

Probabilités et Statistique
Notes du Cours 2020–2021

Anton Thalmaier

UNITÉ DE RECHERCHE EN MATHÉMATIQUES
UNIVERSITÉ DU LUXEMBOURG
CAMPUS BELVAL – MAISON DU NOMBRE
L-4364 ESCH-SUR-ALZETTE

Email address: `anton.thalmaier@uni.lu`
URL: `math.uni.lu/thalmaier`

Table des matières

Chapter 1. Phénomènes aléatoires et modèles probabilistes	1
1.1. Espaces probabilisés discrets	1
1.2. Probabilités sur un espace dénombrable	4
1.3. Modèles d'urnes	9
1.4. La loi hypergéométrique et la loi binômiale	12
Chapter 2. Probabilités conditionnelles et indépendance	15
2.1. Espaces de probabilités conditionnels	15
2.2. La ruine du joueur	19
2.3. Événements indépendants	21
2.4. Produits d'espaces de probabilité	22
Chapter 3. Variables aléatoires et espérance	25
3.1. La loi d'une variable aléatoire	25
3.2. Espérance, variance et covariance	26
3.3. Indépendance de variables aléatoires	33
3.4. Approximation poissonnienne	37
3.5. Interprétation de la covariance: Régression simple et multiple	39
Chapter 4. Inégalités et grandes déviations	47
4.1. Les inégalités de Markov, Tchebychev, Bernstein et Hoeffding	47
4.2. La loi faible des grands nombres	50
4.3. Quelques applications	52
Chapter 5. Axiomes probabilités et lois continues	55
5.1. Limitations de modèles discrets	55
5.2. Les axiomes de Kolmogorov (1933)	56
5.3. Variables aléatoires et loi d'une variable aléatoire	58
5.4. Propriétés de la fonction de répartition	60
5.5. Quelques lois classiques	61
5.6. Loi et densité commune	66
Chapter 6. L'espérance de variables aléatoires continues	69
6.1. L'espérance d'une variable continue	69
6.2. La loi d'une somme de variables aléatoires indépendantes	72
6.3. La loi gamma et le paradoxe de l'autobus	75
Chapter 7. La loi du tout ou rien et la loi forte des grands nombres	81
7.1. Lemme de Borel-Cantelli	81
7.2. Convergence de variables aléatoires	84
7.3. La loi forte des grands nombres	85

Chapter 8. Le théorème de la limite centrale	89
8.1. Fonctions caractéristiques	89
8.2. Le Théorème de Moivre-Laplace	92
8.3. Applications	93
Chapter 9. Statistiques: l'estimation et tests d'hypothèses	95
9.1. Modèles statistiques et estimateurs	95
9.2. La loi du chi-carré et de Student	99
9.3. Intervalles de confiance	101
9.4. Tests d'hypothèses	102
Index	107
Notations	109

CHAPTER 1

Phénomènes aléatoires et modèles probabilistes

Stochastique (en grecque et latin: $\sigma\tau\omicron\chi\alpha\zeta\epsilon\sigma\theta\alpha\iota \hat{=} \text{conjecturare}$ ou $\sigma\tau\omicron\chi\alpha\sigma\tau\iota\kappa\eta \tau\acute{\epsilon}\chi\nu\eta \hat{=} \text{ars coniectandi}$) est la théorie des phénomènes aléatoires et a pour but la description mathématique du hasard.

Une *expérience aléatoire* est une expérience dont on ne connaît pas le résultat à l'avance.

1.1. Espaces probabilisés discrets

NOTATION 1.1. Pour une expérience aléatoire on note Ω l'ensemble des *résultats possibles*. Supposons que Ω soit dénombrable.

DÉFINITION 1.2 (distribution de probabilité). Une *distribution de probabilité* sur Ω est une fonction $p: \Omega \rightarrow [0, 1]$ telle que

$$\sum_{\omega \in \Omega} p(\omega) = 1.$$

REMARQUE 1.3. Si Ω est infini dénombrable ($|\Omega| = \infty$) et $\Omega = \{\omega_1, \omega_2, \dots\}$ un dénombrement de Ω , alors la série

$$\sum_{\omega \in \Omega} p(\omega) = \lim_{n \rightarrow \infty} \sum_{i=1}^n p(\omega_i)$$

est indépendante du dénombrement, car $p(\omega) \geq 0$ pour tout $\omega \in \Omega$.

La première étape de la modélisation mathématique d'une expérience aléatoire consiste à spécifier l'ensemble Ω des *résultats possibles* de cette expérience. Quand on parle de choix «au hasard» sur un ensemble fini, on sous-entend souvent que ce choix est fait au moyen de la probabilité uniforme, c.à.d. en donnant à chaque élément de l'ensemble Ω les mêmes chances d'être choisi.

EXEMPLES 1.4.

- (1) Lancer d'une pièce de monnaie: $\Omega = \{P, F\}$ et $p(P) = p(F) = 1/2$.
- (2) Lancer d'un dé: $\Omega = \{1, 2, 3, 4, 5, 6\}$ et $p(i) = 1/6$ pour tout $i \in \Omega$.
- (3) Lancer d'un dé jusqu'à l'obtention d'un 6: $\Omega = \{1, 2, \dots\} = \mathbb{N}^*$, $p(i) = ?$
Naturellement on a

$$p(i) = (1 - p)^{i-1} p \quad \text{pour } i = 1, 2, \dots \quad (\text{avec } p = 1/6).$$

Il faut vérifier que $\sum_{i \geq 1} p(i) = 1$.

En effet,

$$\begin{aligned} \sum_{i \geq 1} p(i) &= \sum_{i \geq 1} (1-p)^{i-1} p \\ &= p \sum_{i \geq 0} (1-p)^i = p \frac{1}{1-(1-p)} = p \cdot \frac{1}{p} = 1. \end{aligned}$$

(4) Lancer deux fois un dé: $\Omega = \{(1,1), (1,2), \dots, (6,6)\}$ et $p(i,j) = 1/36$.

DÉFINITION 1.5 (événement). Soit Ω l'ensemble des résultats possibles d'une expérience aléatoire. On appelle *événement* toute partie A de Ω , c.à.d.,

$$A \subset \Omega \quad \text{ou} \quad A \in \mathcal{P}(\Omega) := \{B : B \subset \Omega\}.$$

EXEMPLE 1.6. (Nombre de clients connectés à un serveur à un instant) On pose $\Omega = \mathbb{N} = \{0, 1, 2, \dots\}$ et

$$A = \{n \in \mathbb{N} \mid 100 \leq n \leq 1000\} \quad \text{ou} \quad B = \{n \in \mathbb{N} \mid n \geq 10.000\}$$

sont des événements.

EXEMPLES 1.7. Dans les Exemples 1.4 on peut regarder les événements suivants:

- (1) « On obtient face »: $A = \{F\}$.
- (2) « On obtient un nombre pair »: $A = \{2, 4, 6\}$.
- (3) « Le premier 6 n'apparaît pas avant le 3-ième coup »: $A = \{3, 4, 5, \dots\}$.
- (4) « Deux fois le même nombre »: $A = \{(1,1), (2,2), \dots, (6,6)\}$.

DÉFINITION 1.8 (espace de probabilité). Soit Ω dénombrable, p une distribution de probabilité sur Ω et $\mathcal{A} := \mathcal{P}(\Omega) = \{A \mid A \subset \Omega\}$ le système des événements. Alors pour $A \in \mathcal{A}$ on appelle

$$\mathbb{P}(A) := \sum_{\omega \in A} p(\omega)$$

la *probabilité* de A . On dit que l'application \mathbb{P} est une *probabilité* sur (Ω, \mathcal{A}) et on appelle le triple $(\Omega, \mathcal{A}, \mathbb{P})$ *espace de probabilité (discret)* ou *espace probabilisé (discret)*.

Attention On a $p: \Omega \rightarrow [0, 1]$, mais $\mathbb{P}: \mathcal{P}(\Omega) \rightarrow [0, 1]$. Donc $p(\omega)$, mais $\mathbb{P}(\{\omega\})$.

EXEMPLE 1.9. (Lancer d'un dé) Soit $\Omega = \{1, 2, \dots, 6\}$. L'événement d'obtenir un nombre pair est représenté par le sous-ensemble $A = \{2, 4, 6\}$. On obtient:

$$\mathbb{P}(A) = p(2) + p(4) + p(6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

En termes d'événements (Terminologie probabiliste)	En termes des ensembles (Terminologie ensembliste)
resultats possibles	$\omega \in \Omega$ (points de Ω)
A, B événements	$A, B \subset \Omega$ (sous-ensembles de Ω)
A et B	$A \cap B$
A ou B	$A \cup B$
non A (contraire de A ; A ne se produit pas)	$A^c = \Omega \setminus A$
A et B sont incompatibles	$A \cap B = \emptyset$
A implique B	$A \subset B$
A l'événement certain	$A = \Omega$
A l'événement impossible	$A = \emptyset = \Omega^c$

THÉORÈME 1.1. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité (discret). Alors on a les deux conditions suivantes:

$$(A1) \quad \mathbb{P}(\Omega) = 1$$

$$(A2) \quad \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) \text{ pour tous } A_1, A_2, \dots \text{ 2 à 2 disjoints dans } \Omega$$

(c.à.d. $A_i \cap A_j = \emptyset$ pour $i \neq j$).

DÉMONSTRATION. La condition (A1) est évidente, car

$$\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} p(\omega) = 1.$$

Pour (A2) on a

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{\omega \in \bigcup_{i=1}^{\infty} A_i} p(\omega) \stackrel{(*)}{=} \sum_{i=1}^{\infty} \sum_{\omega \in A_i} p(\omega) = \sum_{i=1}^{\infty} \mathbb{P}(A_i),$$

où pour (*) on utilise le fait que $A_i \cap A_j = \emptyset$ si $i \neq j$. □

REMARQUE 1.10. Soit $\mathcal{A} = \mathcal{P}(\Omega)$ avec Ω dénombrable. Alors toute application $\mathbb{P}: \mathcal{A} \rightarrow [0, 1]$ satisfaisant les conditions (A1) et (A2) définit une distribution de probabilité p via

$$p(\omega) = \mathbb{P}(\{\omega\}).$$

Autrement dit:

- (1) Toute distribution p de probabilité sur Ω , c.à.d. toute application $p: \Omega \rightarrow \mathbb{R}_+$ telle que $\sum_{\omega \in \Omega} p(\omega) = 1$, définit une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ via

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

- (2) Toute probabilité \mathbb{P} sur (Ω, \mathcal{A}) , c.à.d. toute application $\mathbb{P}: \mathcal{A} \rightarrow \mathbb{R}_+$ avec (A1) et (A2), peut être obtenue de cette façon, à partir de la fonction

$$p(\omega) = \mathbb{P}(\{\omega\}).$$

COROLLAIRE 1.11. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé discret, c.à.d. Ω un ensemble dénombrable, $\mathcal{A} = \mathcal{P}(\Omega)$ et $\mathbb{P}: \mathcal{A} \rightarrow [0, 1]$ avec (A1) et (A2).

- (1) Si $A \subset \Omega$, alors $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$; en particulier, $\mathbb{P}(\emptyset) = \mathbb{P}(\Omega^c) = 1 - \mathbb{P}(\Omega) = 0$.
 (2) On a toujours

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i), \quad \text{et "=" si } A_i \cap A_j = \emptyset \text{ pour } i \neq j.$$

En particulier, lorsque $A_{n+1} = A_{n+2} = \dots = \emptyset$,

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n \mathbb{P}(A_i), \quad \text{et "=" si } A_i \cap A_j = \emptyset \text{ pour } i \neq j.$$

- (3) Si $A \subset B$, alors $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$; en particulier $\mathbb{P}(A) \leq \mathbb{P}(B)$.
 (4) On a $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

DÉMONSTRATION. (1) Pour $A \subset \Omega$ on a

$$\sum_{\omega \in A} p(\omega) + \sum_{\omega \in A^c} p(\omega) = \sum_{\omega \in \Omega} p(\omega), \quad \text{c.à.d. } \mathbb{P}(A) + \mathbb{P}(A^c) = \mathbb{P}(\Omega) = 1.$$

(2) évident

(3) Pour $A \subset B$ on a $B = A \cup (B \setminus A)$ avec $A \cap (B \setminus A) = \emptyset$. En utilisant (2), on trouve $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$.

(4) On a les décompositions disjointes suivantes:

$$A \cup B = (A \setminus B) \cup B \quad \text{et} \quad A = (A \setminus B) \cup (A \cap B),$$

et donc,

$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}(A \setminus B) + \mathbb{P}(B) \quad \text{et} \\ \mathbb{P}(A) &= \mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B). \end{aligned}$$

On obtient (4) par soustraction. □

1.2. Probabilités sur un espace dénombrable

PROBLÈME 1.12. Comment déterminer les probabilités?

(1) *Approche subjective*

$$\text{l'odds ratio} \equiv \frac{\mathbb{P}(A)}{\mathbb{P}(A^c)} = \frac{\mathbb{P}(A)}{1 - \mathbb{P}(A)}.$$

(2) *Approche fréquentielle*

On répète une expérience n fois de façon indépendante:

$$f_n(A) = \frac{\text{nombre de fois où } A \text{ se réalise}}{n} \quad (\text{fréquence empirique de } A).$$

Alors

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

(3) *Approche combinatoire* (e.g. Ω fini)

Modèle de Laplace (loi uniforme sur Ω): $p(\omega) = \text{const} \implies p(\omega) = \frac{1}{|\Omega|}$.

Par conséquent,

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{\text{nombre de cas favorable}}{\text{nombre de cas possible}}, \quad A \subset \Omega.$$

(4) *Approche axiomatique* (Kolmogorov 1933)

On prend (A1) et (A2) dans Théorème 1.1 comme axiomes et on s'intéresse aux conséquences mathématiques.

EXEMPLES 1.13.

(1) *Lancer 2 fois un dé*

On pose

$$\Omega = \{1, 2, \dots, 6\}^2 = \{(i, j) \mid 1 \leq i, j \leq 6\},$$

$\mathcal{A} = \mathcal{P}(\Omega)$, et $\mathbb{P}(\{\omega\}) = 1/|\Omega|$ pour tout $\omega \in \Omega$, autrement dit,

$$\forall A \in \mathcal{A}, \quad \mathbb{P}(A) = |A|/|\Omega|.$$

Alors, on a pour

$$A := \text{“le total fait 4”} = \{(1, 3), (3, 1), (2, 2)\},$$

$$B := \text{“le total fait 6”} = \{(1, 5), (5, 1), (2, 4), (4, 2), (3, 3)\},$$

les probabilités suivantes:

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{3}{36} = \frac{1}{12} \quad \text{et} \quad \mathbb{P}(B) = \frac{5}{36}.$$

Attention La modélisation par une *loi uniforme* sur

$$\Omega := \{\text{les totaux possibles}\} = \{2, 3, \dots, 12\}$$

n'est pas convenable.

(2) *Le paradoxe de Méré* (Chevalier de Méré, 1610–1685)

Discuté dans une correspondance entre Blaise Pascal et Pierre de Fermat (29.7.1654)

(a) Soit $A :=$ “l'apparition d'au moins un 6 en lançant 4 fois un dé”. Alors

$$\mathbb{P}(A) = \frac{6^4 - 5^4}{6^4} = 1 - \left(\frac{5}{6}\right)^4 \approx 0,518.$$

Il est avantageux de parier sur A .

(b) Soit $B :=$ “l'apparition d'au moins un double-six en lançant 24 fois deux dés”.

Alors

$$\mathbb{P}(A) = \frac{36^{24} - 35^{24}}{36^{24}} = 1 - \left(\frac{35}{36}\right)^{24} \approx 0,492.$$

Le deuxième jeu n'est pas avantageux.

(3) *Anniversaire avec n personnes*

Paul donne une fête à la célébration de son anniversaire. On cherche la probabilité que parmi les invités une autre personne soit née le même jour de l'année que Paul. On note $\mathbb{P}(A)$ cette probabilité.

Donc, on a n personnes: Paul et les $n - 1$ invités. Soit

$$\Omega = \{1, 2, \dots, 365\}^{n-1} = \{(i_1, \dots, i_{n-1}) \mid 1 \leq i_k \leq 365, k = 1, \dots, n-1\},$$

et soit i_n l'anniversaire de Paul. Les cas défavorables sont alors donnés par:

$$D = \{(i_1, \dots, i_{n-1}) \mid 1 \leq i_k \leq 365, i_k \neq i_n, k = 1, \dots, n-1\}.$$

On trouve

$$\mathbb{P}(A) = \frac{|\Omega| - |D|}{|\Omega|} = 1 - \left(\frac{364}{365}\right)^{n-1}.$$

$n - 1$	5	10	15	20	40	253
$\mathbb{P}(A)$	0,01	0,03	0,04	0,05	0,1	0,5

(4) *Anniversaire avec n personnes (suite)*

Dans un amphi il y a n étudiants. Quelle est la probabilité qu'au moins deux d'entre eux soient nés le même jour? On note $\mathbb{P}(A_n)$ cette probabilité.

Maintenant on a

$$\Omega = \{1, 2, \dots, 365\}^n$$

(sans restriction: $n \leq 365$, car $\mathbb{P}(A_n) = 1$ pour $n > 365$). Alors on trouve:

$$\begin{aligned} \mathbb{P}(A_n) &= 1 - \mathbb{P}(A_n^c) = 1 - \frac{365 \cdot 364 \cdot \dots \cdot (365 - n + 1)}{365^n} \\ &= 1 - \left(1 - \frac{1}{365}\right) \left(1 - \frac{2}{365}\right) \dots \left(1 - \frac{n-1}{365}\right). \end{aligned}$$

n	5	10	15	20	23	40	50	55
$\mathbb{P}(A_n)$	0,03	0,12	0,25	0,41	0,51	0,89	0,97	0,99

(5) *La voiture et les chèvres: "Ask Marilyn" (Marilyn vos Savant)*

Marilyn vos Savant tient une rubrique dans Parade Magazine qui est un supplément dominical diffusé à plusieurs millions d'exemplaires aux états-unis. Dans cette rubrique, elle répond aux questions des lecteurs.

Question. Il y a trois portes. On note 1 la porte avec la voiture, 2 et 3 les deux autres. Alors le déroulement du jeu est donné par une liste de 4 éléments (u, v, w, x) où

u := la porte choisie par le joueur

v := la porte choisie par l'animateur

w := la porte pour laquelle le joueur change ou bien qu'il garde

x := le resultat: G gain ou P perte.

(a) *Stratégie A:* Ne pas changer.

Les déroulements possibles sont les suivantes:

$$(1, 2, 1, G), \quad (1, 3, 1, G), \quad (2, 3, 2, P), \quad (3, 2, 3, P)$$

Sous l'hypothèse que chaque porte soit choisie avec même probabilité $1/3$, on trouve:

$$\mathbb{P}\{\text{"gain de la voiture"}\} = \frac{1}{3}.$$

(b) *Stratégie B:* Toujours changer.

