \varphi(3) = \text{jaune}$,

$\varphi(4) = \varphi(5) = \text{rouge}$, $\psi(a) = \psi(b) = \text{vert}$ et $\psi(c) = \text{bleu}$. On a donc $\mathbb{P}(X = \text{jaune}) = 3/5$, $\mathbb{P}(X = \text{rouge}) = 2/5$ ainsi que $\mathbb{P}(Y = \text{vert}) = 2/3$, $\mathbb{P}(Y = \text{bleu}) = 1/3$.

Si de plus ces tirages sont indépendants (au sens habituel du terme), on n'avantage aucun couple de boules au détriment d'autres : la loi de (X, Y) est uniforme sur $\{1, 2, 3, 4, 5\} \times \{a, b, c\}$. On constate qu'alors X et Y sont des variables aléatoires indépendantes au sens mathématique. En effet, pour tous $A \subset \{1, 2, 3, 4, 5\}$ et $B \subset \{a, b, c\}$,

$$\begin{aligned}\mathbb{P}((X, Y) \in A \times B) &= \frac{\#(A \times B)}{\#(\{1, 2, 3, 4, 5\} \times \{a, b, c\})} \\ &= \frac{\#(A) \times \#(B)}{\#(\{1, 2, 3, 4, 5\}) \times \#(\{a, b, c\})} \\ &= \frac{\#(A)}{5} \times \frac{\#(B)}{3} \\ &= \mathbb{P}(X \in A) \mathbb{P}(Y \in B)\end{aligned}$$

En particulier, en prenant $A = \varphi^{-1}(\text{jaune}) = \{1, 2, 3\}$ et $B = \psi^{-1}(\text{vert}) = \{a, b\}$ on obtient

$$\begin{aligned}\mathbb{P}(U = \text{jaune}, V = \text{vert}) &= \mathbb{P}((X, Y) \in \{1, 2, 3\} \times \{a, b\}) \\ &= \mathbb{P}(X \in \{1, 2, 3\}) \mathbb{P}(Y \in \{a, b\}) \\ &= \mathbb{P}(U = \text{jaune}) \mathbb{P}(V = \text{vert})\end{aligned}$$

et de même pour les autres couleurs. Ce qui prouve l'indépendance mathématique de U et V . Mais il est clair que si les tirages dans les deux urnes sont indépendants (au sens habituel) il en est de même pour les couleurs des boules tirées.

EXERCICE 6.7. Soient X et Y deux variables aléatoires indépendantes de fonctions de répartition F_X et F_Y . Déterminer les lois de $U = \max(X, Y)$ et $V = \min(X, Y)$.

SOLUTION. Du fait que pour tout $t \in \mathbb{R}$, $\max(x, y) \leq t \iff (x \leq t \text{ et } y \leq t)$,

$$\begin{aligned}F_U(t) &= \mathbb{P}(\max(X, Y) \leq t) \\ &= \mathbb{P}(\{X \leq t\} \cap \{Y \leq t\}) \\ &= \mathbb{P}(X \leq t) \mathbb{P}(Y \leq t) \\ &= F_X(t) F_Y(t)\end{aligned}$$

où l'on a fait usage de l'indépendance dans l'avant-dernière égalité.

De même, pour tout $t \in \mathbb{R}$, $\min(x, y) > t \iff (x > t) \text{ et } (y > t)$, donc

$$\begin{aligned}1 - F_V(t) &= \mathbb{P}(\min(X, Y) > t) \\ &= \mathbb{P}(\{X > t\} \cap \{Y > t\}) \\ &= \mathbb{P}(X > t) \mathbb{P}(Y > t) \\ &= [1 - F_X(t)][1 - F_Y(t)]\end{aligned}$$

d'où

$$F_V(t) = 1 - [1 - F_X(t)][1 - F_Y(t)], \quad t \in \mathbb{R}.$$

ce qui détermine la loi de V . \square

EXEMPLE 6.8. On se donne deux variables aléatoires X et Y indépendantes de lois exponentielles $\mathcal{E}(\lambda)$ et $\mathcal{E}(\mu)$. Calculons à l'aide de l'exercice précédent les lois de $U = \max(X, Y)$ et $V = \min(X, Y)$.

Nous avons pour tout $t \leq 0$, $F_X(t) = F_Y(t) = 0$ et pour tout $t \geq 0$, $F_X(t) = 1 - e^{-\lambda t}$, $F_Y(t) = 1 - e^{-\mu t}$. Par conséquent pour tout $t > 0$,

$$\begin{aligned} f_U(t) &= F'_U(t) = f_X(t)F_Y(t) + F_X(t)f_Y(t) \\ &= \lambda e^{-\lambda t}(1 - e^{-\mu t}) + \mu e^{-\mu t}(1 - e^{-\lambda t}) \end{aligned}$$

et

$$\begin{aligned} 1 - F_V(t) &= [1 - F_X(t)][1 - F_Y(t)] \\ &= e^{-\lambda t}e^{-\mu t} = e^{-(\lambda+\mu)t} \end{aligned}$$

Pour tout $t \leq 0$, $F_U(t) = F_V(t) = 0$.

On constate que $V = \min(X, Y)$ admet la loi exponentielle $\mathcal{E}(\lambda + \mu)$.

6.4. Couples discrets

Soit un couple de variables aléatoires (X, Y) prenant ses valeurs dans l'ensemble produit $\mathcal{X} \times \mathcal{Y}$ avec $\mathcal{X} = \{x_1, \dots, x_L\}$ et $\mathcal{Y} = \{y_1, \dots, y_K\}$. Pour tout indice $n = (l, k) \in N := \{1, \dots, L\} \times \{1, \dots, K\}$, on note $z_n = (x_l, y_k)$. Cet ensemble étant fini, le couple $Z = (X, Y)$ est une variable aléatoire discrète à valeurs dans $\mathcal{X} \times \mathcal{Y}$. Elle est donc de la forme $P_{X,Y} = P_Z = \sum_{n \in N} p_n \delta_{z_n} = \sum_{1 \leq l \leq L, 1 \leq k \leq K} p_{l,k} \delta_{(x_l, y_k)}$ avec $p_{l,k} = \mathbb{P}((X, Y) = (x_l, y_k)) = \mathbb{P}(X = x_l \text{ et } Y = y_k)$. Pour plus de clarté, on note $p_{l,k} = p_{X,Y}(x_l, y_k)$ et on peut regrouper l'ensemble de ces probabilités élémentaires en un tableau matriciel :

	y_1	y_2	\cdots	y_K	$\leftarrow Y$
x_1	$p_{X,Y}(x_1, y_1)$	$p_{X,Y}(x_1, y_2)$	\cdots	$p_{X,Y}(x_1, y_K)$	$p_X(x_1)$
x_2	$p_{X,Y}(x_2, y_1)$	$p_{X,Y}(x_2, y_2)$	\cdots	$p_{X,Y}(x_2, y_K)$	$p_X(x_2)$
\vdots	\vdots	\vdots		\vdots	\vdots
x_L	$p_{X,Y}(x_L, y_1)$	$p_{X,Y}(x_L, y_2)$	\cdots	$p_{X,Y}(x_L, y_K)$	$p_X(x_L)$
$X \uparrow$	$p_Y(y_1)$	$p_Y(y_2)$	\cdots	$p_Y(y_K)$	1

dont l'intérieur décrit la loi jointe de (X, Y) . Les lois marginales sont données par $P_X = \sum_{1 \leq l \leq L} p_X(x_l) \delta_{x_l}$ et $P_Y = \sum_{1 \leq k \leq K} p_Y(y_k) \delta_{y_k}$ avec

$$\begin{aligned} p_X(x_l) &= \sum_{1 \leq k \leq K} p_{X,Y}(x_l, y_k), \quad 1 \leq l \leq L \\ p_Y(y_k) &= \sum_{1 \leq l \leq L} p_{X,Y}(x_l, y_k), \quad 1 \leq k \leq K \end{aligned}$$

puisque $p_X(x_l) = \mathbb{P}(X = x_l) = \mathbb{P}(X = x_l \text{ et } Y \in \mathcal{Y}) = \mathbb{P}((X, Y) \in \{x_l\} \times \mathcal{Y}) = \sum_{1 \leq k \leq K} \mathbb{P}(X = x_l \text{ et } Y = y_k)$ et de même pour $p_Y(y_k)$.

Par conséquent la dernière ligne du tableau est constituée des sommes par colonnes et la dernière colonne des sommes par lignes : les *marges* du tableau spécifient les lois *marginales* P_X et P_Y .

De façon plus générale, soient X et Y deux variables aléatoires à valeurs dans des ensembles dénombrables \mathcal{X} et \mathcal{Y} . Alors le couple (X, Y) est à valeurs dans l'ensemble

dénombrable $\mathcal{X} \times \mathcal{Y}$ (voir la Proposition A.4) et sa loi jointe est de la forme

$$P_{X,Y} = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p_{X,Y}(x, y) \delta_{(x,y)}.$$

et on montre comme précédemment la

PROPOSITION 6.9. *Les lois marginales sont $P_X = \sum_{x \in \mathcal{X}} p_X(x) \delta_x$ et $P_Y = \sum_{y \in \mathcal{Y}} p_Y(y) \delta_y$ avec*

$$\begin{aligned} p_X(x) &= \sum_{y \in \mathcal{Y}} p_{X,Y}(x, y), \quad x \in \mathcal{X} \\ p_Y(y) &= \sum_{x \in \mathcal{X}} p_{X,Y}(x, y), \quad y \in \mathcal{Y}. \end{aligned}$$

EXEMPLE 6.10. Considérons les deux lois jointes spécifiées par les tableaux suivants :

	1	3	$\leftarrow Y$
-1	0,1	0,2	0,3
2	0,45	0,25	0,7
$X \uparrow$	0,55	0,45	1

	1	3	$\leftarrow Y$
-1	0,2	0,1	0,3
2	0,35	0,35	0,7
$X \uparrow$	0,55	0,45	1

On constate que ces deux lois jointes sont distinctes bien qu'elles possèdent les mêmes lois marginales. Par conséquent la loi jointe $P_{X,Y}$ n'est pas spécifiée par la donnée des deux lois marginales P_X et P_Y . Il y a plus d'information dans l'intérieur du tableau que sur les marges.

PROPOSITION 6.11. *Soit (X, Y) de loi $P_{X,Y} = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p_{X,Y}(x, y) \delta_{(x,y)}$. Les variables X et Y sont indépendantes si et seulement s'il existe deux fonctions $q : \mathcal{X} \rightarrow [0, 1]$ et $r : \mathcal{Y} \rightarrow [0, 1]$ telles que pour tous $x \in \mathcal{X}$ et $y \in \mathcal{Y}$ nous avons $p_{X,Y}(x, y) = q(x)r(y)$. Dans ce cas, nous avons aussi*

$$p_{X,Y}(x, y) = p_X(x)p_Y(y), \quad x \in \mathcal{X}, y \in \mathcal{Y}.$$

DÉMONSTRATION. C'est une conséquence directe de la Proposition 6.5 en prenant $A = \{x\}$ et $B = \{y\}$ avec $x \in \mathcal{X}$ et $y \in \mathcal{Y}$.

Notons aussi que lorsque $p_{X,Y}(x, y) = q(x)r(y)$, $p_X(x) = aq(x)$ pour tout x avec $a = \sum_{y \in \mathcal{Y}} r(y)$. De même pour tout y , $p_Y(y) = br(y)$ avec $1 = \sum_{y \in \mathcal{Y}} p_Y(y) = b \sum_{y \in \mathcal{Y}} r(y) = ab$. Finalement, $r(x)q(y) = p_X(x)p_Y(y)/(ab) = p_X(x)p_Y(y)$. \square

EXEMPLE 6.12. Considérons la loi jointe spécifiée par le tableau

	1	3	$\leftarrow Y$
-1	0,165	0,135	0,3
2	0,385	0,315	0,7
$X \uparrow$	0,55	0,45	1

On constate qu'il possède la structure produit $p_{X,Y}(x, y) = p_X(x)p_Y(y)$, $\forall x, y$. Les variables X et Y sont donc indépendantes. On note que les lois marginales P_X et P_Y sont les mêmes que celles de l'Exemple 6.10.

Puisque le couple discret (X, Y) est une variable discrète à valeurs dans l'ensemble dénombrable $\mathcal{X} \times \mathcal{Y}$ (voir la Proposition A.4) l'espérance de $\varphi(X, Y)$ est donnée par le Théorème 3.10 qui dans ce cas précis s'écrit

$$(6.13) \quad \mathbb{E}\varphi(X, Y) = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \varphi(x, y) p_{X,Y}(x, y)$$

et qui est correctement définie dès lors que

$$\mathbb{E}|\varphi(X, Y)| = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} |\varphi(x, y)| p_{X,Y}(x, y) < \infty.$$

On obtient immédiatement la

PROPOSITION 6.14 (Linéarité et croissance).

(1) *En particulier, avec $\varphi(x, y) = ax + by$, nous obtenons la linéarité de l'espérance*

$$\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y, \quad a, b \in \mathbb{R}$$

pour toute variables aléatoires X et Y telles que $\mathbb{E}|X| < \infty$ et $\mathbb{E}|Y| < \infty$.

Plus généralement pour toutes fonctions φ et ψ telles que $\mathbb{E}|\varphi(X, Y)| < \infty$ et $\mathbb{E}|\psi(X, Y)| < \infty$ et tous réels a, b , nous avons

$$\mathbb{E}[a\varphi(X, Y) + b\psi(X, Y)] = a\mathbb{E}\varphi(X, Y) + b\mathbb{E}\psi(X, Y).$$

(2) *Si les fonctions $\varphi, \psi : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ sont telles que $\varphi \leq \psi$, alors $\mathbb{E}\varphi(X, Y) \leq \mathbb{E}\psi(X, Y)$.*

DÉFINITION 6.15. Nous définissons la *covariance* de (X, Y) par

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

c'est-à-dire

$$\text{Cov}(X, Y) = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} (x - \mathbb{E}X)(y - \mathbb{E}Y) p_{X,Y}(x, y).$$

On dit que X et Y sont *décorellées* si $\text{Cov}(X, Y) = 0$.

Noter que, tout comme l'espérance, la covariance n'est pas toujours définie. Il faut pour cela que $\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} |(x - \mathbb{E}X)(y - \mathbb{E}Y)| p_{X,Y}(x, y) < \infty$. On montrera au Corollaire 6.37 qu'une condition suffisante est que $\mathbb{E}(X^2) < \infty$ et $\mathbb{E}(Y^2) < \infty$.

Un simple calcul nous mène à

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

PROPOSITION 6.16. *Soient X et Y deux variables aléatoires discrètes indépendantes.*

(1) *Pour toutes fonctions φ sur \mathcal{X} et ψ sur \mathcal{Y} telles que $\mathbb{E}|\varphi(X)| < \infty$ et $\mathbb{E}|\psi(Y)| < \infty$, nous avons*

$$\mathbb{E}[\varphi(X)\psi(Y)] = \mathbb{E}[\varphi(X)]\mathbb{E}[\psi(Y)].$$

(2) *Si $\mathbb{E}|X| < \infty$ et $\mathbb{E}|Y| < \infty$ alors $\text{Cov}(X, Y) = 0$.*

DÉMONSTRATION. • Preuve de (1). Avec la Proposition 6.11 nous avons

$$\begin{aligned} \mathbb{E}[\varphi(X)\psi(Y)] &= \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \varphi(x)\psi(y) p_X(x)p_Y(y) \\ &= \sum_{x \in \mathcal{X}} \varphi(x) p_X(x) \sum_{y \in \mathcal{Y}} \psi(y) p_Y(y) \\ &= \mathbb{E}[\varphi(X)]\mathbb{E}[\psi(Y)] \end{aligned}$$

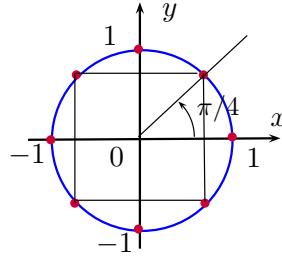
qui est le résultat annoncé.

- Preuve de (2). Grâce à (1), nous avons $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ c'est-à-dire $\text{Cov}(X, Y) = 0$. \square

L'exercice suivant montre que la réciproque de l'assertion (2) de cette proposition est fausse.

EXERCICE 6.17.

- On considère le couple aléatoire (X, Y) dont la loi est uniforme sur les quatre points du plan $(1, 0), (0, 1), (-1, 0)$ et $(0, -1)$. Montrer que $\text{Cov}(X, Y) = 0$ mais que X et Y ne sont pas indépendantes.
- On considère le couple aléatoire (X, Y) dont la loi est uniforme sur les huit points du plan d'affixes $e^{ik\pi/4}$, $0 \leq k \leq 7$.



Montrer que $\text{Cov}(X, Y) = 0$ mais que X et Y ne sont pas indépendantes.

SOLUTION. Nous ne donnons que la solution de (a). Nous avons $P_X = P_Y = \frac{1}{4}\delta_{-1} + \frac{1}{2}\delta_0 + \frac{1}{4}\delta_1$ de sorte que $\mathbb{E}X = \mathbb{E}Y = 0$. De plus $XY = 0$, donc $\mathbb{E}XY = 0$ et $\text{Cov}(X, Y) = 0$. D'autre part X et Y ne sont pas des variables indépendantes puisque $\mathbb{P}(X = 1)\mathbb{P}(Y = 0) = \frac{1}{4} \times \frac{1}{2} = 1/8 \neq 1/4 = \mathbb{P}((X, Y) = (1, 0))$. \square

6.5. Couples continus

Par analogie avec les variables aléatoires continues, nous introduisons la notion suivante.

DÉFINITION 6.18. Un *couple aléatoire* (X, Y) de fonction de répartition jointe $F_{X,Y}$ est dit *continu*, s'il existe une fonction intégrable $f_{X,Y} : \mathbb{R}^2 \rightarrow [0, \infty[$ telle que

$$F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(s, t) ds dt, \quad \forall x, y \in \mathbb{R}.$$

Dans ce cas, la fonction $f_{X,Y}$ est appelée *fonction de densité jointe* du couple aléatoire (X, Y) .

On déduit de cette définition que si $F_{X,Y}$ est continûment dérivable alors

$$(6.19) \quad f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y).$$

PROPOSITION 6.20. *Les lois marginales P_X et P_Y admettent les densités*

$$\begin{aligned} f_X(x) &= \int_{\mathbb{R}} f_{X,Y}(x, y) dy, \quad x \in \mathbb{R} \\ f_Y(y) &= \int_{\mathbb{R}} f_{X,Y}(x, y) dx, \quad y \in \mathbb{R} \end{aligned}$$

DÉMONSTRATION. Nous avons vu que les fonctions de répartition marginales de X et de Y sont $F_X(x) = F_{X,Y}(x, \infty)$ et $F_Y(y) = F_{X,Y}(\infty, y)$. En d'autres termes, $F_X(x) = \int_{-\infty}^x (\int_{\mathbb{R}} f_{X,Y}(s, y) dy) ds$ d'où il vient que $f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y) dy$. De la même manière, nous obtenons que la fonction de densité marginale de Y est $f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y) dx$. \square

DÉFINITION 6.21. Par analogie avec (6.13) et la définition (3.18) qui est justifiée par le Théorème C.10, nous définissons (sans plus de justification cette fois-ci) l'espérance de la variable aléatoire $\varphi(X, Y)$ par

$$\mathbb{E}\varphi(X, Y) := \iint_{\mathbb{R}^2} \varphi(x, y) f_{X,Y}(x, y) dx dy$$

pour toute fonction $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ telle que $|\varphi| f_{X,Y}$ soit intégrable et $\iint_{\mathbb{R}^2} |\varphi(x, y)| f_{X,Y}(x, y) dx dy < \infty$.

On déduit immédiatement de cette définition la

PROPOSITION 6.22 (Linéarité et croissance).

(1) *En particulier, avec $\varphi(x, y) = ax + by$, nous obtenons la linéarité de l'espérance*

$$\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y, \quad a, b \in \mathbb{R}$$

pour toute variables aléatoires X et Y telles que $\mathbb{E}|X| < \infty$ et $\mathbb{E}|Y| < \infty$.

Plus généralement pour toutes fonctions φ et ψ telles que $\mathbb{E}|\varphi(X, Y)| < \infty$ et $\mathbb{E}|\psi(X, Y)| < \infty$, nous avons

$$\mathbb{E}[\varphi(X, Y) + \psi(X, Y)] = \mathbb{E}\varphi(X, Y) + \mathbb{E}\psi(X, Y).$$

(2) *Si les fonctions $\varphi, \psi : \mathbb{R}^2 \rightarrow \mathbb{R}$ sont telles que $\varphi \leq \psi$, alors $\mathbb{E}\varphi(X, Y) \leq \mathbb{E}\psi(X, Y)$.*

Comme pour les couples discrets nous définissons la covariance de (X, Y) par

$$\begin{aligned} \text{Cov}(X, Y) &:= \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] \\ &= \iint_{\mathbb{R}^2} (x - \mathbb{E}X)(y - \mathbb{E}Y) f_{X,Y}(x, y) dx dy. \end{aligned}$$

Noter que, tout comme l'espérance, la covariance n'est pas toujours définie. Nous verrons au Corollaire 6.37 qu'il suffit pour cela $\mathbb{E}(X^2), \mathbb{E}(Y^2) < \infty$.

Comme le montre la proposition suivante, la fonction de densité jointe d'un couple aléatoire continu de variables indépendantes a une forme produit.

PROPOSITION 6.23.

(1) *Soit (X, Y) un couple aléatoire continu de fonction de densité jointe $f_{X,Y}$. S'il existe des fonctions g et h telles que*

$$f_{X,Y}(x, y) = g(x)h(y), \quad x, y \in \mathbb{R},$$

alors X et Y sont des variables aléatoires indépendantes. De plus, la fonction de densité jointe s'écrit alors : $f_{X,Y}(x, y) = f_X(x)f_Y(y)$.

(2) *Soient X et Y des variables aléatoires indépendantes qui admettent des fonctions de densité f_X et f_Y continues par morceaux. Alors la fonction de densité jointe de (X, Y) est*

$$f_{X,Y}(x, y) = f_X(x)f_Y(y), \quad x, y \in \mathbb{R}.$$

DÉMONSTRATION. • Preuve de (1). La première partie de la proposition est presque immédiate. La forme $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ s'obtient par un raisonnement analogue à celui de la preuve de la Proposition 6.11.

• Preuve de (2). Du fait des hypothèses, F_X et F_Y sont des fonctions dérivables partout sauf en un nombre fini de points. De ce fait, la fonction de répartition jointe $F_{X,Y}(x, y) = F_X(x)F_Y(y)$ est partout dérivable, sauf sur la réunion d'un nombre fini de droites (dont l'aire est nulle et que l'on peut exclure des intégrales doubles). En dehors de cet ensemble, on peut appliquer (6.19) qui nous donne $f_{X,Y}(x, y) = F'_X(x)F'_Y(y) = f_X(x)f_Y(y)$. Ce qui achève la preuve. \square

Le résultat suivant est une conséquence immédiate de la proposition précédente.

COROLLAIRE 6.24. *Soit (X, Y) un couple aléatoire continu de variables indépendantes.*

- (1) *Si $\mathbb{E}|X|, \mathbb{E}|Y| < \infty$, alors $\text{Cov}(X, Y) = 0$.*
- (2) *Si $\mathbb{E}|\varphi(X)|, \mathbb{E}|\psi(Y)| < \infty$, alors $\mathbb{E}[\varphi(X)\psi(Y)] = \mathbb{E}\varphi(X)\mathbb{E}\psi(Y)$.*

DÉMONSTRATION. Immédiate. \square

Attention : Il existe des couples aléatoires continus (X, Y) de covariance nulle dont les composantes X et Y ne sont pas indépendantes.