Les déroulements possibles sont les suivantes:

$$(1, 2, 3, P), \quad (1, 3, 2, P), \quad (2, 3, 1, G), \quad (3, 2, 1, G)$$

On obtient:

$$\mathbb{P}\{\text{"gain de la voiture"}\} = \frac{2}{3}.$$

THÉORÈME 1.2 (Formule de Poincaré). *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé discret, et soient $A_1, \dots, A_n \in \mathcal{A}$. Alors on a la formule suivante:*

$$(1.1) \quad \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{\emptyset \neq I \subset \{1, \dots, n\}} (-1)^{|I|+1} \mathbb{P}\left(\bigcap_{i \in I} A_i\right).$$

Autrement dit:

$$(1.2) \quad \mathbb{P}(A_1 \cup \dots \cup A_n) = \sum_{k=1}^n (-1)^{k+1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}),$$

c.à.d.

$$\begin{aligned} \mathbb{P}(A_1 \cup \dots \cup A_n) &= \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{1 \leq i < j \leq n} \mathbb{P}(A_i \cap A_j) + \\ &+ \sum_{1 \leq i < j < k \leq n} \mathbb{P}(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} \mathbb{P}(A_1 \cap \dots \cap A_n). \end{aligned}$$

DÉMONSTRATION. On démontre cette propriété par récurrence. La formule est triviale pour $n = 1$ et vraie pour $n = 2$:

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2), \quad \text{voir Corollaire 1.11 (4).}$$

Supposons-la vraie à l'ordre n . Alors on a pour toute suite A_1, \dots, A_{n+1} d'éléments de \mathcal{A} ,

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^{n+1} A_i\right) &= \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) + \mathbb{P}(A_{n+1}) - \mathbb{P}\left(\left(\bigcup_{i=1}^n A_i\right) \cap A_{n+1}\right) \\ &= \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) + \mathbb{P}(A_{n+1}) - \mathbb{P}\left(\bigcup_{i=1}^n (A_i \cap A_{n+1})\right). \end{aligned}$$

Il reste à appliquer deux fois l'hypothèse de récurrence:

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^{n+1} A_i\right) &= \left[\sum_{i=1}^n \mathbb{P}(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2}) + \right. \\ &\quad \left. + \sum_{1 \leq i_1 < i_2 < i_3 \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \dots \right] + \mathbb{P}(A_{n+1}) \\ &\quad - \left[\sum_{j=1}^n \mathbb{P}(A_j \cap A_{n+1}) - \sum_{1 \leq j_1 < j_2 \leq n} \mathbb{P}(A_{j_1} \cap A_{j_2} \cap A_{n+1}) + \dots \right. \\ &\quad \left. + (-1)^{n+1} \mathbb{P}(A_1 \cap \dots \cap A_n \cap A_{n+1}) \right]. \end{aligned}$$

En regroupant les termes on obtient:

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^{n+1} A_i\right) &= \sum_{i=1}^{n+1} \mathbb{P}(A_i) - \sum_{1 \leq i_1 < i_2 \leq n+1} \mathbb{P}(A_{i_1} \cap A_{i_2}) + \dots \\ &\quad + (-1)^{n+2} \mathbb{P}(A_1 \cap \dots \cap A_{n+1}). \quad \square \end{aligned}$$

EXEMPLE 1.14. On lance 3 fois un dé et on veut calculer

$$\mathbb{P}\{\text{au moins un } 6\}.$$

Pour cela on note

$$A_i = \text{l'événement que le } "i^{\text{ième}} \text{ jet donne } 6", \quad i = 1, 2, 3.$$

Alors

$$\begin{cases} \mathbb{P}(A_1) = \mathbb{P}(A_2) = \mathbb{P}(A_3) = \frac{1}{6}; \\ \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_2 \cap A_3) = \mathbb{P}(A_1 \cap A_3) = \frac{1}{36}; \\ \mathbb{P}(A_1 \cap A_2 \cap A_3) = \frac{1}{216}, \end{cases}$$

et donc

$$\mathbb{P}(A_1 \cup A_2 \cup A_3) = 3 \cdot \frac{1}{6} - 3 \cdot \frac{1}{36} + \frac{1}{216} = \frac{91}{216}.$$

Autre solution Soit $\Omega = \{1, \dots, 6\}^3$ (donc $|\Omega| = 6^3 = 216$) et

$$A = \{\text{au moins un } 6\}.$$

Alors $A^c = \{1, \dots, 5\}^3$ (donc $|A^c| = 5^3 = 125$), et par conséquent,

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - \frac{5^3}{6^3} = \frac{91}{216}.$$

EXEMPLE 1.15. (*Problème du vestiaire*; Montmort 1708)

Considérons la situation suivante: n personnes laissent leur manteau au vestiaire; lorsqu'elles viennent les chercher, la dame du vestiaire distribue au hasard les n manteaux parmi les n invités. Quelle est la probabilité qu'aucun d'entre eux n'obtienne son propre manteau à la sortie?

On prend pour Ω l'ensemble de permutations de $\{1, \dots, n\}$, c.à.d.

$$\Omega = \{\omega: \{1, \dots, n\} \rightarrow \{1, \dots, n\} \mid \omega \text{ bijection}\},$$

puis $\mathcal{A} = \mathcal{P}(\Omega) := \{A : A \subset \Omega\}$ (tribu des événements) et pour probabilité \mathbb{P} sur (Ω, \mathcal{A}) la lois uniforme, c.à.d.

$$\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|} = \frac{1}{n!}.$$

Pour $i = 1, 2, \dots, n$ soit

$$\begin{aligned} A_i &= \{\omega \in \Omega \mid \omega(i) = i\} \\ &= \{\text{"personne } i \text{ obtient son propre manteau"}\}. \end{aligned}$$

On cherche la probabilité $\mathbb{P}(A)$ où

$$A := \left(\bigcup_{i=1}^n A_i \right)^c = \bigcap_{i=1}^n A_i^c.$$

Par la formule de Poincaré, on obtient

$$\begin{aligned}
 \mathbb{P}(A^c) &= \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) \\
 &= \sum_{k=1}^n (-1)^{k+1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) \\
 &= \sum_{k=1}^n (-1)^{k+1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \frac{(n-k)!}{n!} \\
 &= \sum_{k=1}^n (-1)^{k+1} \binom{n}{k} \frac{(n-k)!}{n!} \\
 &= \sum_{k=1}^n (-1)^{k+1} \frac{n!}{k!(n-k)!} \frac{(n-k)!}{n!} = - \sum_{k=1}^n \frac{(-1)^k}{k!},
 \end{aligned}$$

et par conséquent

$$\begin{aligned}
 \mathbb{P}(A) &= 1 - \mathbb{P}(A^c) \\
 &= 1 + \sum_{k=1}^n \frac{(-1)^k}{k!} \\
 &= \sum_{k=0}^n \frac{(-1)^k}{k!} \xrightarrow{n \rightarrow \infty} e^{-1} \approx 0,3679.
 \end{aligned}$$

1.3. Modèles d'urnes

Situation: Soient n boules dans une urne, numérotées de 1 à n ; on tire k boules au hasard. On note

$$U_n = \{1, \dots, n\}$$

l'urne et $\omega_1, \dots, \omega_k$ avec $\omega_i \in U_n$ un "échantillon de taille k ".

Analyse combinatoire

- (1) L'ordre du choix est-il important ?
 $(\omega_1, \dots, \omega_k)$ choix avec ordre (*arrangement*)
 e.g. $(1, 5, 3, 4) \neq (3, 5, 4, 1)$
- (2) L'ordre du choix est hors de propos
 $[\omega_1, \dots, \omega_k]$ choix sans ordre (*combinaison*)
 e.g. $[1, 5, 3, 4] = [3, 5, 4, 1]$
- (3) Tirage avec remise
 $\omega_1, \dots, \omega_k$ où $\omega_i \in U_n$
- (4) Tirage sans remise
 $\omega_1, \dots, \omega_k$ où $\omega_i \in U_n, \omega_i \neq \omega_j$ pour $i \neq j$

Conclusion: Il faut distinguer 4 cas, c.à.d. pour un échantillon à k éléments parmi un ensemble de n éléments ($k \leq n$) on a 4 types de dénombrement.

- (1) *Arrangements sans répétition* (avec ordre + sans remise)
 Soit $\Omega_1 := \{\omega = (\omega_1, \dots, \omega_k) \mid \omega_i \in U_n = \{1, \dots, n\}, \omega_i \neq \omega_j \text{ pour } i \neq j\}$.
 On a

$$|\Omega_1| = n(n-1) \cdot \dots \cdot (n-k+1) = \frac{n!}{(n-k)!}.$$

Cas particulier $n = k$:

$\Omega_1 = \{\text{permutations de } U_n\} = \{\omega: U_n \rightarrow U_n \mid \omega \text{ bijective}\}$,
et donc $|\Omega_1| = n!$

(2) *Arrangements avec répétition* (avec ordre + avec remise)

$$\Omega_2 := \{\omega = (\omega_1, \dots, \omega_k) \mid \omega_1, \dots, \omega_k \in U_n\}.$$

On a $|\Omega_2| = |U_n|^k = n^k$.

(3) *Combinaison sans répétition* (sans ordre + sans remise)

$$\Omega_3 := \{\omega = [\omega_1, \dots, \omega_k] \mid \omega_i \in U_n = \{1, \dots, n\}, \omega_i \neq \omega_j \text{ pour } i \neq j\}.$$

Considérons sur Ω_1 une relation d'équivalence (égalité mod permutation)

$$(\omega_1, \dots, \omega_k) \sim (\omega_1^*, \dots, \omega_k^*) \Leftrightarrow \exists \text{ permutation } \pi \text{ de } \{1, \dots, k\} \text{ telle que} \\ \omega_i^* = \omega_{\pi(i)} \quad \forall i = 1, \dots, k.$$

Alors

$$\Omega_3 = \Omega_1 / \sim$$

et toute classe d'équivalence a $k!$ éléments, donc

$$|\Omega_1| = k! |\Omega_3| \Rightarrow |\Omega_3| = \frac{n!}{k!(n-k)!} = \binom{n}{k} = C_n^k.$$

En particulier, $\binom{n}{0} = 1$, car $0! = 1$. On remarque que

$$\Omega_3 = \{\omega = (\omega_1, \dots, \omega_k) \mid \omega_1, \dots, \omega_k \in U_n \text{ et } \omega_1 < \dots < \omega_k\}.$$

(4) *Combinaisons avec répétition* (sans ordre + avec remise)

$$\Omega_4 = \{\omega = [\omega_1, \dots, \omega_k] \mid \omega_1, \dots, \omega_k \in U_n\} = \Omega_2 / \sim \\ = \{(\omega_1, \dots, \omega_k) \mid \omega_1, \dots, \omega_k \in U_n, \omega_1 \leq \omega_2 \leq \dots \leq \omega_k\}.$$

Considérons la fonction

$$\begin{array}{ccc} (\omega_1, \dots, \omega_k) & \xrightarrow{\varphi} & (\omega_1^*, \dots, \omega_k^*) \\ \omega_1 \leq \omega_2 \leq \dots \leq \omega_k & & = (\omega_1, \omega_2+1, \omega_3+2, \dots, \omega_k+k-1) \end{array}$$

où $\omega_i^* := \omega_i + i - 1$. Alors, on a une bijection

$$\Omega_4 \xrightarrow[\varphi]{\sim} \Omega_4^* := \{(\omega_1^*, \dots, \omega_k^*) \mid \omega_1^*, \dots, \omega_k^* \in U_n^*, \omega_1^* < \dots < \omega_k^*\}$$

où $U_n^* = \{1, 2, \dots, n+k-1\}$,

$$\stackrel{(3)}{\implies} |\Omega_4| = \binom{n+k-1}{k} = C_{n+k-1}^k.$$

Résumé: Pour un échantillon à k éléments parmi un ensemble de n éléments ($k \leq n$) il existe 4 types de dénombrement.

Choix	sans remise	avec remise
avec ordre (= arrangements)	(1) $\frac{n!}{(n-k)!} = A_n^k$	(2) n^k
sans ordre (= combinaisons)	(3) $\binom{n}{k} = C_n^k$	(4) $\binom{n+k-1}{k}$

EXEMPLE 1.16. $n = 3, k = 2$

Choix	sans remise	avec remise
avec ordre	(1, 2)(1, 3)	(1, 1)(1, 2)(1, 3)
	(2, 1)(2, 3)	(2, 1)(2, 2)(2, 3)
	(3, 1)(3, 2)	(3, 1)(3, 2)(3, 3)
sans ordre	[1, 2][1, 3]	[1, 1][1, 2][1, 3]
	[2, 3]	[2, 2][2, 3]
		[3, 3]

Interprétation alternative

Répartition de k objets dans n boîtes

$$\begin{aligned}
 & 2 \text{ types de placements } \begin{cases} \text{au plus un objet par boîte} \\ \text{plusieurs objets par boîte} \end{cases} \\
 & 2 \text{ types d'objets } \begin{cases} \text{discernable} \\ \text{non discernable} \end{cases}
 \end{aligned}$$

	au plus un objet par boîte	plusieurs objets par boîte
discernable	$\frac{n!}{(n-k)!} = A_n^k$	n^k
non discernable	$\binom{n}{k} = C_n^k$	$\binom{n+k-1}{k}$

EXEMPLE 1.17.

- (1) On jette quatre dés. Trouver la probabilité pour l'événement E que tous les chiffres soient différents. On peut écrire

$$\mathbb{P}(E) = \frac{|E|}{|\Omega|} = \frac{|\text{cas favorable}|}{|\text{cas possible}|}$$

avec

$$\begin{aligned}
 E &= \{(\omega_1, \dots, \omega_4) \mid \omega_i \in \{1, \dots, 6\}, \omega_i \neq \omega_j \text{ pour } i \neq j\} \\
 \Omega &= \{(\omega_1, \dots, \omega_4) \mid \omega_i \in \{1, \dots, 6\}\} = \{1, \dots, 6\}^4,
 \end{aligned}$$

et par conséquent,

$$\begin{aligned}
 |\Omega| &= 6^4 \quad (\text{cas (2) avec } n = 6, k = 4) \\
 |E| &= \frac{6!}{(6-4)!} = A_6^4 \quad (\text{cas (1) avec } n = 6, k = 4),
 \end{aligned}$$

Donc on a:

$$\mathbb{P}(E) = \frac{|E|}{|\Omega|} = \frac{5}{18}.$$

- (2) *Le loto*

On a $U_{49} = \{1, \dots, 49\}$ et

$$\Omega = \{(\omega_1, \dots, \omega_6) \mid \omega_i \in U_{49}, \omega_i \neq \omega_j \text{ pour } i \neq j\},$$

donc

$$|\Omega| = \binom{49}{6} = C_{49}^6 \quad (\text{cas (3) avec } n = 49, k = 6),$$

c.à.d.

$$\mathbb{P}\{\text{"6 chiffres corrects"}\} = \frac{1}{|\Omega|} = \frac{1}{\binom{49}{6}} = \frac{1}{13.983.816}.$$

(3) *L'Eurolotto*

On a $U_{50} = \{1, \dots, 50\}$ et $U_9 = \{1, \dots, 9\}$. Par conséquent,

$$\mathbb{P}\{\text{"Jack-Pot"}\} = \frac{1}{\binom{50}{5} \binom{9}{2}} = \frac{1}{2.118.760 \times 36} = \frac{1}{76.275.360}.$$

1.4. La loi hypergéométrique et la loi binômiale

Problème Une urne contient n boules (m blanches et $n - m$ noires). On tire k boules ($k \leq n$). Soit A_ℓ l'événement que l'on tire exactement ℓ boules blanches (et donc $k - \ell$ boules noires). On utilise un modèle de Laplace.

(1) *La loi hypergéométrique*

Pour un *tirage sans remise* la probabilité de A_ℓ est donnée par

$$\mathbb{P}(A_\ell) = \mathcal{H}(k; m, n)(\ell) := \frac{\binom{m}{\ell} \binom{n-m}{k-\ell}}{\binom{n}{k}}, \quad \ell = 0, \dots, \min(m, k).$$

En effet, on peut supposer qu'on tire simultanément les k boules; donc chaque résultat est une partie à k éléments de $\{1, \dots, n\}$.

Soit Ω l'ensemble de toutes les parties à k éléments de $\{1, \dots, n\}$, muni de la loi uniforme. Rappelons que $|\Omega| = \binom{n}{k}$.

Pour réaliser A_ℓ , il faut tirer

- ℓ boules parmi les m boules blanches: $\binom{m}{\ell}$ possibilités, et
- $k - \ell$ boules parmi les $n - m$ boules noires: $\binom{n-m}{k-\ell}$ possibilités.

On obtient

$$\mathbb{P}(A_\ell) = \frac{\binom{m}{\ell} \binom{n-m}{k-\ell}}{\binom{n}{k}}.$$

(2) *La loi binômiale*

Pour un *tirage avec remise* la probabilité de A_ℓ est donnée par

$$\mathbb{P}(A_\ell) = \mathcal{B}(k, p)(\ell) := \binom{k}{\ell} p^\ell (1-p)^{k-\ell}, \quad \ell = 0, \dots, k, \quad \text{avec } p = \frac{m}{n}.$$

En effet, il y a n^k tirages possibles; il y a $m^\ell (n - m)^{k-\ell}$ possibilités pour tirer l'échantillon

$$\underbrace{(b, b, \dots, b)}_{\ell} \underbrace{(n, \dots, n)}_{k-\ell}.$$

Le même nombre de possibilités existe pour chaque autre échantillon *favorable* à k boules (ℓ fois blanche et $k - \ell$ fois noir). Au total, on a $\binom{k}{\ell}$ possibilités de placer les ℓ boules blanches. On trouve

$$\begin{aligned} \mathbb{P}(A_\ell) &= \frac{\binom{k}{\ell} m^\ell (n - m)^{k-\ell}}{n^k} \\ &= \binom{k}{\ell} \frac{m^\ell}{n^\ell} \frac{(n - m)^{k-\ell}}{n^{k-\ell}} = \binom{k}{\ell} p^\ell (1 - p)^{k-\ell}. \end{aligned}$$

EXEMPLE 1.18. Dans un troupeau comprenant n animaux il y a m animaux malades. On examine un échantillon de k animaux. Quelle est la probabilité que la maladie soit détectée?

On note A_ℓ l'événement "l'échantillon contient exactement ℓ animaux malades". Alors la probabilité cherchée est

$$\begin{aligned} \mathbb{P}\{\text{"la maladie soit détectée"}\} &= 1 - \mathbb{P}(A_0) \\ &= 1 - \mathcal{H}(k; m, n)(0) \\ &= 1 - \frac{\binom{m}{0} \binom{n-m}{k}}{\binom{n}{k}}. \end{aligned}$$

REMARQUE 1.19.

- (1) (*Approximation de loi hypergéométrique par la loi binomiale*)
On a

$$\mathcal{H}(k; m, n)(\ell) = \frac{\binom{m}{\ell} \binom{n-m}{k-\ell}}{\binom{n}{k}} \xrightarrow[\text{tels que } m/n \rightarrow p]{n \rightarrow \infty, m \rightarrow \infty} \binom{k}{\ell} p^\ell (1-p)^{k-\ell} = \mathcal{B}(k; p)(\ell).$$

- (2) (*Approximation poissonnienne de la loi binomiale*)
On a

$$\mathcal{B}(k, p_k)(\ell) = \binom{k}{\ell} p_k^\ell (1-p_k)^{k-\ell} \xrightarrow[\text{tels que } kp_k \rightarrow \lambda > 0]{k \rightarrow \infty, p_k \rightarrow 0} e^{-\lambda} \frac{\lambda^\ell}{\ell!} =: \pi_\lambda(\ell).$$

On note π_λ "la loi de Poisson de paramètre $\lambda > 0$ ".

DÉMONSTRATION. (1) Pour k, ℓ fixe, on a

$$\begin{aligned} \frac{\binom{m}{\ell} \binom{n-m}{k-\ell}}{\binom{n}{k}} &= \frac{m!}{(m-\ell)! \ell!} \frac{(n-m)!}{(n-m-k+\ell)! (k-\ell)!} \\ &= \frac{k!}{(k-\ell)! \ell!} \frac{m!}{(m-\ell)!} \frac{(n-m)!}{(n-m-k+\ell)!} \frac{(n-k)!}{n!} \\ &= \binom{k}{\ell} \frac{m(m-1) \cdots (m-\ell+1)}{n(n-1) \cdots (n-k+1)} \frac{(n-m)(n-m-1) \cdots (n-m-k+\ell+1)}{m^\ell} \\ &= \binom{k}{\ell} \left(\frac{m}{n}\right)^\ell \left(\frac{n-m}{n}\right)^{k-\ell} \frac{m(m-1) \cdots (m-\ell+1)}{m^\ell} \\ &\quad \times \frac{(n-m) \cdots (n-m-k+\ell+1)}{(n-m)^{k-\ell}} \frac{n^k}{n \cdots (n-k+1)} \\ &\rightarrow \binom{k}{\ell} p^\ell (1-p)^{k-\ell}, \quad \text{quand } n, m \rightarrow \infty \text{ tels que } \frac{m}{n} \rightarrow p. \end{aligned}$$

(2) On pose $\lambda_k := kp_k$. Alors on a

$$\begin{aligned}
 B(k, p_k)(\ell) &= \binom{k}{\ell} p_k^\ell (1 - p_k)^{k-\ell} = \frac{k!}{(k-\ell)! \ell!} \left(\frac{\lambda_k}{k}\right)^\ell \left(1 - \left(\frac{\lambda_k}{k}\right)^{\ell-k}\right) \\
 &= \underbrace{\frac{k!}{(k-\ell)! k^\ell}}_{\frac{k}{k} \frac{k-1}{k} \dots \frac{k-\ell+1}{k} \rightarrow 1} \underbrace{\frac{\lambda_k^\ell}{\ell!}}_{\rightarrow \frac{\lambda^\ell}{\ell!}} \underbrace{\frac{1}{\left(1 - \frac{\lambda_k}{k}\right)^\ell}}_{\rightarrow 1} \underbrace{\left(1 - \frac{\lambda_k}{k}\right)^k}_{\rightarrow e^{-\lambda}} \rightarrow \frac{\lambda^\ell}{\ell!} e^{-\lambda}.
 \end{aligned}$$

□

DÉFINITION 1.20.

(1) La *loi de Poisson* de paramètre $\lambda > 0$ est la probabilité \mathbb{P} sur \mathbb{N} définie par

$$p(n) = \frac{e^{-\lambda} \lambda^n}{n!}, \quad n = 0, 1, 2, \dots$$

(2) La *loi géométrique* de paramètre $p \in]0, 1]$ est la probabilité \mathbb{P} sur \mathbb{N}^* définie par

$$p(n) = (1 - p)^{n-1} p, \quad n = 1, 2, \dots$$

Probabilités conditionnelles et indépendance

2.1. Espaces de probabilités conditionnels

EXEMPLE 2.1. On note

Ω = les étudiants du cours “probabilités et statistiques”, et
 A = les étudiants du cours “qui passent l’examen”.

La modélisation par un modèle de Laplace donne

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

On note maintenant

B = les étudiants qui font régulièrement les exercices,

et on trouve pour la “probabilité de A sous la condition B ”

$$\mathbb{P}(A|B) \equiv \frac{|A \cap B|}{|B|} = \frac{|A \cap B|}{|\Omega|} \bigg/ \frac{|B|}{|\Omega|} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

DÉFINITION 2.2. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité discret.

(1) Soient A, B deux événements avec $\mathbb{P}(B) > 0$. On appelle

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

la *probabilité conditionnelle de A sachant B* .

(2) Deux événements A, B sont dits *indépendants* s'ils vérifient

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

REMARQUE 2.3. Sous l'hypothèse que $\mathbb{P}(B) > 0$ deux événements A, B sont indépendants si et seulement si

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

PROPOSITION 2.4. *Pour les probabilités conditionnelles on a les propriétés suivantes:*

(1) $\mathbb{P}(\Omega|B) = 1$

(2) Soient $(A_i)_{i \geq 1}$ 2 à 2 disjoints. Alors

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

DÉMONSTRATION.

$$\begin{aligned}
 \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i \mid B\right) &= \frac{1}{\mathbb{P}(B)} \mathbb{P}\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B\right) \\
 &= \frac{1}{\mathbb{P}(B)} \mathbb{P}\left(\bigcup_{i=1}^{\infty} (A_i \cap B)\right) \\
 &= \frac{1}{\mathbb{P}(B)} \sum_{i=1}^{\infty} \mathbb{P}(A_i \cap B) \\
 &= \sum_{i=1}^{\infty} \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} \\
 &= \sum_{i=1}^{\infty} \mathbb{P}(A_i \mid B). \quad \square
 \end{aligned}$$

COROLLAIRE 2.5. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et $B \in \mathcal{A}$ avec $\mathbb{P}(B) > 0$. Alors $(\Omega, \mathcal{A}, \mathbb{P}(\cdot|B))$ est aussi un espace de probabilité. On appelle $\mathbb{P}(\cdot|B)$ probabilité conditionnelle sachant B .

Attention: On a toujours

$$\begin{aligned}
 \mathbb{P}(A|B) + \mathbb{P}(A^c|B) &= 1, \quad \text{mais} \\
 \mathbb{P}(A|B) + \mathbb{P}(A|B^c) &\neq 1 \quad \text{en général.}
 \end{aligned}$$

EXEMPLE 2.6. Une urne contient n billets de loterie avec m billets gagnants ($n \geq 2$). On tire 2 billets (sans remise) et on note

$$\begin{aligned}
 A &= \text{“le 1}^{\text{er}} \text{ billet est un billet gagnant”} \quad (\Rightarrow \mathbb{P}(A) = \frac{m}{n}), \\
 B &= \text{“le 2}^{\text{ème}} \text{ billet est un billet gagnant”}.
 \end{aligned}$$

Alors on trouve:

$$\begin{aligned}
 \mathbb{P}(B|A) &= \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} = \frac{\frac{m(m-1)}{n(n-1)}}{\frac{m}{n}} = \frac{m-1}{n-1}, \\
 \mathbb{P}(B|A^c) &= \frac{\mathbb{P}(B \cap A^c)}{\mathbb{P}(A^c)} = \frac{\frac{(n-m)m}{n(n-1)}}{\frac{n-m}{n}} = \frac{m}{n-1}.
 \end{aligned}$$

On remarque que

$$\begin{aligned}
 \mathbb{P}(B) &= \mathbb{P}(B|A) \mathbb{P}(A) + \mathbb{P}(B|A^c) \mathbb{P}(A^c) \\
 &= \frac{m-1}{n-1} \frac{m}{n} + \frac{m}{n-1} \frac{n-m}{n} \\
 &= \frac{m}{n} \underbrace{\left(\frac{m-1}{n-1} + \frac{n-m}{n-1}\right)}_{=1} = \frac{m}{n}.
 \end{aligned}$$

EXEMPLE 2.7.

(1) On jette deux dés et on regarde les événements

A_ℓ = “le chiffre du premier dé est égale à ℓ ” et

B_k = “la somme totale est égale à k ”

pour $\ell = 1, \dots, 6$ et $k = 2, \dots, 11$.

Pour $k = 7$, on trouve:

$$\mathbb{P}(A_\ell|B_7) = \frac{\mathbb{P}(A_\ell \cap B_7)}{\mathbb{P}(B_7)} = \frac{1/36}{1/6} = \frac{1}{6} = \mathbb{P}(A_\ell),$$

c.à.d. l’information de plus “la somme totale est 7” ne change pas la probabilité de A_ℓ .

Pour $k = 12$ on trouve:

$$\mathbb{P}(A_\ell|B_{12}) = \begin{cases} 1 & \text{pour } \ell = 6, \\ 0 & \text{pour } \ell \neq 6. \end{cases}$$

(2) A. Une famille a deux enfants dont une fille. Quelle est la probabilité que l’autre enfant soit un garçon? On a

$$\Omega = \{(f, g), (f, f), (g, f), (g, g)\}$$

avec la loi uniforme \mathbb{P} , et on cherche

$$\mathbb{P}(A|B) \quad \text{où } A = \{(f, g), (g, f)\} \text{ et} \\ B = \{(f, g), (f, f), (g, f)\}.$$

Alors

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{2/4}{3/4} = \frac{2}{3}.$$

B. Une autre famille a deux enfants. Le plus jeune est une fille. Quelle est la probabilité P que l’aîné enfant soit un garçon? $P = 1/2!$

THÉORÈME 2.8 (Formule des probabilités totales). Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et (B_n) une partition finie ou dénombrable de Ω , c.à.d.

$$B_n \in \mathcal{A}, \quad B_n \cap B_m = \emptyset \text{ si } n \neq m, \quad \bigcup_n B_n = \Omega.$$

Pour tout $A \in \mathcal{A}$ on a

$$\mathbb{P}(A) = \sum_n \mathbb{P}(A|B_n)\mathbb{P}(B_n),$$

où par définition $\mathbb{P}(A|B_n)\mathbb{P}(B_n) = 0$ si $\mathbb{P}(B_n) = 0$.

DÉMONSTRATION. En utilisant la décomposition

$$A = A \cap \Omega = A \cap \left(\bigcup_n B_n \right) = \bigcup_n (A \cap B_n),$$

on a

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_n (A \cap B_n) \right) = \sum_n \mathbb{P}(A \cap B_n) = \sum_n \mathbb{P}(A|B_n)\mathbb{P}(B_n).$$

□

THÉORÈME 2.9 (Formule de Bayes ou “probabilité des causes”). *Soit (B_n) une partition finie ou dénombrable de Ω . Alors, pour tout $A \in \mathcal{A}$ tel que $\mathbb{P}(A) > 0$, on a*

$$\mathbb{P}(B_n|A) = \frac{\mathbb{P}(A|B_n)\mathbb{P}(B_n)}{\sum_m \mathbb{P}(A|B_m)\mathbb{P}(B_m)}, \quad n = 1, 2, \dots$$

DÉMONSTRATION. On obtient

$$\mathbb{P}(B_n|A) = \frac{\mathbb{P}(B_n \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_n)\mathbb{P}(B_n)}{\mathbb{P}(A)} \stackrel{(*)}{=} \frac{\mathbb{P}(A|B_n)\mathbb{P}(B_n)}{\sum_m \mathbb{P}(A|B_m)\mathbb{P}(B_m)}$$

où pour $(*)$ on utilise la formule des probabilités totales. \square

Interprétation. Soit $\Omega = B_1 \dot{\cup} B_2 \dot{\cup} \dots$ une décomposition de Ω (les B_n jouent la rôle des “hypothèses”). Alors

$\mathbb{P}(B_n) \equiv$ la probabilité *a priori* pour B_n , et

$\mathbb{P}(B_n|A) \equiv$ la probabilité *a posteriori* pour B_n sachant A .

EXEMPLE 2.10 (Examen de dépistage). Une maladie M affecte une personne sur 2000. On dispose d’un test sanguin avec une efficacité de 95% (= probabilité que le test soit positif pour une personne atteinte de M), et une probabilité de fausse alarme de 10% (= probabilité que le test soit positif pour une personne non atteinte). Quelle est la probabilité qu’une personne soit réellement malade lorsqu’elle a un test positif ?

Solution. On cherche la probabilité

$$\mathbb{P}(\text{“personne malade”} | \text{“test positif”}).$$

Pour cela on note

$A =$ “test positif”

$B_1 =$ “personne malade”

$B_2 =$ “de bonne santé”.

Alors $\Omega = B_1 \dot{\cup} B_2$ et on obtient avec la formule de Bayes:

$$\begin{aligned} \mathbb{P}(B_1|A) &= \frac{\mathbb{P}(A|B_1)\mathbb{P}(B_1)}{\mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2)} \\ &= \frac{0,95 \times 0,0005}{0,95 \times 0,0005 + 0,1 \times 0,9995} = 0,00473. \end{aligned}$$

Interprétation. Le test a environ 995 chances sur 1000 d’effrayer inutilement une personne qui a une réaction positive. En plus, 5% de cas de la maladie ne sont pas détectés.

EXEMPLE 2.11 (Discrimination sexuelle à Berkeley; automne 1973).
Parmi les 2900 candidatures masculines 1050 ($\sim 36\%$) ont été rejetées; parmi les 4400 candidatures féminines 2190 ($\sim 50\%$) ont été rejetées.

	masculin			féminin		
	candidaté	rejeté	en %	candidaté	rejeté	en %
Math	800	160	20%	400	40	10%
Physique	700	280	40%	300	90	30%
Chimie	900	270	30%	400	80	20%
Sciences de l'éducation	300	240	80%	2200	1540	70%
Littérature	200	100	50%	1100	440	40%
Total	2900	1050	36%	4400	2190	50%

Observation. On a

$$\forall \text{ matière } S_i, \quad \mathbb{P}(\text{rejeté} \mid S_i + \text{féminin}) < \mathbb{P}(\text{rejeté} \mid S_i + \text{masculin}), \text{ mais} \\ \mathbb{P}(\text{rejeté} \mid \text{féminin}) > \mathbb{P}(\text{rejeté} \mid \text{masculin}).$$

Attention. Il convient de se montrer extrêmement prudent en ce qui concerne les relations de causalité à partir des fréquences relatives et probabilités conditionnelles.

2.2. La ruine du joueur

EXEMPLE 2.12 (La ruine du joueur). Un joueur mise 1 euro sur un événement “ R ” (par exemple, la sortie du rouge à la roulette). On note

$$p = \mathbb{P}(R) \in]0, 1[.$$

Le joueur a k euros, son but est d'augmenter son capital jusqu'à K euros ($0 \leq k \leq K$). Il jouera jusqu'à ce qu'il ait gagné ce capital, ou qu'il soit ruiné.

On pose

$$\begin{aligned} k &= \text{le capital initial,} \\ K &= \text{le capital but,} \\ p_k &= \mathbb{P}\{\text{“atteindre le but } K \text{ avec un capital initial de } k \text{ euros”}\}, \\ q_k &= \mathbb{P}\{\text{“être ruiné avec un capital initial de } k \text{ euros”}\} =: \mathbb{P}(A_k). \end{aligned}$$

Pour $0 < k < K$, on a

$$\begin{aligned} \mathbb{P}(A_k) &= \mathbb{P}(A_k \cap \text{“succès au 1}^{\text{er}} \text{ coup”}) + \mathbb{P}(A_k \cap \text{“échec au 1}^{\text{er}} \text{ coup”}) \\ &= \mathbb{P}(A_k \cap R) + \mathbb{P}(A_k \cap \text{non } R) \\ &= \underbrace{\mathbb{P}(A_k \mid R)}_{= \mathbb{P}(A_{k+1})} \times \underbrace{\mathbb{P}(R)}_{= p} + \underbrace{\mathbb{P}(A_k \mid R^c)}_{= \mathbb{P}(A_{k-1})} \times \underbrace{\mathbb{P}(R^c)}_{= 1-p}. \end{aligned}$$

$$\text{Par conséquent, } \begin{cases} q_k = pq_{k+1} + (1-p)q_{k-1}, & 0 < k < K, \\ q_0 = 1, \quad q_K = 0. \end{cases}$$

Autrement dit, on obtient

$$\left. \begin{aligned} q_2 &= q_1 \frac{1}{p} - q_0 \frac{1-p}{p} \\ q_3 &= q_2 \frac{1}{p} - q_1 \frac{1-p}{p} \\ q_4 &= q_3 \frac{1}{p} - q_2 \frac{1-p}{p} \\ &\dots \end{aligned} \right\} \text{(formules de récurrence),}$$

c.à.d. on peut calculer les q_k sous la condition que q_0 et q_1 seront connus.

Pour le calcul de d_k on pose

$$r := \frac{1-p}{p}, \quad d_k := q_k - q_{k+1}.$$

La relation

$$\underbrace{q_k}_{pq_k + (1-p)q_k} = p q_{k+1} + (1-p) q_{k-1}$$

donne

$$\underbrace{q_k - q_{k+1}}_{= d_k} = \underbrace{\frac{1-p}{p}}_{= r} \underbrace{(q_{k-1} - q_k)}_{= d_{k-1}},$$

d'où

$$d_k = r d_{k-1} = \dots = r^k d_0.$$

On obtient

$$\begin{aligned} 1 &= \underbrace{q_0}_{=1} - \underbrace{q_K}_{=0} \\ &= \sum_{k=0}^{K-1} (q_k - q_{k+1}) = \sum_{k=0}^{K-1} d_k \\ &= \sum_{k=0}^{K-1} r^k d_0 = \begin{cases} K d_0 & \text{si } p = \frac{1}{2} \\ \frac{1-r^K}{1-r} d_0 & \text{si } p \neq \frac{1}{2}. \end{cases} \end{aligned}$$

Autrement dit,

$$(1) \quad d_0 = \begin{cases} 1/K & \text{si } p = \frac{1}{2} \quad (\Leftrightarrow r = 1) \\ \frac{1-r}{1-r^K} & \text{si } p \neq \frac{1}{2} \quad (\Leftrightarrow r \neq 1). \end{cases}$$

De la même façon, on a

$$(2) \quad \begin{aligned} q_k &= q_k - q_K \\ &= \sum_{i=k}^{K-1} \underbrace{(q_i - q_{i+1})}_{=d_i} \\ &= \sum_{i=k}^{K-1} r^i d_0 = \begin{cases} (K-k) d_0 & \text{si } p = \frac{1}{2} \\ \frac{r^k - r^K}{1-r} & \text{si } p \neq \frac{1}{2}. \end{cases} \end{aligned}$$

Par substitution de (1) dans (2) on obtient, pour $0 < k < K$,

$$q_k = \begin{cases} (K - k)/K & \text{si } p = \frac{1}{2} \\ \frac{r^k - r^K}{1 - r^K} & \text{si } p \neq \frac{1}{2}. \end{cases}$$

Pour le calcul de

$$p_k = \mathbb{P}\{\text{“atteindre } K \text{ euros avec un capital initial de } k \text{ euros”}\}$$

on peut faire comme avant en passant,

$$\begin{aligned} p &\mapsto 1 - p \quad (\text{c.à.d. } r \mapsto \frac{1}{r}), \\ k &\mapsto K - k. \end{aligned}$$

On obtient

$$p_k = \begin{cases} \frac{K - (K - k)}{K} = \frac{k}{K} & \text{si } p = \frac{1}{2} \\ \frac{1/r^{K-k} - 1/r^K}{1 - 1/r^K} = \frac{1 - r^k}{1 - r^K} & \text{si } p \neq \frac{1}{2}. \end{cases}$$

Observation On note que pour tout k ,

$$p_k + q_k = 1,$$

c.à.d. le jeu prend fin avec “ruine” ou “succès”.

EXEMPLE 2.13 (Roulette). On a 18 résultats rouge; 18 résultats noir et 2 résultats vert. Le joueur mise sur la sortie du “rouge” (R), c.à.d.

$$p = \frac{18}{38} = 0,4736 \quad \text{et} \quad 1 - p = \frac{20}{38} = 0,5264.$$

Pour $k = 50$, $K = 100$, on trouve

$$p_k = \frac{1 - \left(\frac{20}{18}\right)^{50}}{1 - \left(\frac{20}{18}\right)^{100}} = 0,005127 \quad \text{et} \quad q_k = 0,994873.$$

2.3. Événements indépendants

DÉFINITION 2.14. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité.

- (1) Une famille d'événements $(A_i)_{i \in I}$ est dite *indépendante* si pour toute partie $\emptyset \neq J \subset I$ finie:

$$\mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i),$$

c.à.d. on a $\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_r}) = \mathbb{P}(A_{i_1}) \cdot \dots \cdot \mathbb{P}(A_{i_r})$, $\forall J = \{i_1, \dots, i_r\} \subset I$.

- (2) Une famille $(A_i)_{i \in I}$ est dite *indépendante deux à deux* si

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j) \quad \forall i, j \in I, i \neq j.$$

REMARQUE 2.15.

(1) Indépendance deux à deux $\not\Rightarrow$ indépendance.

Par exemple, on jette un dé deux fois et on considère

$$A = \text{“la somme des points est 7”} \quad (\Rightarrow \mathbb{P}(A) = 1/6)$$

$$B = \text{“le premier jet amène 3 points”} \quad (\Rightarrow \mathbb{P}(B) = 1/6)$$

$$C = \text{“le second jet amène 5 points”} \quad (\Rightarrow \mathbb{P}(C) = 1/6).$$

De plus,

$$\mathbb{P}(A \cap B) = \frac{1}{36}, \quad \mathbb{P}(A \cap C) = \frac{1}{36}, \quad \mathbb{P}(B \cap C) = \frac{1}{36}.$$

Donc, A, B, C sont indépendants deux à deux, mais

$$\underbrace{\mathbb{P}(A \cap B \cap C)}_{=\mathbb{P}(\emptyset)=0} \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).$$

(2) On note que

$$A_1, \dots, A_n \text{ indépendants} \Rightarrow \mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \dots \mathbb{P}(A_n).$$

Mais attention: “ \Leftarrow ” n’est pas vrai en général, e.g.

soient $A_1 = \emptyset, A_2 = A_3$ avec $0 < \mathbb{P}(A_2) < 1$.

2.4. Produits d’espaces de probabilité

Pour tout $1 \leq i \leq n$, soit $(\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$ un espace probabilité (discret). On pose

$$\Omega = \Omega_1 \times \dots \times \Omega_n \quad \text{et} \quad \mathcal{A} = \mathcal{P}(\Omega).$$

Alors la formule

$$\mathbb{P}(\{(\omega_1, \dots, \omega_n)\}) = \mathbb{P}_1(\{\omega_1\}) \dots \mathbb{P}_n(\{\omega_n\})$$

définit une probabilité sur (Ω, \mathcal{A}) .

En fait,

$$\sum_{(\omega_1, \dots, \omega_n) \in \Omega} \mathbb{P}(\{(\omega_1, \dots, \omega_n)\}) = \underbrace{\sum_{\omega_1 \in \Omega_1} \mathbb{P}_1(\{\omega_1\})}_{=1} \dots \underbrace{\sum_{\omega_n \in \Omega_n} \mathbb{P}_n(\{\omega_n\})}_{=1} = 1.$$

DÉFINITION 2.16. On appelle

$$(\Omega, \mathcal{A}, \mathbb{P}) = \bigotimes_{i=1}^n (\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$$

espace probabilisé produit de $(\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$, $i = 1, \dots, n$.