EXERCICE 6.25. Montrer, sans calculs explicites, que c'est le cas pour le tirage aléatoire uniforme d'un point (X, Y) du disque unité.

Au fait, quelle peut bien être la fonction de densité jointe de ce couple aléatoire ?

EXEMPLE 6.26 (L'aiguille de Buffon). Les lignes d'équations $y = n$ ($n \in \mathbb{Z}$), sont tracées sur un plan et une aiguille de longueur unité est jetée sur ce plan. Quelle est la probabilité qu'elle intersecte l'une des lignes ? On suppose que l'aiguille n'a pas de préférence de position ni de direction.

Cherchons la solution de ce problème. Soient (X, Y) les coordonnées du centre de l'aiguille et Θ l'angle, modulo π , de l'aiguille avec l'axe des x . On note $Z = Y - \lfloor Y \rfloor$ ($\lfloor Y \rfloor$ est la partie entière de Y) la distance du centre de l'aiguille à la ligne immédiatement en-dessous de lui. Nos hypothèses se traduisent par

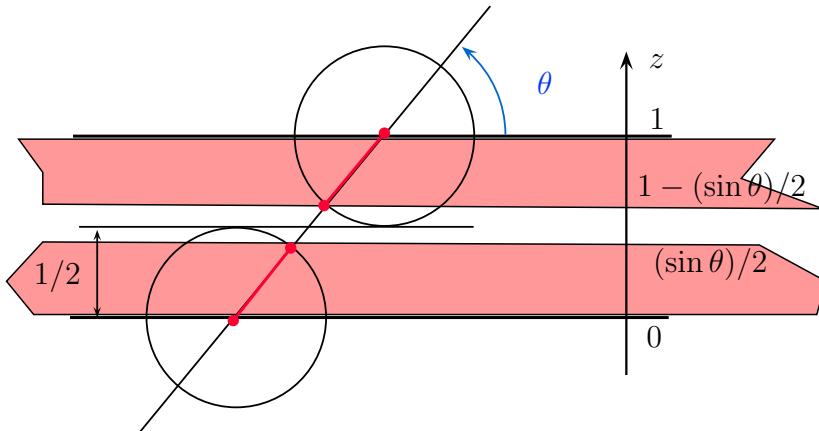
- (a) Z est distribué uniformément sur $[0, 1]$: $f_Z = \mathbf{1}_{[0,1]}$.
- (b) Θ est distribué uniformément sur $[0, \pi]$: $f_\Theta = \frac{1}{\pi}\mathbf{1}_{[0,\pi]}$.
- (c) Z et Θ sont indépendantes : $f_{Z,\Theta}(z, \theta) = f_Z(z)f_\Theta(\theta)$.

Par conséquent, (Z, Θ) a pour fonction de densité jointe

$$f(z, \theta) = \frac{1}{\pi}\mathbf{1}_{(0 \leq z \leq 1, 0 \leq \theta \leq \pi)}.$$

A l'aide d'un dessin, on constate qu'il y a intersection si et seulement si $Z \in I$ avec

$$I = \left\{ (z, \theta) \in [0, 1] \times [0, \pi]; z \leq \frac{1}{2} \sin \theta \text{ ou } 1 - z \leq \frac{1}{2} \sin \theta \right\}.$$



Le lieu des centres possibles de l'aiguille impliquant une intersection est en rouge.

Par conséquent,

$$\begin{aligned}
 \mathbb{P}(\text{intersection}) &= \iint_I f(z, \theta) dz d\theta \\
 &= \frac{1}{\pi} \int_0^\pi \left(\int_0^{\frac{1}{2} \sin \theta} dz + \int_{1 - \frac{1}{2} \sin \theta}^1 dz \right) d\theta \\
 &= 2/\pi.
 \end{aligned}$$

Buffon a effectivement mis en place cette expérience pour obtenir une valeur approchée de π .

EXEMPLE 6.27 (Loi normale bivariée). Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ la fonction définie par

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right)$$

où $-1 < \rho < 1$. On vérifie que f est bien une fonction de densité jointe, c'est-à-dire : $f(x, y) \geq 0$ et $\iint_{\mathbb{R}^2} f(x, y) dx dy = 1$.

EXERCICE 6.28.

- Vérifier que $\iint_{\mathbb{R}^2} f(x, y) dx dy = 1$.
- Montrer que les lois marginales de X et de Y sont des lois normales centrées réduites.
- Montrer que $\text{Cov}(X, Y) = \iint_{\mathbb{R}^2} xy f(x, y) dx dy = \rho$.

La fonction de densité jointe d'une *loi normale bivariée* générale est plus compliquée. On dit que (X, Y) suit une loi normale bivariée de moyennes μ_1 et μ_2 , de variances σ_1^2 et σ_2^2 et de corrélation ρ avec $-1 < \rho < 1$, si sa fonction de densité jointe est donnée par

$$(6.29) \quad f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2}Q(x, y)\right]$$

où $\sigma_1, \sigma_2 > 0$ et Q est la forme quadratique :

$$Q(x, y) = \frac{1}{1-\rho^2} \left[\left(\frac{x - \mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x - \mu_1}{\sigma_1} \right) \left(\frac{y - \mu_2}{\sigma_2} \right) + \left(\frac{y - \mu_2}{\sigma_2} \right)^2 \right].$$

EXERCICE 6.30. Montrer que

- (a) $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ et $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$,
- (b) $\text{Cov}(X, Y) = \rho\sigma_1\sigma_2$.

A l'aide de la Définition 6.36 plus bas du coefficient de corrélation $\text{Cor}(X, Y)$, l'énoncé de (b) est $\text{Cor}(X, Y) = \rho$.

PROPOSITION 6.31. *Soit (X, Y) un couple aléatoire normal. Si $\text{Cov}(X, Y) = 0$ alors X et Y sont des variables aléatoires indépendantes.*

Ce résultat est remarquable car en général la décorrélation (covariance nulle) n'implique pas l'indépendance, voir l'Exercice 6.17. C'est une propriété spécifique des couples aléatoires normaux.

DÉMONSTRATION. Compte tenu de l'exercice précédent, nous avons $\rho = 0$. En injectant $\rho = 0$ dans la formule (6.29), on obtient $f(x, y) = f_X(x)f_Y(y)$ (avec $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ et $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$) et on conclut avec la Proposition 6.23. \square

EXERCICE 6.32. Soit (X, Y) un couple aléatoire de fonction de densité jointe

$$f(x, y) = \mathbf{1}_{\{x,y>0\}} \frac{1}{y} \exp\left(-y - \frac{x}{y}\right), \quad x, y \in \mathbb{R}.$$

Trouver la loi marginale de Y .

SOLUTION. Pour tout $y \leq 0$, $f_Y(y) = \int_{\mathbb{R}} f(x, y) dx = 0$ et pour tout $y > 0$,

$$f_Y(y) = \int_{\mathbb{R}} f(x, y) dx = \int_0^\infty \frac{1}{y} \exp\left(-y - \frac{x}{y}\right) dx = e^{-y}$$

puisque l'on reconnaît que $x \mapsto \mathbf{1}_{\{x>0\}} \frac{1}{y} \exp\left(-\frac{x}{y}\right)$ est la fonction de densité d'une loi (exponentielle). Par conséquent $Y \sim \mathcal{E}(1)$. \square

6.6. Fonctions caractéristiques

On les définit de façon analogue aux transformées de Laplace et de Fourier des variables réelles, voir la Définition 5.6.

DÉFINITIONS 6.33.

- (1) La *transformée de Laplace* de la loi de (X, Y) est définie par

$$(s, t) \in \mathbb{R}^2 \mapsto L_{X,Y}(s, t) = \mathbb{E}e^{sX+tY} \in [0, \infty]$$

- (2) La *fonction caractéristique* de la loi de (X, Y) est définie par

$$(s, t) \in \mathbb{R}^2 \mapsto \phi_{X,Y}(s, t) = \mathbb{E}e^{i(sX+tY)} \in \mathbb{C}$$

où i est le nombre imaginaire tel que $i^2 = -1$.

On peut montrer, mais cette preuve est au delà du niveau de ce cours, que la fonction caractéristique *caractérise* la loi $P_{X,Y}$. C'est-à-dire que si nous connaissons $\phi_{X,Y}$, on peut calculer $P_{X,Y}$ et qu'il n'y a qu'une seule loi $P_{X,Y}$ qui admet $\phi_{X,Y}$ comme fonction caractéristique. Un résultat analogue est valide pour la transformée de Laplace sous l'hypothèse que $L_{X,Y}$ est finie sur un voisinage ouvert de $(0, 0)$.

PROPOSITION 6.34. *Soient (X, Y) un couple discret ou continu.*

(1) *Les variables X et Y sont indépendantes si et seulement si la fonction caractéristique de (X, Y) satisfait*

$$\phi_{XY}(s, t) = \phi_X(s)\phi_Y(t), \quad s, t \in \mathbb{R}.$$

(2) *Si les transformées de Laplace L_X et L_Y sont finies au voisinage de zéro, alors X et Y sont indépendantes si et seulement si*

$$L_{XY}(s, t) = L_X(s)L_Y(t), \quad s, t \in \mathbb{R}.$$

DÉMONSTRATION. • Preuve de (1). Soient X et Y indépendantes. À l'aide de la Proposition 6.16 et du Corollaire 6.24, on obtient $\phi_{XY}(s, t) = \mathbb{E}e^{i(sX+tY)} = \mathbb{E}[e^{isX}e^{itY}] = \mathbb{E}e^{isX}\mathbb{E}e^{itY} = \phi_X(s)\phi_Y(t)$.

Montrons la réciproque. On se donne (X, Y) tel que $\phi_{XY}(s, t) = \phi_X(s)\phi_Y(t)$ pour tous s, t . Soit (U, V) un couple de variables indépendantes telles que $U \stackrel{d}{=} X$ et $V \stackrel{d}{=} Y$. Ceci implique bien sûr que $\phi_U = \phi_X$ et $\phi_V = \phi_Y$. D'après ce que nous venons de montrer, nous avons $\phi_{U,V}(s, t) = \phi_U(s)\phi_V(t) = \phi_X(s)\phi_Y(t)$. Donc, $\phi_{U,V} = \phi_{X,Y}$. Mais puisque les fonctions caractéristiques caractérisent les lois (résultat admis), ceci implique $(X, Y) \stackrel{d}{=} (U, V)$. D'où le résultat annoncé.

• Preuve de (2). Analogue à celle de (1). □

6.7. Inégalité de Cauchy-Schwarz

Cette inégalité permet de contrôler en espérance les fluctuations jointes de (X, Y) à l'aide des variances individuelles de X et Y , voir le Corollaire 6.37 plus bas.

THÉORÈME 6.35 (Inégalité de Cauchy-Schwarz). *Pour tout couple aléatoire discret ou continu (X, Y) nous avons*

$$[\mathbb{E}(XY)]^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$$

avec égalité si et seulement s'il existe $a, b \in \mathbb{R}$ dont l'un au moins est non nul tels que $\mathbb{P}(aX = bY) = 1$.

Il est entendu que dans l'énoncé de ce théorème que $\mathbb{E}|XY| < \infty$ de sorte que les intégrales qui interviennent sont bien définies, éventuellement à valeurs infinie.

DÉMONSTRATION. On peut supposer sans perte de généralité que $\mathbb{E}(X^2), \mathbb{E}(Y^2) < \infty$.

Pour tous $a, b \in \mathbb{R}$, l'espérance de la variable positive $(aX - bY)^2$ est positive. Donc

$$\mathbb{E}((aX - bY)^2) = a^2\mathbb{E}(X^2) - 2ab\mathbb{E}(XY) + b^2\mathbb{E}(Y^2) \geq 0$$

Si $\mathbb{P}(X = 0) = 1$, l'assertion est évidente.

Si $\mathbb{P}(X = 0) < 1$, alors $\mathbb{E}(X^2) > 0$ et l'inégalité ci-dessus peut être vue comme une inéquation du second degré en a , à b fixé. Ceci implique que le discriminant réduit : $b^2([\mathbb{E}(XY)]^2 - \mathbb{E}(X^2)\mathbb{E}(Y^2))$ est strictement négatif (si $a^2\mathbb{E}(X^2) - 2ab\mathbb{E}(XY) + b^2\mathbb{E}(Y^2) > 0$ pour tout a) ou nul (s'il existe un a tel que $a^2\mathbb{E}(X^2) - 2ab\mathbb{E}(XY) + b^2\mathbb{E}(Y^2) = \mathbb{E}((aX - bY)^2) = 0$).

En choisissant $b \neq 0$, on obtient $[\mathbb{E}(XY)]^2 < \mathbb{E}(X^2)\mathbb{E}(Y^2)$ dans le premier cas et $[\mathbb{E}(XY)]^2 = \mathbb{E}(X^2)\mathbb{E}(Y^2)$ lorsque $\mathbb{E}((aX - bY)^2) = 0$, c'est-à-dire lorsque $\mathbb{P}(aX - bY = 0) = 1$. \square

DÉFINITION 6.36. Le *coefficient de corrélation* de (X, Y) est défini par

$$\text{Cor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

Pour que cette définition soit valide, il est nécessaire que $\mathbb{E}(X^2) < \infty$ et $\mathbb{E}(Y^2) < \infty$ et que $\text{Var}X, \text{Var}Y > 0$.

Une conséquence simple de l'inégalité de Cauchy-Schwarz est le

COROLLAIRE 6.37.

(1) *Pour que $\text{Cov}(X, Y)$ soit défini, il suffit que $\mathbb{E}(X^2), \mathbb{E}(Y^2) < \infty$.*

(2) *Soit (X, Y) tel que $0 < \text{Var}(X), \text{Var}(Y) < \infty$. Alors*

$$-1 \leq \text{Cor}(X, Y) \leq 1.$$

DÉMONSTRATION. • Preuve de (1). C'est une conséquence immédiate du Théorème 6.35 et du Corollaire 3.35.

• Preuve de (2). On applique le Théorème 6.35 avec $X - \mathbb{E}X$ et $Y - \mathbb{E}Y$ à la place de X et Y . \square

CHAPITRE 7

Fonctions d'un couple aléatoire

7.1. Quelques exercices corrigés

EXERCICE 7.1. Soient X et Y deux variables aléatoires indépendantes de lois normales $\mathcal{N}(0, 1)$. Calculer la fonction de densité de $W = X^2 + Y^2$.

SOLUTION. Pour tout $w \geq 0$,

$$\begin{aligned} \mathbb{P}(W \leq w) &= \iint_{\{x^2+y^2 \leq w\}} \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2+y^2)\right) dx dy \\ &\stackrel{(a)}{=} \int_0^{\sqrt{w}} \int_0^{2\pi} \frac{1}{2\pi} \exp(-r^2/2) r dr d\theta \\ &\stackrel{(b)}{=} \int_0^{w/2} e^{-u} du \end{aligned}$$

avec le changement de variable en coordonnées polaires en (a) et en posant $u = r^2/2$ en (b). On constate que W admet la fonction de densité $f(u) = \mathbf{1}_{(u \geq 0)} \frac{e^{-u/2}}{2}$. C'est-à-dire que W suit une loi exponentielle de paramètre $1/2$. \square

Attention. Ce n'est pas parce que X est une variable aléatoire continue qu'il en est de même pour $Y = \varphi(X)$. Par exemple, considérer $\varphi(x) = 3, \forall x \in \mathbb{R}$.

EXERCICE 7.2. On se donne un couple aléatoire (X_1, X_2) de fonction de densité jointe f_{X_1, X_2} et on considère le couple aléatoire (Y_1, Y_2) tel que

$$\begin{aligned} X_1 &= aY_1 + bY_2 \\ X_2 &= cY_1 + dY_2 \end{aligned}$$

avec $ad - bc \neq 0$. Cherchons la loi de (Y_1, Y_2) .

SOLUTION. Pour cela, évaluons pour tout ensemble $B \subset \mathbb{R}^2$ (suffisamment régulier) la probabilité $\mathbb{P}((Y_1, Y_2) \in B)$. Soit A l'image de B par $T(y_1, y_2) = (ay_1 + by_2, cy_1 + dy_2)$ qui est une bijection du fait de l'hypothèse $ad - bc \neq 0$.

$$\begin{aligned} \mathbb{P}((Y_1, Y_2) \in B) &= \mathbb{P}((X_1, X_2) \in A) \\ &= \iint_A f_{X_1, X_2}(x_1, x_2) dx_1 dx_2 \\ &= \iint_B f_{X_1, X_2}(ay_1 + by_2, cy_1 + dy_2) |ad - bc| dy_1 dy_2 \end{aligned}$$

où $|ad - bc|$ est la valeur absolue du jacobien de la transformation T . On en déduit que (Y_1, Y_2) est un couple aléatoire continu de fonction de densité jointe :

$$f_{Y_1, Y_2}(y_1, y_2) = |ad - bc| f_{X_1, X_2}(ay_1 + by_2, cy_1 + dy_2)$$

ce qui achève l'exercice. \square

En fait, le procédé est général pour toute transformation bijective T .

EXERCICE 7.3. Soient (X, Y) deux variables aléatoires indépendantes exponentielles de paramètre λ . Trouver la fonction de densité jointe de

$$U = X + Y, \quad V = X/Y$$

et montrer que ce sont des variables aléatoires indépendantes.

SOLUTION. On considère la transformation S donnée par

$$S(x, y) = (x + y, x/y), \quad x, y > 0.$$

Elle est bijective et son inverse S^{-1} donnée par

$$(x, y) = S^{-1}(u, v) = \left(\frac{uv}{1+v}, \frac{u}{1+v} \right), \quad u, v > 0$$

a pour jacobien

$$J(u, v) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{vmatrix} = -\frac{u}{(1+v)^2}.$$

Par conséquent, avec la formule de changement de variables

$$dxdy = |J(u, v)|dudv,$$

nous obtenons pour tout $B \subset \mathbb{R}^2$ (suffisamment régulier),

$$\begin{aligned} \mathbb{P}((U, V) \in B) &= \mathbb{P}(S^{-1}(U, V) \in S^{-1}(B)) \\ &= \mathbb{P}((X, Y) \in S^{-1}(B)) \\ &= \iint_{S^{-1}(B)} \mathbf{1}_{(x>0, y>0)} \lambda^2 \exp(-\lambda(x+y)) dxdy \\ &= \iint_B \mathbf{1}_{(u>0, v>0)} \lambda^2 \exp(-\lambda u) \frac{u}{(1+v)^2} dudv \end{aligned}$$

Par conséquent, (U, V) admet la densité

$$\begin{aligned} f_{U,V}(u, v) &= \mathbf{1}_{(u>0, v>0)} \lambda^2 \exp(-\lambda u) \frac{u}{(1+v)^2} \\ &= [\lambda^2 \mathbf{1}_{(u>0)} u \exp(-\lambda u)] \left[\mathbf{1}_{(v>0)} \frac{1}{(1+v)^2} \right] \end{aligned}$$

où la forme produit de la densité nous indique l'indépendance de U et V . \square

7.2. Somme de deux variables aléatoires indépendantes

Soient X et Y deux variables aléatoires continues indépendantes de fonctions de densité f_X et f_Y . Déterminons la loi de $S = X + Y$. Pour cela nous effectuons le changement

de variables $\begin{cases} s = x + y \\ t = x \end{cases} \iff \begin{cases} x = t \\ y = s - t \end{cases}$ qui nous donne $dsdt = dxdy$ et

$$\begin{aligned} F_S(u) &= \mathbb{P}(X + Y \leq u) \\ &= \iint_{\mathbb{R}^2} \mathbf{1}_{(x+y \leq u)} f_X(x) f_Y(y) dxdy \\ &= \iint_{\mathbb{R}^2} \mathbf{1}_{(s \leq u)} f_X(t) f_Y(s - t) \\ &= \int_{-\infty}^u \left[\int_{\mathbb{R}} f_X(t) f_Y(s - t) dt \right] ds \end{aligned}$$

cette dernière égalité est dûe au théorème de Fubini. Par conséquent, S est une variable aléatoire continue de fonction de densité

$$f_{X+Y}(s) = \int_{\mathbb{R}} f_X(x) f_Y(s - x) dx.$$

DÉFINITION 7.4. Soient f et g deux fonctions numériques,

$$f * g(s) = \int_{\mathbb{R}} f(x) g(s - x) dx, \quad s \in \mathbb{R}$$

est la *convoluée* de f et g (si cette intégrale est bien définie). L'opération $*$ est le *produit de convolution*.

On constate facilement que $f * g = g * f$. On vient de montrer le résultat suivant.

PROPOSITION 7.5. Soient X et Y deux variables aléatoires continues indépendantes de fonctions de densité f_X et f_Y . Alors la somme $X + Y$ est une variable aléatoire continue de fonction de densité

$$f_{X+Y} = f_X * f_Y$$

EXERCICE 7.6. Soient X et Y deux variables aléatoires indépendantes de lois respectives $\mathcal{N}(0, \sigma^2)$ et $\mathcal{N}(0, \tau^2)$. Montrer que $X + Y$ suit une loi $\mathcal{N}(0, \sigma^2 + \tau^2)$.

SOLUTION. Pour tout $s \in \mathbb{R}$,

$$\begin{aligned} f_{X+Y}(s) &= f_X * f_Y(s) \\ &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/(2\sigma^2)} \frac{1}{\sqrt{2\pi\tau^2}} e^{-(s-x)^2/(2\tau^2)} dx \\ &= \int_{\mathbb{R}} \frac{1}{2\pi\sigma\tau} \exp\left(-\frac{1}{2}[x^2/\sigma^2 + (s-x)^2/\tau^2]\right) dx \end{aligned}$$

Or, $x^2/\sigma^2 + (s-x)^2/\tau^2 = \frac{\sigma^2 + \tau^2}{\sigma^2\tau^2}(x - \frac{\sigma^2}{\sigma^2 + \tau^2}s)^2 + \frac{s^2}{\sigma^2 + \tau^2}$. Par conséquent,

$$\begin{aligned} f_{X+Y}(s) &= \frac{1}{2\pi\sigma\tau} \exp\left(-\frac{1}{2} \frac{s^2}{\sigma^2 + \tau^2}\right) \int_{\mathbb{R}} \exp\left(-\frac{1}{2} \left[\frac{\sigma^2 + \tau^2}{\sigma^2\tau^2}(x - \frac{\sigma^2}{\sigma^2 + \tau^2}s)^2 \right]\right) dx \\ &= \frac{1}{\sqrt{2\pi(\sigma^2 + \tau^2)}} \exp\left(-\frac{s^2}{2(\sigma^2 + \tau^2)}\right) \end{aligned}$$

puisque

$$\frac{1}{\sqrt{2\pi \frac{\sigma^2 + \tau^2}{\sigma^2\tau^2}}} \int_{\mathbb{R}} \exp\left(-\frac{1}{2} \left[\frac{\sigma^2 + \tau^2}{\sigma^2\tau^2}(x - \frac{\sigma^2}{\sigma^2 + \tau^2}s)^2 \right]\right) dx = 1$$

en tant que fonction de densité de la loi $\mathcal{N}(\frac{\sigma^2}{\sigma^2+\tau^2}s, \frac{\sigma^2\tau^2}{\sigma^2+\tau^2})$. \square

On en déduit le résultat suivant.

PROPOSITION 7.7. *Soient X_1 et X_2 des variables aléatoires indépendantes de lois respectives $\mathcal{N}(\mu_1, \sigma_1^2)$ et $\mathcal{N}(\mu_2, \sigma_2^2)$, alors $X_1 + X_2$ suit une loi $\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.*

DÉMONSTRATION. La loi de (X_1, X_2) est égale à celle de $(\mu_1 + \sigma_1 Z_1, \mu_2 + \sigma_2 Z_2)$ où (Z_1, Z_2) est un couple aléatoire normal standard. Ce que nous écrivons rapidement

$$(X_1, X_2) \stackrel{\mathcal{L}}{=} (\mu_1 + \sigma_1 Z_1, \mu_2 + \sigma_2 Z_2).$$

Par conséquent, $X_1 + X_2 \stackrel{\mathcal{L}}{=} (\mu_1 + \mu_2) + \sigma_1 Z_1 + \sigma_2 Z_2$. Mais, nous venons de montrer que $\sigma_1 Z_1 + \sigma_2 Z_2 \stackrel{\mathcal{L}}{=} \sqrt{\sigma_1^2 + \sigma_2^2} Z$ avec $Z \sim \mathcal{N}(0, 1)$. Ce qui achève la preuve. \square

THÉORÈME 7.8. *Soient X et Y deux variables aléatoires (discrètes ou continues) indépendantes de fonctions caractéristiques ϕ_X , ϕ_Y et de transformées de Laplace L_X et L_Y .*

(1) *La fonction caractéristique de $X + Y$ est*

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t), \quad t \in \mathbb{R}.$$

(2) *Si L_X et L_Y sont finies au voisinage de zéro, la transformée de Laplace de $X + Y$ est*

$$L_{X+Y}(t) = L_X(t)L_Y(t), \quad t \in \mathbb{R}.$$

DÉMONSTRATION. D'après la Proposition 6.34, $\phi_{X+Y}(t) = \phi_{X,Y}(t, t) = \phi_X(t)\phi_Y(t)$ et $L_{X+Y}(t) = L_{X,Y}(t, t) = L_X(t)L_Y(t)$. \square

EXERCICE 7.9 (Suite de l'Exercice 7.6). On reprend l'Exercice 7.6 à l'aide du Théorème 7.8.

SOLUTION. Grâce à la Proposition 5.11, $\phi_X(t) = e^{-\sigma^2 t^2/2}$ et $\phi_Y(t) = e^{-\tau^2 t^2/2}$. Le Théorème 7.8 nous donne $\phi_{X+Y}(t) = e^{-\sigma^2 t^2/2} e^{-\tau^2 t^2/2} = e^{-(\sigma^2 + \tau^2)t^2/2}$ qui est la fonction caractéristique de $\mathcal{N}(0, \sigma^2 + \tau^2)$. \square

CHAPITRE 8

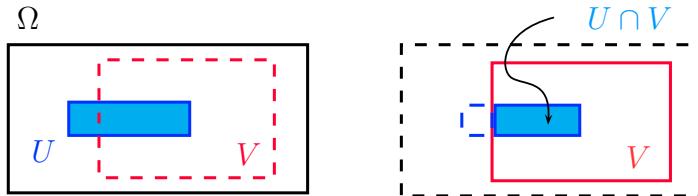
Conditionnement

8.1. Probabilité conditionnelle

Soit $V \in \mathcal{A}$ tel que $\mathbb{P}(V) > 0$. La probabilité de U conditionnelle à V est définie par la formule de Bayes

$$\mathbb{P}(U|V) := \frac{\mathbb{P}(U \cap V)}{\mathbb{P}(V)}, \quad U \in \mathcal{A}.$$

Puisque $\mathbb{P}(V|V) = 1$, l'univers de $\mathbb{P}(\cdot|V)$ est restreint à $V \subset \Omega$.



PROPOSITION 8.1. *La fonction d'ensemble $U \mapsto \mathbb{P}(U|V)$ est une mesure de probabilité sur la tribu \mathcal{A} ainsi que sur la tribu $\mathcal{A}_V := \{U \cap V; U \in \mathcal{A}\}$, trace de \mathcal{A} sur V . De plus, $\mathcal{A}_V \subset \mathcal{A}$.*

DÉMONSTRATION. En effet, $\mathbb{P}(\Omega|V) = \mathbb{P}(V|V) = 1$ et si $(U_n)_{n \geq 1}$ est une suite de \mathcal{U} telle que $\bigsqcup_{n \geq 1} U_n = \Omega$, nous avons $\bigsqcup_{n \geq 1} (U_n \cap V) = V$ et d'après la Définition 1.9 d'une mesure de probabilité, $\mathbb{P}(\bigsqcup_{n \geq 1} U_n|V) = \mathbb{P}(\bigsqcup_{n \geq 1} (U_n \cap V))/\mathbb{P}(V) = \sum_{n \geq 1} \mathbb{P}(U_n \cap V)/\mathbb{P}(V) = \sum_{n \geq 1} \mathbb{P}(U_n|V)$; ce qui prouve que $\mathbb{P}(\cdot|V)$ est une mesure de probabilité. \square

Puisque $\mathbb{P}(\cdot|V)$ est une mesure de probabilité, on peut définir la loi de (X, Y) sous $\mathbb{P}(\cdot|V)$ par

$$P_{X,Y|V}(C) := \mathbb{P}((X, Y) \in C|V)$$

pour C dans la tribu de Borel de \mathbb{R}^2 , ainsi qu'une espérance par rapport à $\mathbb{P}(\cdot|V)$

$$\mathbb{E}(\varphi(X, Y)|V) := \int_{\mathbb{R}} \varphi(x, y) P_{X,Y|V}(dx dy).$$

On voit aisément que

(a) lorsque (X, Y) est un couple aléatoire discret de loi

$$P_{X,Y} = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p_{X,Y}(x, y) \delta_{(x,y)}$$

on a

$$\begin{aligned} P_{X,Y|V} &= \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \frac{\mathbf{1}_{\{(x,y) \in X(V) \times Y(V)\}}}{\mathbb{P}(V)} p_{X,Y}(x, y) \delta_{(x,y)} \\ \mathbb{E}(\varphi(X, Y)|V) &= \sum_{x \in X(V), y \in Y(V)} \varphi(x, y) \frac{p_{X,Y}(x, y)}{\mathbb{P}(V)}; \end{aligned}$$

(b) lorsque (X, Y) est un couple aléatoire continu de loi

$$P_{X,Y}(dxdy) = f_{X,Y}(x, y) dxdy$$

on a

$$\begin{aligned} P_{X,Y|V}(dxdy) &= \frac{\mathbf{1}_{\{x \in X(V), y \in Y(V)\}}}{\mathbb{P}(V)} f_{X,Y}(x, y) dxdy \\ \mathbb{E}(\varphi(X, Y)|V) &= \iint_{X(V) \times Y(V)} \varphi(x, y) \frac{f_{X,Y}(x, y)}{\mathbb{P}(V)} dxdy. \end{aligned}$$

On note $P_{X|V}(dx)$ et $P_{Y|V}(dy)$ les lois marginales de $P_{X,Y|V}(dxdy)$.

8.2. Conditionnement dans le cas discret

Soit (X, Y) un couple aléatoire discret de loi $P_{X,Y} = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p_{X,Y}(x, y) \delta_{(x,y)}$. En prenant $V = \{Y = y\}$ avec $y \in \mathcal{Y}$ tel que $p_Y(y) > 0$, on obtient $X(V) \times Y(V) = \mathcal{X} \times \{y\}$ et

$$P_{X,Y|Y=y} = \sum_{x \in \mathcal{X}} \frac{p_{X,Y}(x, y)}{p_Y(y)} \delta_{(x,y)}$$

de sorte que

$$\begin{aligned} P_{X|Y=y} &= \sum_{x \in \mathcal{X}} p_{X|Y=y}(x) \delta_x \quad \text{avec} \\ (8.2) \quad p_{X|Y=y}(x) &= \frac{p_{X,Y}(x, y)}{p_Y(y)} = \mathbb{P}(X = x|Y = y) \quad \text{et} \\ (8.3) \quad \mathbb{E}(\varphi(X)|Y = y) &= \sum_{x \in \mathcal{X}} \varphi(x) p_{X|Y=y}(x). \end{aligned}$$

De façon analogue, on montre que pour tout $x \in \mathcal{X}$ tel que $p_X(x) > 0$,

$$\begin{aligned} P_{Y|X=x} &= \sum_{y \in \mathcal{Y}} p_{Y|X=x}(y) \delta_y \quad \text{avec} \\ (8.4) \quad p_{Y|X=x}(y) &= \frac{p_{X,Y}(x, y)}{p_X(x)} = \mathbb{P}(Y = y|X = x) \quad \text{et} \\ (8.5) \quad \mathbb{E}(\psi(Y)|X = x) &= \sum_{y \in \mathcal{Y}} \psi(y) p_{Y|X=x}(y). \end{aligned}$$

On remarque qu'il suffit que $\mathbb{E}|\varphi(X)| < \infty$ et $\mathbb{E}|\psi(Y)| < \infty$ pour que ces sommes soient absolument convergentes.

EXEMPLE 8.6. On reprend la loi jointe de l'Exemple 6.10 :

	1	3	$\leftarrow Y$
-1	0,1	0,2	0,3
2	0,45	0,25	0,7
$X \uparrow$	0,55	0,45	1

On voit que $P_{X|Y=1} = \frac{0,1}{0,55} \delta_{-1} + \frac{0,45}{0,55} \delta_2 = 0,1818 \delta_{-1} + 0,8182 \delta_2$ et que $P_{Y|X=2} = \frac{0,45}{0,7} \delta_1 + \frac{0,25}{0,7} \delta_3 = 0,6429 \delta_1 + 0,3571 \delta_3$.

On a aussi $\mathbb{E}(X^2|Y = 1) = 0,1818 \cdot (-1)^2 + 0,8182 \cdot 2^2 = 3,4546$ et $\mathbb{E}(Y|X = 2) = 0,6429 \cdot 1 + 0,3571 \cdot 3 = 1,7142$.

DÉFINITION 8.7. Pour toutes fonctions φ et ψ telles que $\mathbb{E}|\varphi(X)| < \infty$ et $\mathbb{E}|\psi(Y)| < \infty$, on définit les variables aléatoires

$$\begin{aligned}\mathbb{E}(\varphi(X)|Y) &= \sum_{y \in \mathcal{Y}} \mathbf{1}_{\{Y=y\}} \mathbb{E}(\varphi(X)|Y=y) \\ \mathbb{E}(\psi(Y)|X) &= \sum_{x \in \mathcal{X}} \mathbf{1}_{\{X=x\}} \mathbb{E}(\psi(Y)|X=x)\end{aligned}$$

et on les appelle espérance de $\varphi(X)$ sachant Y et espérance de $\psi(Y)$ sachant X .

On note que $\mathbb{E}(\varphi(X)|Y) = \alpha(Y)$ est la fonction de Y qui vaut $\mathbb{E}(\varphi(X)|Y = y)$ lorsque $Y = y$ et $\mathbb{E}(\psi(Y)|X) = \beta(X)$ est la fonction de X qui vaut $\mathbb{E}(\psi(Y)|X = x)$ lorsque $X = x$.

PROPOSITION 8.8. *Pour toutes fonctions φ et ψ telles que $\mathbb{E}|\varphi(X)| < \infty$ et $\mathbb{E}|\psi(Y)| < \infty$, nous avons*

$$\mathbb{E}[\mathbb{E}(\varphi(X)|Y)] = \mathbb{E}\varphi(X) \quad \text{et} \quad \mathbb{E}[\mathbb{E}(\psi(Y)|X)] = \mathbb{E}\psi(Y).$$

DÉMONSTRATION. Nous avons

$$\begin{aligned}\mathbb{E}[\mathbb{E}(\varphi(X)|Y)] &= \sum_{y \in \mathcal{Y}} p_Y(y) \mathbb{E}(\varphi(X)|Y=y) \\ &= \sum_{y \in \mathcal{Y}} p_Y(y) \sum_{x \in \mathcal{X}} \varphi(x) p_{X|Y=y}(x) \\ &= \sum_{y \in \mathcal{Y}} p_Y(y) \sum_{x \in \mathcal{X}} \varphi(x) \frac{p_{X,Y}(x,y)}{p_Y(y)} \\ &= \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} \varphi(x) p_{X,Y}(x,y) \\ &\stackrel{(a)}{=} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \varphi(x) p_{X,Y}(x,y) \\ &= \sum_{x \in \mathcal{X}} \varphi(x) \sum_{y \in \mathcal{Y}} p_{X,Y}(x,y) \\ &\stackrel{(b)}{=} \sum_{x \in \mathcal{X}} \varphi(x) p_X(x) \\ &= \mathbb{E}\varphi(X)\end{aligned}$$

Nous avons pu commuter les sommes en (a) car la série est absolument convergente. En (b), nous avons fait usage de la Proposition 6.9. La seconde égalité se prouve de façon analogue. \square

8.3. Conditionnement dans le cas continu

Soit (X, Y) un couple aléatoire continu de loi $P_{X,Y}(dxdy) = f_{X,Y}(x, y) dxdy$. On ne peut plus considérer aussi simplement que dans le cas discret le conditionnement par $Y = y$ car pour tout y nous avons $\mathbb{P}(Y = y) = 0$ du fait que Y est une variable continue.

Nous allons donc introduire des notions analogues aux quantités discrètes sans les justifier dans un premier temps. Nous en donnerons une justification un peu plus bas.

Pour tout y réel tel que $f_Y(y) > 0$, on définit les lois, densités et espérance conditionnelles

$$(8.9) \quad \begin{aligned} P_{X|Y=y}(dx) &= f_{X|Y=y}(x) dx \quad \text{avec} \\ f_{X|Y=y}(x) &:= \frac{f_{X,Y}(x,y)}{f_Y(y)} \quad \text{et} \end{aligned}$$

$$(8.10) \quad \mathbb{E}(\varphi(X)|Y=y) := \int_{\mathbb{R}} \varphi(x) f_{X|Y=y}(x) dx.$$

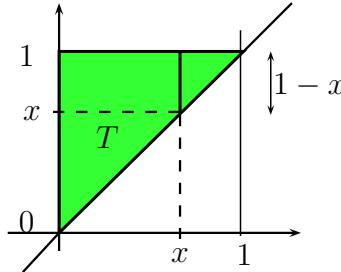
De façon analogue, on définit pour tout x réel tel que $f_X(x) > 0$,

$$(8.11) \quad \begin{aligned} P_{Y|X=x}(dy) &= f_{Y|X=x}(y) dy \quad \text{avec} \\ f_{Y|X=x}(y) &:= \frac{f_{X,Y}(x,y)}{f_X(x)} \quad \text{et} \end{aligned}$$

$$(8.12) \quad \mathbb{E}(\psi(Y)|X=x) := \int_{\mathbb{R}} \psi(y) f_{Y|X=x}(y) dy.$$

On remarque qu'il suffit que $\mathbb{E}|\varphi(X)| < \infty$ et $\mathbb{E}|\psi(Y)| < \infty$ pour que ces intégrales soient absolument convergentes.

EXEMPLE 8.13. Le couple (X, Y) suit la loi uniforme sur le domaine $T = \{(x, y) \in \mathbb{R}^2; 0 \leq x \leq y \leq 1\}$, c'est-à-dire que sa loi est $P_{X,Y}(dxdy) = f_{X,Y}(x,y) dxdy$ avec $f_{X,Y}(x,y) = 2 \cdot \mathbf{1}_T(x,y)$ puisque $\int_T dxdy = \int_0^1 \left[\int_x^1 dy \right] dx = \int_0^1 (1-x) dx = [x - x^2/2]_0^1 = 1/2$: l'aire du triangle T vaut $1/2$.



Calculons la densité marginale f_X . Pour tout x , $f_X(x) = 2 \int_{\mathbb{R}} \mathbf{1}_{\{(x,y) \in T\}} dy$. Donc, pour $x \notin [0, 1]$, $(x, y) \notin T, \forall y \in \mathbb{R}$ et $f_X(x) = 0$. Alors que pour tout $0 \leq x \leq 1$, $(x, y) \in T \Leftrightarrow x \leq y \leq 1$ et $f_X(x) = 2 \int_x^1 dy = 2(1-x)$. On a donc $f_X(x) = \mathbf{1}_{\{0 \leq x \leq 1\}} 2(1-x)$, $x \in \mathbb{R}$. Par conséquent, si $0 \leq x < 1$, $f_{Y|X=x}(y) = \frac{\mathbf{1}_{\{x \leq y \leq 1\}}}{2(1-x)}$, $y \in \mathbb{R}$. La loi de Y sachant $X = x$ est donc la loi uniforme sur $[x, 1]$. On en déduit que pour $0 \leq x < 1$, $\mathbb{E}(Y|X = x) = (1+x)/2$.

DÉFINITION 8.14. Pour toutes fonctions φ et ψ telles que $\mathbb{E}|\varphi(X)| < \infty$ et $\mathbb{E}|\psi(Y)| < \infty$, on définit les variables aléatoires

$$\begin{aligned} \mathbb{E}(\varphi(X)|Y) &= \alpha(Y) \quad \text{où } \alpha(y) = \mathbb{E}(\varphi(X)|Y = y), y \in \mathbb{R} \\ \mathbb{E}(\psi(Y)|X) &= \beta(X) \quad \text{où } \beta(x) = \mathbb{E}(\psi(Y)|X = x), x \in \mathbb{R} \end{aligned}$$

et on les appelle espérance de $\varphi(X)$ sachant Y et espérance de $\psi(Y)$ sachant X .

PROPOSITION 8.15. *Pour toutes fonctions φ et ψ telles que $\mathbb{E}|\varphi(X)| < \infty$ et $\mathbb{E}|\psi(Y)| < \infty$, nous avons*

$$\mathbb{E}[\mathbb{E}(\varphi(X)|Y)] = \mathbb{E}\varphi(X) \quad \text{et} \quad \mathbb{E}[\mathbb{E}(\psi(Y)|X)] = \mathbb{E}\psi(Y).$$

DÉMONSTRATION. Nous avons

$$\begin{aligned} \mathbb{E}[\mathbb{E}(\varphi(X)|Y)] &= \int_y f_Y(y) \mathbb{E}(\varphi(X)|Y=y) dy \\ &= \int_y f_Y(y) \left[\int_x \varphi(x) f_{X|Y=y}(x) dx \right] dy \\ &= \int_y f_Y(y) \left[\int_x \varphi(x) \frac{f_{X,Y}(x,y)}{f_Y(y)} dx \right] dy \\ &= \iint_{\mathbb{R}^2} \varphi(x) f_{X,Y}(x,y) dx dy \\ &= \int_x \varphi(x) \left[\int_y p_{X,Y}(x,y) dy \right] dx \\ &= \int_x \varphi(x) f_X(x) dx \\ &= \mathbb{E}\varphi(X) \end{aligned}$$

Nous avons pu commuter les intégrales à l'aide de leur convergence absolue. La seconde égalité se prouve de façon analogue. \square

L'ensemble des définitions introduites en (8.9), (8.10), (8.11) et (8.12) est justifié par l'obtention de la Proposition 8.15 dont l'énoncé est analogue à celui de la Proposition 8.8.

EXEMPLE 8.16 (Suite de l'Exemple 8.13). En appliquant la Proposition 8.15, on obtient $\mathbb{E}Y = \mathbb{E}[\mathbb{E}(Y|X)] = \int_{\mathbb{R}} (1+x)/2 f_X(x) dx = \int_0^1 \frac{1+x}{2} 2(1-x) dx \int_0^1 (1-x^2) dx = [x - x^3/3]_0^1 = 2/3$.

D'autre part, par symétrie on voit que $f_Y(y) = f_X(1-y) = 2y \mathbf{1}_{\{0 \leq y \leq 1\}}$ de sorte qu'on retrouve $\mathbb{E}Y = \int_{\mathbb{R}} y f_Y(y) dy = \int_0^1 2y^2 dy = [2y^3/3]_0^1 = 2/3$.

CHAPITRE 9

Indépendance (revisitée)

Nous revenons dans ce chapitre sur la notion importante d'indépendance que nous avons déjà abordée au Chapitre 6.

Lorsque je lance deux fois de suite une pièce de monnaie en la faisant à chaque fois tourner sur elle-même un grand nombre de fois, je peux me dire avec confiance que ces deux expériences sont indépendantes l'une de l'autre. En revanche, si en guise de second lancer je me contente de retourner la pièce à l'issue du premier lancer, il est clair que les deux expériences ne sont pas indépendantes.

Je lance maintenant ma pièce n fois conséutivement de sorte que je peux de prendre pour univers de l'expérience $\Omega = \{\text{p,f}\}^n$. On suppose que chaque lancer est indépendant des autres, au sens habituel du terme. Ceci se traduit par le fait que chaque suite de lancers $\omega \in \Omega$ a la même chance de se produire qu'une autre. On fait ici un *raisonnement intuitif* liant la notion ressentie d'indépendance à celle de symétrie. Ce raisonnement n'est pas mathématique, mais il s'impose à notre entendement. Nous devons traduire l'indépendance des lancers en travaillant, mathématiquement cette fois-ci, avec la probabilité \mathbb{P} qui est *uniforme* sur Ω : $\mathbb{P}(\{\omega\}) = 2^{-n}$, $\omega \in \Omega$.

EXEMPLE 9.1. J'ai une pièce de monnaie et un dé. Je lance d'abord la pièce, puis le dé. L'univers de l'expérience est $\Omega = \{\text{p,f}\} \times \{1, 2, \dots, 6\}$. On suppose que ces deux lancers sont indépendants l'un de l'autre de sorte que la probabilité \mathbb{P} est uniforme sur Ω : $\mathbb{P}((\text{p}, 1)) = \dots = \mathbb{P}((\text{f}, 6)) = 1/12$. On construit les variables aléatoires X et Y comme suit :

$$X(\omega) = \begin{cases} 0 & \text{si } \omega \in \{\text{p}\} \times \{1, 2, \dots, 6\} \\ 1 & \text{si } \omega \in \{\text{f}\} \times \{1, 2, \dots, 6\} \end{cases} ; \quad Y(\omega) = \begin{cases} 0 & \text{si } \omega \in \{\text{p}\} \times \{1, 2, 3, 4\} \\ 1 & \text{sinon} \end{cases} .$$

On voit que $\{X = 0\}$ et $\{X = 1\}$ sont respectivement les événements qui correspondent à l'obtention de **pile** et **face**. La loi de X est $\frac{1}{2}(\delta_0 + \delta_1)$ et celle de Y est $\frac{4}{12}\delta_0 + \frac{8}{12}\delta_1$.

Les variables X et Y ne sont pas indépendantes.

En effet, si je sais que $Y(\omega) = 0$, et qu'on me demande de parier sur la valeur de $X(\omega)$, j'aurai avantage à parier sur **pile**, c'est-à-dire sur $X(\omega) = 0$. Ceci car $Y(\omega) = 0$ implique que j'ai obtenu **pile**. Par conséquent, l'information $Y(\omega) = 0$ m'a permis d'obtenir une information sur $X(\omega)$.

Voici une autre manière de voir que X et Y ne sont pas indépendantes. On me demande de parier sur la valeur de $Y(\omega)$. Sans information supplémentaire, j'ai intérêt à parier sur 1, puisque la loi de Y est $\frac{4}{12}\delta_0 + \frac{8}{12}\delta_1$. En revanche, si je sais que j'ai obtenu **pile**, Y vaudra 0 si mon dé me donne 1,2,3 ou 4, soit 4 chances sur 6. J'ai donc intérêt à parier sur $Y(\omega) = 0$. Une information sur X m'a permis de modifier mon pari concernant Y . Ces deux variables ne sont donc pas indépendantes.

9.1. Définition

À l'aide de l'exemple suivant, nous allons justifier la définition mathématique de l'indépendance de deux variables aléatoires X et Y .