REMARQUE 2.17. L’espace probabilisé produit

$$(\Omega, \mathcal{A}, \mathbb{P}) = \bigotimes_{i=1}^n (\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$$

décrit un *couplage indépendant* d’expériences $(\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$ avec $i = 1, \dots, n$.

Plus précisément: Soient A_i ($i = 1, \dots, n$) des événements de type

$$A_i = \Omega_1 \times \dots \times \Omega_{i-1} \times B_i \times \Omega_{i+1} \times \dots \times \Omega_n \quad \text{avec} \quad B_i \subset \Omega_i, \quad i = 1, \dots, n.$$

Alors la famille (A_1, \dots, A_n) est indépendante.

DÉMONSTRATION. Il faut démontrer que $\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_r}) = \mathbb{P}(A_{i_1}) \dots \mathbb{P}(A_{i_r})$. Pour $i \neq i_1, \dots, i_r$, on pose $B_i = \Omega_i$, c.à.d. $A_i = \Omega_1 \times \dots \times \Omega_n = \Omega$. Alors on a

$$\begin{aligned} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_r}) &= \mathbb{P}(A_1 \cap \dots \cap A_n) = \sum_{\omega_1 \in B_1, \dots, \omega_n \in B_n} \mathbb{P}(\{(\omega_1, \dots, \omega_n)\}) \\ &= \sum_{\omega_1 \in B_1, \dots, \omega_n \in B_n} \mathbb{P}_1(\{\omega_1\}) \cdot \dots \cdot \mathbb{P}_n(\{\omega_n\}) \\ &= \mathbb{P}_1(B_1) \cdot \dots \cdot \mathbb{P}_n(B_n) \\ &= \mathbb{P}(A_1) \cdot \dots \cdot \mathbb{P}(A_n) \\ &= \mathbb{P}(A_{i_1}) \cdot \dots \cdot \mathbb{P}(A_{i_r}). \end{aligned} \quad \square$$

EXEMPLE 2.18. Pour $i = 1, \dots, n$ soit

$$\Omega_i = \{+, -\} \equiv \{\text{succès, échec}\}$$

avec $\mathbb{P}_i(\{+\}) = p$. On pose

$$\Omega = \Omega_1 \times \dots \times \Omega_n.$$

Alors pour $\omega = (\omega_1, \dots, \omega_n) \in \Omega$ avec $\omega_i = +$ où $-$, on a

$$\begin{aligned} \mathbb{P}(\{\omega\}) &= \mathbb{P}_1(\{\omega_1\}) \cdot \dots \cdot \mathbb{P}_n(\{\omega_n\}) \\ &= p^k (1-p)^{n-k}, \quad k := \text{nombre de "+" dans } \omega_1, \dots, \omega_n. \end{aligned}$$

On constate que le couplage décrit une expérience de Bernoulli. Par exemple, si

$$A_k = \text{"}k \text{ fois succès au total"}, \quad k = 0, 1, \dots, n,$$

alors

$$\begin{aligned} \mathbb{P}(A_k) &= \sum_{\omega \in A_k} \mathbb{P}(\{\omega\}) = \sum_{\omega \in A_k} p^k (1-p)^{n-k} \\ &= \binom{n}{k} p^k (1-p)^{n-k} = \mathcal{B}(n, p)(k) \end{aligned}$$

(loi binomiale de paramètres p et n).

EXEMPLE 2.19. Un événement E se produit avec une probabilité très petite (e.g. $p = 1/n$ avec $n \in \mathbb{N}$ grand), par exemple,

$$E = \text{"crash d'un avion de ligne survolant un réacteur nucléaire"}.$$

On répète l'expérience n fois de façon indépendante.

Avec quelle probabilité l'événement se produit au moins une fois? (Réponse $\approx 63\%$)

En effet, soit $\Omega_i = \{+, -\}$ avec

$$\begin{aligned} \text{"+" pour "l'événement se produit", et} \\ \text{"-" pour "l'événement se ne produit pas".} \end{aligned}$$

Alors

$$\mathbb{P}_i(\{\omega_i\}) = p(\omega_i) = \begin{cases} \frac{1}{n}, & \text{si } \omega_i = + \\ 1 - \frac{1}{n}, & \text{si } \omega_i = - \end{cases} \quad (1 \leq i \leq n).$$

Considérons $\Omega = \Omega_1 \times \dots \times \Omega_n$ avec $\mathbb{P} =$ mesure produit de \mathbb{P}_i , c.à.d.

$$\mathbb{P}(\{(\omega_1, \dots, \omega_n)\}) = p(\omega_1) \cdot \dots \cdot p(\omega_n).$$

Soit $A_i = \{(\omega_1, \dots, \omega_n) \in \Omega \mid \omega_i = +\}$ pour $i = 1, \dots, n$. Alors

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) &= 1 - \mathbb{P}\left(\left(\bigcup_{i=1}^n A_i\right)^c\right) = 1 - \mathbb{P}(A_1^c \cap \dots \cap A_n^c) \\ &\stackrel{(*)}{=} 1 - \mathbb{P}(A_1^c) \dots \mathbb{P}(A_n^c) = 1 - \prod_{i=1}^n \underbrace{(1 - \mathbb{P}(A_i))}_{1/n} \\ &= 1 - \left(1 - \frac{1}{n}\right)^n \rightarrow 1 - \frac{1}{e} \approx 0,63. \end{aligned}$$

Pour (*) on utilise (voir feuille 5, Ex 6):

$$(A_1, \dots, A_n) \text{ indépendant} \Rightarrow (A_1^c, \dots, A_n^c) \text{ indépendant.}$$

CHAPTER 3

Variables aléatoires et espérance

3.1. La loi d'une variable aléatoire

DÉFINITION 3.1. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité (discret). On appelle *variable aléatoire* une fonction

$$X: \Omega \rightarrow [-\infty, \infty].$$

Interprétation: “La valeur $X(\omega)$ dépend du hasard $\omega \in \Omega$ ”.

Pour $A \subset [-\infty, \infty]$ on note les événements

$$\begin{aligned} X^{-1}(A) &= \{\omega \in \Omega \mid X(\omega) \in A\} = \{X \in A\} \quad (\text{“}X \text{ prend des valeurs dans } A\text{”}), \\ X^{-1}(\{x\}) &= \{\omega \in \Omega \mid X(\omega) = x\} = \{X = x\} \quad (\text{“}X \text{ prend la valeur } x\text{”}), \\ X^{-1}([-\infty, x]) &= \{\omega \in \Omega \mid X(\omega) \leq x\} = \{X \leq x\} \quad (\text{“}X \text{ prend des valeurs } \leq x\text{”}). \end{aligned}$$

EXEMPLE 3.2.

- (1) Soit $\Omega = \{+, -\} \hat{=} \{\text{succès, échec}\}$ avec $\mathbb{P}(\{+\}) = p$. Considérons la variable aléatoire

$$X: \Omega \rightarrow \{0, 1\}, \quad \text{définie par } X(+):= 1 \text{ et } X(-):= 0.$$

Alors

$$\mathbb{P}(\{X = 1\}) = p = 1 - \mathbb{P}(\{X = 0\}).$$

- (2) Soit $\Omega_i = \{+, -\}$ avec $\mathbb{P}_i(\{+\}) = p$ pour $i = 1, \dots, n$, et

$$(\Omega, \mathcal{A}, \mathbb{P}) = \bigotimes_{i=1}^n (\Omega_i, \mathcal{A}_i, \mathbb{P}_i).$$

Considérons les variables aléatoires

$$X_i: \Omega \rightarrow \{0, 1\}, \quad X_i(\omega) = \begin{cases} 0 & \text{si } \omega_i = - \\ 1 & \text{si } \omega_i = +. \end{cases}$$

Alors

$$X = \sum_{i=1}^n X_i$$

est le nombre de succès dans une expérience de Bernoulli de longueur n , en particulier,

$$\mathbb{P}\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, \dots, n.$$

- (3) (Somme totale à deux jets d'un dé) Soit $\Omega = \{1, \dots, 6\}^2$ avec \mathbb{P} la loi uniforme. Considérons la variable aléatoire

$$X : \Omega \rightarrow \mathbb{R}, \quad X(i, j) = i + j \quad (\text{somme totale}).$$

Alors e.g.

$$\mathbb{P}\{X = 3\} = \mathbb{P}\{(1, 2), (2, 1)\} = \frac{2}{36} = \frac{1}{18},$$

et

$$\mathbb{P}\{X \leq 4\} = \mathbb{P}\{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (3, 1)\} = \frac{1}{6}.$$

DÉFINITION 3.3. Soit X une variable aléatoire définie sur $(\Omega, \mathcal{A}, \mathbb{P})$. Alors

$$\mathbb{P}_X(A) := \mathbb{P}\{X \in A\} \equiv \mathbb{P}\{\omega \in \Omega \mid X(\omega) \in A\}$$

définit une probabilité sur $X(\Omega)$, appelée *la loi* de X ou *la distribution* de X . La loi \mathbb{P}_X est déterminée par les poids

$$\mathbb{P}_X(\{x\}) = \mathbb{P}\{X = x\}, \quad x \in X(\Omega).$$

Rappel. Si Ω est dénombrable, alors $X(\Omega)$ est dénombrable aussi. De plus, on a

$$\Omega = \bigcup_{x \in X(\Omega)} \{X = x\} \quad \text{2 à 2 disjoints,}$$

et donc

$$1 = \mathbb{P}(\Omega) = \sum_{x \in X(\Omega)} \mathbb{P}\{X = x\} = \sum_{x \in X(\Omega)} \mathbb{P}_X(\{x\}).$$

Notation On dit que

- X est *binomiale* (de paramètres n et p) si $\mathbb{P}_X = \mathcal{B}(n, p)$, c.à.d.

$$\mathbb{P}(\{X = k\}) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, \dots, n.$$

- X est de *Poisson* (de paramètre λ) si $\mathbb{P}_X = \mathcal{P}(\lambda)$, c.à.d.

$$\mathbb{P}(\{X = n\}) = e^{-\lambda} \frac{\lambda^n}{n!} \quad (n = 0, 1, 2, \dots),$$

etc.

3.2. Espérance, variance et covariance

DÉFINITION 3.4. On appelle espérance de X le nombre

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\}) = \sum_{\omega \in \Omega} X(\omega) p(\omega),$$

lorsqu'il est bien défini.

REMARQUE 3.5 (précision de la définition). On utilise que $\mathbb{E}[X]$ est toujours bien défini si $X \geq 0$ ($\infty \cdot 0 = 0$), le cas $\mathbb{E}[X] = +\infty$ n'est pas exclu. En général, on écrit $X = X_+ - X_-$ où

$$X_+ = \begin{cases} X(\omega) & \text{si } X(\omega) \geq 0 \\ 0 & \text{sinon;} \end{cases}$$

$$X_- = \begin{cases} -X(\omega) & \text{si } X(\omega) \leq 0 \\ 0 & \text{sinon.} \end{cases}$$

Alors $\mathbb{E}[X] = \mathbb{E}[X_+] - \mathbb{E}[X_-]$, si $\mathbb{E}[X_+] < \infty$ ou $\mathbb{E}[X_-] < \infty$.

EXEMPLE 3.6. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ une expérience de Laplace, c.à.d. $\mathbb{P}(A) = |A|/|\Omega|$. Alors

$$\mathbb{E}[X] = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} X(\omega).$$

Donc $\mathbb{E}[X]$ est la moyenne arithmétique des valeurs $X(\omega)$, $\omega \in \Omega$.

REMARQUE 3.7.

(1) L'espérance $\mathbb{E}[X]$ dépend seulement de la loi \mathbb{P}_X de X , plus précisément,

$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} x \mathbb{P}_X(\{x\}) = \sum_{x \in X(\Omega)} x \mathbb{P}\{X = x\}.$$

En fait, sans restriction $X \geq 0$. Alors on a

$$\begin{aligned} \mathbb{E}[X] &= \sum_{\omega \in \Omega} X(\omega) p(\omega) \\ &= \sum_{x \in X(\Omega)} \sum_{\omega: X(\omega)=x} X(\omega) p(\omega) \quad (\text{changement d'ordre de sommation}) \\ &= \sum_{x \in X(\Omega)} x \sum_{\omega: X(\omega)=x} p(\omega) = \sum_{x \in X(\Omega)} x \mathbb{P}\{X = x\}. \end{aligned}$$

(2) L'espérance est linéaire et positive, c.à.d.

$$\mathbb{E} \left[\sum_{i=1}^n a_i X_i \right] = \sum_{i=1}^n a_i \mathbb{E}[X_i], \quad a_i \in \mathbb{R},$$

et un opérateur positif, c.à.d.

$$X \geq 0 \implies \mathbb{E}[X] \geq 0.$$

REMARQUE 3.8. Soit X une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ et

$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} x \mathbb{P}\{X = x\}.$$

(1) $\mathbb{E}[X]$ existe (dans $\mathbb{R} \cup \{\pm\infty\}$) $\iff \mathbb{E}[X_+] < \infty$ ou $\mathbb{E}[X_-] < \infty$.

(2) $\mathbb{E}|X| < \infty \iff \mathbb{E}[X]$ existe dans \mathbb{R} , c.à.d. $\mathbb{E}[X_+] < \infty$ et $\mathbb{E}[X_-] < \infty$.

(3) On a toujours

$$\mathbb{E}|X| = \mathbb{E}[X_+] + \mathbb{E}[X_-].$$

(4) $\mathbb{E}[X^2] < \infty \implies \mathbb{E}|X| < \infty$, car $0 \leq 2|X| \leq X^2 + 1$.

EXEMPLE 3.9.

- (1) Soit $X = \mathbf{1}_A$ = l'indicatrice d'un événement A . Rappelons que par définition

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

Alors on a

$$\mathbb{E}[X] = 1 \cdot \mathbb{P}(A) + 0 \cdot \mathbb{P}(A^c) = \mathbb{P}(A).$$

- (2) (Contrat d'assurance simple) Soit A le sinistre couvert par un contrat. Alors la variable aléatoire

$$X(\omega) = c \mathbf{1}_A(\omega) = \begin{cases} c & \text{si } \omega \in A \\ 0 & \text{sinon} \end{cases}$$

décrit "l'indemnisation d'assurance". On cherche la contrepartie de l'assuré (prime d'assurance). Une *prime équitable* est donnée par

$$\mathbb{E}[X] = c \cdot \mathbb{P}(A).$$

- (3) À un contrat d'assurance de type

$$X = \sum_{i=1}^n c_i \mathbf{1}_{A_i}$$

correspond la prime équitable suivante

$$\mathbb{E}[X] = \sum_{i=1}^n c_i \mathbb{P}(A_i).$$

EXEMPLE 3.10.

1. Soit X de Poisson de paramètre λ , c.à.d.

$$\mathbb{P}\{X = n\} = \frac{\lambda^n}{n!} e^{-\lambda}, \quad n = 0, 1, 2, 3, \dots$$

Alors

$$\begin{aligned} \mathbb{E}[X] &= \sum_{n=0}^{\infty} n \mathbb{P}\{X = n\} = \sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda} = \lambda e^{-\lambda} \underbrace{\sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!}}_{= \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = e^\lambda} = \lambda. \end{aligned}$$

2. Soit X est de Bernoulli de paramètre $p \in [0, 1]$, c.à.d.

$$\mathbb{P}\{X = 1\} = p = 1 - \mathbb{P}\{X = 0\}.$$

Alors

$$\mathbb{E}[X] = 1 \cdot \mathbb{P}\{X = 1\} + 0 \cdot \mathbb{P}\{X = 0\} = p.$$

3. Soit X binomiale de paramètre p et n , c.à.d.

$$\mathbb{P}\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

Alors

$$\mathbb{E}[X] = \sum_{k=0}^n k \mathbb{P}\{X = k\} = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} \stackrel{(*)}{=} np.$$

Il reste à vérifier (*): On réalise n expériences successives pour une variable de Bernoulli (avec le même paramètre p). Soit

X_i = le résultat de la $i^{\text{ième}}$ expérience,

c.à.d.

$$X_i = \begin{cases} 1 & \text{si succès,} \\ 0 & \text{sinon.} \end{cases}$$

Alors

$$X = X_1 + \dots + X_n$$

est $\mathcal{B}(p; n)$, et donc

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i] = \sum_{i=1}^n \mathbb{P}\{X_i = 1\} = \sum_{i=1}^n p = np.$$

4. Supposons que X suit une loi géométrique de paramètre $p \in]0, 1]$, c.à.d.

$$\mathbb{P}\{X = n\} = (1 - p)^{n-1}p, \quad n = 1, 2, 3, \dots$$

On réalise des expériences de Bernoulli (avec le même paramètre p). Soit X le numéro de l'expérience au cours de laquelle l'événement se réalise pour la première fois. Alors

$$\mathbb{E}[X] = \sum_{n=1}^{\infty} n \mathbb{P}\{X = n\} = p \sum_{n=1}^{\infty} n(1 - p)^{n-1} = \frac{1}{p}.$$

En fait, pour $|x| < 1$, on a

$$f(x) = \sum_{n=0}^{\infty} x^n = \frac{1}{1 - x} \quad (\text{série géométrique}), \text{ et}$$

$$f'(x) = \sum_{n=1}^{\infty} n x^{n-1} = \frac{1}{(1 - x)^2}.$$

Avec $x = 1 - p$, on obtient

$$\sum_{n=1}^{\infty} n(1 - p)^{n-1} = \frac{1}{p^2}.$$

5. (La loi hypergéométrique)

Une urne contient N boules (par exemple, N poissons dans un lac, N électeurs avant une élection, ...). Parmi les N boules, il y a

- K boules rouges (e.g. K poissons marqués) et
- $N - K$ boules noires (e.g. $N - K$ poissons non-marqués).

On tire un échantillon de n boules (sans remise). Soit X le nombre de boules rouges dans l'échantillon. On a

$$\mathbb{P}\{X = k\} = \mathcal{H}(n; K, N)(k) = \frac{\binom{K}{k} \binom{N - K}{n - k}}{\binom{N}{n}}, \quad k = 0, \dots, \min(n, k).$$

Alors

$$\mathbb{E}[X] = \sum_{k=0}^{\min(n,K)} k \mathbb{P}\{X = k\} = \sum_{k=0}^{\min(n,K)} k \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}} = ?$$

Soit

$$S = S_{\text{rouge}} \cup S_{\text{noir}}$$

l'ensemble de boules et soit

$$\Omega = \{\omega = (\omega_1, \dots, \omega_n) \mid \omega_i \in S, \omega_i \neq \omega_j \ (i \neq j)\}$$

avec \mathbb{P} la loi uniforme sur Ω . On note

$$X_i(\omega) = \begin{cases} 1 & \text{si } \omega_i \in S_{\text{rouge}} \quad (\text{“la } i^{\text{ème}} \text{ boule tirée est rouge”}) \\ 0 & \text{sinon.} \end{cases}$$

Pour déterminer $\mathbb{P}\{X_i = 1\}$, on remarque que

$$|\Omega| = N(N-1)\dots(N-n+1) = \frac{N!}{(N-n)!}$$

et

$$\begin{aligned} |\{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in S_{\text{rouge}}\}| &= K \frac{(N-1)!}{((N-1)-(n-1))!} \\ &= K \frac{(N-1)!}{(N-n)!}. \end{aligned}$$

On a donc

$$\mathbb{P}\{X_i = 1\} = K \frac{(N-1)!}{((N-1)-(n-1))!} \bigg/ \frac{N!}{(N-n)!} = \frac{K}{N},$$

et puis

$$\mathbb{E}[X_i] = \mathbb{P}\{X_i = 1\} = \frac{K}{N}.$$

Comme la variable

$$X := X_1 + \dots + X_n$$

suit la loi $\mathbb{P}_X = \mathcal{H}(n; K, N)$, on trouve que

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i] = n \frac{K}{N}.$$

THÉORÈME 3.11. *Soit X une variable aléatoire et $f: X(\Omega) \rightarrow \mathbb{R}$ une fonction. Rappelons que*

$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} x \mathbb{P}\{X = x\} \quad (\text{lorsqu'il est bien défini}).$$

Pour la variable aléatoire $Y = f(X)$ l'on a

$$\mathbb{E}[Y] = \sum_{x \in X(\Omega)} f(x) \mathbb{P}\{X = x\} \quad (\text{lorsqu'il est bien défini}).$$

DÉMONSTRATION. On obtient

$$\begin{aligned}
 \mathbb{E}[Y] &= \sum_y y \mathbb{P}\{Y = y\} = \sum_y y \mathbb{P}\{f(X) = y\} \\
 &= \sum_y y \sum_{x:f(x)=y} \mathbb{P}\{X = x\} \\
 &= \sum_y \sum_{x:f(x)=y} f(x) \mathbb{P}\{X = x\} \\
 &= \sum_x f(x) \mathbb{P}\{X = x\}. \quad \square
 \end{aligned}$$

Plus généralement: Soient X_1, \dots, X_n des variables aléatoires sur $(\Omega, \mathcal{A}, \mathbb{P})$ et soit

$$f : X_1(\Omega) \times \dots \times X_n(\Omega) \longrightarrow \mathbb{R}$$

une fonction. Alors $Y = f(X_1, \dots, X_n)$ est aussi une variable aléatoire et

$$\mathbb{E}|Y| < \infty \iff \sum_{x_1 \in X_1(\Omega), \dots, x_n \in X_n(\Omega)} |f(x_1, \dots, x_n)| \mathbb{P}\{X_1 = x_1, \dots, X_n = x_n\} < \infty.$$

Dans ce cas, l'on a

$$\mathbb{E}[Y] = \sum_{x_1 \in X_1(\Omega), \dots, x_n \in X_n(\Omega)} f(x_1, \dots, x_n) \mathbb{P}\{X_1 = x_1, \dots, X_n = x_n\}.$$

DÉFINITION 3.12.