EXEMPLE 9.2. On joue $n + m$ fois à pile ou face. L'univers de notre expérience est donc $\Omega = \{\text{p,f}\}^{n+m}$ et l'on note $\omega_i \in \{\text{p,f}\}$ le résultat du i -ème lancer ainsi que $\omega = (\omega_1, \dots, \omega_{n+m}) \in \Omega$, la description complète de l'expérience. Une notation bien pratique est celle fournie par les variables aléatoires $Z_i : \omega \in \Omega \mapsto Z_i(\omega) = \omega_i \in \{\text{p,f}\}$, $1 \leq i \leq n + m$ ainsi que $Z = (Z_i)_{1 \leq i \leq n+m}$. On a évidemment $Z(\omega) = \omega$ pour tout $\omega \in \Omega$ et Z_i est le résultat du i -ème lancer.

On prend $n = 3$ et $m = 10$. Les variables aléatoires X et Y sont définies par

$$X = 1 + \sum_{i=1}^3 \mathbf{1}_{\{Z_i=\text{p}\}} 2^{i-1} \quad \text{et} \quad Y = 1 + \sum_{j=4}^{13} \mathbf{1}_{\{Z_j=\text{p}\}} 2^{j-1}$$

de sorte que X est une variable discrète uniforme sur $\{1, \dots, 8\}$ et Y est uniforme sur $\{1, \dots, 1024\}$. Puisque X et Y sont construites respectivement sur des tirages distincts, les trois premiers pour X et les autres pour Y , ces variables sont indépendantes (au sens intuitif). La définition mathématique de l'indépendance devra donc être cohérente avec cette constatation.

Calculons

$$\mathbb{P}(X \in A \text{ et } Y \in B)$$

avec $A \subset \{1, \dots, 8\}$ et $B \subset \{1, \dots, 1024\}$. L'espace Ω est $\Omega = \{\text{p,f}\}^{3+10} = \{\text{p,f}\}^{13}$ et toutes les réalisations ont même probabilité : $\mathbb{P}(\omega) = 2^{-13}$, pour tout $\omega \in \Omega$. L'événement $(X = 3)$ est égal à $(Z_1 = \text{f}, Z_2 = \text{p}, Z_3 = \text{f}, Z_4, \dots, Z_{14} \in \{\text{p,f}\})$. De même, $(Y = 6) = (Z_4 = \text{p}, Z_5 = \text{f}, Z_6 = \text{p}, Z_7 = \dots = Z_{14} = \text{f})$. Et en explicitant tous les tirages, nous voyons que

$$\begin{aligned} (X = 3) &= (Z_1 = \text{f}, Z_2 = \text{p}, Z_3 = \text{f}, Z_4, \dots, Z_{14} \in \{\text{p,f}\}) \\ (Y = 6) &= (Z_1, Z_2, Z_3 \in \{\text{p,f}\}, Z_4 = \text{p}, Z_5 = \text{f}, Z_6 = \text{p}, Z_7 = \dots = Z_{14} = \text{f}) \end{aligned}$$

On en déduit immédiatement que

$$(X = 3, Y = 6) = (Z_1 = \text{f}, Z_2 = \text{p}, Z_3 = \text{f}, Z_4 = \text{p}, Z_5 = \text{f}, Z_6 = \text{p}, Z_7 = \dots = Z_{14} = \text{f}).$$

Par conséquent nous avons $\mathbb{P}(X = 3) = 2^{-3}$, $\mathbb{P}(Y = 6) = 2^{-10}$ et $\mathbb{P}(X = 3, Y = 6) = 2^{-13}$. Il en est de même pour tous les événements élémentaires $(X = x, Y = y)$, de sorte qu'en notant $\#A$ et $\#B$ les cardinaux de A et B , on obtient

$$\begin{aligned} \mathbb{P}(X \in A, Y \in B) &= (\#A \times \#B) \times 2^{-13} \\ (9.3) \quad &= (\#A \times 2^{-3}) \times (\#B \times 2^{-10}) \\ &= \mathbb{P}(X \in A) \mathbb{P}(Y \in B). \end{aligned}$$

Maintenant, considérons deux fonctions $s : \{1, \dots, 8\} \rightarrow \mathbb{R}$ et $t : \{1, \dots, 1024\} \rightarrow \mathbb{R}$ ainsi que les nouvelles variables élémentaires $S = s(X)$ et $T = t(Y)$. Puisque S ne dépend que des trois premiers tirages et T que des autres tirages, ces deux variables aléatoires sont indépendantes (au sens habituel du terme). Pour tous $C, D \subset \mathbb{R}$, en posant $A = s^{-1}(C)$ et $B = t^{-1}(D)$, nous obtenons $(S \in C) = (X \in A)$ et $(T \in D) = (Y \in B)$. De sorte que

$$\begin{aligned} \mathbb{P}(S \in C, T \in D) &= \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B) \\ &= \mathbb{P}(S \in C) \mathbb{P}(T \in D) \end{aligned}$$

où la deuxième égalité est (9.3).

Cet exemple a préparé le chemin pour la définition mathématique suivante.

DÉFINITION 9.4. (Variables indépendantes)

- (1) Deux variables aléatoires X et Y sont dites *indépendantes sous la probabilité \mathbb{P}* , si pour toutes les réunions dénombrables d'intervalles A, B de \mathbb{R} ,

$$(9.5) \quad \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

- (2) Plus généralement, k variables aléatoires X_1, \dots, X_k sont dites *mutuellement indépendantes sous la probabilité \mathbb{P}* , si pour toutes les réunions dénombrables d'intervalles A_1, \dots, A_k de \mathbb{R} ,

$$(9.6) \quad \mathbb{P}(X_1 \in A_1, \dots, X_k \in A_k) = \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_k \in A_k).$$

On omettra en général de rappeler que des variables qui sont indépendantes le sont *sous \mathbb{P}* . Mais il convient de garder à l'esprit que l'indépendance n'est pas une propriété qui ne concerne que les variables aléatoires, mais en fait leur lien sous une probabilité \mathbb{P} donnée.

Revenons maintenant à l'Exemple 9.1. Puisque $(X = 0, Y = 0) = \{\mathbf{p}\} \times \{1, 2, \dots, 6\}$, nous avons $\mathbb{P}(X = 0, Y = 0) = 4/12$. D'autre part $\mathbb{P}(X = 0) = 1/2$ et $\mathbb{P}(Y = 0) = 4/12$, de sorte que $\mathbb{P}(X = 0, Y = 0) \neq \mathbb{P}(X = 0)\mathbb{P}(Y = 0)$. On retrouve le fait que X et Y ne sont pas indépendantes. En effet, il suffit pour cela que (9.5) soit invalidé pour *un* couple A, B .

Nous aurons besoin par la suite du résultat préliminaire suivant.

LEMME 9.7. *Pour que des variables aléatoires X_1, \dots, X_k soient mutuellement indépendantes sous la probabilité \mathbb{P} , il suffit que (9.6) soit satisfait pour des intervalles A_1, \dots, A_k de \mathbb{R} .*

On peut même choisir ces intervalles de la forme $A_i =]-\infty, a_i]$ avec $a_i \in \mathbb{R}$, $1 \leq i \leq k$.

On admet ce lemme dont la preuve est une jonglerie abstraite au sujet de la notion de tribu.

9.2. Propriétés élémentaires

Nous revisitons ici la Proposition 6.5 et sa preuve. Nous commençons par remarquer que des fonctions de variables indépendantes restent des variables indépendantes.

PROPOSITION 9.8. *Soient X et Y des variables indépendantes ainsi que deux fonctions $\varphi, \psi : \mathbb{R} \rightarrow \mathbb{R}$ suffisamment régulières (continues par morceaux, par exemple) pour que $S = \varphi(X)$ et $T = \psi(Y)$ soient des variables aléatoires. Alors S et T sont des variables indépendantes.*

DÉMONSTRATION. Soient C et D deux intervalles de \mathbb{R} . On a pris φ et ψ suffisamment régulières pour que $\varphi^{-1}(C) \subset \mathbb{R}$ et $\psi^{-1}(D) \subset \mathbb{R}$ puissent être approchés par des réunions finies d'intervalles disjoints. À savoir que (nous devons l'admettre au niveau de ce cours,

mais ce qui suit est très naturel) :

$$\begin{aligned}\mathbb{P}(S \in C) &= \mathbb{P}(X \in \varphi^{-1}(C)) = \lim_{K \rightarrow \infty} \mathbb{P}(X \in \sqcup_{k \leq K} I_k^K), \\ \mathbb{P}(T \in D) &= \mathbb{P}(Y \in \psi^{-1}(D)) = \lim_{L \rightarrow \infty} \mathbb{P}(Y \in \sqcup_{l \leq L} J_l^L), \\ \mathbb{P}(S \in C, T \in D) &= \lim_{K, L \rightarrow \infty} \mathbb{P}(X \in \sqcup_{k \leq K} I_k^K, Y \in \sqcup_{l \leq L} J_l^L).\end{aligned}$$

On a donc

$$\begin{aligned}\mathbb{P}(S \in C, T \in D) &= \lim_{K, L \rightarrow \infty} \mathbb{P}(X \in \sqcup_{k \leq K} I_k^K, Y \in \sqcup_{l \leq L} J_l^L) \\ &\stackrel{(a)}{=} \lim_{K, L \rightarrow \infty} \sum_{k \leq K, l \leq L} \mathbb{P}(X \in I_k^K, Y \in J_l^L) \\ &\stackrel{(b)}{=} \lim_{K, L \rightarrow \infty} \sum_{k \leq K, l \leq L} \mathbb{P}(X \in I_k^K) \mathbb{P}(Y \in J_l^L) \\ &= \lim_{K \rightarrow \infty} \sum_{k \leq K} \mathbb{P}(X \in I_k^K) \lim_{L \rightarrow \infty} \sum_{l \leq L} \mathbb{P}(Y \in J_l^L) \\ &= \mathbb{P}(S \in C) \mathbb{P}(T \in D).\end{aligned}$$

L'égalité (a) est satisfaite car les intervalles sont disjoints et l'égalité (b) est vérifiée grâce à l'indépendance de X et Y . Ce qui prouve l'indépendance sous \mathbb{P} de S et T . \square

PROPOSITION 9.9. *Soient X_1, \dots, X_{m+n} des variables mutuellement indépendantes ainsi que $\varphi : \mathbb{R}^m \rightarrow \mathbb{R}$ et $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$ deux fonctions suffisamment régulières (continues par morceaux, par exemple) pour que $S = \varphi(X_1, \dots, X_m)$ et $T = \psi(X_{m+1}, \dots, X_{m+n})$ soient des variables aléatoires. Alors S et T sont indépendantes.*

DÉMONSTRATION. La preuve de cette proposition est analogue à celle de la proposition précédente, en un peu plus technique. Les intervalles I_k^K et J_l^L doivent être remplacés par des produits cartésiens d'intervalles. Nous omettons les détails. \square

On rappelle maintenant le contenu des Propositions 6.16, 6.24 et 6.37.

PROPOSITION 9.10. *Soient X et Y deux variables indépendantes, discrètes ou continues.*

- (1) *Alors pour toutes fonctions φ et ψ telles que $\mathbb{E}|\varphi(X)| < \infty$ et $\mathbb{E}|\psi(Y)| < \infty$, nous avons $\mathbb{E}|\varphi(X)\psi(Y)| < \infty$ et $\mathbb{E}[\varphi(X)\psi(Y)] = \mathbb{E}[\varphi(X)]\mathbb{E}[\psi(Y)]$.*
- (2) *Si $\mathbb{E}|X|^2 < \infty$ et $\mathbb{E}|Y|^2 < \infty$ alors $\text{Cov}(X, Y) = 0$.*

PROPOSITION 9.11. *Soient X_1, \dots, X_{m+n} des variables mutuellement indépendantes ainsi que $\varphi : \mathbb{R}^m \rightarrow \mathbb{R}$ et $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$ deux fonctions telles que $\mathbb{E}|\varphi(X_1, \dots, X_m)| < \infty$ et $\mathbb{E}|\psi(X_{m+1}, \dots, X_{m+n})| < \infty$. Alors, $\mathbb{E}(|\varphi(X_1, \dots, X_m)||\psi(X_{m+1}, \dots, X_{m+n})|) < \infty$ et $\mathbb{E}[\varphi(X_1, \dots, X_m)\psi(X_{m+1}, \dots, X_{m+n})] = \mathbb{E}\varphi(X_1, \dots, X_m)\mathbb{E}\psi(X_{m+1}, \dots, X_{m+n})$.*

DÉMONSTRATION. C'est une conséquence directe des Propositions 9.9 et 9.10. \square

PROPOSITION 9.12. *Soient X et Y deux variables aléatoires indépendantes telles que $\mathbb{E}|X|^2 < \infty$ et $\mathbb{E}|Y|^2 < \infty$. Alors, $\mathbb{V}\text{ar}(X + Y) = \mathbb{V}\text{ar}(X) + \mathbb{V}\text{ar}(Y)$.*

De façon plus générale, si X_1, \dots, X_n sont des variables aléatoires mutuellement indépendantes telles que $\mathbb{E}|X_i|^2 < \infty$ pour tout $1 \leq i \leq n$, alors $\mathbb{V}\text{ar}(X_1 + \dots + X_n) = \mathbb{V}\text{ar}(X_1) + \dots + \mathbb{V}\text{ar}(X_n)$.

DÉMONSTRATION. Il suffit de prouver la première partie car la seconde s'en déduit aisément. Puisque X et Y sont indépendantes, $\tilde{X} := X - \mathbb{E}X$ et $\tilde{Y} := Y - \mathbb{E}Y$ sont indépendantes par la Proposition 9.9. On a donc

$$\begin{aligned}\mathbb{V}\text{ar}(X + Y) &= \mathbb{E}(\tilde{X} + \tilde{Y})^2 \\ &= \mathbb{E}(\tilde{X})^2 + 2\mathbb{E}(\tilde{X}\tilde{Y}) + \mathbb{E}(\tilde{Y})^2 \\ &\stackrel{(a)}{=} \mathbb{E}(\tilde{X})^2 + 2\mathbb{E}(\tilde{X})\mathbb{E}(\tilde{Y}) + \mathbb{E}(\tilde{Y})^2 \\ &\stackrel{(b)}{=} \mathbb{E}(\tilde{X})^2 + \mathbb{E}(\tilde{Y})^2 \\ &= \mathbb{V}\text{ar}(X) + \mathbb{V}\text{ar}(Y)\end{aligned}$$

où nous avons invoqué l'indépendance de \tilde{X} et \tilde{Y} à l'égalité (a) et $\mathbb{E}(\tilde{X}) = \mathbb{E}(\tilde{Y}) = 0$ à l'égalité (b). \square

9.3. Échantillons

On se donne une loi de variable aléatoire déterminée par la fonction de répartition F ainsi que X une variable aléatoire suivant cette loi.

DÉFINITIONS 9.13.

- (1) On appelle *copie* de X toute variable aléatoire X' ayant la même loi que X , c'est-à-dire telle que $X \stackrel{\mathcal{L}}{=} X'$.
- (2) On dit d'une suite finie (X_1, \dots, X_n) qu'elle est indépendante pour signifier que X_1, \dots, X_n sont mutuellement indépendantes.
- (3) On dit d'une suite infinie $(X_i)_{i \geq 1}$ qu'elle est indépendante pour signifier que pour tout $n \geq 2$, la suite finie (X_1, \dots, X_n) est indépendante.

DÉFINITIONS 9.14.

- (1) On dit d'une suite finie (X_1, \dots, X_n) qu'elle est un *n-échantillon* de (la loi de) X si c'est une suite indépendante de copies de X .
- (2) On dit d'une suite infinie $(X_i)_{i \geq 1}$ qu'elle est un *échantillon* de (la loi de) X si pour tout $n \geq 2$, (X_1, \dots, X_n) est un *n-échantillon* de X .
- (3) On appelle *moyenne empirique* de (X_1, \dots, X_n) la variable aléatoire

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i.$$

PROPOSITION 9.15. Soit $(X_i)_{i \geq 1}$ un échantillon de la variable X telle que $\mathbb{E}|X|^2 < \infty$. Nous avons pour tout n ,

$$\mathbb{E}\bar{X}_n = \mathbb{E}X \quad \text{et} \quad \mathbb{V}\text{ar}\bar{X}_n = \frac{\mathbb{V}\text{ar}X}{n}.$$

DÉMONSTRATION. Par linéarité de l'espérance,

$$\mathbb{E}\bar{X}_n = \frac{1}{n} \sum_{i=1}^n \mathbb{E}X_i = \frac{1}{n} n \mathbb{E}X = \mathbb{E}X.$$

D'autre part, avec les Propositions 3.32 et 9.12, nous voyons que

$$\mathbb{V}\text{ar}\bar{X}_n = \frac{1}{n^2} \mathbb{V}\text{ar} \left(\sum_{i=1}^n X_i \right) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}\text{ar} X_i = \frac{n}{n^2} \mathbb{V}\text{ar} X = \frac{\mathbb{V}\text{ar} X}{n}.$$

Ce qui achève la preuve. \square

Bien que simple, le lemme suivant a des conséquences importantes en théorie des probabilités.

LEMME 9.16.

(1) Soit Y une variable aléatoire positive. Alors, pour tout $a > 0$,

$$\mathbb{P}(Y \geq a) \leq \frac{\mathbb{E}Y}{a}.$$

(2) Soit X une variable aléatoire de variance σ^2 finie. On note $\mu = \mathbb{E}X$. Pour tout $\delta > 0$,

$$\mathbb{P}(|X - \mu| > \delta) \leq \sigma^2 / \delta^2.$$

DÉMONSTRATION. • Preuve de (1). Du fait que $Y \geq 0$, nous avons $a\mathbf{1}_{\{Y \geq a\}} \leq Y$. En prenant l'espérance, nous obtenons $\mathbb{E}[a\mathbf{1}_{\{Y \geq a\}}] \leq \mathbb{E}Y$, c'est-à-dire $a\mathbb{P}(Y \geq a) \leq \mathbb{E}Y$, qui est le résultat annoncé.

• Preuve de (2). Puisque $\mathbb{P}(|X - \mu| > \delta) = \mathbb{P}(|X - \mu|^2 > \delta^2)$, c'est une application directe de (1) avec $Y = |X - \mu|^2$, de sorte que $\mathbb{E}Y = \sigma^2$ et $a = \delta^2$. \square

THÉORÈME 9.17 (Loi faible des grands nombres). Soit $(X_i)_{i \geq 1}$ un échantillon de la variable X de variance σ^2 finie. On note $\mu = \mathbb{E}X$. Pour tout $\delta > 0$ et tout $n \geq 1$,

$$\mathbb{P}(|\bar{X}_n - \mu| > \delta) \leq \frac{\sigma^2}{n\delta^2}.$$

En particulier, pour tout $\delta > 0$,

$$(9.18) \quad \mathbb{P}(|\bar{X}_n - \mu| > \delta) \xrightarrow{n \rightarrow \infty} 0.$$

DÉMONSTRATION. L'inégalité est une conséquence immédiate de la Proposition 9.15 et du Lemme 9.16. La limite s'en déduit. \square

En passant au complémentaire, on voit que (9.18) équivaut à

$$\mathbb{P}(|\bar{X}_n - \mu| \leq \delta) \xrightarrow{n \rightarrow \infty} 1, \quad \forall \delta > 0.$$

Puisque $\delta > 0$ peut être choisi arbitrairement petit, ceci nous dit que lorsque n tend vers l'infini, la moyenne empirique \bar{X}_n , qui est une variable aléatoire, tend vers la moyenne théorique $\mu = \mathbb{E}X$, qui est un nombre non-aléatoire. Ce résultat théorique est fondamental, on l'appelle la *loi des grands nombres*.

Il permet entre autre, sur la base de l'observation d'un grand échantillon de X d'estimer la moyenne théorique $\mu = \mathbb{E}X$ que l'on suppose inconnue à l'aide de la moyenne empirique observée $\bar{X}_n(\omega)$. C'est le principe de l'inférence en statistique mathématique.

En fait, l'observation d'un grand échantillon de X permet aussi d'estimer la loi de X et à la limite, l'observation d'un échantillon infini de X permettrait (en théorie, bien sûr) de reconstruire des approximations arbitrairement fines de la loi de X . C'est ce qu'énonce le résultat suivant.

THÉORÈME 9.19. Soit $(X_i)_{i \geq 1}$ un échantillon de la variable aléatoire X sans aucune hypothèse supplémentaire (pas besoin de variance finie, ni même de $\mathbb{E}|X| < \infty$). Considérons K intervalles $I^{(1)}, \dots, I^{(K)}$, par exemple une partition dont la réunion recouvre les valeurs possibles de X . On note pour tout $1 \leq k \leq K$ et tout $n \geq 1$,

$$\hat{p}_n^{(k)} = \bar{Y}_n^{(k)} = \frac{\#\{1 \leq i \leq n; X_i \in I^{(k)}\}}{n}$$

la proportion observée de valeurs de l'échantillon "tombées" dans $I^{(k)}$. Nous avons la loi des grands nombres suivante :

$$\mathbb{P} \left(\max_{1 \leq k \leq K} |\hat{p}_n^{(k)} - \mathbb{P}(X \in I^{(k)})| \leq \delta \right) \xrightarrow{n \rightarrow \infty} 1, \quad \forall \delta > 0.$$

DÉMONSTRATION. On note pour tout $1 \leq k \leq K$ et tout $i \geq 1$,

$$Y_i^{(k)} = \mathbf{1}_{I^{(k)}}(X_i) = \begin{cases} 1 & \text{si } X_i \in I^{(k)} \\ 0 & \text{sinon} \end{cases}.$$

À k fixé, la suite $(Y_i^{(k)})_{i \geq 1}$ est un échantillon de la variable $Y^{(k)}$ qui suit la loi de Bernoulli $\mathcal{B}(p^{(k)})$ avec $p^{(k)} = \mathbb{P}(X \in I^{(k)}) = \mathbb{E}(Y^{(k)})$. D'autre part, $\hat{p}_n^{(k)} = \bar{Y}_n^{(k)}$ est la moyenne empirique des $Y_i^{(k)}$, elle obéit donc à la loi des grands nombres énoncée au Théorème 9.17. Par conséquent, pour tout k ,

$$\mathbb{P}(|\hat{p}_n^{(k)} - \mathbb{P}(X \in I^{(k)})| > \delta) \xrightarrow{n \rightarrow \infty} 0, \quad \forall \delta > 0.$$

Or, $\left(\max_{1 \leq k \leq K} |\hat{p}_n^{(k)} - \mathbb{P}(X \in I^{(k)})| > \delta \right) = \bigcup_{1 \leq k \leq K} \left(|\hat{p}_n^{(k)} - \mathbb{P}(X \in I^{(k)})| > \delta \right)$. Donc,

$$\begin{aligned} \mathbb{P} \left(\max_{1 \leq k \leq K} |\hat{p}_n^{(k)} - \mathbb{P}(X \in I^{(k)})| > \delta \right) &\leq \sum_{1 \leq k \leq K} \mathbb{P}(|\hat{p}_n^{(k)} - \mathbb{P}(X \in I^{(k)})| > \delta) \\ &\xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Ce qui achève la preuve de la proposition. \square

Les Théorèmes 9.17 et 9.19 qui sont des lois faibles des grands nombres, admettent une amélioration dont la preuve dépasse le niveau de ce cours. Il s'agit de la loi forte des grands nombres.