- (1) Soit X une variable aléatoire telle que $\mathbb{E}|X| < \infty$ et soit $n \in \mathbb{N}$. On appelle
- $\mathbb{E}[X^n]$ *moment d'ordre n* de X , et
 - $\mathbb{E}[(X - \mathbb{E}[X])^n]$ *moment centré d'ordre n* de X (lorsque bien définis).

En particulier,

$$\text{var}(X) \equiv \sigma^2(X) := \mathbb{E}[(X - \mathbb{E}[X])^2]$$

est appelée *variance* de la variable aléatoire X , et

$$\sigma(X) = \sqrt{\text{var}(X)}$$

l'*écart type* de X .

- (2) Soient X et Y des variables aléatoires telles que $\mathbb{E}|X| < \infty$ et $\mathbb{E}|Y| < \infty$. On appelle *covariance* de X et Y le nombre

$$\text{cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \quad (\text{lorsque bien défini}),$$

et

$$\rho(X, Y) := \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

coefficient de corrélation de X et Y (avec la convention $\rho(X, Y) := 0$ si $\sigma(X) = 0$ ou $\sigma(Y) = 0$).

PROPOSITION 3.13 (Inégalité de Jensen). *Soit X une variable aléatoire telle que $\mathbb{E}|X| < \infty$ et soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction convexe, c.à.d.*

$$\forall x_0 \in \mathbb{R} \exists \lambda(x_0) \in \mathbb{R}, \quad f(x) \geq f(x_0) + \lambda(x_0)(x - x_0) \quad \forall x \in \mathbb{R}.$$

Alors

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)].$$

DÉMONSTRATION. Avec $x_0 = \mathbb{E}[X]$ on a

$$f(X) \geq f(\mathbb{E}[X]) + \lambda(\mathbb{E}[X])(X - \mathbb{E}[X]),$$

et puis

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X]) + \lambda(\mathbb{E}[X]) \underbrace{\mathbb{E}[X - \mathbb{E}[X]]}_{=\mathbb{E}[X] - \mathbb{E}[X] = 0}. \quad \square$$

Rappel Si $f: \mathbb{R} \rightarrow \mathbb{R}$ est de classe C^2 , alors f est convexe si $f'' \geq 0$. Dans ce cas, on a $\lambda(x_0) = f'(x_0)$.

EXEMPLE 3.14. On a toujours $(\mathbb{E}|X|)^2 \leq \mathbb{E}[X^2]$.

PROPOSITION 3.15. Soit X une variable aléatoire telle que $\mathbb{E}|X| < \infty$. On note

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

$$\sigma(X) = \sqrt{\text{var}(X)}.$$

(1) On a

$$\text{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

En particulier, $\text{var}(X) < \infty \iff \mathbb{E}[X^2] < \infty$. En fait,

$$\begin{aligned} \text{var}(X) &= \mathbb{E}[(X - \mu)^2] \quad \text{où } \mu = \mathbb{E}[X] \\ &= \mathbb{E}[X^2 - 2\mu X + \mu^2] \\ &= \mathbb{E}[X^2] - 2\mu \mathbb{E}[X] + \mu^2 \\ &= \mathbb{E}[X^2] - \mu^2. \end{aligned}$$

(2) Pour tout $a, b \in \mathbb{R}$, on a

$$\text{var}(aX + b) = a^2 \text{var}(X).$$

En particulier,

$$\sigma(aX + b) = |a| \sigma(X).$$

(3) On a $\text{var}(X) = 0$ si et seulement si X est constante presque sûrement, c.à.d.

$$\mathbb{P}\{X = c\} = 1 \quad \text{avec } c = \mathbb{E}[X].$$

REMARQUE 3.16. Soient X et Y des variables aléatoires telles que $\mathbb{E}|X| < \infty$ et $\mathbb{E}|Y| < \infty$. La covariance

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

est bien définie si en plus $\mathbb{E}|XY| < \infty$. On dit que:

$$X, Y \text{ corrélées positivement} \iff \text{cov}(X, Y) > 0 \iff \rho(X, Y) > 0,$$

$$X, Y \text{ corrélées négativement} \iff \text{cov}(X, Y) < 0 \iff \rho(X, Y) < 0,$$

$$X, Y \text{ non-corrélées} \iff \text{cov}(X, Y) = 0 \iff \rho(X, Y) = 0.$$

PROPOSITION 3.17 (L'inégalité de Cauchy). Soient X et Y des variables aléatoires telles que $\mathbb{E}[X^2] < \infty$ et $\mathbb{E}[Y^2] < \infty$. Alors

$$\mathbb{E}|XY| < \infty,$$

et l'on a

$$|\mathbb{E}[XY]| \leq \mathbb{E}|XY| \leq \sqrt{\mathbb{E}[X^2]} \sqrt{\mathbb{E}[Y^2]}.$$

En particulier,

$$|\text{cov}(X, Y)| \leq \sigma(X)\sigma(Y), \quad \text{c.à.d. } |\rho(X, Y)| \leq 1.$$

DÉMONSTRATION. Sans restrictions on peut se ramener au cas où $\mathbb{E}[X^2] > 0$ et $\mathbb{E}[Y^2] > 0$. On pose

$$\tilde{X} := \frac{X}{\sqrt{\mathbb{E}[X^2]}} \quad \text{et} \quad \tilde{Y} := \frac{Y}{\sqrt{\mathbb{E}[Y^2]}}.$$

Comme $(|\tilde{X}| - |\tilde{Y}|)^2 \geq 0$, on a $2|\tilde{X}\tilde{Y}| \leq \tilde{X}^2 + \tilde{Y}^2$, et puis

$$2\mathbb{E}|\tilde{X}\tilde{Y}| \leq \mathbb{E}[\tilde{X}^2] + \mathbb{E}[\tilde{Y}^2] = 2,$$

autrement dit

$$|\mathbb{E}[\tilde{X}\tilde{Y}]| \leq \mathbb{E}|\tilde{X}\tilde{Y}| \leq 1. \quad \square$$

3.3. Indépendance de variables aléatoires

DÉFINITION 3.18. Soit $(X_i)_{i \in I}$ une famille de variables aléatoires sur le même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. Les variables aléatoires $(X_i)_{i \in I}$ sont *indépendantes* si pour toute famille $(A_i)_{i \in I}$ où $A_i \subset [-\infty, +\infty]$, $i \in I$, les événements

$$\{X_i \in A_i\}, \quad i \in I,$$

sont indépendants, c.à.d. pour tout $\emptyset \neq J \subset I$ fini, on a

$$\mathbb{P} \left(\bigcap_{i \in J} \{X_i \in A_i\} \right) = \prod_{i \in J} \mathbb{P}\{X_i \in A_i\}.$$

REMARQUE 3.19. Soit $(X_i)_{i \in I}$ une famille finie de variables aléatoires (c.à.d. avec I fini). Alors $(X_i)_{i \in I}$ est indépendante si et seulement si

$$\mathbb{P} \left(\bigcap_{i \in I} \{X_i \in A_i\} \right) = \prod_{i \in I} \mathbb{P}\{X_i \in A_i\}, \quad \forall A_i \subset [-\infty, \infty], \quad i \in I.$$

DÉMONSTRATION. Il faut démontrer que la condition est suffisante. Pour cela soit $\emptyset \neq J \subset I$ et $(B_i)_{i \in J}$ une famille quelconque d'ensembles $B_i \subset [-\infty, \infty]$. On définit,

$$A_i = \begin{cases} B_i, & i \in J, \\ [-\infty, \infty], & i \notin J. \end{cases}$$

Comme

$$\{X_i \in A_i\} = \Omega \quad \text{pour } i \notin J,$$

on obtient

$$\mathbb{P} \left(\bigcap_{i \in J} \{X_i \in B_i\} \right) = \mathbb{P} \left(\bigcap_{i \in I} \{X_i \in A_i\} \right) = \prod_{i \in I} \mathbb{P}\{X_i \in A_i\} = \prod_{i \in J} \mathbb{P}\{X_i \in B_i\}.$$

\square

PROPOSITION 3.20. Soient X_1, \dots, X_n des variables aléatoires. Les X_1, \dots, X_n sont indépendantes si et seulement si pour tous les $x_i \in X_i(\Omega)$, $i = 1, \dots, n$,

$$\mathbb{P}\{X_1 = x_1, \dots, X_n = x_n\} = \mathbb{P}\{X_1 = x_1\} \cdot \dots \cdot \mathbb{P}\{X_n = x_n\}.$$

DÉMONSTRATION. La nécessité de la condition est évidente avec $A_i = \{x_i\}$. Pour démontrer que la condition est suffisante, soient $A_i \subset [-\infty, +\infty]$. Il faut vérifier que

$$\mathbb{P}\left(\bigcap_{i=1}^n \{X_i \in A_i\}\right) = \prod_{i=1}^n \mathbb{P}\{X_i \in A_i\}.$$

Comme

$$\{X_i \in A_i\} = \bigcup_{x_i \in X_i(\Omega) \cap A_i} \{X_i = x_i\},$$

on obtient

$$\begin{aligned} \mathbb{P}\left(\bigcap_{i=1}^n \{X_i \in A_i\}\right) &= \mathbb{P}\left(\bigcup_{(x_1, \dots, x_n) \in (X_1(\Omega) \cap A_1) \times \dots \times (X_n(\Omega) \cap A_n)} \bigcap_{i=1}^n \{X_i = x_i\}\right) \\ &= \sum_{(x_1, \dots, x_n) \in (X_1(\Omega) \cap A_1) \times \dots \times (X_n(\Omega) \cap A_n)} \mathbb{P}\left(\bigcap_{i=1}^n \{X_i = x_i\}\right) \\ &= \sum_{(x_1, \dots, x_n) \in (X_1(\Omega) \cap A_1) \times \dots \times (X_n(\Omega) \cap A_n)} \prod_{i=1}^n \mathbb{P}\{X_i = x_i\} \\ &= \prod_{i=1}^n \sum_{x_i \in X_i(\Omega) \cap A_i} \mathbb{P}\{X_i = x_i\} = \prod_{i=1}^n \mathbb{P}\{X_i \in A_i\}. \quad \square \end{aligned}$$

COROLLAIRE 3.21. *On a*

$$X_1, \dots, X_n \text{ indépendantes} \iff \mathbb{P}_{(X_1, \dots, X_n)} = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n},$$

c.à.d. la loi commune de X_1, \dots, X_n est le produit des lois de X_1, \dots, X_n .

REMARQUE 3.22. Soient X_1, \dots, X_n indépendantes et soient $f_i: X_i(\Omega) \rightarrow \mathbb{R}$ pour $i = 1, \dots, n$. Alors les variables

$$Y_i = f_i \circ X_i, \quad i = 1, \dots, n,$$

sont aussi indépendantes.

DÉMONSTRATION. Pour $y_1, \dots, y_n \in \mathbb{R}$, on a

$$\{Y_i = y_i\} = \{X_i \in A_i\} \quad \text{avec } A_i = \{x_i \in X_i(\Omega) \mid f_i(x_i) = y_i\},$$

et donc

$$\begin{aligned} \mathbb{P}\{Y_1 = y_1, \dots, Y_n = y_n\} &= \mathbb{P}\{X_1 \in A_1, \dots, X_n \in A_n\} \\ &= \mathbb{P}\{X_1 \in A_1\} \cdot \dots \cdot \mathbb{P}\{X_n \in A_n\} \\ &= \mathbb{P}\{Y_1 = y_1\} \cdot \dots \cdot \mathbb{P}\{Y_n = y_n\}. \quad \square \end{aligned}$$

PROPOSITION 3.23. *Soient X, Y deux variables aléatoires indépendantes telles que $\mathbb{E}|X| < \infty$ et $\mathbb{E}|Y| < \infty$. Alors aussi $\mathbb{E}|XY| < \infty$, et l'on a*

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

En particulier, si X, Y sont indépendantes alors

$$\text{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0.$$

DÉMONSTRATION. Avec $f(x, y) = |xy|$, on obtient

$$\begin{aligned}
 \mathbb{E}|XY| &= \mathbb{E}[f(X, Y)] \\
 &= \sum_{x \in X(\Omega), y \in Y(\Omega)} \underbrace{f(x, y)}_{=|xy|} \mathbb{P}\{X = x, Y = y\} \\
 &= \sum_{x, y} |x| |y| \mathbb{P}\{X = x\} \mathbb{P}\{Y = y\} \quad (\text{car } X, Y \text{ sont indépendantes}) \\
 &= \underbrace{\left(\sum_x |x| \mathbb{P}\{X = x\} \right)}_{= \mathbb{E}|X|} \underbrace{\left(\sum_y |y| \mathbb{P}\{Y = y\} \right)}_{= \mathbb{E}|Y|} < \infty.
 \end{aligned}$$

Donc $\mathbb{E}|XY| < \infty$, et par conséquent, $\mathbb{E}[XY] < \infty$. Le même calcul sans $|\cdot|$ donne alors $\mathbb{E}[XY] = \mathbb{E}[X] \mathbb{E}[Y]$. \square

REMARQUE 3.24. Plus généralement, soient X_1, \dots, X_n indépendantes telles que

$$\mathbb{E}|X_i| < \infty, \quad i = 1, \dots, n.$$

Alors

$$\mathbb{E}|X_1 \cdot \dots \cdot X_n| < \infty \quad \text{et} \quad \mathbb{E}[X_1 \cdot \dots \cdot X_n] = \mathbb{E}[X_1] \cdot \dots \cdot \mathbb{E}[X_n].$$

PROPOSITION 3.25. Soient X_1, \dots, X_n des variables aléatoires telles que

$$\mathbb{E}|X_i Y_j| < \infty \quad \text{pour } 1 \leq i, j \leq n.$$

Alors

$$\text{var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{var}(X_i) + 2 \sum_{\substack{i, j=1 \\ i < j}}^n \text{cov}(X_i, X_j).$$

En particulier, si les variables X_1, \dots, X_n sont indépendantes, alors on a l'égalité de Bienaymé:

$$\text{var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{var}(X_i).$$

DÉMONSTRATION. On a

$$\begin{aligned}
 \text{var} \left(\sum_{i=1}^n X_i \right) &= \mathbb{E} \left[\left(\sum_1^n X_i - \mathbb{E} \left[\sum_1^n X_i \right] \right)^2 \right] \\
 &= \mathbb{E} \left[\left(\sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \right)^2 \right] \\
 &= \sum_{i, j=1}^n \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])] \\
 &= \sum_{i=1}^n \text{var}(X_i) + \sum_{\substack{i, j=1 \\ i \neq j}}^n \text{cov}(X_i, X_j). \quad \square
 \end{aligned}$$

EXEMPLE 3.26.

(1) Soit X binomiale $\mathcal{B}(n, p)$, c.à.d.

$$\mathbb{P}\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, \dots, n.$$

Sans restrictions, on peut supposer que $X = X_1 + \dots + X_n$ avec

$$X_i = \begin{cases} 1 & \text{si succès pour l'expérience } i, \\ 0 & \text{sinon,} \end{cases}$$

et que les variables X_1, \dots, X_n sont indépendantes. Comme

$$\mathbb{E}[X_i^2] = \mathbb{E}[X_i] = p,$$

on a

$$\mathbb{E}[X] = np.$$

D'autre part,

$$\text{var}(X_i) = \mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2 = p - p^2 = p(1-p).$$

En utilisant l'indépendance de X_1, \dots, X_n , on obtient

$$\begin{aligned} \text{var}(X) &= \sum_{i=1}^n \text{var}(X_i) = np(1-p), \\ \sigma(X) &= \sqrt{np(1-p)}. \end{aligned}$$

(2) Soit X géométrique de paramètre $p \in]0, 1]$, c.à.d.

$$\mathbb{P}\{X = n\} = (1-p)^{n-1}p, \quad n = 1, 2, \dots$$

On remarque que, pour $|x| < 1$,

$$\begin{aligned} \sum_{n=0}^{\infty} x^n &= \frac{1}{1-x}, \\ \sum_{n=1}^{\infty} nx^{n-1} &= \frac{1}{(1-x)^2}, \\ \sum_{n=2}^{\infty} n(n-1)x^{n-2} &= \frac{2}{(1-x)^3}. \end{aligned}$$

Par conséquent, avec $x = 1-p \in [0, 1[$, on obtient

$$\begin{aligned} \mathbb{E}[X(X-1)] &= \sum_{n=1}^{\infty} n(n-1) \underbrace{\mathbb{P}\{X = n\}}_{=(1-p)^{n-1}p} \\ &= p(1-p) \underbrace{\sum_{n \geq 2} n(n-1)(1-p)^{n-2}}_{=2p^{-3}} = 2 \frac{1-p}{p^2}, \end{aligned}$$

et donc

$$\begin{aligned}
 \text{var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\
 &= \mathbb{E}[X(X-1) + X] - (\mathbb{E}[X])^2 \\
 &= \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2 \\
 &= 2\frac{1-p}{p^2} + \frac{1}{p} - \frac{1}{p^2} \\
 &= \frac{2-2p+p-1}{p^2} = \frac{1-p}{p^2}.
 \end{aligned}$$

(3) Soit X de Poisson de paramètre $\lambda > 0$, c.à.d.

$$\mathbb{P}\{X = k\} = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

Alors on a

$$\mathbb{E}[X] = \lambda \quad \text{et} \quad \text{var}(X) = \lambda.$$

En effet,

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

D'autre part, on a

$$\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = \lambda^2 + \lambda.$$

Par conséquent, $\text{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = (\lambda^2 + \lambda) - \lambda^2 = \lambda$.

3.4. Approximation poissonnienne

Approximation de la loi binomiale par la loi de Poisson

Soient X_1, \dots, X_n des variables aléatoires indépendantes telles que

$$\mathbb{P}\{X_i = 1\} = 1 - \mathbb{P}\{X_i = 0\} = p_n, \quad i = 1, \dots, n.$$

On note

$$S_n = X_1 + \dots + X_n.$$

Alors,

$$\mathbb{P}\{S_n = k\} = \binom{n}{k} p_n^k (1-p_n)^{n-k}, \quad k = 0, \dots, n.$$

En particulier, on a

$$\mathbb{E}[S_n] = np_n \quad \text{et} \quad \text{var}(S_n) = np_n(1-p_n).$$

Considérons la situation

n très grand et p_n très petit.

(1) Quand $n \rightarrow \infty$ et $p_n \rightarrow 0$ tel que $np_n \equiv \mathbb{E}[S_n] \rightarrow \lambda$, on a :

$$P\{S_n = k\} \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \dots$$

(2) Rappel: une variable aléatoire X suit la loi de Poisson de paramètre λ , si

$$\mathbb{P}\{X = k\} = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \dots$$

On remarque que

- $E[S_n] = np_n \rightarrow \lambda = \mathbb{E}[X]$ (donc l'interprétation de λ comme "nombre attendu de succès" est valide aussi pour la limite),
- $\text{var}(S_n) = np_n(1 - p_n) \rightarrow \lambda = \text{var}(X)$.

EXEMPLE 3.27. *Connexions erronées dans un réseau téléphonique*

(F. Thorndike, Applications of Poisson's Probability Summation, *Bell System Technical Journal* 5 (1926), 604–624).