THÉORÈME 9.20 (Loi forte des grands nombres). Soit $(X_i)_{i \geq 1}$ un échantillon de la variable aléatoire X telle que $\mathbb{E}|X| < \infty$. Alors il existe une partie $N \in \mathcal{A}$ telle que $\mathbb{P}(N) = 0$ (dite \mathbb{P} -négligeable) telle que

$$\lim_{n \rightarrow \infty} \bar{X}^n(\omega) = \mathbb{E}X, \quad \text{pour tout } \omega \in \Omega \setminus N.$$

En particulier, sans supposer que $\mathbb{E}|X| < \infty$, en notant pour tout $n \geq 1$,

$$\hat{p}_n(\omega) = \frac{\#\{1 \leq i \leq n; X_i(\omega) \in I\}}{n}$$

la proportion observée de valeurs de l'échantillon "tombées" dans un intervalle donné I , il existe un ensemble \mathbb{P} -négligeable N tel que

$$\lim_{n \rightarrow \infty} \hat{p}_n(\omega) = \mathbb{P}(X \in I), \quad \text{pour tout } \omega \in \Omega \setminus N.$$

CHAPITRE 10

Construction d'une variable aléatoire réelle générale

Donnons-nous une fonction F candidate à être une fonction de répartition, c'est-à-dire qui satisfait les conditions (1), (2) et (4) de la Proposition 2.8. Nous allons décrire un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et construire explicitement une variable aléatoire dont la fonction de répartition est effectivement F . Nous commençons par le cas particulier d'une répartition uniforme sur $[0, 1]$.

10.1. Construction d'une variable aléatoire continue uniforme

Soit X une variable aléatoire uniforme sur l'ensemble des chiffres : $\{0, 1, \dots, 9\}$. Construisons un échantillon $(X_n)_{n \geq 1}$ de X , c'est-à-dire une suite $(X_n)_{n \geq 1}$ de copies indépendantes de X . Pour cela, on prend pour Ω l'ensemble des suites $\omega = (\omega_1, \omega_2, \dots)$ à valeurs dans $\{0, 1, \dots, 9\}$ et on définit

$$X_n(\omega) = \omega_n \in \{0, 1, \dots, 9\}, \quad \omega \in \Omega, n \geq 1$$

qui représente le résultat du n -ième tirage. On prend pour \mathcal{A} la plus petite tribu qui contient toutes les parties de Ω de la forme

$$\bigcap_{i=1}^n \{X_i \in A_i\}, \quad n \geq 1, A_i \subset \{0, \dots, 9\}, 1 \leq i \leq n$$

et on choisit une mesure de probabilité \mathbb{P} qui satisfait

$$\mathbb{P} \left(\bigcap_{i=1}^n \{X_i \in A_i\} \right) = \prod_{i=1}^n \frac{\#(A_i)}{10}, \quad \forall n \geq 1, A_1, \dots, A_n \subset \{0, \dots, 9\}.$$

Cette situation est celle de l'équiprobabilité des événements élémentaires, puisqu'il y a $\prod_{i=1}^n \#(A_i)$ nombres parmi les 10^n nombres entiers de $[0, 10^n - 1]$ dont le i -ème chiffre est dans A_i pour tout $1 \leq i \leq n$.

On admet qu'une telle mesure de probabilité sur (Ω, \mathcal{A}) existe et est unique.

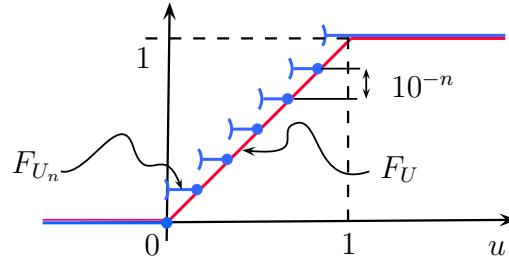
Pour tout $n \geq 1$, on définit la variable aléatoire

$$\begin{aligned} U_n(\omega) &= 0, \omega_1 \dots \omega_n \quad (\text{développement décimal}) \\ &= \sum_{i=1}^n \omega_i 10^{-i} \end{aligned}$$

Il est clair que U_n peut prendre 10^n valeurs dans $[0, 1[$. Calculons sa fonction de répartition. Bien sûr, $F_{U_n}(u) = 0$, si $u < 0$ et $F_{U_n}(u) = 1$ si $u \geq 1$. Soit maintenant $0 \leq u < 1$.

En notant $u = 0, x_1 x_2 \dots$ son développement décimal,

$$\begin{aligned}
 F_{U_n}(u) &= \mathbb{P}(U_n \leq u) \\
 &= \mathbb{P}\left(\{\omega \in \Omega; 0, \omega_1 \dots \omega_n \leq 0, x_1 \dots x_n x_{n+1} \dots\}\right) \\
 &= \mathbb{P}\left(\{X_1 \leq x_1 - 1\} \sqcup [\{X_1 = x_1\} \cap \{X_2 \leq x_2 - 1\}] \sqcup \dots \right. \\
 &\quad \left. \sqcup [\{X_1 = x_1\} \cap \dots \cap \{X_{n-1} = x_{n-1}\} \cap \{X_n \leq x_n - 1\}] \right. \\
 &\quad \left. \sqcup [\{X_1 = x_1\} \cap \dots \cap \{X_n = x_n\}]\right) \\
 &= 10^{-1} x_1 + 10^{-2} x_2 + \dots + 10^{-n} x_n + 10^{-n} \\
 &= 0, x_1 x_2 \dots x_n + 10^{-n}.
 \end{aligned}$$



Par conséquent, $\lim_{n \rightarrow \infty} F_{U_n}(u) = G(u) := \begin{cases} 0 & \text{si } u \leq 0 \\ u & \text{si } 0 \leq u \leq 1 \\ 1 & \text{si } u \geq 1 \end{cases}, \forall u \in \mathbb{R}$. Posons

$$(10.1) \quad U(\omega) = \lim_{n \rightarrow \infty} U_n(\omega) = 0, \omega_1 \omega_2 \dots, \omega \in \Omega.$$

Puisque $\sup_{\omega \in \Omega} |U_n(\omega) - U(\omega)| \leq 10^{-n}$, pour tout $\varepsilon > 0$ et tout entier n suffisamment grand pour que $10^{-n} \leq \varepsilon$, nous avons : $\{U_n \leq u - \varepsilon\} \subset \{U \leq u\} \subset \{U_n \leq u + \varepsilon\}$. D'où il vient que $F_{U_n}(u - \varepsilon) \leq F_U(u) \leq F_{U_n}(u + \varepsilon)$. Ce qui en faisant tendre n vers l'infini nous donne $G(u - \varepsilon) \leq F_U(u) \leq G(u + \varepsilon)$, puis en faisant tendre ε vers zéro, nous donne $F_U = G$. Soit

$$F_U(u) = \begin{cases} 0 & \text{si } u \leq 0 \\ u & \text{si } 0 \leq u \leq 1 \\ 1 & \text{si } u \geq 1 \end{cases}, \quad u \in \mathbb{R}.$$

La loi de U , spécifiée par sa fonction de répartition F_U , est appelée *loi uniforme* sur $[0, 1]$. Sa fonction de densité est donnée par

$$f_U(u) = \begin{cases} 1 & \text{si } u \in [0, 1] \\ 0 & \text{sinon} \end{cases}, \quad u \in \mathbb{R}.$$

On vient de construire U à l'aide d'une infinité dénombrable de tirages indépendants uniformes dans $\{0, \dots, 9\}$.

REMARQUE 10.2. Lors de la preuve de la Proposition A.8, on montre que le procédé de construction (10.1) atteint tous les réels de $[0, 1]$ une seule fois à l'exception de certains qui sont atteints deux fois : les éléments de \mathbb{D} , l'ensembles des nombres dans $[0, 1]$ qui admettent un développement décimal fini. Or, il est aussi prouvé que \mathbb{D} est *dénombrable* de sorte que $\mathbb{P}(U \in \mathbb{D}) = \sum_{x \in \mathbb{D}} \mathbb{P}(U = x) = \sum_{x \in \mathbb{D}} 0 = 0$ où l'avant-dernière égalité vient de $\mathbb{P}(U = x) = 0$ pour tout x et la dernière a du sens car \mathbb{D} étant *dénombrable*, $\sum_{x \in \mathbb{D}}$ est une série numérique.

10.2. Construction d'une variable aléatoire réelle générale

La variable aléatoire U va nous permettre de construire toutes les autres variables aléatoires sur $(\Omega, \mathcal{A}, \mathbb{P})$. Le procédé de construction est le suivant.

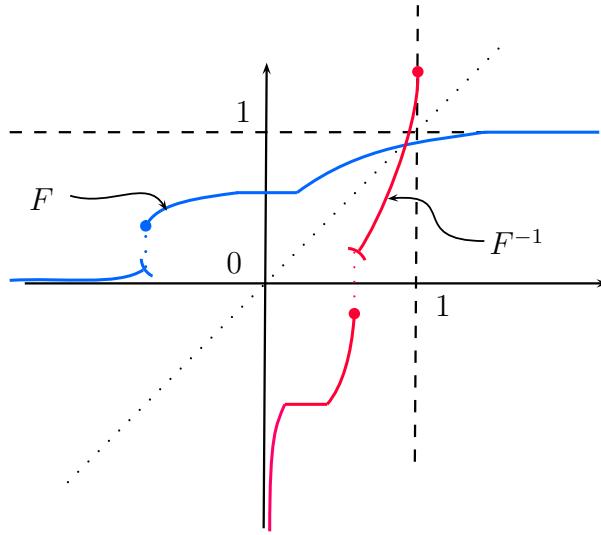
THÉORÈME 10.3. *Soit une fonction $F : \mathbb{R} \rightarrow [0, 1]$, croissante et continue à gauche telle que $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow \infty} F(x) = 1$. On définit son inverse sur $]0, 1[$ par*

$$(10.4) \quad F^{-1}(u) := \inf\{x \in \mathbb{R}; F(x) \geq u\}, \quad u \in]0, 1[.$$

On considère $U \sim \mathcal{U}(0, 1)$ une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ de loi uniforme sur $]0, 1[$. Alors

$$(10.5) \quad X = F^{-1}(U)$$

est une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ de fonction de répartition F .



PREUVE DU THÉORÈME 10.3. Rappelons que pour tout $0 \leq u \leq 1$, $F_U(u) = \mathbb{P}(U \leq u) = \mathbb{P}(U < u) = u$.

Si x est un point de continuité de F , alors $F^{-1}(u) \leq x \iff u \leq F(x)$, de sorte que

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) = \mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) \\ &= F_U(F(x)) = F(x) \end{aligned}$$

On note $F(x^+)$ et $F(x^-)$ les limites à droite et à gauche de F en x (ces limites existent puisque F est supposée croissante). Si x est un point de discontinuité de F , alors $F(x^-) < F(x) = F(x^+)$, $F^{-1}(u) < x \iff u < F(x^-)$ et $F^{-1}(u) = x \iff F(x^-) \leq u \leq F(x)$. Donc,

$$\begin{aligned} F_X(x) &= \mathbb{P}(F^{-1}(U) < x) + \mathbb{P}(F^{-1}(U) = x) \\ &= \mathbb{P}(U < F(x^-)) + \mathbb{P}(F(x^-) \leq U \leq F(x)) \\ &= F(x^-) + [F(x) - F(x^-)] = F(x). \end{aligned}$$

Ce qui achève la preuve de $F_X = F$ et donc de la proposition. \square

Remarquons que nous avons déjà montré à la Proposition 2.8 que toute fonction de répartition jouit des propriétés imposées à F dans le Théorème 10.3. Nous en déduisons le résultat suivant.

COROLLAIRE 10.6. *Une fonction F est la fonction de répartition d'une variable aléatoire si et seulement si $F : \mathbb{R} \rightarrow [0, 1]$ est croissante, continue à gauche et satisfait $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow \infty} F(x) = 1$.*

EXEMPLES 10.7.

- (a) Loi de Bernoulli $\mathcal{B}(p)$. Nous avons $F(x) = q\mathbf{1}_{[0,1]}(x) + p\mathbf{1}_{[1,\infty]}(x)$ avec $p + q = 1$, dont l'inverse est $F^{-1}(u) = \mathbf{1}_{[q,1]}(u)$, $0 \leq u \leq 1$.
 Par conséquent $X = \begin{cases} 0 & \text{si } U \in [0, q] \\ 1 & \text{si } U \in]q, 1] \end{cases}$ suit la loi $\mathcal{B}(p)$. On remarque que la longueur de $[0, q]$ est $q = \mathbb{P}(X = 0)$ et que celle de $]q, 1]$ est $1 - q = p = \mathbb{P}(X = 1)$.
- (b) Loi exponentielle $\mathcal{E}(\lambda)$. Nous avons $F(x) = \mathbf{1}_{\{x \geq 0\}}(1 - e^{-\lambda x})$ de sorte que $F^{-1}(u) = -\ln(1 - u)/\lambda$, $u \in [0, 1[$. On voit donc que $X = -\ln(1 - U)/\lambda$ suit la loi $\mathcal{E}(\lambda)$. Or $U \stackrel{d}{=} 1 - U$, donc $X = -\ln(U)/\lambda \sim \mathcal{E}(\lambda)$.

Attention, dans (10.5) F^{-1} n'est pas l'inverse traditionnel de F mais seulement son inverse généralisé. En particulier il n'est pas vrai en général que $F(X) = U$, c'est-à-dire que $F(X)$ soit une variable aléatoire uniforme sur $(0, 1)$.

EXERCICE 10.8.

- (a) Soit $X \sim \mathcal{B}(2, 1/2)$ la variable aléatoire de l'Exemple 2.1, montrer que $F(X)$ n'est pas uniforme sur $(0, 1)$.
 Calculer sa loi.
- (b) Soit X une variable aléatoire continue de fonction de répartition F , montrer que $F(X)$ est uniforme sur $(0, 1)$.

SOLUTION. • Solution de (a). Puisque $\#(X(\Omega)) = \#(\{0, 1, 2\}) = 3$ et $\#(U(\Omega)) = \#([0, 1]) = \infty$, $\#(F(X(\Omega))) \leq 3$ donc $F(X)$ ne peut pas avoir la même loi que U .
 Plus précisément, $P_X = \frac{1}{4}\delta_{F(0)} + \frac{1}{2}\delta_{F(1)} + \frac{1}{4}\delta_{F(2)} = \frac{1}{4}\delta_{1/4} + \frac{1}{2}\delta_{3/4} + \frac{1}{4}\delta_1$.

• Solution de (b). Au début de la preuve du Théorème 10.3, nous avons vu que si x est un point de continuité de F , alors pour tout $0 \leq u \leq 1$, $F^{-1}(u) \leq x \iff u \leq F(x)$. Or, sous notre hypothèse, F est continue partout, donc pour tout $0 \leq u \leq 1$,

$$\begin{aligned} \mathbb{P}(F(X) \geq u) &= \mathbb{P}(X \geq F^{-1}(u)) \\ &\stackrel{(a)}{=} \mathbb{P}(X > F^{-1}(u)) \\ &\stackrel{(b)}{=} 1 - F(F^{-1}(u)) \\ &\stackrel{(c)}{=} 1 - u \end{aligned}$$

où l'égalité (a) est vraie car X est une variable continue, (b) vient de la définition de la fonction de répartition F et (c) se vérifie comme suit.

Pour tout $0 \leq u \leq 1$, $F(F^{-1}(u)) = F(\inf\{x; F(x) \geq u\}) = \lim_{x \rightarrow \alpha^-} F(x) := F(\alpha^-)$ où α est l'unique nombre tel que $F(\alpha^-) \leq u \leq F(\alpha)$. Or F est supposée continue, donc $F(\alpha^-) = F(\alpha)$, ce qui implique que $F(\alpha^-) = u$ et $F(F^{-1}(u)) = u$.

On en déduit que $\mathbb{P}(F(X) \leq u) = 1 - \lim_{v \rightarrow u^-} (1 - v) = 1 - (1 - u) = u$ pour tout $0 \leq u \leq 1$, ce qui montre que $F(X)$ suit une loi uniforme sur $(0, 1)$. \square

CHAPITRE 11

Simulation d'une variable aléatoire

Il existe des algorithmes qui génèrent des suites de tirages pseudo-aléatoires indépendants de loi $\mathcal{U}(0, 1)$ uniforme sur $[0, 1]$. La plupart des calculettes permettent d'exécuter de tels programmes, souvent baptisés RAND¹. En général leur conception repose sur des propriétés arithmétiques de certaines suites récurrentes. Ces algorithmes sont déterministes, c'est-à-dire qu'il n'ont rien d'aléatoire. Si vous utilisez le même algorithme avec la même donnée initiale, il vous donnera toujours la même suite de nombres. De plus, ces suites de tirages de valeurs numériques sont périodiques, mais avec une période extrêmement grande. C'est la raison pour laquelle ces générateurs sont appelés *pseudo*-aléatoires plutôt qu'aléatoires.

11.1. Description rapide de certains générateurs

Une famille de générateurs populaire est celle des générateurs congruentiels linéaires. Ils génèrent des suites de nombres entiers $(x_n)_{n \geq 1}$ dans l'ensemble $\{0, \dots, m - 1\}$ où m est un grand nombre. Il suffit ensuite de prendre $u_n = x_n/m$ pour obtenir une suite de tirages $(u_n)_{n \geq 1}$ dans $[0, 1]$ dont les valeurs sont des nombres arrondis avec une précision de l'ordre de $1/m$. La suite $(x_n)_{n \geq 1}$ est solution de l'équation de récurrence

$$x_{n+1} = ax_n + b \pmod{m}, \quad n \geq 0$$

en partant d'une donnée initiale entière x_0 . On rappelle que $x = r$ modulo m signifie que r est le reste de la division euclidienne (celle de la petite école) de x par m . En d'autres termes $x = qm + r$ avec un quotient q entier et $0 \leq r \leq m - 1$. On constate immédiatement qu'une telle suite est périodique (de période au plus m). Il faut donc que m soit très grand. En choisissant intelligemment a et b , cette période est effectivement m . D'autre part il faut aussi choisir adéquatement les nombres a, b et m pour que la suite simule correctement de très longues séquences (de l'ordre de $m/10$) de tirages *uniformes* et *indépendants*. En fait, le choix de ces paramètres est loin d'être évident et est encore l'objet de recherche. La fonction `rand` de Scilab utilise les valeurs $m = 2^{31}$, $a = 843314861$ et $b = 453816693$. La fonction `grand` de Scilab est basée sur un type de générateur déterministe plus performant dont la période $2^{19937} - 1$ est fabuleuse. La plupart des générateurs utilise la date et l'heure de votre ordinateur pour décider de la valeur initiale x_0 .

11.2. Simulation. Principe et applications

Nous appellerons U le résultat d'un tirage de loi $\mathcal{U}(0, 1)$. Puisque les ordinateurs ont une précision finie, les valeurs u_n que nous fournit notre générateur sont des tirages

¹En anglais, *au hasard* se dit *at random* qui vient de l'ancienne expression française "aller à randon" qui signifie avancer de façon désordonnée et que l'on retrouve dans *randomnée*.

uniformes sur un ensemble de grand cardinal et nous utilisons en fait une approximation U_m de la variable U dans le même esprit que (10.1).

Principe général de la simulation. Ce principe est une application directe du Théorème 10.3. Soit U_1, U_2, \dots un échantillon de la loi uniforme $\mathcal{U}(0, 1)$. Alors, grâce au Théorème 10.3, on sait que, F^{-1} désignant l'inverse généralisé de la fonction de répartition F de la loi de X , voir (10.4),

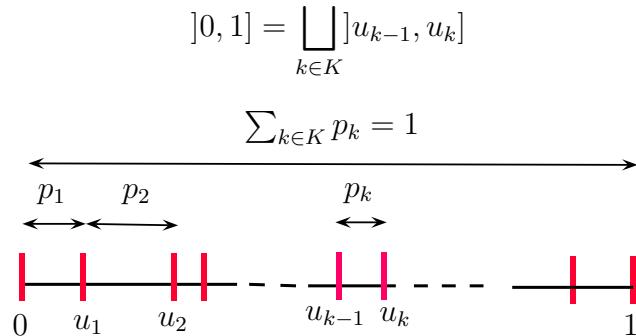
$$X_i := F^{-1}(U_i), \quad i \geq 1$$

définit un échantillon de la loi de X . C'est-à-dire une famille de copies indépendantes de X . Ce principe s'applique donc lorsqu'on connaît une expression de F^{-1} .

Variables discrètes. Dans le cas d'une variable discrète, le principe précédent correspond à une manipulation intuitivement claire que nous allons décrire. La méthode est simple.

La variable discrète X que nous souhaitons simuler prend ses valeurs dans $\{x_k; k \in K\}$ avec $K \subset \{1, 2, \dots\}$. Sa loi s'écrit $\sum_{k \in K} p_k \delta_{x_k}$. On suppose sans perte de généralité que $p_k > 0$ pour tout k .

On partitionne l'intervalle $]0, 1]$ de sorte que



avec $u_0 = 0$ et $u_k = \sum_{1 \leq i \leq k} p_i$, $k \in K$. La probabilité que la variable U de loi uniforme sur $(0, 1)$ tombe dans k -ième boîte $B_k =]u_{k-1}, u_k]$ est

$$\mathbb{P}(U \in B_k) = \mathbb{P}(u_{k-1} < U \leq u_k) = u_k - u_{k-1} = p_k, \quad k \in K.$$

La variable

$$(11.1) \quad X = \sum_{k \in K} x_k \mathbf{1}_{\{U \in B_k\}}$$

qui vaut x_k si et seulement si $U \in B_k$, $k \in K$ a pour loi $\sum_{k \in K} p_k \delta_{x_k}$.

EXERCICE 11.2. Montrer que la variable X définie par (11.1) satisfait l'égalité (10.5) : $X = F^{-1}(U)$, du Théorème 10.3.

EXEMPLES 11.3.

- Pour simuler un tirage du jeu de pile ou face il suffit de décider pile si $U \in [0, 1/2[$ et face si $U \in [1/2, 1[$.
- Pour simuler la variable aléatoire X de l'Exemple 2.6, on décide par exemple : $X(\omega) = 0$ si $U(\omega) \in [0, 1/4[$, $X(\omega) = 1$ si $U(\omega) \in [1/4, 3/4[$, $X(\omega) = 2$ si $U(\omega) \in [3/4, 1[$.

Ou bien, $X(\omega) = 0$ si $U(\omega) \in [0, 1/8[\cup [3/4, 7/8[$, $X(\omega) = 1$ si $U(\omega) \in [1/8, 1/2[\cup [7/8, 1[$, $X(\omega) = 2$ si $U(\omega) \in [1/2, 3/4[$. Mais c'est moins pratique.

- (c) Pour simuler le premier instant X d'apparition de pile lors d'une suite de lancers indépendants d'une pièce que nous avons rencontré à l'Exemple 2.12-(b), on peut inverser la fonction de répartition : $X(\omega) = 0$ si $U(\omega) \in [0, 1/2[$, $X(\omega) = 1$ si $U(\omega) \in [1/2, 3/4[$, $X(\omega) = 2$ si $U(\omega) \in [3/4, 7/8[$, ...

Ou bien on peut décomposer $U(\omega)$ en base 2 et choisir pour $X(\omega)$ la place de la première apparition de 1 dans cette décomposition.

Variables exponentielles et variables de Poisson. Nous avons vu à l'Exemple 10.7-(b) que

$$(11.4) \quad T = -\ln(U)/\lambda$$

suit une loi exponentielle $\mathcal{E}(\lambda)$ lorsque U est une variable uniforme sur $[0, 1]$. Or le générateur `rand` produit des réalisations indépendantes U_1, U_2, \dots de variables de loi $\mathcal{U}(0, 1)$ uniforme sur $[0, 1]$. Par conséquent $(T_i)_{i \geq 1}$, où $T_i = -\ln(U_i)/\lambda$, est une suite de variables indépendantes de loi $\mathcal{E}(\lambda)$. La suite croissante $(S_n)_{n \geq 1}$ définie par

$$S_n = \sum_{i=1}^n T_i$$

décrit ce qu'on appelle un processus de Poisson de paramètre λ . Les S_n sont les instants de réalisations de certains événements alors que les T_i sont les temps d'attente entre deux événements consécutifs.