Dans 267 séries de chacune 515 appels on a observé le nombre des connexions erronées (dans chaque série: beaucoup d'appels, c.à.d. n grand, mais une probabilité petite pour de connexions erronées).

Soit N_k = nombre de séries avec k connexions erronées.

k	0-2	3	4	5	6	7	8	9	10	11	12	13	14	15	≥ 16
N_k	1	5	11	14	22	43	31	40	35	20	18	12	7	6	2
$Ne^{-\lambda} \frac{\lambda^k}{k!}$	2	5	10	18	26	33	36	35	31	24	18	12	7	4	5

Au total on a observé

$$\begin{aligned} \sum_k kN_k &= 2334 \text{ connexions erronées (dans } N = 267 \text{ séries), c.à.d.} \\ &\approx \frac{2334}{267} = 8,74 \text{ par série.} \end{aligned}$$

En plus,

$$\frac{N_k}{N} \approx \mathbb{P}\{\text{"}k\text{ connexions erronées dans une série"}\} = \binom{n}{k} p^k (1-p)^{n-k}$$

où

$$n = 515 \text{ et}$$

$$p = \mathbb{P}\{\text{"connexion erronée"}\} \approx \frac{2334}{267 \times 515}.$$

Par l'approximation poissonnienne, on obtient

$$\frac{N_k}{N} \approx e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{où } \lambda = np = 8,74.$$

On remarque que

- $e^{-\lambda} \frac{\lambda^k}{k!}$ est la valeur théorique de la fréquence relative N_k/N de séries avec exactement k connexions erronées, $k = 0, 1, 2, \dots$, et
- $Ne^{-\lambda} \frac{\lambda^k}{k!}$ est la valeur théorique de la fréquence N_k .

EXEMPLE 3.28. Décès par ruades de cheval dans la cavalerie prussienne (Ladislav von Bortkiewicz, 1898). Livre “loi des petits nombres”.

3.5. Interprétation de la covariance: Régression simple et multiple

Soient X et Y deux variables aléatoires telles que $\mathbb{E}[X^2] < \infty$ et $\mathbb{E}[Y^2] < \infty$. Alors la covariance

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

est bien définie.

EXEMPLE 3.29. Par exemple, il est raisonnable de supposer que si X dénote la “pression atmosphérique” et Y la “température”, alors $\text{cov}(X, Y) > 0$; si X dénote la “consommation de cigarettes par jour” et Y la “capacité respiratoire”, alors $\text{cov}(X, Y) < 0$.

3.5.1. Problème de régression linéaire.

Situation Y une variable numérique à expliquer (ou estimer),
 X_1, \dots, X_p des variables explicatives,

$$Y = \underbrace{\Phi(X_1, \dots, X_p)}_{\text{estimation de } Y} + \varepsilon$$

où $\Phi: \mathbb{R}^p \rightarrow \mathbb{R}$ est une fonction à déterminer.

Objectif général

On cherche une approximation \hat{Y} de Y comme fonction de X_1, \dots, X_p :

$$\hat{Y} = \Phi(X_1, \dots, X_p).$$

La différence $\varepsilon = Y - \hat{Y}$ s'appelle *l'erreur d'estimation*.

De plus, on cherche une fonction Φ “simple” telle que l'erreur ε soit “petite”. Il faut adopter un compromis entre les objectifs de *conservation de l'information* et de *simplicité descriptive*.

Typiquement *Modèles de régression linéaire*

$$\hat{Y} \hat{=} \Phi(X_1, \dots, X_p) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

où $\beta_0, \beta_1, \dots, \beta_p$ sont des constantes.

Alors, on cherche des paramètres $\beta_0, \beta_1, \dots, \beta_p$ tels que

$$\mathbb{E} \left[\underbrace{(Y - (\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p))^2}_{= \hat{Y}} \right] \rightarrow \text{minimum.}$$

3.5.2. Le Cas Univarié ($p = 1$).

Soient X et Y des variables aléatoires telles que $\mathbb{E}[X^2] < \infty$ et $\mathbb{E}[Y^2] < \infty$. On s'intéresse à la "meilleure" approximation de Y comme fonction affine de X :

$$\hat{Y} = aX + b.$$

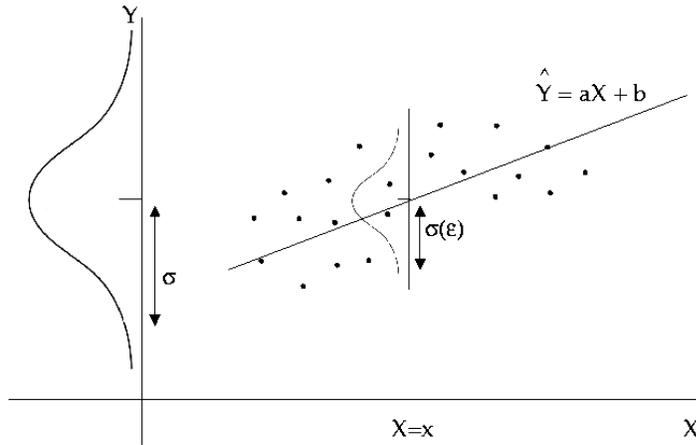
Question: $\sigma(\varepsilon) \leq \sigma(Y)$ où $\varepsilon = Y - \hat{Y}$?

Trouver $a^*, b^* \in \mathbb{R}$ tels que

$$Y = \underbrace{(a^*X + b^*)}_{=\hat{Y}} + \varepsilon \quad \text{et} \quad \mathbb{E}|Y - \hat{Y}|^2 \equiv \mathbb{E}[\varepsilon^2] \rightarrow \text{minimum.}$$

Problème de minimisation

$$\min_{a, b \in \mathbb{R}} F(a, b) \quad \text{où} \quad F(a, b) = \mathbb{E}[(Y - (aX + b))^2]$$



On obtient (comme condition nécessaire):

$$\begin{aligned} \frac{\partial}{\partial a} F(a, b) &= -2 \mathbb{E}[(Y - (aX + b))X] = 0 \\ (3.1) \quad &\iff \mathbb{E}[XY] - a \mathbb{E}[X^2] - b \mathbb{E}[X] = 0 \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial b} F(a, b) &= -2 \mathbb{E}[Y - (aX + b)] = 0 \\ (3.2) \quad &\iff \mathbb{E}[Y] - a \mathbb{E}[X] - b = 0. \end{aligned}$$

En substituant (3.2) dans (3.1), on trouve

$$\mathbb{E}[XY] - a \mathbb{E}[X^2] - (\mathbb{E}[Y] - a \mathbb{E}[X]) \mathbb{E}[X] = 0,$$

et on conclut que les paramètres a^* , b^* optimaux sont donnés par:

$$\begin{aligned} a^* &= \frac{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}{\mathbb{E}[X^2] - \mathbb{E}[X]^2} = \frac{\text{cov}(X, Y)}{\text{var}(X)} \\ b^* &= \mathbb{E}[Y] - \frac{\text{cov}(X, Y)}{\text{var}(X)} \mathbb{E}[X]. \end{aligned}$$

Conclusion La droite optimale est donnée par

$$(3.3) \quad a^*X + b^* = \frac{\text{cov}(X, Y)}{\text{var}(X)} (X - \mathbb{E}[X]) + \mathbb{E}[Y]$$

Interprétation

1. $\text{cov}(X, Y) = 0 \implies$ l'estimation optimale est $a^*X + b^* = \mathbb{E}[Y]$.
Plus grande est la valeur de $\frac{|\text{cov}(X, Y)|}{\text{var}(X)}$, plus la valeur de X compte dans le calcul de l'estimation.
2. $\text{cov}(X, Y) > 0 \implies$ la pente de la droite est positive;
 $\text{cov}(X, Y) < 0 \implies$ la pente de la droite est négative.
3. Soit $\varepsilon = Y - (a^*X + b^*) \hat{=} Y - \hat{Y}$ l'erreur de prédiction.
Alors on a

$$(3.4) \quad \mathbb{E}[\varepsilon] = 0 \quad \text{et} \quad \text{var}(\varepsilon) = \text{var}(Y) [1 - \rho^2(X, Y)]$$

où $\rho(X, Y)$ design le *coefficient de corrélation* défini par

$$\rho(X, Y) = \begin{cases} \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}, & \text{si } \sigma(X) > 0 \text{ et } \sigma(Y) > 0, \\ 0, & \text{sinon.} \end{cases}$$

En effet,

$$\begin{aligned} \text{var}(\varepsilon) &= \text{var}(Y) + \text{var}(\hat{Y}) - 2\text{cov}(Y, \hat{Y}) \\ &= \text{var}(Y) + a^{*2}\text{var}(X) - 2a^*\text{cov}(X, Y) \\ &= \text{var}(Y) + \frac{\text{cov}^2(X, Y)}{\text{var}^2(X)}\text{var}(X) - 2\frac{\text{cov}(X, Y)}{\text{var}(X)}\text{cov}(X, Y) \\ &= \text{var}(Y) - \frac{\text{cov}^2(X, Y)}{\text{var}(X)} \\ &= \text{var}(Y) [1 - \rho^2(X, Y)]. \end{aligned}$$

4. La variance de Y se décompose comme:

$$(3.5) \quad \underbrace{\text{var}(Y)}_{\text{variance totale}} = \underbrace{\text{var}(\hat{Y})}_{\text{variance expliquée}} + \underbrace{\text{var}(\varepsilon)}_{\text{variance résiduelle}}.$$

En utilisant que $\hat{Y} = a^*X + b^*$ avec $a^* = \frac{\text{cov}(X, Y)}{\text{var}(X)}$, on a

$$\text{var}(\hat{Y}) = \frac{\text{cov}^2(X, Y)}{\text{var}^2(X)}\text{var}(X) = \frac{\text{cov}^2(X, Y)}{\text{var}(X)},$$

et donc

$$\frac{\text{var}(\hat{Y})}{\text{var}(Y)} = \rho^2(X, Y).$$

En particulier, on a

$$\text{cov}(\hat{Y}, \varepsilon) = 0.$$

5. On note que

$$\rho(\hat{Y}, Y) = \frac{\text{cov}(\hat{Y}, Y)}{\sigma(\hat{Y}) \sigma(Y)} = \frac{\text{var}(\hat{Y})}{\sigma(\hat{Y}) \sigma(Y)} = \frac{\sigma(\hat{Y})}{\sigma(Y)}.$$

Le quotient

$$\frac{\text{var}(\hat{Y})}{\text{var}(Y)} = \frac{\text{variance expliquée}}{\text{variance totale}} \quad (\hat{=} \text{ pourcentage de variance expliquée})$$

est un indicateur de *qualité de la régression*. On a la relation fondamentale

$$(3.6) \quad \boxed{\frac{\text{var}(\hat{Y})}{\text{var}(Y)} = \rho^2(\hat{Y}, Y) = \rho^2(X, Y).}$$

On note que

$$(3.7) \quad \boxed{\frac{\text{var}(\varepsilon)}{\text{var}(Y)} = 1 - \rho^2(X, Y).}$$

3.5.3. Le Cas de la Régression Multiple ($p > 1$).

Soit $\hat{Y} = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$ l'estimation de Y comme fonction linéaire de X_1, \dots, X_p . On cherche des paramètres $\beta_0, \beta_1, \dots, \beta_p$ tels que

$$F(\beta_0, \beta_1, \dots, \beta_p) = \mathbb{E}[(Y - \hat{Y})^2] \longrightarrow \text{minimum.}$$

Pour $F(\beta_0, \beta_1, \dots, \beta_p) = \mathbb{E}[(Y - (\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p))^2]$ on obtient

$$\begin{aligned} \frac{\partial}{\partial \beta_0} F &= -2 \mathbb{E}[Y - (\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)] \quad \text{et} \\ \frac{\partial}{\partial \beta_i} F &= -2 \mathbb{E}[(Y - (\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)) X_i], \quad 1 \leq i \leq p. \end{aligned}$$

Par conséquent,

$$(E_0) \quad \beta_0 + \beta_1 \mathbb{E}[X_1] + \dots + \beta_p \mathbb{E}[X_p] = \mathbb{E}[Y]$$

$$(E_i) \quad \beta_0 \mathbb{E}[X_i] + \beta_1 \mathbb{E}[X_i X_1] + \dots + \beta_p \mathbb{E}[X_i X_p] = \mathbb{E}[X_i Y].$$

En substituant (E_0) ,

$$\beta_0 = \mathbb{E}[Y] - \beta_1 \mathbb{E}[X_1] - \dots - \beta_p \mathbb{E}[X_p],$$

dans (E_i) , on obtient pour $1 \leq i \leq p$:

$$\begin{aligned} \beta_1 \underbrace{(\mathbb{E}[X_i X_1] - \mathbb{E}[X_i] \mathbb{E}[X_1])} + \dots + \beta_p \underbrace{(\mathbb{E}[X_i X_p] - \mathbb{E}[X_i] \mathbb{E}[X_p])} &= \underbrace{\mathbb{E}[X_i Y] - \mathbb{E}[X_i] \mathbb{E}[Y]} \\ = \text{cov}(X_i, X_1) & \qquad \qquad \qquad = \text{cov}(X_i, X_p) & \qquad \qquad \qquad = \text{cov}(X_i, Y) \end{aligned}$$

Par conséquent, car $\text{cov}(X_i, X_i) = \text{var}(X_i)$, on a

$$\begin{cases} \text{var}(X_1) \beta_1 + \text{cov}(X_1, X_2) \beta_2 + \dots + \text{cov}(X_1, X_p) \beta_p = \text{cov}(X_1, Y) \\ \text{cov}(X_2, X_1) \beta_1 + \text{var}(X_2) \beta_2 + \dots + \text{cov}(X_2, X_p) \beta_p = \text{cov}(X_2, Y) \\ \dots \qquad \qquad \qquad \dots \qquad \qquad \qquad \dots \\ \text{cov}(X_p, X_1) \beta_1 + \text{cov}(X_p, X_2) \beta_2 + \dots + \text{var}(X_p) \beta_p = \text{cov}(X_p, Y), \end{cases}$$

et sous forme matricielle

$$\underbrace{\begin{pmatrix} \text{cov}(X_1, X_1) & \text{cov}(X_1, X_2) & \dots & \text{cov}(X_1, X_p) \\ \text{cov}(X_2, X_1) & \text{cov}(X_2, X_2) & \dots & \text{cov}(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_p, X_1) & \text{cov}(X_p, X_2) & \dots & \text{cov}(X_p, X_p) \end{pmatrix}}_{= \Sigma_{X,X}} \underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}}_{= \beta} = \underbrace{\begin{pmatrix} \text{cov}(X_1, Y) \\ \text{cov}(X_2, Y) \\ \vdots \\ \text{cov}(X_p, Y) \end{pmatrix}}_{= \Sigma_{X,Y}}.$$

DÉFINITION 3.30. On appelle la matrice

$$\Sigma_{X,X} = (\text{cov}(X_i, X_j))_{1 \leq i, j \leq p} \in \text{M}_{p \times p}(\mathbb{R})$$

matrice de covariance.

Conclusion On a

$$\Sigma_{X,X} \cdot \beta = \Sigma_{X,Y}$$

et, à condition que $\Sigma_{X,X}$ soit *invertible*, on obtient

$$(3.8) \quad \boxed{\beta = \Sigma_{X,X}^{-1} \cdot \Sigma_{X,Y}}$$

et ensuite,

$$(3.9) \quad \boxed{\beta_0 = \mathbb{E}[Y] - \beta_1 \mathbb{E}[X_1] - \dots - \beta_p \mathbb{E}[X_p].}$$

DÉFINITION 3.31. On appelle

$$R := \rho(Y, \hat{Y}) \hat{=} \frac{\text{cov}(Y, \hat{Y})}{\sigma(Y) \sigma(\hat{Y})}, \quad \text{resp. } R^2 = \rho^2(Y, \hat{Y}),$$

le *coefficient de corrélation multiple*, resp. *coefficient de détermination*.

Question Comment calculer R ?

On note

$$\begin{aligned} \hat{Y} &= \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p \\ &= (\mathbb{E}[Y] - \beta_1 \mathbb{E}[X_1] - \dots - \beta_p \mathbb{E}[X_p]) + \beta_1 X_1 + \dots + \beta_p X_p \\ &= \mathbb{E}[Y] + \sum_{i=1}^p \beta_i (X_i - \mathbb{E}[X_i]); \end{aligned}$$

en particulier,

$$\mathbb{E}[Y] = \mathbb{E}[\hat{Y}] \quad \text{et} \quad \hat{Y} - \mathbb{E}[Y] = \sum_{i=1}^p \beta_i (X_i - \mathbb{E}[X_i]).$$

On trouve que

$$\begin{aligned} \text{cov}(Y, \hat{Y}) &= \mathbb{E}[(Y - \mathbb{E}[Y])(\hat{Y} - \mathbb{E}[\hat{Y}])] \\ &= \sum_{i=1}^p \beta_i \mathbb{E}[(Y - \mathbb{E}[Y])(X_i - \mathbb{E}[X_i])] \\ &= \sum_{i=1}^p \beta_i \text{cov}(X_i, Y) = \beta^t \Sigma_{X,Y} \end{aligned}$$

et

$$\text{var}(\hat{Y}) = \sum_{i,j=1}^p \beta_i \beta_j \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])] = \beta^t \Sigma_{X,X} \beta.$$

En utilisant que $\Sigma_{X,X} \beta = \Sigma_{X,Y}$, on a donc

$$\text{var}(\hat{Y}) = \beta^t \Sigma_{X,Y} = \text{cov}(Y, \hat{Y}),$$

et parce que $Y = \hat{Y} + \varepsilon$ et $\text{cov}(\hat{Y}, \hat{Y}) = \text{var}(\hat{Y})$,

$$\text{cov}(\varepsilon, \hat{Y}) = 0.$$

Conclusion Comme dans le cas univarié, on a la relation fondamentale:

(3.10)

$$\boxed{\text{var}(Y) = \text{var}(\hat{Y}) + \text{var}(\varepsilon)}$$

où $\text{var}(\hat{Y}) = \beta^t \Sigma_{X,Y} = (\Sigma_{X,X}^{-1} \Sigma_{X,Y})^t \Sigma_{X,Y} = \Sigma_{X,Y}^t \Sigma_{X,X}^{-1} \Sigma_{X,Y}$, et en plus,

1. (*pourcentage de variance expliquée*)

$$R^2 = \rho^2(Y, \hat{Y}) = \frac{\text{cov}^2(Y, \hat{Y})}{\sigma(Y) \sigma(\hat{Y})} = \frac{\text{var}(\hat{Y})}{\text{var}(Y)}.$$

2. (*pourcentage de variance résiduelle*)

$$1 - R^2 = \frac{\text{var}(Y) - \text{var}(\hat{Y})}{\text{var}(Y)} = \frac{\text{var}(\varepsilon)}{\text{var}(Y)}.$$

REMARQUE 3.32 (Calcul en termes des coefficients de corrélation).

Définissons

$$R_{X,X} = (\rho(X_i, X_j))_{1 \leq i, j \leq p} = \begin{pmatrix} 1 & \rho(X_1, X_2) & \dots & \rho(X_1, X_p) \\ \rho(X_2, X_1) & 1 & \dots & \rho(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(X_p, X_1) & \rho(X_p, X_2) & \dots & 1 \end{pmatrix} \in M_{p \times p}(\mathbb{R})$$

$$R_{X,Y} = (\text{var}(X_i, Y))_{1 \leq i \leq p} = \begin{pmatrix} \rho(X_1, Y) \\ \rho(X_2, Y) \\ \vdots \\ \rho(X_p, Y) \end{pmatrix} \in M_{p \times 1}(\mathbb{R}) \quad \text{et}$$

$$\Lambda = \begin{pmatrix} \text{var}(X_1) & 0 & \dots & 0 \\ 0 & \text{var}(X_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \text{var}(X_p) \end{pmatrix} \in M_{p \times p}(\mathbb{R}).$$

On remarque que

$$R_{X,X} = \Lambda^{-1/2} \Sigma_{X,X} \Lambda^{-1/2} \quad \text{et} \quad R_{X,Y} = \frac{\Lambda^{-1/2} \Sigma_{X,Y}}{\sigma(Y)},$$

où $\Lambda^{-1/2}$ est la matrice diagonale d'éléments $1/\sigma(X_i)$.

Posons $b_i := \beta_i \frac{\sigma(X_i)}{\sigma(Y)}$ et $b^t = (b_1, \dots, b_p)$, on a

$$\beta^t = \sigma(Y) b^t \Lambda^{-1/2}$$

et donc

$$R^2 = \frac{\beta^t \Sigma_{X,Y}}{\text{var}(Y)} = \frac{b^t \Lambda^{-1/2} \Sigma_{X,Y}}{\text{var}(Y)} = b^t R_{X,Y};$$

d'autre part, l'égalité $\Sigma_{X,X} \beta = \Sigma_{X,Y}$ s'écrit comme $\Sigma_{X,X} \Lambda^{-1/2} b = \frac{\Sigma_{X,Y}}{\sigma(Y)}$, c.à.d.

$$R_{X,X} b = R_{X,Y}.$$

Conclusion

$$(3.11) \quad \boxed{R^2 = \sum_{i=1}^p b_i \rho(X_i, Y) \quad \text{où} \quad b = R_{X,X}^{-1} R_{X,Y} \quad \text{si} \quad R_{X,X} \text{ est inversible.}}$$

Inégalités et grandes déviations

4.1. Les inégalités de Markov, Tchebychev, Bernstein et Hoeffding

PROPOSITION 4.1 (L'inégalité de Markov). *Soit X une variable aléatoire et $h : [-\infty, +\infty] \rightarrow [0, +\infty]$ une fonction croissante. On a alors*

$$\mathbb{P}\{X \geq a\} \leq \frac{\mathbb{E}[h(X)]}{h(a)}, \quad \forall a \in \mathbb{R} \text{ tels que } h(a) > 0.$$

DÉMONSTRATION.

$$\mathbb{E}[h(X)] \geq \mathbb{E}[h(X) \mathbf{1}_{\{X \geq a\}}] \geq h(a) \mathbb{E}[\mathbf{1}_{\{X \geq a\}}] = h(a) \mathbb{P}\{X \geq a\}. \quad \square$$

Cas particuliers

(1) Soit $h(x) = x \vee 0 := \max(x, 0)$. Alors on a

$$\mathbb{P}\{X \geq a\} \leq \mathbb{P}\{|X| \geq a\} \leq \frac{\mathbb{E}|X|}{a} \quad \forall a > 0.$$

(2) (L'inégalité de Tchebychev-Markov)

Soit $h(x) = (x \vee 0)^2$. Alors on a

$$\mathbb{P}\{|X| \geq a\} \leq \frac{\mathbb{E}|X|^2}{a^2} \quad \forall a > 0.$$

En particulier,

$$\mathbb{P}\{|X - \mathbb{E}[X]| \geq a\} \leq \frac{\text{var}(X)}{a^2} \quad \forall a > 0.$$

(3) (L'inégalité de Bernstein)

Soit $h(x) = e^{tx}$ avec $t > 0$ choisi appropriément. Alors on a

$$\mathbb{P}\{X \geq a\} \leq \frac{\mathbb{E}[e^{tX}]}{e^{ta}} \quad \forall a \in \mathbb{R}.$$

EXEMPLE 4.2. On jette un dé non pipé 1000 fois. Quelle est la probabilité d'obtenir une somme totale au moins égale à 4000?

Soit X_i = le résultat du $i^{\text{ème}}$ jet ($i = 1, \dots, 1000$). Alors les variables aléatoires X_1, \dots, X_{1000} sont indépendantes et

$$\mathbb{P}\{X_i = k\} = \frac{1}{6}, \quad k = 1, 2, \dots, 6.$$

Pour la somme totale

$$S = \sum_{i=1}^{1000} X_i$$

on trouve

$$\begin{aligned}\mathbb{E}[S] &= \sum_{i=1}^{1000} \mathbb{E}[X_i] = 1000 \times \mathbb{E}[X_i] \\ \text{var}(S) &= \text{var}\left(\sum_{i=1}^{1000} X_i\right) = \sum_{i=1}^{1000} \text{var}(X_i) + \sum_{\substack{i,j=1 \\ i \neq j}}^{1000} \underbrace{\text{cov}(X_i, X_j)}_{=0} = 1000 \times \text{var}(X_i).\end{aligned}$$

Comme

$$\begin{aligned}\mathbb{E}[X_i] &= \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = \frac{7}{2} \\ \mathbb{E}[X_i^2] &= \frac{1+4+9+\dots+36}{6} = \frac{91}{6},\end{aligned}$$

on a

$$\text{var}(X_i) = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 = \frac{91}{6} - \frac{49}{4} = \frac{182 - 147}{12} = \frac{35}{12}.$$

Donc, avec l'inégalité de Tchebychev-Markov, on obtient

$$\begin{aligned}\mathbb{P}\{S \geq 4000\} &= \mathbb{P}\{S - \mathbb{E}[S] \geq 500\} \\ &= \frac{1}{2} \mathbb{P}\{|S - \mathbb{E}[S]| \geq 500\} \\ &\leq \frac{1}{2} \frac{\text{var}(S)}{500^2} = \frac{1}{2} \frac{1000}{500^2} \text{var}(X_1) \\ &= \frac{1}{2} \frac{\text{var}(X_1)}{250} = \frac{1}{2} \frac{35}{12 \times 250} = \frac{7}{1200}.\end{aligned}$$

LEMME 4.3. Soit X une variable aléatoire telle que $|X| \leq a$ et $\mathbb{E}[X] = 0$. Alors, pour tout $t \in \mathbb{R}$,

$$\mathbb{E}[e^{tX}] \leq e^{t^2 a^2 / 2}.$$

DÉMONSTRATION. Sans restrictions (en remplaçant X par tX), on peut supposer que $t = 1$ et en plus $a > 0$. On remarque que la fonction exponentielle $x \mapsto f(x) = e^x$ est convexe, c.à.d.

$$f(\lambda u + (1 - \lambda)v) \leq \lambda f(u) + (1 - \lambda)f(v), \quad \forall u, v \in \mathbb{R}, \lambda \in [0, 1].$$

Avec $u = a$, $v = -a$ et $\lambda = \frac{a+x}{2a}$ (c.à.d. $1 - \lambda = \frac{a-x}{2a}$), on obtient

$$\forall x \in [-a, a], \quad f(x) \leq \frac{a+x}{2a} f(a) + \frac{a-x}{2a} f(-a),$$

c.à.d.

$$e^x \leq \frac{a+x}{2a} e^a + \frac{a-x}{2a} e^{-a}.$$

Par conséquent, comme $\mathbb{E}[X] = 0$, on obtient

$$\mathbb{E}[e^X] \leq \underbrace{\frac{e^a + e^{-a}}{2}}_{\cosh(a)} \leq \exp\left(\frac{a^2}{2}\right).$$

On utilise que

$$\cosh(a) = \sum_{n=0}^{\infty} \frac{a^{2n}}{(2n)!} \leq \sum_{n=0}^{\infty} \frac{(a^2)^n}{2^n n!} = \sum_{n=0}^{\infty} \frac{(a^2/2)^n}{n!} = \exp(a^2/2). \quad \square$$

THÉORÈME 4.4 (L'inégalité de Hoeffding). *Soient X_1, \dots, X_n des variables aléatoires indépendantes telles que pour tout i ,*

$$a_i \leq X_i \leq b_i \quad \text{et} \quad \mathbb{E}[X_i] = 0.$$

On pose

$$S_n = X_1 + \dots + X_n.$$

Alors on a pour tout $c > 0$,

$$\mathbb{P}\{S_n \geq c\} \leq \exp\left(-\frac{c^2}{2m}\right)$$

où $m := \sum_{i=1}^n \max\{a_i^2, b_i^2\}$.

DÉMONSTRATION. Pour $t > 0$, on a

$$\begin{aligned} \mathbb{P}\{S_n \geq c\} &= \mathbb{P}\{e^{tS_n} \geq e^{tc}\} \\ &\leq \frac{\mathbb{E}[\exp(tS_n)]}{\exp(tc)} \quad (\text{par l'inégalité de Markov}) \\ &= \frac{\mathbb{E}\left[\prod_{i=1}^n e^{tX_i}\right]}{\exp(tc)} = \frac{\prod_{i=1}^n \mathbb{E}[e^{tX_i}]}{\exp(tc)} \quad (X_1, \dots, X_n \text{ sont indépendantes}) \\ &\leq \frac{\prod_{i=1}^n \exp(t^2 \max(a_i^2, b_i^2)/2)}{\exp(tc)} \quad (\text{par le lemme}) \\ &= \exp\left(\frac{t^2 m}{2} - tc\right) =: \varphi(t). \end{aligned}$$

Par conséquent,

$$\mathbb{P}\{S_n \geq c\} \leq \min_t \varphi(t) = \varphi\left(\frac{c}{m}\right) = \exp\left(-\frac{c^2}{2m}\right). \quad \square$$

EXEMPLE 4.5. Retour au l'exemple 4.2. On jette 1000-fois un dé non pipé. Quelle est la probabilité d'obtenir une somme totale au moins égale à 4000?

On note

$$S = \sum_{i=1}^{1000} X_i.$$

(1) Estimation par l'inégalité de Tchebychev-Markov:

$$\mathbb{P}\{S \geq 4000\} = \mathbb{P}\{S - \mathbb{E}[S] \geq 500\} \leq \frac{1}{2} \frac{\text{var}(S)}{500^2} = \frac{7}{1200}.$$

(2) Estimation par l'inégalité de Hoeffding:

Comme

$$\underbrace{-2,5}_{=a_i} \leq X_i - \underbrace{\mathbb{E}[X_i]}_{3,5} \leq \underbrace{2,5}_{=b_i} \quad \text{et} \quad m = \sum_{i=1}^{1000} a_i^2 \vee b_i^2 = 1000 \frac{25}{4},$$

on obtient

$$\begin{aligned} \mathbb{P}\{S \geq 4000\} &= \mathbb{P}\{S - \mathbb{E}[S] \geq 500\} \\ &= \mathbb{P}\left\{\sum_{i=1}^{1000} (X_i - \mathbb{E}[X_i]) \geq 500\right\} \\ &\leq \exp\left(-\frac{500^2}{2 \cdot 1000 \cdot \frac{25}{4}}\right) \\ &= e^{-20} \approx 2 \times 10^{-9}. \end{aligned}$$

4.2. La loi faible des grands nombres

THÉORÈME 4.6 (Loi faible des grands nombres; weak LLN = “weak law of large numbers”). Soient X_1, \dots, X_n des variables aléatoires sur $(\Omega, \mathcal{A}, \mathbb{P})$. Supposons que

$$\mathbb{E}|X_i|^2 < \infty, \quad \text{cov}(X_i, X_j) = 0 \quad \forall i \neq j \quad \text{et} \quad \frac{1}{n^2} \sum_{i=1}^n \text{var}(X_i) \rightarrow 0.$$

Alors

$$\forall \varepsilon > 0, \quad \mathbb{P}\left\{\left|\frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}[X_i])\right| \geq \varepsilon\right\} \rightarrow 0, \quad n \rightarrow \infty.$$

En particulier, si

$$\mathbb{E}[X_i] = \mathbb{E}[X_1] = \mu \quad \forall i,$$

alors

$$\mathbb{P}\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \geq \varepsilon\right\} = \mathbb{P}\left\{\left|\frac{1}{n} S_n - \mu\right| \geq \varepsilon\right\} \rightarrow 0, \quad n \rightarrow \infty,$$

avec $S_n := \sum_{i=1}^n X_i$.

DÉMONSTRATION. Sans restrictions on peut supposer que $\mathbb{E}[X_i] = 0$ pour tout i . En utilisant l'inégalité de Tchebychev-Markov, on obtient

$$\mathbb{P}\left\{\left|\frac{S_n}{n}\right| \geq \varepsilon\right\} \leq \frac{\text{var}(S_n)}{n^2 \varepsilon^2} = \frac{1}{n^2 \varepsilon^2} \sum_{i=1}^n \text{var}(X_i) \rightarrow 0, \quad n \rightarrow \infty. \quad \square$$

COROLLAIRE 4.7. Pour tout $n = 1, 2, \dots$ soient X_1, \dots, X_n des variables aléatoires, deux à deux non corrélées, de même espérance et variance, c.à.d. $\mathbb{E}[X_i] = \mu$ et $\text{var}(X_i) = \sigma^2$, $\forall i = 1, \dots, n$. On pose

$$S_n = \sum_{i=1}^n X_i \quad \text{et} \quad \bar{S}_n = \frac{1}{n} S_n.$$

Alors on a

$$\lim_{n \rightarrow \infty} \mathbb{P}\{|\bar{S}_n - \mu| \geq \varepsilon\} = 0, \quad \forall \varepsilon > 0.$$

DÉMONSTRATION. Par Tchebychev-Markov on a

$$\mathbb{P}\{|\bar{S}_n - \mu| \geq \varepsilon\} \leq \frac{\text{var}(\bar{S}_n)}{\varepsilon^2} = \frac{\text{var}(S_n)}{n^2 \varepsilon^2} = \frac{n\sigma^2}{n^2 \varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0, \quad n \rightarrow +\infty. \quad \square$$

EXEMPLE 4.8. Soient X_1, \dots, X_n des variables aléatoires i.i.d. telles que

$$\mathbb{P}\{X_i = 1\} = 1 - \mathbb{P}\{X_i = 0\} = p = \mathbb{E}[X_i].$$

Alors $S_n = X_1 + \dots + X_n \sim \mathcal{B}(n, p)$, et donc

$$\mathbb{E}[S_n] = np, \quad \text{var}(S_n) = np(1-p).$$

Par la loi faible, on obtient

$$\begin{aligned} \mathbb{P}\left\{\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right\} &= \mathbb{P}\{|S_n - np| \geq n\varepsilon\} \\ &= \sum_{k: |k-np| \geq n\varepsilon} \mathbb{P}\{S_n = k\} \\ &= \sum_{k: |k-np| \geq n\varepsilon} \binom{n}{k} p^k (1-p)^{n-k} \rightarrow 0, \quad n \rightarrow \infty. \end{aligned}$$

REMARQUE 4.9 (Application typique). Soit $(\Omega, \mathcal{A}, \mathbb{P})$ une expérience aléatoire. On répète cette expérience n fois indépendamment. Soit

$$\Omega^n = \Omega \times \dots \times \Omega = \{\omega = (\omega_1, \dots, \omega_n) \mid \omega_i \in \Omega\}, \quad \mathcal{A}^n := \mathcal{P}(\Omega^n),$$

et \mathbb{P}^n la mesure produit. On pose

$$R_i : \Omega^n \rightarrow \Omega, \quad R_i(\omega) = \omega_i \quad (\text{résultat de l'expérience } i).$$

Pour $A \subset \Omega$ on note

$$X_i = \mathbb{1}_A \circ R_i = \begin{cases} 1 & \text{si l'événement } A \text{ se produit au } i^{\text{ème}} \text{ coup,} \\ 0 & \text{sinon.} \end{cases}$$

Alors on a

$$\begin{aligned} \mathbb{E}[X_i] &= \mathbb{E}[X_i^2] = \mathbb{P}(A), \quad \text{et} \\ \text{var}(X_i) &= \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 = \mathbb{P}(A)(1 - \mathbb{P}(A)) \leq \frac{1}{4}. \end{aligned}$$

La loi faible des grands nombres implique que

$$\begin{aligned} \forall \varepsilon > 0, \quad \mathbb{P}^n \left\{ \left| \frac{1}{n} \sum_{i=1}^n X_i - \mathbb{P}(A) \right| > \varepsilon \right\} \\ &= \mathbb{P}^n \left\{ \omega \in \Omega^n : \left| \frac{1}{n} \sum_{i=1}^n X_i - \mathbb{P}(A) \right| > \varepsilon \right\} \\ &\leq \frac{n \cdot \text{var}(X_1)}{n^2 \varepsilon^2} \leq \frac{1}{4n\varepsilon^2}. \end{aligned}$$

Interprétation

$$\begin{aligned} \frac{1}{n} S_n(\omega) &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_A(\omega_i) && \text{la fréquence empirique;} \\ p &= \mathbb{P}(A) && \text{la probabilité exacte;} \\ \Delta_n &= \frac{1}{n} S_n - p && \text{la déviation (variable aléatoire).} \end{aligned}$$

Par la *loi faible des grands nombres* on a:

$$\forall \varepsilon > 0, \quad \mathbb{P}\{|\Delta_n| > \varepsilon\} \leq \frac{1}{4n\varepsilon^2} \rightarrow 0, \quad n \rightarrow \infty,$$

e.g. la déviation

$$\Delta_n = \frac{1}{n}S_n - p$$

est plus grande (en valeur absolue) que $5/\sqrt{n}$ seulement dans 1% de cas.

Typiquement, on fixe $\alpha > 0$ (petit), alors:

$$\forall \varepsilon > 0, \quad \mathbb{P}\left\{\left|\frac{S_n}{n} - p\right| \leq \varepsilon\right\} \geq 1 - \alpha, \quad \text{si } n \geq \frac{1}{4\varepsilon^2\alpha}.$$

4.3. Quelques applications

EXEMPLE 4.10 (Un jeu profitable, où l'on perd à long terme). Soit Z_n une suite indépendante des variables aléatoires Bernoulli telle que

$$\underbrace{\mathbb{P}\{Z_n = 0\}}_{\text{“pile”}} = \underbrace{\mathbb{P}\{Z_n = 1\}}_{\text{“face”}} = \frac{1}{2} \quad \text{pour tout } n.$$

On définit un jeu de la façon suivante:

- K_0 soit la fortune initiale;
- K_n la fortune après le $n^{\text{ième}}$ coup ($n \geq 1$)

$$K_n = \begin{cases} \frac{1}{2} K_{n-1} & \text{si } Z_n = 0, \\ \frac{5}{3} K_{n-1} & \text{si } Z_n = 1. \end{cases}$$

a) *Le jeu est profitable*, c.à.d. $\mathbb{E}[K_n] > \mathbb{E}[K_{n-1}]$ pour tout $n \geq 1$.

En effet, comme

$$K_n = \frac{1}{2} K_{n-1} \mathbf{1}_{\{Z_n=0\}} + \frac{5}{3} K_{n-1} \mathbf{1}_{\{Z_n=1\}},$$

on a

$$\begin{aligned} \mathbb{E}[K_n] &= \frac{1}{2} \mathbb{E}[K_{n-1}] \underbrace{\mathbb{P}\{Z_n = 0\}}_{=1/2} + \frac{5}{3} \mathbb{E}[K_{n-1}] \underbrace{\mathbb{P}\{Z_n = 1\}}_{=1/2} \\ &= \left(\frac{1}{2} + \frac{5}{3}\right) \mathbb{E}[K_{n-1}] \frac{1}{2} = \frac{13}{12} \mathbb{E}[K_{n-1}]. \end{aligned}$$

Pour $i = 1, 2, \dots$ on note

$$X_i = \begin{cases} \frac{1}{2} & \text{si “pile” au } i^{\text{ième}} \text{ coup,} \\ \frac{5}{3} & \text{si “face” au } i^{\text{ième}} \text{ coup.} \end{cases}$$

Alors les X_i sont des variables aléatoires et

$$K_n = \underbrace{K_0}_{=1} X_1 X_2 \cdots X_n.$$

Comme $\mathbb{E}[X_i] = 13/12$ pour tout i , on obtient

$$\mathbb{E}[K_n] = \mathbb{E}[X_1] \cdots \mathbb{E}[X_n] = \left(\frac{13}{12}\right)^n \rightarrow \infty, \quad n \rightarrow \infty.$$

b) Par la *loi faible des grands nombres*, on a

$$\forall \varepsilon > 0, \quad \mathbb{P} \left\{ \left| \frac{\log X_1 + \dots + \log X_n}{n} - \mathbb{E}[\log X_1] \right| > \varepsilon \right\} \rightarrow 0, \quad n \rightarrow \infty.$$

Soit

$$\mu := \mathbb{E}[\log X_1] = \frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{5}{3} = \frac{1}{2} \log \frac{5}{6} < 0.$$

Alors on a, pour tout $\varepsilon > 0$,

$$\mathbb{P} \left\{ \left| \frac{\log X_1 + \dots + \log X_n}{n} - \mu \right| \leq \varepsilon \right\} \rightarrow 1, \quad n \rightarrow \infty.$$

En particulier, pour $\varepsilon = -\mu/2$, on obtient

$$\mathbb{P} \left\{ \left| \frac{\log K_n}{n} - \mu \right| \leq -\frac{\mu}{2} \right\} \leq \mathbb{P} \left\{ \frac{\log K_n}{n} \leq \frac{\mu}{2} \right\} \rightarrow 1, \quad n \rightarrow \infty.$$

Mais

$$\mathbb{P} \left\{ \frac{\log K_n}{n} \leq \frac{\mu}{2} \right\} = \mathbb{P} \left\{ K_n \leq e^{\mu n/2} \right\} = \mathbb{P} \left\{ K_n \leq e^{\lambda n} \right\}, \quad \lambda := \mu/2 < 0.$$

On remarque que $e^{\lambda n} \rightarrow 0$, quand $n \rightarrow \infty$.