Par exemple, les instants de désintégration d'un corps constitué d'un élément radioactif de composition pure sont très bien décrits par une telle suite aléatoire. Le paramètre de fréquence λ est alors proportionnel à la masse du corps et inversement proportionnel à la période de demi-vie de l'élément.

Soit N le nombre d'occurrences d'événement pendant l'intervalle de temps $[0, 1]$. En d'autres termes, N est spécifié par :

$$(11.5) \quad S_N \leq 1 < S_{N+1}.$$

On peut montrer que N est une variable aléatoire de Poisson de paramètre λ . De façon plus générale, le nombre d'événements pendant un intervalle de temps $[s, t]$ est une variable de Poisson de paramètre $(t-s)\lambda$. Cette propriété permet de simuler une variable N de Poisson $\mathcal{P}(\lambda)$. En effet, (11.5) équivaut à

$$\prod_{i=1}^{N+1} U_i < e^{-\lambda} \leq \prod_{i=1}^N U_i.$$

De sorte que $N+1$ est le nombre de fois qu'il faut multiplier entre eux des $U_i \sim \mathcal{U}(0, 1)$ indépendants, pour passer pour la première fois en dessous de $e^{-\lambda}$.

Cette méthode de simulation d'une variable de Poisson est plus performante que celle basée sur le principe général que nous avons présentée à la Section 11.2.

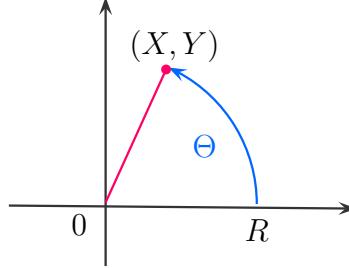
Variables normales. On appelle *couple* aléatoire normal standard un couple (X, Y) de variables aléatoires indépendantes normales standard $X, Y \sim \mathcal{N}(0, 1)$. L'application directe du Théorème 10.3 est compromise par le fait qu'il n'existe pas d'*expression analytique de la fonction de répartition de $\mathcal{N}(0, 1)$* . A fortiori, nous n'avons pas d'*expression*

explicite de sa fonction réciproque. Nous allons toutefois contourner ce problème en résolvant l'exercice suivant.

EXERCICE 11.6. Soit (X, Y) un couple normal standard. On définit (R, Θ) comme étant les coordonnées polaires de (X, Y) , c'est-à-dire

$$\begin{cases} X = R \cos \Theta \\ Y = R \sin \Theta \end{cases}$$

avec $R \geq 0$ et $0 \leq \Theta < 2\pi$.



Montrer que R et Θ sont des variables indépendantes telles que $R^2 \sim \mathcal{E}(1/2)$ et $\Theta \sim \mathcal{U}(0, 2\pi)$.

SOLUTION. La densité de la loi de (X, Y) est $f_{X,Y}(x, y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}$ et notons $g(r, \theta)$ celle de (R, Θ) , si elle existe. Soit T la transformation inverse de $(r, \theta) \mapsto (x, y) = (r \cos \theta, r \sin \theta)$ de sorte que $(R, \Theta) = T(X, Y)$.

On se donne φ une fonction bornée régulière quelconque sur $[0, \infty[\times [0, 2\pi[$. Nous avons

$$\begin{aligned} \mathbb{E}\varphi(R, \Theta) &= \mathbb{E}\varphi(T(X, Y)) \\ &= \iint_{\mathbb{R}^2} \varphi(T(x, y)) \frac{1}{2\pi} e^{-(x^2+y^2)/2} dx dy \\ &= \iint_{[0, \infty[\times [0, 2\pi[} \varphi(r, \theta) \frac{1}{2\pi} e^{-r^2/2} r dr d\theta \\ &= \iint_{\mathbb{R}^2} \varphi(r, \theta) g(r, \theta) dr d\theta \end{aligned}$$

avec $g(r, \theta) = g_\Theta(\theta)g_R(r)$ où $g_\Theta(\theta) = \frac{1}{2\pi} \mathbf{1}_{[0, 2\pi[}(\theta)$ et $g_R(r) = \mathbf{1}_{[0, \infty[}(r)re^{-r^2/2}$, en effectuant un changement de variables en coordonnées polaires à l'avant-dernière égalité. Puisque g a la forme produit, R et Θ sont indépendantes de densité g_R et g_Θ . Les variables R^2 et Θ sont donc aussi indépendantes. Clairement, $\Theta \sim \mathcal{U}(0, 2\pi)$ et pour tout $t \geq 0$, $\mathbb{P}(R^2 \leq t) = \mathbb{P}(R \leq \sqrt{t}) = \int_0^{\sqrt{t}} e^{-r^2/2} r dr = \int_0^t e^{-s/2} ds/2$ en faisant le changement de variable $s = r^2$. On voit donc que la densité de la loi de $S = R^2$ est $\mathbf{1}_{[0, \infty[}(s) \frac{1}{2} e^{-s/2}$, c'est-à-dire $R^2 \sim \mathcal{E}(1/2)$. \square

Il suffit maintenant de simuler (R, Θ) à l'aide d'un couple (U, V) de variables indépendantes distribuées uniformément sur $[0, 1]$ dont la réalisation est donnée par deux valeurs consécutives du programme RAND. On prend alors

$$\begin{cases} R = \sqrt{-2 \ln U} \\ \Theta = 2\pi V \end{cases}$$

où l'on a utilisé (11.4) dans le calcul de R et (4.6) dans celui de Θ . Finalement, nous venons de montrer que le couple (X, Y) donné par

$$\begin{cases} X = \sqrt{-2 \ln U} \cos(2\pi V) \\ Y = \sqrt{-2 \ln U} \sin(2\pi V) \end{cases}$$

est un couple normal standard. Bien sûr, avec un échantillon $(U_i)_{i \geq 1}$ de $\mathcal{U}(0, 1)$,

$$\left(\sqrt{-2 \ln U_1} \cos(2\pi U_2), \sqrt{-2 \ln U_1} \sin(2\pi U_2), \sqrt{-2 \ln U_3} \cos(2\pi U_4), \sqrt{-2 \ln U_3} \sin(2\pi U_4), \dots \right)$$

forme un échantillon de $\mathcal{N}(0, 1)$.

D'autre part, si $Z \sim \mathcal{N}(0, 1)$, on sait que $X = m + \sigma Z$ suit la loi normale $\mathcal{N}(m, \sigma^2)$. On en déduit que

$$\left(m + \sigma \sqrt{-2 \ln U_1} \cos(2\pi U_2), m + \sigma \sqrt{-2 \ln U_1} \sin(2\pi U_2), m + \sigma \sqrt{-2 \ln U_3} \cos(2\pi U_4), m + \sigma \sqrt{-2 \ln U_3} \sin(2\pi U_4), \dots \right)$$

forme un échantillon de $\mathcal{N}(m, \sigma^2)$.

11.3. Histogrammes

Un générateur `rand` parfait devrait produire une suite de réalisations de variables aléatoires

- (1) de loi $\mathcal{U}(0, 1)$
- (2) qui sont mutuellement indépendantes.

Mais qu'est-ce que cela signifie et comment s'en assurer ? En ce qui concerne l'indépendance, le problème est assez délicat et nous ne l'aborderons pas ici. Disons seulement qu'il existe des tests statistiques d'indépendance et qu'il est recommandé que les générateurs pseudo-aléatoires passent ces tests avec de faibles erreurs de première et seconde espèces.

Revenons au premier point, à savoir que la loi du pseudo-échantillon soit bien uniforme. Puisque nous ne sommes pas en mesure de produire un argument de symétrie comme lors d'un jeu de pile ou face, notre seule façon de comprendre ce que signifie *suivre une loi donnée* (ici, uniforme) est de se référer à une interprétation fréquentielle. À savoir que si l'on est face à un très grand nombre de réalisations consécutives, ces tirages se laisseront classés avec des proportions observées qui sont proches des proportions théoriques attendues. Par exemple, si l'on découpe le segment $[0, 1]$ en 100 sous-intervalles de même longueur et qu'on observe 40 000 de réalisations, on s'attend à ce qu'il y ait à peu près $40\ 000/100=400$ nombres dans chacun des sous-intervalles, et ce avec des fluctuations typiques de ce que la théorie des probabilités prévoit, ici de l'ordre de ± 20 ; on pense en particulier au théorème central limite qui quantifie ces fluctuations lorsque la taille n de l'échantillon est grande.

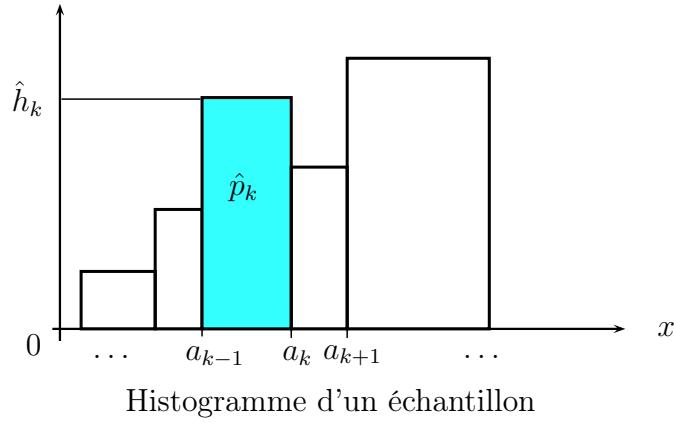
Dans le sous-intervalle $[a, b]$ on attend donc une proportion $(b - a)/(1 - 0) = b - a$ de tirages $\mathcal{U}(0, 1)$ lorsque n est grand. C'est ce dont nous assure la *loi des grands nombres* et que allons tester en construisant des *histogrammes*.

On partitionne un intervalle contenant toutes les valeurs possibles de la variable X en un nombre fini K de sous-intervalles $[a_{k-1}, a_k[, 1 \leq k \leq K$. Les sous-intervalles (ou classes) sur les bords pouvant éventuellement ne pas être bornés. Par exemple, si X est une variable binomiale $\mathcal{B}(m, p)$, elle prend a priori les valeurs $\{0, 1, \dots, m\}$ et on pourra considérer les classes $[-0.5, 0.5[, [0.5, 1.5[, \dots, [m - 0.5, m + 0.5]$ qui encadrent de façon symétrique les valeurs effectives de X .

On observe un échantillon de X de taille n , c'est-à-dire les réalisations $x_1 = X_1(\omega), \dots, x_n = X_n(\omega)$ de la suite de copies indépendantes X_1, \dots, X_n de X . On note \hat{p}_k la proportion de x_i dans la k -ième classe $[a_{k-1}, a_k[,$ soit

$$\hat{p}_k(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{x_i \in [a_{k-1}, a_k[\}} , \quad 1 \leq k \leq K.$$

Par définition, l'histogramme des observations est la figure suivante, où \hat{h}_k est calculé de telle sorte que l'aire au-dessus de la k -ième classe soit \hat{p}_k .



C'est-à-dire

$$(11.7) \quad \hat{h}_k(x_1, \dots, x_n) = \frac{\hat{p}_k(x_1, \dots, x_n)}{a_k - a_{k-1}}, \quad 1 \leq k \leq K.$$

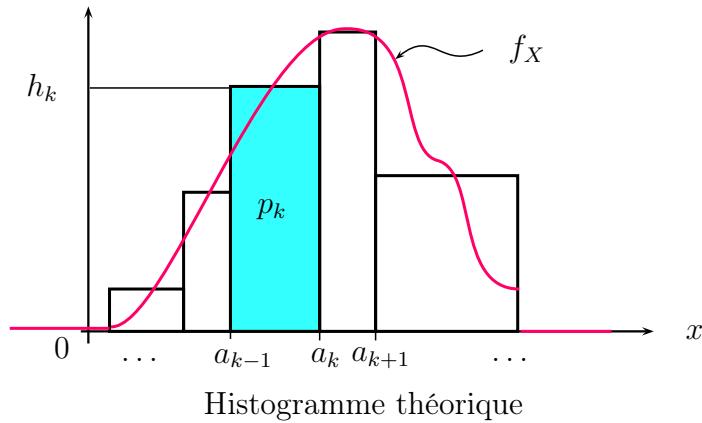
Supposons que X soit une variable de densité f_X . On sait que

$$p_k := \mathbb{P}(X \in [a_{k-1}, a_k[) = \int_{a_{k-1}}^{a_k} f_X(x) dx,$$

de sorte que $p_k = (a_k - a_{k-1})h_k$ en posant

$$(11.8) \quad h_k = \frac{\int_{a_{k-1}}^{a_k} f_X(x) dx}{a_k - a_{k-1}}, \quad 1 \leq k \leq K,$$

qui n'est autre que la valeur moyenne de f_X sur la classe $[a_{k-1}, a_k[$. En traçant le graphe des h_k en fonction des classes $[a_{k-1}, a_k[,$ on obtient l'histogramme théorique suivant.



La similarité des formules (11.7) et (11.8) justifie le mode de construction des histogrammes d'échantillon. En effet, la courbe de l'histogramme théorique h est une simplification de la courbe de densité f_X qui ne retient que l'information d'appartenance aux classes $[a_{k-1}, a_k]$. D'autre part, avec la loi forte des grands nombres énoncée au Théorème 9.20, on sait que pour tout $1 \leq k \leq K$, et \mathbb{P} -presque toute réalisation ω ,

$$\lim_{n \rightarrow \infty} \hat{h}_k(X_1(\omega), \dots, X_n(\omega)) = h_k.$$

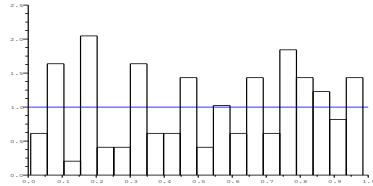
Par conséquent, lorsque n et K sont grands, l'histogramme observé

$$x \mapsto \hat{h}(X_1, \dots, X_n)(x) = \sum_{1 \leq k \leq K} \hat{h}^k(X_1, \dots, X_n) \mathbf{1}_{[a_{k-1}, a_k]}(x)$$

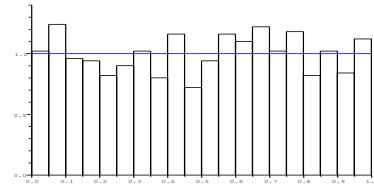
est proche de la densité théorique $x \mapsto f_X(x)$.

On voit donc que si l'on sait que les $(X_i)_{1 \leq i \leq n}$ sont bien des copies indépendantes de la loi de X , l'histogramme donne une approximation raisonnable de la densité f_X lorsque K et n sont grands.

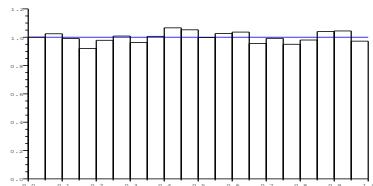
Les figures suivantes sont les histogrammes à 20 classes équilibrées de 100, 1000, 10 000 et 100 000 tirages uniformes effectués à l'aide du générateur `rand` de Scilab.



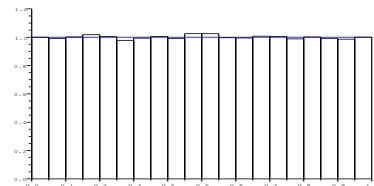
$n = 100$



$n = 1000$



$n = 10000$



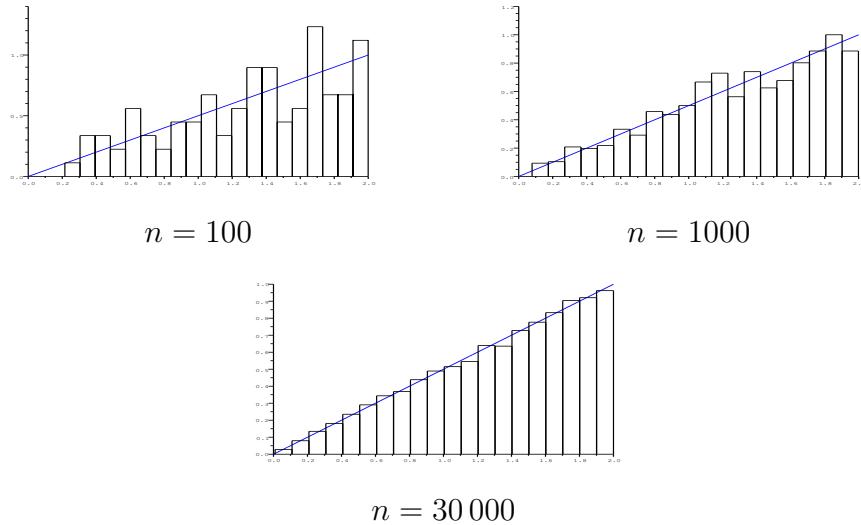
$n = 100000$

La ligne horizontale est à l'altitude 1, c'est la densité théorique de $\mathcal{U}(0, 1)$. Attention, les échelles verticales diffèrent d'une figure à l'autre.

EXEMPLE 11.9. Soit la variable aléatoire X à valeurs dans $[0, 2]$ de densité

$$f_X(x) = \mathbf{1}_{[0,2]}(x)x/2, \quad x \in \mathbb{R}.$$

Sa fonction de répartition vaut $F_X(x) = x^2/4$ pour $0 \leq x \leq 2$ et sa fonction réciproque est $F_X^{-1}(u) = 2\sqrt{u}$, $0 \leq u \leq 1$. De ce fait, avec $U \sim \mathcal{U}(0, 1)$, la variable aléatoire $2\sqrt{U}$ a même loi que X , ce qui s'écrit $X \stackrel{d}{=} 2\sqrt{U}$. Les histogrammes suivants de 100, 1000 et 30 000 copies indépendantes de X ont été obtenus avec `rand`.



On constate à nouveau que plus n est grand, plus l'histogramme est proche du graphe de la densité f_X , qui est ici représenté par le segment de droite oblique d'équation $y = x/2$.

CHAPITRE 12

Convergence des variables aléatoires

CHAPITRE 13

Inégalités de convexité

On s'intéresse ici à un lien entre les probabilités et les fonctions convexes. Les notions de base concernant la convexité sont rappelées à l'Annexe D.

Soient $x, y \in \mathbb{R}^d$ et $0 \leq t \leq 1$. La mesure de probabilité sur \mathbb{R}^d : $(1-t)\delta_x + t\delta_y$ est la loi de $Z_t = \begin{cases} x & \text{avec la probabilité } (1-t) \\ y & \text{avec la probabilité } t \end{cases}$, voir les Remarques 3.7-(2&3) au sujet des variables discrètes à valeurs dans un espace vectoriel. On a $\mathbb{E}(Z_t) = (1-t)x + ty$, de sorte que la définition (D.3) de la convexité de la fonction φ sur la partie convexe C de \mathbb{R}^d se réécrit

$$\varphi(\mathbb{E}Z_t) \leq \mathbb{E}\varphi(Z_t),$$

pour tout $0 \leq t \leq 1$. Cette inégalité est en fait un cas particulier du résultat général énoncé plus bas en (13.4).

LEMME 13.1 (Variable discrète). *Soit X une variable aléatoire discrète à valeurs dans une partie convexe C de \mathbb{R}^d telle que $\mathbb{E}\|X\| < \infty$. Si de plus l'une des propriétés suivantes est satisfaite*

- C est un ouvert
 - C est un fermé
 - X prend un nombre fini de valeurs
- alors, $\mathbb{E}X \in C$.

DÉMONSTRATION. Si X prend un nombre fini de valeurs, $\mathbb{E}X = \sum_{n \in N} p_n x_n$ est une combinaison linéaire finie et on montre par récurrence à l'aide de la définition (D.2) que $\sum_{n \in N} p_n x_n \in C$. Par exemple avec $N = \{1, 2, 3\}$,

$$p_1 x_1 + p_2 x_2 + p_3 x_3 = p_1 x_1 + (p_2 + p_3) \underbrace{\left[\underbrace{\frac{p_2}{p_2 + p_3} x_2 + \frac{p_3}{p_2 + p_3} x_3}_{\in C} \right]}_{\in C}$$

et ainsi de suite pour un nombre fini de valeurs. Lorsque $N = \{1, 2, \dots\}$ est infini, nous avons en posant $\pi_m = \sum_{n=1}^m p_n$, $\mathbb{E}X = \sum_{n \geq 1} p_n x_n = \sum_{n=1}^m p_n x_n + \sum_{n > m} p_n x_n = \pi_m \sum_{n=1}^m \frac{p_n}{\pi_m} x_n + \sum_{n > m} p_n x_n$. Or, $\sum_{n=1}^m \frac{p_n}{\pi_m} x_n \in C$ puisque $\sum_{n=1}^m \frac{p_n}{\pi_m} = 1$, $\lim_{m \rightarrow \infty} \pi_m = 1$ et $\lim_{m \rightarrow \infty} \sum_{n > m} p_n x_n = 0$. Donc, $\mathbb{E}X$ appartient à la fermeture de C dans \mathbb{R}^d .

Si C est fermé, nous venons de montrer que $\mathbb{E}X \in C$.

Si C est ouvert, il est égal à son intérieur. Donc x_1 est dans l'intérieur de C . On en déduit que $\mathbb{E}X = p_1 x_1 + \sum_{n > 1} p_n x_n$ est dans l'intérieur de C ; donc dans C . \square

EXERCICE 13.2. Justifier les dernières lignes de la preuve précédente.

PROPOSITION 13.3 (Inégalité de Jensen). *Soient $\varphi : C \rightarrow \mathbb{R}$ une fonction convexe différentiable sur la partie ouverte convexe C de \mathbb{R}^d et X une variable aléatoire à valeurs*

dans C telle que $\mathbb{E}|\varphi(X)| < \infty$ et $\mathbb{E}\|X\| < \infty$. Alors,

$$(13.4) \quad \varphi(\mathbb{E}X) \leq \mathbb{E}\varphi(X).$$

DÉMONSTRATION. Du fait des hypothèses $\mathbb{E}|\varphi(X)| < \infty$ et $\mathbb{E}\|X\| < \infty$ les espérances que nous considérons sont bien définies. Nous avons avec la Proposition D.5 : $\varphi(x) \geq \varphi(a) + \langle \varphi'(a), x - a \rangle$ pour tous $x, a \in C$. Puisque C est un ensemble convexe, le Lemme 13.1 nous dit que $\mathbb{E}X$ appartient aussi à C . En prenant $a = \mathbb{E}X$ dans l'inégalité précédente, nous obtenons $\varphi(X) \geq \varphi(\mathbb{E}X) + \langle \varphi'(\mathbb{E}X), X - \mathbb{E}X \rangle$. En prenant les espérances, la linéarité et la croissance de l'espérance nous assurent de $\mathbb{E}\varphi(X) \geq \varphi(\mathbb{E}X) + \langle \varphi'(\mathbb{E}X), \mathbb{E}(X - \mathbb{E}X) \rangle = \varphi(\mathbb{E}X)$ puisque $\mathbb{E}(X - \mathbb{E}X) = 0$. Ce qui achève la démonstration. \square

REMARQUES 13.5.

- (1) Le Lemme 13.1 reste vrai pour toute partie convexe C de \mathbb{R}^d . La preuve de cette extension nécessite une étude des propriétés élémentaires des ensembles convexes de \mathbb{R}^d que nous ne ferons pas ici.
- (2) L'inégalité de Jensen reste vraie lorsque la fonction convexe φ n'est pas différentiable et C n'est pas un ouvert. Il suffit pour cela de tenir compte de la remarque (1) précédente et de remplacer $\varphi(x) \geq \varphi(a) + \langle \varphi'(a), x - a \rangle$ par $\varphi(x) \geq \varphi(a) + \langle \lambda, x - a \rangle$ où $\lambda = \varphi(a) + \langle \lambda, u - a \rangle$, avec $\lambda \in \mathbb{R}^d$, est l'équation en $(u, \lambda) \in \mathbb{R}^d \times \mathbb{R}$ d'un hyperplan "tangent" au graphe de φ en a . C'est-à-dire un hyperplan passant par $(a, \varphi(a))$ et tel que le graphe de φ soit entièrement dans le demi-espace "supérieur" délimité par cet hyperplan.

En dimension 1 avec $\varphi(x) = x^2$, on retrouve $\mathbb{E}(X^2) \geq (\mathbb{E}X)^2$, c'est-à-dire $\text{Var}(X) \geq 0$. Avec $\varphi(x) = e^{ax}$, on obtient $\ln \mathbb{E}e^{aX} \geq a\mathbb{E}X$, $a \in \mathbb{R}$.

En appliquant l'inégalité de Jensen à la fonction convexe $\varphi(x) = \|x\|^p$, $x \in \mathbb{R}^d$ avec $p \geq 1$ (voir l'Exercice D.7), on obtient $\|\mathbb{E}X\|^p \leq \mathbb{E}[\|X\|^p]$, $p \geq 1$. Avec $p = 1$, nous avons $\|\mathbb{E}X\| \leq \mathbb{E}\|X\|$ et en regroupant ces résultats :

$$\|\mathbb{E}X\| \leq \mathbb{E}\|X\| \leq \mathbb{E}[\|X\|^p]^{1/p}, \quad p \geq 1.$$

COROLLAIRE 13.6. Soient $0 < p \leq q$ et X une variable aléatoire sur \mathbb{R}^d telle que $\mathbb{E}[\|X\|^q] < \infty$. Alors,

$$\mathbb{E}[\|X\|^p]^{1/p} \leq \mathbb{E}[\|X\|^q]^{1/q}.$$

DÉMONSTRATION. La fonction $\varphi(y) = y^{q/p}$, $y \geq 0$ est convexe puisque $q/p \geq 1$. Avec $Y = \|X\|^p$, nous avons $\|X\|^q = \varphi(Y)$ et avec l'inégalité de Jensen : $\mathbb{E}[\|X\|^p]^{q/p} = \varphi(\mathbb{E}Y) \leq \mathbb{E}\varphi(Y) = \mathbb{E}[\|X\|^{p \times q/p}] = \mathbb{E}[\|X\|^q]$ qui est le résultat annoncé. \square

En particulier, avec $1 = p \leq q$ nous retrouvons $\mathbb{E}\|X\| \leq \mathbb{E}[\|X\|^q]^{1/q}$.

ANNEXE A

Dénombrabilité

Un ensemble est dénombrable si on peut le dénombrer, c'est-à-dire coller un numéro distinct sur chacun de ses éléments. L'ensemble de tous les numéros possibles étant l'ensemble \mathbb{N} des entiers naturels, nous arrivons à la définition abstraite suivante.

DÉFINITION A.1. Un ensemble E est dit dénombrable s'il existe une injection de E dans \mathbb{N} .

REMARQUES A.2.

- (1) Appelons $\iota : E \rightarrow \mathbb{N}$ une telle injection. Alors son application réciproque $\iota^{-1} : \iota(E) \rightarrow E$ est une bijection, c'est l'application qui à tout numéro pris dans $\iota(E) \subset \mathbb{N}$ associe un élément unique de E .
- (2) Bien sûr, tout ensemble fini est dénombrable et \mathbb{N} est dénombrable.
- (3) De même, tout sous-ensemble d'un ensemble dénombrable est dénombrable et par contraposition, tout ensemble contenant une partie non-dénombrable est non-dénombrable.
- (4) Si deux ensembles sont en bijection, ils sont soit dénombrables tous les deux, soit non-dénombrables tous les deux.

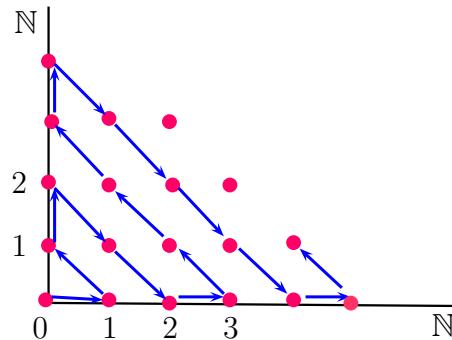
EXERCICE A.3. Montrer que \mathbb{Z} est dénombrable.

SOLUTION. On numérote les entiers relatifs dans l'ordre suivant :

$0, 1, -1, 2, -2, 3, \dots, n, -n, \dots$ Il s'agit de l'application $f : \mathbb{Z} \rightarrow \mathbb{N}_* := \{1, 2, \dots\}$ définie par $f(n) = 2n$ et $f(-n) = 2n + 1$ pour tout $n \geq 1$ et $f(0) = 1$. Elle est bijective de \mathbb{Z} sur \mathbb{N}_* . \square

PROPOSITION A.4. *Le produit cartésien d'un nombre fini d'ensembles dénombrables est dénombrable.*

DÉMONSTRATION. Par récurrence, il suffit de montrer ce résultat pour le produit de deux ensembles dénombrables. Compte tenu de la définition de la dénombrabilité, il suffit pour cela de montrer que \mathbb{N}^2 est dénombrable. Le procédé de numérotation de \mathbb{N}^2 suivant



permet de voir que \mathbb{N}^2 est en bijection avec \mathbb{N} . \square

L'exercice et la proposition précédents nous permettent de voir que pour tout $d \geq 1$, \mathbb{Z}^d est dénombrable. On en déduit que l'ensemble des nombres rationnels \mathbb{Q} est aussi dénombrable. En effet, à tout $x \in \mathbb{Q}$ on associe le couple d'entiers $(p, q) \in \mathbb{Z} \times \mathbb{N}_*$ tels que $x = p/q$ soit une fraction irréductible. Cette application est clairement une injection de \mathbb{Q} dans $\mathbb{Z} \times \mathbb{N}_* \subset \mathbb{Z}^2$ qui est dénombrable.

PROPOSITION A.5. *Une réunion dénombrable d'ensembles dénombrables est dénombrable.*

DÉMONSTRATION. Soient $(E_i)_{i \in I}$ une collection dénombrable (l'ensemble I des indices est dénombrable) d'ensembles dénombrables. On peut sans perte de généralité prendre $I \subset \mathbb{N}$. D'autre part chacun des E_i est en injection dans \mathbb{N} : on peut décrire $E_i = \{x_j^i; j \in J(i)\}$ avec $J(i) \subset \mathbb{N}$. Par conséquent $\bigcup_{i \in I} E_i = \{x_j^i; (i, j) : i \in I, j \in J(i)\}$. L'application qui à tout x de $\bigcup_{i \in I} E_i$ associe un couple (i, j) tel que $x_j^i = x$ est une injection de $\bigcup_{i \in I} E_i$ dans $\{(i, j) : i \in I, j \in J(i)\} \subset \mathbb{N}^2$. Puisque, d'après la Proposition A.4, \mathbb{N}^2 est dénombrable, il en est de même pour $\bigcup_{i \in I} E_i$. \square

Nous allons voir à la Proposition A.8 plus bas qu'aucun intervalle réel d'intérieur non-vide n'est dénombrable. Pour cela nous aurons besoin du résultat préliminaire suivant.

LEMME A.6. *Soit \mathcal{X} un ensemble non vide et $2^\mathcal{X}$ l'ensemble de toutes les parties de \mathcal{X} . Il n'existe pas d'injection de $2^\mathcal{X}$ dans \mathcal{X} .*

DÉMONSTRATION. On fait une preuve par l'absurde. Supposons qu'il existe une injection de $2^\mathcal{X}$ dans \mathcal{X} . Alors, il existe une partie \mathcal{Y} de \mathcal{X} et une application $P : \mathcal{Y} \rightarrow 2^\mathcal{X}$ qui est bijective. L'application P permet de nommer les parties de \mathcal{X} à l'aide des éléments du sous-ensemble \mathcal{Y} de \mathcal{X} .

Considérons la partie

$$A = \{y \in \mathcal{Y}; y \notin P(y)\}$$

ainsi que l'élément $z = P^{-1}(A) \in \mathcal{Y}$.

- Soit $z \in A = P(z)$, mais ceci est impossible par définition de A ;
- Soit $z \notin A = P(z)$ et par définition de A : $z \in P(z)$, ce qui est contradictoire.

Les deux cas sont exclus, par conséquent notre hypothèse de départ est impossible : il n'existe donc aucune injection de $2^\mathcal{X}$ dans \mathcal{X} . \square

Cette preuve est due à Bertrand Russel, philosophe, humaniste et grand mathématicien britannique du XX-ième siècle. Elle est basée sur le paradoxe suivant, énoncé par lui : "Le barbier rase tous les hommes de son village qui ne se rasant pas eux-mêmes".

LEMME A.7. *Soit A un ensemble fini contenant au moins deux éléments.*

- (1) *L'ensemble des suites finies composées d'éléments de A est dénombrable.*
- (2) *L'ensemble $A^\mathbb{N}$ des suites infinies composées d'éléments de A est non-dénombrable.*

On peut voir A comme un alphabet : un ensemble de lettres et toute suite finie comme un mot de taille finie composé avec cet alphabet. Les suites infinies sont des mots de taille infinie. ce sont toutes les applications de \mathbb{N}_* dans A .

DÉMONSTRATION. • Preuve de (1). En notant S_n l'ensemble des suites de longueur n et S_f l'ensemble des suites finies, on a $S_f = \bigcup_{n \geq 1} S_n$ qui est dénombrable d'après la Proposition A.5, puisque réunion dénombrable d'ensembles finis : $\#(S_n) = \#(A)^n < \infty$.

• Preuve de (2). Du fait que $\#(A) \geq 2$, il suffit de montrer que l'ensemble $\{0, 1\}^{\mathbb{N}}$ des suites infinies composées de 0 et de 1 n'est pas dénombrable. En effet, en choisissant deux éléments distincts a_0 et a_1 de A , on voit immédiatement que l'application qui à la suite $(\epsilon_n)_{n \in \mathbb{N}}$ dans $\{0, 1\}^{\mathbb{N}}$ associe la suite $(a_{\epsilon_n})_{n \in \mathbb{N}}$ dans $\{a_0, a_1\}^{\mathbb{N}}$ est une bijection de $\{0, 1\}^{\mathbb{N}}$ sur $\{a_0, a_1\}^{\mathbb{N}}$. C'est donc une injection de $\{0, 1\}^{\mathbb{N}}$ dans $A^{\mathbb{N}}$.

Or $\{0, 1\}^{\mathbb{N}}$ est en bijection avec l'ensemble $2^{\mathbb{N}}$ des parties de \mathbb{N} : à la suite $(\epsilon_n)_{n \in \mathbb{N}}$ on associe la partie $\{n \in \mathbb{N}; \epsilon_n = 1\}$. Mais on a vu au Lemme A.6 que $2^{\mathbb{N}}$ n'est pas dénombrable, donc $\{0, 1\}^{\mathbb{N}}$ ne l'est pas non plus. \square

Nous sommes maintenant en mesure de prouver la

PROPOSITION A.8. *Tout intervalle d'intérieur non-vide (i.e. de la forme (a, b) avec $-\infty \leq a < b \leq +\infty$) est non-dénombrable. En particulier, \mathbb{R} n'est pas dénombrable.*

DÉMONSTRATION. Il suffit de montrer que le segment $[0, 1]$ n'est pas dénombrable. Car alors la bijection $x \in [0, 1] \mapsto \alpha + (\beta - \alpha)x \in [\alpha, \beta]$ nous assure qu'il en est de même pour $[\alpha, \beta]$. Tout intervalle d'intérieur non-vide (a, b) contient un tel segment $[\alpha, \beta]$ (il suffit pour cela que $a < \alpha < \beta < b$) et est de ce fait non-dénombrable.

Montrons que $[0, 1]$ n'est pas dénombrable. Tout $x \in [0, 1]$ admet un développement décimal $x = 0, x_1 x_2 x_3 \dots$ infini (avec éventuellement $x_n = 0$ pour tout n à partir d'un certain rang) où l'on adopte la convention que si le développement se termine par une succession infinie de 9, c'est-à-dire si $x = a_1 \dots a_k 9999 \dots$ avec $0 \leq a_k \leq 8$, on remplace ce développement décimal par $0, a_1 \dots a_{k-1} (a_k + 1) 0000 \dots$ En effet $0, 9999 \dots = 9 \sum_{n \geq 1} (1/10)^n = 9 \frac{1/10}{1-1/10} = 1 = 1, 0000 \dots$ On note $D(x) = (x_1, x_2, \dots) \in \{0, \dots, 9\}^{\mathbb{N}^*}$ la suite correspondant à ce développement décimal unique.

Notons G l'ensemble des suites finies (a_1, \dots, a_k) d'éléments de $\{0, 1, \dots, 9\}$ dont le dernier terme a_k est différent de 9. L'ensemble des x concernés par la modification précédente du développement décimal est l'ensemble des x de la forme $x = a_1 \dots a_k 9999 \dots$. Il est clairement en bijection avec G . Par conséquent, $D : [0, 1] \rightarrow \{0, \dots, 9\}^{\mathbb{N}^*} \setminus G$ est une bijection. Or, d'après la Proposition A.7, $\{0, \dots, 9\}^{\mathbb{N}^*}$ est non-dénombrable et G est dénombrable (en tant que sous-ensemble des suites finies) donc $\{0, \dots, 9\}^{\mathbb{N}^*} \setminus G$ est non-dénombrable et il en est de même pour $[0, 1]$. \square

ANNEXE B

Éléments de théorie de l'intégration

Nous reprenons la notion d'espérance en introduisant (sans preuves) les résultats fondamentaux de la théorie de l'intégrale de Lebesgue.

Notations. Nous avons déjà rencontré les espérances des variables aléatoires discrètes

$$\mathbb{E}(X) = \sum_{x \in \mathcal{X}} x p_X(x)$$

et des variables aléatoires continues

$$\mathbb{E}(X) = \int_{\mathbb{R}} x f_X(x) dx.$$

Dans les deux cas, la fonction de répartition F_X permet le calcul :

$$\begin{aligned} \mathbb{E}(X) &= \sum x \Delta F_X(x) \text{ où } \Delta F_X(x) = F_X(x) - F_X(x^-) \\ \mathbb{E}(X) &= \int_{\mathbb{R}} x dF_X(x) \text{ où } dF_X(x) = f_X(x) dx \end{aligned}$$

Ceci nous suggère la notation unifiée

$$\mathbb{E}(X) = \int_{\mathbb{R}} x dF_X(x).$$

Ainsi, nous obtenons aussi

$$\mathbb{E}(\varphi(X)) = \int \varphi(x) dF_X(x).$$

Intégration abstraite. L'espérance de X est déterminée par la fonction de répartition F_X et puisque F_X est elle-même spécifiée par la donnée de X et de $(\Omega, \mathcal{A}, \mathbb{P})$ on s'attend à ce qu'une notion générale d'espérance de X puisse être définie à partir des données $(\Omega, \mathcal{A}, \mathbb{P})$ et $X : \Omega \rightarrow \mathbb{R}$.

La variable aléatoire $X : \Omega \rightarrow \mathbb{R}$ est dite *simple* si elle prend un nombre fini de valeurs. Les variables simples s'écrivent donc

$$X = \sum_{i=1}^n x_i \mathbf{1}_{A_i}$$

où A_1, \dots, A_n est une partition de Ω . On définit l'intégrale de X , notée $\mathbb{E}(X)$, par

$$\mathbb{E}(X) = \sum_{i=1}^n x_i \mathbb{P}(A_i).$$

Toute variable aléatoire positive $X : \Omega \rightarrow [0, \infty[$ est limite croissante d'une suite $(X_n)_{n \geq 1}$ de variables aléatoires simples. C'est-à-dire : $X_n(\omega) \uparrow X(\omega)$ pour tout $\omega \in \Omega$. On définit alors l'intégrale de X par

$$\mathbb{E}(X) = \lim_{n \rightarrow \infty} \mathbb{E}(X_n) \in [0, \infty].$$

Cette quantité, qui est éventuellement infinie, existe en tant que limite d'une suite croissante et est non-ambigüe : on peut montrer qu'elle ne dépend pas de la suite croissante approximante $(X_n)_{n \geq 1}$.

Pour toute variable aléatoire X , notons pour tout $\omega \in \Omega$,

$$X^+(\omega) = \max(X(\omega), 0) \text{ et } X^-(\omega) = \max(-X(\omega), 0)$$

de sorte $X = X^+ - X^-$ avec $X^+, X^- \geq 0$.

Si $\mathbb{E}(X^+)$ et $\mathbb{E}(X^-)$ ne sont pas infinis simultanément, on définit

$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-) \in [-\infty, +\infty].$$

C'est en particulier le cas lorsque

$$\mathbb{E}(|X|) = \mathbb{E}(X^+ + X^-) < \infty.$$

En théorie de la mesure on note

$$\mathbb{E}(X) = \int_{\Omega} X(\omega) \mathbb{P}(d\omega) = \int_{\Omega} X d\mathbb{P}.$$

L'opération \mathbb{E} est donc un opérateur qui agit sur l'ensemble des variables aléatoires X telles que $\mathbb{E}(|X|) < \infty$. On montre que pour de telles variables aléatoires X, Y et pour tous $a, b \in \mathbb{R}$,

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$$

c'est-à-dire que l'ensemble des variables aléatoires X telles que $\mathbb{E}(|X|) < \infty$ est un espace vectoriel et que \mathbb{E} est une forme linéaire qui agit sur cet espace vectoriel.

Les propriétés de continuité de l'espérance mathématique sont les suivantes.

THÉORÈME B.1 (Théorèmes de continuité de \mathbb{E}). *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires qui converge simplement vers $X : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$, pour tout $\omega \in \Omega$, alors*

(1) (convergence monotone) *si $(X_n)_{n \geq 1}$ est une suite positive et croissante, alors*

$$\lim_{n \rightarrow \infty} \mathbb{E}(X_n) = \mathbb{E}(X) \in [0, \infty];$$

(2) (convergence dominée) *si $|X_n(\omega)| \leq Y(\omega)$, pour tout $\omega \in \Omega$ et $\mathbb{E}(Y) < \infty$, alors*

$$\lim_{n \rightarrow \infty} \mathbb{E}(X_n) = \mathbb{E}(X) \in \mathbb{R};$$

(3) (convergence bornée) *s'il existe $c \in \mathbb{R}$ tel que $|X_n(\omega)| \leq c$, pour tout $\omega \in \Omega$, alors*

$$\lim_{n \rightarrow \infty} \mathbb{E}(X_n) = \mathbb{E}(X) \in \mathbb{R}.$$

La convergence bornée est bien sûr un cas particulier de convergence dominée.

Des conséquences directes du théorème de convergence dominée sont les deux résultats suivants.

THÉORÈME B.2 (Continuité par rapport au paramètre). *Soit $X(t, \omega)$ une fonction sur $\mathbb{R} \times \Omega$ telle que pour tout $t \in \mathbb{R}$, $X(t, \cdot)$ soit \mathbb{P} -intégrable et pour tout ω , $t \mapsto X(t, \omega) \in \mathbb{R}$ soit continue en t_o .*

Si de plus, il existe $\delta > 0$ et une variable aléatoire $Y \geq 0$ telle que $\mathbb{E}(Y) < \infty$ et $\sup_{t \in [t_o - \delta, t_o + \delta]} |X(t, \omega)| \leq Y(\omega)$, pour tout $\omega \in \Omega$, alors

$$t \mapsto \mathbb{E}(X(t, \cdot)) \in \mathbb{R}$$

est continue en t_o .

THÉORÈME B.3 (Dérivation sous le signe somme). *Soient T un ensemble ouvert de \mathbb{R} et $X(t, \omega)$ une fonction sur $T \times \Omega$ telle que pour tout $t \in T$, $X(t, \cdot)$ soit \mathbb{P} -intégrable et pour tout ω , $t \in T \mapsto X(t, \omega) \in \mathbb{R}$ soit dérivable. On note $\frac{d}{dt}X(t, \omega)$ cette dérivée.*

Si de plus, il existe $\delta > 0$ et une variable aléatoire $Y \geq 0$ telle que $\mathbb{E}(Y) < \infty$ et $\sup_{t \in [t_o - \delta, t_o + \delta]} |\frac{d}{dt}X(t, \omega)| \leq Y(\omega)$, pour tout $\omega \in \Omega$, alors

$$G : t \in T \mapsto \mathbb{E}(X(t, \cdot)) \in \mathbb{R}$$

est dérivable en t_o et sa dérivée est donnée par

$$G'(t_o) = \mathbb{E}\left(\frac{d}{dt}X(t, \cdot)\big|_{t=t_o}\right).$$

Intégrale de Lebesgue-Stieltjes. Elle peut apparaître comme le cas particulier de l'intégrale abstraite (de Lebesgue) avec $\Omega = \mathbb{R}$. Plus précisément, soit X une variable aléatoire de fonction de répartition F . On fabrique à partir de F une mesure de probabilité μ_F sur la tribu de Borel de \mathbb{R} comme suit.

- (a) définir $\mu_F([a, b]) = F(b) - F(a)$,
- (b) étendre le domaine de définition de μ_F à la plus petite tribu de \mathbb{R} contenant tous les intervalles : la tribu de Borel \mathcal{B} .

Ainsi, $(\mathbb{R}, \mathcal{B}, \mu_F)$ est un espace de probabilité et

$$\int \varphi d\mu_F$$

est appelée l'intégrale de Lebesgue-Stieltjes de φ par rapport à μ_F . On la note habituellement

$$\int \varphi dF \text{ ou } \int \varphi(x) dF(x).$$

Si X est une variable aléatoire discrète ou continue, on reconnaît alors

$$\mathbb{E}(\varphi(X)) = \int \varphi(x) dF(x).$$

On prend cette égalité comme la définition générale de l'espérance de la variable aléatoire $\varphi(X)$ (que X soit discrète, continue ou autre).

Une notation bien pratique, avec $A \in \mathcal{B}$:

$$\mathbb{E}(\mathbf{1}_{\{X \in A\}} \varphi(X)) = \int_{\Omega} \mathbf{1}_{\{X \in A\}} \varphi(X) d\mathbb{P} = \int_{\{X \in A\}} \varphi(X) d\mathbb{P} = \int_A \varphi(x) dF(x).$$

On remarque en passant que

$$\mathbb{E}(\mathbf{1}_{\{X \in A\}}) = \mathbb{P}(X \in A) = \mu_F(A).$$

ANNEXE C

Espérance mathématique sans théorie de l'intégration

La notion d'espérance mathématique a été introduite sans ambiguïté dans le cadre des variables aléatoires discrètes, voir (6.13). Rappelons que pour tout couple aléatoire discret (X, Y) prenant ses valeurs dans \mathbb{R}^2 et telles que $\sum_{x \in \mathcal{X}} |x| p_X(x) < \infty$ et $\sum_{y \in \mathcal{Y}} |y| p_Y(y) < \infty$, l'espérance mathématique de $aX + bY$ est définie par

$$\mathbb{E}(aX + bY) = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} (ax + by) p_{X,Y}(x, y).$$

Elle possède les propriétés suivantes :

- (C.1) $\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y, a, b \in \mathbb{R}$ (linéarité)
- (C.1') si $X \geq 0, \mathbb{E}X \geq 0$ (positivité)
- (C.1'') $\mathbb{E}(1) = 1$ (normalisation).

Notre but est de construire une extension de l'opérateur : $X \mapsto \mathbb{E}(X)$, à une classe de variables aléatoires X à valeurs réelles plus générale que celle des variables discrètes. Nous allons montrer que lorsqu'on impose à cette extension de satisfaire les propriétés (C.1), elle est unique sur la classe considérée.

Soit $X \mapsto \mathbb{E}(X)$ une extension de l'espérance qui possède les propriétés (C.1). Cet opérateur est croissant au sens où :

$$(C.2) \quad X \leq Y \implies \mathbb{E}(X) \leq \mathbb{E}(Y).$$

En effet, avec (C.1) et (C.1') : $\mathbb{E}(Y) - \mathbb{E}(X) = \mathbb{E}(Y - X) \geq 0$. On en déduit que

$$(C.3) \quad |\mathbb{E}(X)| \leq \mathbb{E}(|X|).$$

Pour décrire la classe sur laquelle l'extension de l'espérance est calculée, nous introduisons l'ensemble fonctionnel suivant.

DÉFINITION C.4. La *classe* Ψ est l'ensemble des fonctions de $]0, 1[$ dans \mathbb{R} qui sont bornées et dont l'ensemble des points de discontinuité est dénombrable et admet un nombre fini de points d'accumulation.

THÉORÈME C.5. Soit $X \mapsto \mathbb{E}(X)$ un opérateur qui prolonge l'espérance mathématique des variables aléatoires discrètes à des variables aléatoires plus générales et qui possède les propriétés (C.1). Soit U une variable aléatoire de loi uniforme sur $[0, 1]$. Alors, pour toute fonction ψ dans Ψ ,

$$\mathbb{E}(\psi(U)) = \int_0^1 \psi(u) du.$$

DÉMONSTRATION. On se replace dans le cadre de la suite des tirages indépendants uniformes sur $\{0, \dots, 9\}$ étudié au Chapitre 10. On considère maintenant les approximations discrètes de U définies pour tout $n \geq 0$ par

$$U_n(\omega) = 0, \omega_1 \dots \omega_n.$$

Cette variable aléatoire est discrète : elle prend chacune des 10^n valeurs $u_{n,k} = 10^{-n}k$, $(0 \leq k \leq 10^n - 1)$ avec la probabilité 10^{-n} . Soit ψ une fonction numérique quelconque sur $]0, 1[$. Son espérance mathématique est

$$\mathbb{E}(\psi(U_n)) = \sum_{0 \leq k \leq 10^n - 1} 10^{-n} \psi(u_{n,k}).$$

Cette somme est l'intégrale de Riemann d'une fonction en escalier qui approxime ψ . On en déduit que si ψ est intégrable au sens de Riemann,

$$(C.6) \quad \lim_{n \rightarrow \infty} \mathbb{E}(\psi(U_n)) = \int_0^1 \psi(u) du.$$

On suppose pour le moment que $\psi :]0, 1[\rightarrow \mathbb{R}$ est continue et bornée. Puisqu'elle admet un prolongement continu sur le compact $[0, 1]$; elle est absolument continue, c'est-à-dire que $w_\psi(\delta) := \sup\{|\psi(u) - \psi(v)|; u, v \text{ tels que } |u - v| < \delta\}$ tend vers zéro lorsque δ décroît vers zéro. D'autre part, puisque $\sup_{n \geq 0} |U - U_n| \leq 10^{-n}$,

$$\begin{aligned} |\mathbb{E}[\psi(U)] - \mathbb{E}[\psi(U_n)]| &= |\mathbb{E}[\psi(U) - \psi(U_n)]| && \text{avec (C.1)} \\ &\leq \mathbb{E}[|\psi(U) - \psi(U_n)|] && \text{avec (C.3)} \\ &\leq \mathbb{E}[w_\psi(\sup_{n \geq 0} |U - U_n|)] && \text{avec (C.2)} \\ &\leq \mathbb{E}[w_\psi(10^{-n})] && \text{avec (C.2)} \\ &= w_\psi(10^{-n}) && \text{avec (C.1) et (C.1'')} \end{aligned}$$

D'où il vient que

$$(C.7) \quad \mathbb{E}[\psi(U)] = \lim_{n \rightarrow \infty} \mathbb{E}[\psi(U_n)].$$

En rapprochant cette identité de (C.6), nous obtenons le résultat désiré lorsque ψ est continue :

$$\mathbb{E}(\psi(U)) = \int_0^1 \psi(u) du.$$

Il reste à étendre cette identité au cas général : $\psi \in \Psi$.

Soit $\psi \in \Psi$. Son ensemble de points de discontinuité est tel que pour tout $\varepsilon > 0$, il existe une réunion finie d'intervalles qui le recouvre, que nous noterons A_ε et dont la longueur totale $|A_\varepsilon|$ est inférieure à ε . Il est clair que la restriction de ψ au complémentaire de A_ε admet un prolongement continu sur $[0, 1]$ (on peut procéder à une série d'interpolations linéaires entre les bornes de A_ε). Notons ψ_ε cette approximation continue de ψ . Puisque ψ est bornée, c'est-à-dire : $\kappa := \sup_{0 \leq u \leq 1} |\psi(u)| < \infty$, on peut choisir ψ_ε de même borne κ que ψ et nous obtenons

$$|\psi(u) - \psi_\varepsilon(u)| \leq 2\kappa \mathbf{1}_{(u \in A_\varepsilon)}, \quad u \in]0, 1[.$$

Par conséquent,

$$\begin{aligned}
\left| \mathbb{E}[\psi(U)] - \mathbb{E}[\psi_\varepsilon(U)] \right| &= \left| \mathbb{E}[\mathbf{1}_{(U \notin A_\varepsilon)}(\psi(U) - \psi_\varepsilon(U))] + \mathbb{E}[\mathbf{1}_{(U \in A_\varepsilon)}(\psi(U) - \psi_\varepsilon(U))] \right| \\
&= \left| \mathbb{E}[\mathbf{1}_{(U \in A_\varepsilon)}(\psi(U) - \psi_\varepsilon(U))] \right| \\
&\leq 2\kappa \mathbb{P}(U \in A_\varepsilon) \\
&= 2\kappa |A_\varepsilon| \\
&\leq 2\kappa \varepsilon
\end{aligned}$$

où l'on a fait usage d'arguments similaires à ceux invoqués lors de la preuve de (C.7), ainsi que de $\mathbb{E}[\mathbf{1}_{(U \in A_\varepsilon)}] = \mathbb{P}(U \in A_\varepsilon)$ ($\mathbf{1}_{(U \in A_\varepsilon)}$ est une variable discrète dont on connaît l'espérance) et de $\mathbb{P}(U \in A_\varepsilon) = |A_\varepsilon|$ (puisque $\mathbb{P}(a \leq U \leq b) = b - a$).

Des arguments analogues nous mènent à

$$\left| \int_0^1 \psi(u) du - \int_0^1 \psi_\varepsilon(u) du \right| \leq 2\kappa \varepsilon,$$

de sorte que pour tout ε ,

$$\begin{aligned}
\left| \mathbb{E}[\psi(U)] - \int_0^1 \psi(u) du \right| &\leq \left| \mathbb{E}[\psi_\varepsilon(U)] - \int_0^1 \psi_\varepsilon(u) du \right| + 4\kappa \varepsilon \\
&= 4\kappa \varepsilon,
\end{aligned}$$

puisque, ψ_ε étant continue, nous avons montré plus haut que $\mathbb{E}[\psi_\varepsilon(U)] = \int_0^1 \psi_\varepsilon(u) du$. La preuve s'achève en faisant tendre ε vers zéro. \square

Nous allons donner plus bas une définition de l'espérance mathématique pour une classe de variables aléatoires continues assez générale. Compte tenu du Théorème 10.3, toute variable aléatoire X admet le même comportement aléatoire (la même loi) que $F_X^{-1}(U)$. Par conséquent, φ étant une fonction numérique, il est loisible d'écrire $\mathbb{E}(\varphi(X)) = \mathbb{E}(\varphi \circ F_X^{-1}(U))$. Le Théorème C.8 plus bas est une conséquence immédiate du Théorème C.5.

Nous sommes en mesure d'énoncer le théorème suivant.

THÉORÈME C.8. *Soit $X \mapsto \mathbb{E}(X)$ un prolongement de l'espérance des variables aléatoires discrètes à une classe plus générale de variables aléatoires qui satisfait les propriétés (C.1). Soit X une variable aléatoire de fonction de répartition F_X et φ une fonction numérique. Si $\varphi \circ F_X^{-1}$ est dans la classe Ψ , alors*

$$\mathbb{E}(\varphi(X)) = \int_0^1 \varphi \circ F_X^{-1}(u) du.$$

C'est en particulier le cas lorsque F_X^{-1} est dans la classe Ψ et φ est bornée et continue par morceaux.

REMARQUE C.9 (Au sujet des points de discontinuité de F_X^{-1}). La fonction F_X^{-1} est croissante et continue à gauche. Nous notons $\langle F_X^{-1}(u) \rangle$ l'intervalle semi-ouvert $[F_X^{-1}(u), F_X^{-1}(u^+)]$. Il est non vide si et seulement si u est un point de discontinuité de F_X^{-1} . Dans ce cas nous disons que $\langle F_X^{-1}(u) \rangle$ est un *intervalle d'absence* de X . Cette terminologie est justifiée par la constatation que lorsque $a_u := F_X^{-1}(u) < F_X^{-1}(u^+) := b_u$, la fonction F_X est plate sur l'intervalle $[a_u, b_u]$, plus précisément : $[a_u, b_u] \subset \{x \in \mathbb{R}; F_X(x) = u\} \subset [a_u, b_u]$. Ceci

implique que $\mathbb{P}(X \in [a_u, b_u]) = 0$, et que pour tout $\alpha > 0$, $\mathbb{P}(X \in]a_u - \alpha, b_u[) > 0$ et $\mathbb{P}(X \in]a_u, b_u + \alpha[) > 0$.

La formule assez générale du Théorème C.8 n'est pas très parlante. Nous allons l'élucider les variables aléatoires continues. Pour une variable aléatoire continue, un intervalle d'absence correspond à un intervalle maximal (composante connexe) de l'ensemble des points d'annulation de f_X . Pour que F_X^{-1} soit dans la classe Ψ , il suffit que X admette un nombre fini d'intervalles d'absence. On en déduit que si l'ensemble $\{x \in \mathbb{R}; f_X(x) = 0\}$ est une réunion finie d'intervalles, la fonction F_X^{-1} est dans la classe Ψ .

Supposons maintenant que X admette une fonction de densité f_X continue par morceaux. Dans ce cas, F_X est partout continue donc $x = F_X^{-1}(u) \iff u = F_X(x)$; de plus, sauf en un nombre fini de points, nous avons $F'_X(x) = f_X(x)$.

La formule de changement de variable dans l'intégrale, nous permet en posant $x = F_X^{-1}(u)$ "d'injecter" $du = F'_X(x)dx = f_X(x)dx$. Ce qui nous donne $\mathbb{E}(\varphi(X)) = \int_0^1 \varphi \circ F_X^{-1}(u) du = \int_{-\infty}^{\infty} \varphi(x) f_X(x) dx$. L'ensemble de ces considérations nous amènent au résultat suivant.

THÉORÈME C.10. *Soit $X \mapsto \mathbb{E}(X)$ un prolongement de l'espérance des variables aléatoires discrètes à une classe plus générale de variables aléatoires qui satisfait les propriétés (C.1). Soit X une variable aléatoire continue dont la densité f_X est continue par morceaux et telle que $\{x \in \mathbb{R}; f_X(x) = 0\}$ est une réunion finie d'intervalles. Soit φ une fonction numérique bornée et continue par morceaux, alors*

$$\mathbb{E}(\varphi(X)) = \int_{\mathbb{R}} \varphi(x) f_X(x) dx.$$

ANNEXE D

Convexité

On se place dans l'espace vectoriel \mathbb{R}^d .

DÉFINITIONS D.1 (Ensemble et fonction convexes). Pour tous $x, y \in \mathbb{R}^d$, on note $[x, y]$ le segment qui relie x et y , c'est-à-dire $[x, y] = \{(1-t)x + ty; 0 \leq t \leq 1\}$.

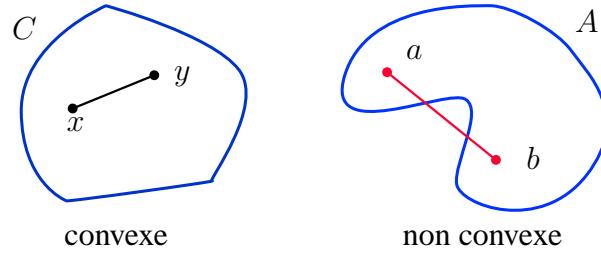
(1) On dit qu'une partie C de \mathbb{R}^d est convexe si

$$(D.2) \quad \forall x, y \in \mathbb{R}^d, x, y \in C \Rightarrow [x, y] \subset C.$$

(2) On dit que la fonction $\varphi : C \rightarrow \mathbb{R}$ est convexe sur l'ensemble convexe C si

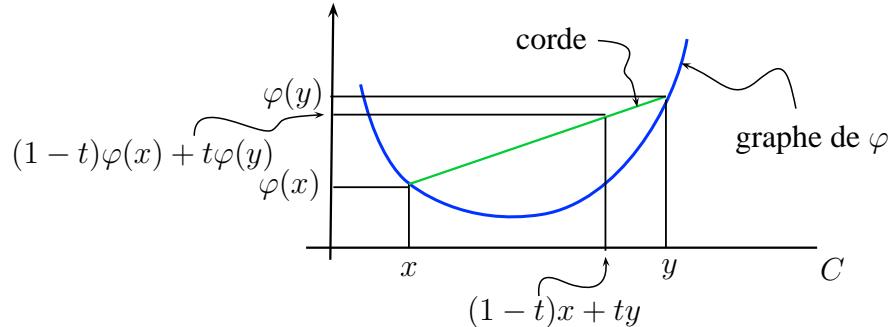
$$(D.3) \quad \forall x, y \in C, \forall 0 \leq t \leq 1, \varphi((1-t)x + ty) \leq (1-t)\varphi(x) + t\varphi(y).$$

Dans la figure suivante, C est une partie convexe du plan alors que A ne l'est pas puisque $[a, b] \not\subset A$ bien que $a, b \in A$:

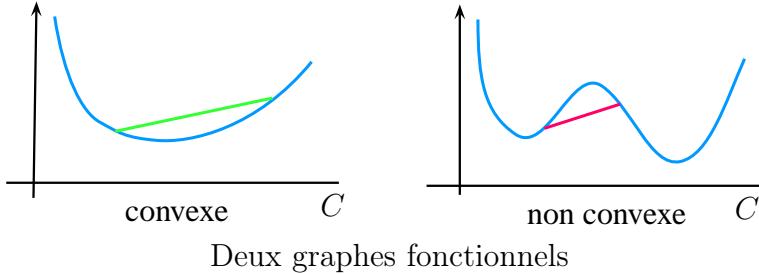


EXERCICE D.4. Montrer que les parties convexes de \mathbb{R} sont les intervalles.

La propriété (D.3) signifie que toutes les cordes liant deux points du graphe de la fonction convexe φ sont situées au-dessus du graphe. C'est ce qu'illustre la figure suivante.



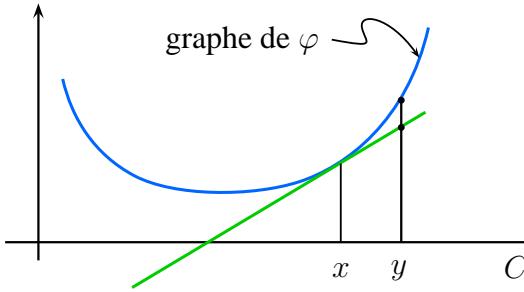
Dans la figure suivante, le graphe de gauche est celui d'une fonction convexe puisque toutes ses cordes sont situées au-dessus, alors que celui de droite est celui d'une fonction non-convexe.



PROPOSITION D.5. Soit $\varphi : C \rightarrow \mathbb{R}$ une fonction dérivable sur une partie ouverte et convexe C de \mathbb{R}^d . Les assertions suivantes sont équivalentes.

- (a) φ est convexe sur C .
 - (b) Pour tous $x, y \in C$, $\varphi(y) \geq \varphi(x) + \langle \varphi'(x), y - x \rangle$
où $\varphi'(x) = (\frac{\partial \varphi}{\partial x_1}(x), \dots, \frac{\partial \varphi}{\partial x_d}(x))$ est le gradient de φ en x et $\langle u, v \rangle$ est le produit scalaire de u et v dans \mathbb{R}^d .
 - (c) Pour tout $x \in I$, $\varphi''(x) \geq 0$.
- Dans le cas particulier où $\mathbb{R}^d = \mathbb{R}$, si de plus φ est une fonction sur un intervalle ouvert $I \subset \mathbb{R}$, deux fois continûment différentiable (de classe C^2), alors les assertions (a) et (b) sont aussi équivalentes à

La partie C est supposée ouverte pour pouvoir définir sans encombre la dérivée de φ . La propriété (b) signifie que le graphe de φ se situe au-dessus de tous ses hyperplans tangents.



DÉMONSTRATION. • Preuve de (a) \Rightarrow (b). (a) exprime que pour tous $x, y \in C$ et tout $0 \leq t \leq 1$, $\varphi(x + t(y - x)) \leq \varphi(x) + t[\varphi(y) - \varphi(x)]$. D'où en prenant $t > 0$, $[\varphi(x + t(y - x)) - \varphi(x)]/t \leq \varphi(y) - \varphi(x)$, et en le faisant tendre vers 0 : $\langle \varphi'(x), y - x \rangle \leq \varphi(y) - \varphi(x)$, c'est-à-dire (b).

• Preuve de (b) \Rightarrow (a). Par l'absurde. Supposons que (b) soit satisfait et que (a) ne le soit pas. Nous allons montrer une contradiction. Puisque (a) n'est pas satisfait, il existe $0 < t < 1$ tel que

$$(D.6) \quad \varphi(x_t) > (1 - t)\varphi(x) + t\varphi(y).$$

L'hyperplan tangent au graphe de φ en $x_t := (1 - t)x + ty$ a pour équation avec les coordonnées $(u, \alpha) \in \mathbb{R}^d \times \mathbb{R}$: $\alpha = \varphi(x_t) + \langle \lambda, u - x_t \rangle$ où $\lambda = \varphi'(x_t) \in \mathbb{R}^d$. Puisque (b) est supposé vrai, nous avons en x :

$$(X) \quad \varphi(x_t) + \langle \lambda, x - x_t \rangle = \varphi(x_t) - \langle \lambda, t(y - x) \rangle \leq \varphi(x)$$

et en y :

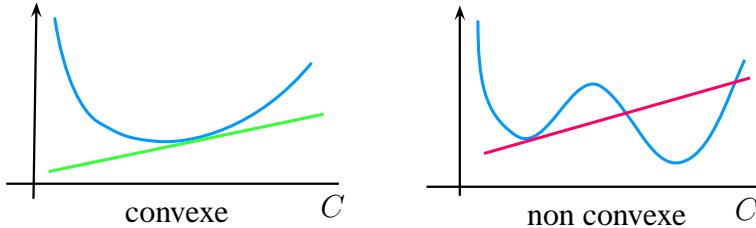
$$(Y) \quad \varphi(x_t) + \langle \lambda, y - x_t \rangle = \varphi(x_t) + \langle \lambda, (1-t)(y-x) \rangle \leq \varphi(y)$$

En faisant $(1-t)(X) + t(Y)$, nous obtenons $\varphi(x_t) \leq (1-t)\varphi(x) + t\varphi(y)$ qui contredit (D.6).

• Preuve de (b) \Rightarrow (c). Prenons $y - x = th$ avec $t > 0$ de sorte que (b) nous donne $\varphi(x + th) - \varphi(x) - \varphi'(x)th \geq 0$. D'autre part, puisque φ est C^2 , il existe $0 \leq \theta \leq 1$ tel que $\varphi(x + th) - \varphi(x) - \varphi'(x)th = \varphi''(x + \theta th)t^2/2$. On en déduit que $\varphi''(x + \theta th) \geq 0$ et en faisant tendre t vers 0, nous obtenons grâce à la continuité de φ'' que $\varphi''(x) \geq 0$.

• Preuve de (c) \Rightarrow (b). Puisque $\varphi'' \geq 0$, φ' est croissante et pour tous $x \leq y$, $\varphi(y) = \varphi(x) + \int_x^y \varphi'(z) dz \geq \varphi(x) + \int_x^y \varphi'(x) dz = \varphi(x) + \varphi'(x)(y - x)$. Lorsque $y \leq x$, $\varphi(y) = \varphi(x) + \int_x^y \varphi'(z) dz = \varphi(x) - \int_y^x \varphi'(z) dz \geq \varphi(x) + \int_y^x \varphi'(x) dz = \varphi(x) + \varphi'(x)(y - x)$. Ce qui prouve (b) et achève la preuve de la proposition. \square

Dans la figure suivante, le graphe de gauche est celui d'une fonction convexe puisque toutes ses tangentes sont situées au-dessous, alors que celui de droite est celui d'une fonction non-convexe.



Deux graphes fonctionnels

EXERCICE D.7. Montrer que les fonctions suivants sont convexes.

(a) $\varphi(x) = ax + b$, $x \in \mathbb{R}$, avec $a, b \in \mathbb{R}$.

(b) $\varphi(x) = |x|^p$, $x \in \mathbb{R}$, avec $p \geq 1$.

(c) $\varphi(x) = -x^p$, $x \in [0, \infty[$, avec $0 \leq p < 1$.

(d) $\varphi(x) = e^{ax}$, $x \in \mathbb{R}$, avec $a \in \mathbb{R}$.

(e) $\varphi(x) = x \ln x - x + 1$, $x > 0$.

(f) $\varphi(x) = -\ln x$, $x > 0$.

(g) $\varphi(x) = \|x\|$, $x \in \mathbb{R}^d$ une norme sur \mathbb{R}^d .

Par exemple, $\|x\| = (x_1^2 + \dots + x_d^2)^{1/2}$ ou $\|x\| = |x_1| + \dots + |x_d|$ ou $\|x\| = \max_{1 \leq i \leq d} |x_i|$.

(h) $\varphi(x) = \psi(\|x\|)$, $x \in \mathbb{R}^d$ où $\|\cdot\|$ est une norme sur \mathbb{R}^d et ψ est une fonction convexe croissante sur $[0, \infty[$.

En particulier, $\varphi(x) = \|x\|^p$, $x \in \mathbb{R}^d$, avec $p \geq 1$.

Index

2^A , ensemble des parties de A , 2
 \mathcal{A} , voir tribu
 $A \setminus B$, différence d'ensembles, 2
 $A \Delta B$, différence symétrique, 2
 $A \sqcup B$, réunion disjointe, 3
 A^c , complémentaire, 1
 δ_x , masse de Dirac en x , Déf. 1.12, 4
 Ω , voir univers
 \mathbb{P} , voir probabilité
complémentaire, voir A^c
différence d'ensembles, voir $A \setminus B$
différence symétrique, voir $A \Delta B$
Dirac (masse de), voir δ_x
espace de probabilité, 4
événement, 1, 3
 certain, 2
 complémentaire, voir A^c
 disjoints, 4
 impossible, 2
 incompatibles, 2
probabilité (mesure de), Déf. 1.9, 3
réalisation, 1
tribu, 3
univers, 1