Conclusion Quand $n \rightarrow \infty$, avec probabilité ≈ 1 , on perd tout l'argent (à vitesse exponentielle).

PROPOSITION 4.11 (Théorème de Weierstraß). *Toute fonction continue sur $[0, 1]$ est limite uniforme de fonctions polynomiales, c.à.d. soit $f \in C([0, 1])$, alors*

$$\forall \varepsilon > 0, \quad \exists P \in \text{Pol}([0, 1]) \quad \text{tel que} \quad \sup_{x \in [0, 1]} |f(x) - P(x)| \leq \varepsilon.$$

DÉMONSTRATION. Pour tout $n \geq 1$ on pose

$$B_n(x) = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}, \quad x \in [0, 1].$$

On fixe $0 < x < 1$. Soient X_1, X_2, X_3, \dots des variables aléatoires indépendantes de loi Bernoulli de paramètre x , c.à.d.

$$\forall i, \quad \mathbb{P}\{X_i = 1\} = 1 - \mathbb{P}\{X_i = 0\} = x.$$

Alors

$$S_n = X_1 + \dots + X_n$$

suit la loi binômiale $\mathcal{B}(n, x)$ de paramètres n et x , c.à.d.

$$\mathbb{P}\{S_n = k\} = \binom{n}{k} x^k (1-x)^{n-k}.$$

On pose

$$\bar{S}_n = \frac{1}{n} S_n.$$

(1) Alors

$$\mathbb{E}[f(\bar{S}_n)] = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k} = B_n(x).$$

(2) Avec Tchebychev-Markov on obtient

$$\mathbb{P}\{|\bar{S}_n - x| \geq \delta\} \leq \frac{\text{var}(\bar{S}_n)}{\delta^2} \stackrel{(*)}{=} \frac{x(1-x)}{n\delta^2} \leq \frac{1}{4} \frac{1}{n\delta^2}.$$

Pour (*) on utilise que

$$\text{var}(S_n) = \sum_{i=1}^n \text{var}(X_i) = nx(1-x) \Rightarrow \text{var}(\bar{S}_n) = \frac{x(1-x)}{n}.$$

Pour $\delta > 0$ on pose

$$a_\delta = \sup\{|f(x) - f(y)| : x, y \in [0, 1], |x - y| \leq \delta\}.$$

On fixe maintenant $\varepsilon > 0$. Choisissons

$$\delta > 0 \text{ tel que } a_\delta < \frac{\varepsilon}{2} \text{ et } n \in \mathbb{N} \text{ tel que } \frac{a_1}{4n\delta^2} < \frac{\varepsilon}{2}.$$

Alors on obtient

$$\begin{aligned} |f(x) - B_n(x)| &= |f(x) - \mathbb{E}[f(\bar{S}_n)]| \\ &= |\mathbb{E}[f(x) - f(\bar{S}_n)]| \\ &\leq \mathbb{E}|f(\bar{S}_n) - f(x)| \\ &= \mathbb{E}\left[|f(\bar{S}_n) - f(x)| \mathbf{1}_{\{|\bar{S}_n - x| < \delta\}}\right] + \mathbb{E}\left[|f(\bar{S}_n) - f(x)| \mathbf{1}_{\{|\bar{S}_n - x| \geq \delta\}}\right] \\ &\leq a_\delta + a_1 \mathbb{P}\{|\bar{S}_n - x| \geq \delta\} \\ &\leq a_\delta + \frac{a_1}{4n\delta^2} \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

Par conséquent,

$$\sup_{x \in [0, 1]} |f(x) - B_n(x)| < \varepsilon. \quad \square$$

Axiomes probabilités et lois continues

Modèles discrets: $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace de probabilité avec Ω un ensemble dénombrable tel que

$$\mathcal{A} = \mathcal{P}(\Omega) \quad \text{et}$$

$$\mathbb{P}(A) = \sum_{\omega \in A} p(\omega), \quad \forall A \in \mathcal{A},$$

où $p(\omega) := \mathbb{P}(\{\omega\})$.

5.1. Limitations de modèles discrets

EXEMPLE 5.1. On tire au hasard un nombre réel x compris entre 0 et 1 ($0 \leq x < 1$). Alors $\Omega = [0, 1[$ est non dénombrable. On veut avoir

$$\mathbb{P}([0, \frac{1}{2}[) = \mathbb{P}([\frac{1}{2}, 1]) = \frac{1}{2},$$

$$\mathbb{P}([0, \frac{1}{4}[) = \mathbb{P}([\frac{1}{4}, \frac{1}{2}[) = \mathbb{P}([\frac{1}{2}, \frac{3}{4}[) = \mathbb{P}([\frac{3}{4}, 1]) = \frac{1}{4},$$

et plus généralement,

$$\forall [a, b[\subset [0, 1[\quad \text{avec } a < b, \quad \mathbb{P}([a, b]) = b - a.$$

Problème On ne peut pas définir \mathbb{P} sur $\mathcal{P}(\Omega)$, car il n'existe pas de mesure de probabilité \mathbb{P} sur $\mathcal{P}(\Omega)$ qui est invariante par translation, c.à.d. telle que

$$\mathbb{P}(\tau_x A) = \mathbb{P}(A), \quad \forall A \subset \Omega, \quad \forall x \in \Omega,$$

où $\tau_x(y) := x + y \bmod 1$. En plus, si $\mathbb{P}(A) > 0$, alors

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{a \in A} \{a\}\right) \neq \sum_{a \in A} \mathbb{P}(\{a\}) = 0.$$

Note Soit $a \in [0, 1[$. Alors $\mathbb{P}(\{a\}) \leq \mathbb{P}([a, a + \frac{1}{n}[) = \frac{1}{n}$ et puis $\mathbb{P}(\{a\}) = 0$.

EXEMPLE 5.2. On répète une expérience de Bernoulli $p(1) = p(-1) = \frac{1}{2}$ infiniment souvent, c.à.d.

$$\Omega = \{\omega = (\omega_1, \omega_2, \dots) : \omega_i \in \{\pm 1\}\} = \{\pm 1\}^{\mathbb{N}}.$$

(1) Par exemple, en fixant $\varepsilon_i \in \{\pm 1\}$, $i = 1, \dots, N$, considérons

$$A = \{\omega \in \Omega : \omega_i = \varepsilon_i, \quad i = 1, \dots, N\}.$$

Alors $\mathbb{P}(A) = p(\varepsilon_1) \cdot p(\varepsilon_2) \cdot \dots \cdot p(\varepsilon_n) = 2^{-N}$, car $p(\varepsilon_i) = 1/2$ pour tout i .
En particulier, pour tout $\omega \in \Omega$,

$$\mathbb{P}(\{\omega\}) \leq 2^{-N} \quad \forall N \Rightarrow \mathbb{P}(\{\omega\}) = 0.$$

- (2) Soit $B = \{\omega \in \Omega \mid \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \omega_i = 0\}$; on veut, par exemple, démontrer que $\mathbb{P}(B) = 1$. Mais ni B , ni Ω n'est dénombrable.

5.2. Les axiomes de Kolmogorov (1933)

DÉFINITION 5.3. On appelle *espace de probabilité* (ou *espace probabilisé*) tout triple $(\Omega, \mathcal{A}, \mathbb{P})$ avec les propriétés suivantes:

- (1) Ω est un ensemble
 (2) \mathcal{A} est une *tribu* sur Ω (“tribu des événements”)

$$\text{c.à.d. } \mathcal{A} \subset \mathcal{P}(\Omega) \text{ tel que } \begin{cases} \emptyset \in \mathcal{A} \\ A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A} \\ A_1, A_2, \dots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}; \end{cases}$$

- (3) \mathbb{P} est une *probabilité* sur \mathcal{A} , c.à.d. $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ tel que:
 a) $\mathbb{P}(\Omega) = 1$
 b) Si $A_1, A_2, \dots \in \mathcal{A}$ tels que $A_i \cap A_j = \emptyset, i \neq j$, alors

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

REMARQUE 5.4. Tout espace probabilisé discret est un espace probabilisé au sens de la Définition 5.3.

REMARQUE 5.5. En général, on a $\mathcal{A} \subsetneq \mathcal{P}(\Omega)$. Typiquement, on fixe \mathcal{A}_0 un ensemble de parties de Ω . Alors

$$\mathcal{A} \equiv \sigma(\mathcal{A}_0) := \bigcap \{\mathcal{B} : \mathcal{B} \text{ tribu, } \mathcal{A}_0 \subset \mathcal{B}\}$$

est une tribu sur Ω , appelée *tribu engendrée* par \mathcal{A}_0 . On remarque que $\mathcal{A} = \sigma(\mathcal{A}_0)$ est la plus petite tribu contenant \mathcal{A}_0

EXEMPLE 5.6. Soit $\Omega = [0, 1[$ et

$$\mathcal{A}_0 = \{[a, b[: 0 \leq a < b < 1\}.$$

Alors $\mathcal{A} = \sigma(\mathcal{A}_0)$ est la *tribu borélienne*, $\mathcal{A} \subsetneq \mathcal{P}(\Omega)$, et il existe une seule probabilité \mathbb{P} sur \mathcal{A} telle que

$$\mathbb{P}([a, b[) = b - a.$$

La mesure \mathbb{P} est appelée *mesure de Lebesgue* sur $[0, 1[$.

Évidemment on a en général:

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{a \in A} \{a\}\right) \neq \sum_{a \in A} \underbrace{\mathbb{P}(\{a\})}_{=0} = 0.$$

REMARQUE 5.7. Une tribu est stable par rapport les opérations dénombrables, i.e., par exemple, si $A_1, A_2, \dots \in \mathcal{A}$ alors aussi

$$\bigcap_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{\infty} A_i^c\right)^c \in \mathcal{A}, \quad \text{et}$$

$$A_{\infty} := \bigcap_n \left(\bigcup_{k \geq n} A_k\right) \in \mathcal{A} \quad (\text{“une infinité d'événements } A_i \text{ se réalisent”}).$$

PROPOSITION 5.8. Soit \mathcal{A} une tribu sur Ω et $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ telle que $\mathbb{P}(\Omega) = 1$ et telle que \mathbb{P} est additive au sens que

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

pour toute famille finie A_1, \dots, A_n d'événements de \mathcal{A} , qui sont 2 à 2 disjoints. Alors les conditions suivantes sont équivalentes:

(1) \mathbb{P} est σ -additive, c.à.d.

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

pour toute famille A_1, A_2, \dots d'événements de \mathcal{A} , 2 à 2 disjoints;

- (2) si $A_n \in \mathcal{A}$ et $A_n \downarrow A$, alors $\mathbb{P}(A_n) \downarrow \mathbb{P}(A)$;
- (3) si $A_n \in \mathcal{A}$ et $A_n \uparrow A$, alors $\mathbb{P}(A_n) \uparrow \mathbb{P}(A)$;
- (4) si $A_n \in \mathcal{A}$ et $A_n \downarrow \emptyset$, alors $\mathbb{P}(A_n) \downarrow \mathbb{P}(\emptyset) = 0$;
- (5) si $A_n \in \mathcal{A}$ et $A_n \uparrow \Omega$, alors $\mathbb{P}(A_n) \uparrow \mathbb{P}(\Omega) = 1$.

DÉMONSTRATION. Comme pour les espaces probabilisés dénombrables. \square

THÉORÈME 5.9 (Expériences de Bernoulli répétées). Soit

$$\Omega = \{\omega = (x_1, x_2, \dots) \mid x_i \in \{0, 1\}\}$$

l'ensemble (non-dénombrable) de 0/1-séquences. Pour $i \in 1, 2, \dots$, on note

$$X_i(\omega) = x_i, \quad \text{si } \omega = (x_1, x_2, \dots).$$

Soit

$\mathcal{A}_n :=$ la tribu engendrée par les événements $\{X_i = 1\}$, $i = 1, 2, \dots, n$, et

$\mathcal{A} :=$ la tribu engendrée par les événements $\{X_i = 1\}$, $i = 1, 2, \dots$

On remarque que $\mathcal{A}_n \subset \mathcal{A}$. La tribu \mathcal{A}_n est appelée "tribu des événements antérieurs à l'instant n " ou "tribu des événements disponibles à l'instant n ".

On fixe $0 \leq p \leq 1$. Alors il existe une probabilité unique \mathbb{P} sur \mathcal{A} telle que:

- (1) $\mathbb{P}\{X_i = 1\} = p$ pour chaque $i = 1, 2, \dots$;
- (2) les événements $\{X_i = 1\}$, $i = 1, 2, \dots$, sont indépendants par rapport à \mathbb{P} .

DÉMONSTRATION. On fixe (x_1, \dots, x_n) avec $x_i \in \{0, 1\}$ et $n \in \{1, 2, \dots\} = \mathbb{N}^*$.

Soit $k := \sum_{i=1}^n x_i$. Alors, en utilisant l'indépendance des événements

$$\{X_i = x_i\}, \quad i = 1, \dots, n,$$

on a nécessairement:

$$\mathbb{P}\{X_1 = x_1, \dots, X_n = x_n\} = \prod_{i=1}^n \mathbb{P}\{X_i = x_i\} = p^k (1-p)^{n-k}.$$

Par conséquent, \mathbb{P} est déterminé sur \mathcal{A}_n , et donc sur $\cup_n \mathcal{A}_n$. Il reste à démontrer que \mathbb{P} , définie sur $\cup_n \mathcal{A}_n$, admet une unique extension ou continuation, notée aussi \mathbb{P} , à la tribu $\sigma(\cup_n \mathcal{A}_n) = \mathcal{A}$ (théorème de Caratheodory; voir cours « Théorie de la Mesure »). \square

EXEMPLE 5.10. (a) Soit $0 < p < 1$. Alors:

$$\mathbb{P}(\{\omega\}) = 0, \quad \forall \omega = (x_1, x_2, \dots) \in \Omega.$$

En effet,

$$\begin{aligned} \mathbb{P}(\{\omega\}) &= \mathbb{P}\left(\bigcap_{i=1}^{\infty} \{X_i = x_i\}\right) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcap_{i=1}^n \{X_i = x_i\}\right) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}\{X_1 = x_1, \dots, X_n = x_n\} \\ &= \lim_{n \rightarrow \infty} p^{k(n)}(1-p)^{n-k(n)} = 0 \quad \text{avec } k(n) = \sum_{i=1}^n x_i. \end{aligned}$$

(b) Soit

$$T : \Omega \rightarrow \mathbb{N}^* \cup \{+\infty\}, \quad T(\omega) := \min\{n \geq 1 : X_n = 1\},$$

le “temps d’attente du premier 1” (avec la convention $\min \emptyset := +\infty$). On fixe $p \in]0, 1]$. Alors

$$\mathbb{P}\{T = n\} = \underbrace{(1-p)^{n-1}}_{n-1 \text{ fois échec}} \underbrace{p}_{\text{succès à l'instant } n},$$

et puis

$$\begin{aligned} \mathbb{P}\{T < \infty\} &= \sum_{n=1}^{\infty} \mathbb{P}\{T = n\} = p \sum_{n=1}^{\infty} (1-p)^{n-1} \\ &= p \sum_{n=0}^{\infty} (1-p)^n = p \frac{1}{1-(1-p)} = 1. \end{aligned}$$

On remarque que T peut prendre la valeur $+\infty$, par exemple pour $\omega = (0, 0, 0, \dots)$, mais

$$\mathbb{P}\{T = \infty\} = 1 - \mathbb{P}\{T < \infty\} = 0.$$

5.3. Variables aléatoires et loi d’une variable aléatoire

Soit \mathcal{B} la tribu borélienne de \mathbb{R} , engendrée par les intervalles $] -\infty, b]$, $b \in \mathbb{R}$. Autrement dit, \mathcal{B} est la plus petite tribu sur \mathbb{R} contenant les intervalles $] -\infty, b]$, $b \in \mathbb{R}$. On note que \mathcal{B} contient aussi les intervalles (pour $a < b$):

$$]a, b] =] -\infty, b] \setminus] -\infty, a],$$

$$]a, b[= \bigcup_n]a, b - \frac{1}{n}],$$

$$[a, b] = \bigcap_n]a - \frac{1}{n}, b].$$

DÉFINITION 5.11. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. On appelle *variable aléatoire* toute fonction mesurable $X : \Omega \rightarrow \mathbb{R}$, c.à.d. telle que

$$X^{-1}(B) \in \mathcal{A}, \quad \forall B \in \mathcal{B}.$$

REMARQUE 5.12. Une application $X \rightarrow \mathbb{R}$ est une variable aléatoire si et seulement si, pour tout $b \in \mathbb{R}$,

$$X^{-1}(]-\infty, b]) = \{X \leq b\} = \{\omega \in \Omega \mid X(\omega) \leq b\} \in \mathcal{A}.$$

En effet, X^{-1} commute avec les réunions et avec les intersections (finies, dénombrables ou infinies, non-dénombrables).

DÉFINITION 5.13. Soit X une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$. La loi \mathbb{P}_X de X est définie par:

$$\mathbb{P}_X(B) := \mathbb{P}\{X \in B\} = \mathbb{P}(X^{-1}(B)) = \{\omega \in \Omega \mid X(\omega) \in B\}, \quad B \in \mathcal{B}.$$

REMARQUE 5.14. La loi \mathbb{P}_X de X est une mesure de probabilité sur $(\mathbb{R}, \mathcal{B})$, c.à.d.

- (1) $\mathbb{P}_X(\mathbb{R}) = \mathbb{P}(X^{-1}(\mathbb{R})) = \mathbb{P}(\Omega) = 1$
- (2) Pour $B_n \in \mathcal{B}$, $B_n \cap B_m = \emptyset$, $n \neq m$, l'on a

$$\begin{aligned} \mathbb{P}_X(\cup_n B_n) &= \mathbb{P}(X^{-1}(\cup_n B_n)) \\ &= \mathbb{P}(\cup_n X^{-1}(B_n)) \\ &= \sum_n \mathbb{P}(X^{-1}(B_n)) \quad (\text{car } \mathbb{P} \text{ est une probabilité}) \\ &= \sum_n \mathbb{P}_X(B_n). \end{aligned}$$

REMARQUE 5.15. La loi \mathbb{P}_X de X est entièrement caractérisée par la *fonction de répartition* de X définie par:

$$F_X: \mathbb{R} \rightarrow [0, 1], \quad F_X(b) = \mathbb{P}_X(]-\infty, b]) = \mathbb{P}\{X \leq b\}.$$

EXEMPLE 5.16. On trouve, par exemple,

$$\begin{aligned} \mathbb{P}\{X \in]a, b]\} &= \mathbb{P}_X(]a, b]) \\ &= \mathbb{P}_X(]-\infty, b] \setminus]-\infty, a]) \\ &= \mathbb{P}_X(]-\infty, b]) - \mathbb{P}_X(]-\infty, a]) \\ &= F_X(b) - F_X(a), \end{aligned}$$

et

$$\begin{aligned} \mathbb{P}\{X = a\} &= \mathbb{P}_X(\{a\}) \\ &= \mathbb{P}_X\left(\bigcap_{n=1}^{\infty}]a - \frac{1}{n}, a]\right) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}_X\left(]a - \frac{1}{n}, a]\right) \\ &= \lim_{n \rightarrow \infty} \left[F_X(a) - F_X\left(a - \frac{1}{n}\right) \right] \\ &= F_X(a) - \lim_{n \rightarrow \infty} F_X\left(a - \frac{1}{n}\right) \\ &= F_X(a) - F_X(a-). \end{aligned}$$

5.4. Propriétés de la fonction de répartition

PROPOSITION 5.17. *Soit X une variable aléatoire et F_X sa fonction de répartition.*

- (1) F_X est croissante, c.à.d. $x \leq y \Rightarrow F_X(x) \leq F_X(y)$;
- (2) F_X est continue à droite avec une limite à gauche en tout point;
- (3) $\lim_{x \rightarrow -\infty} F_X(x) = 0$ et $\lim_{x \rightarrow \infty} F_X(x) = 1$.

DÉMONSTRATION.

- (1) La monotonie est claire.
- (2) Soient $0 < h_n \downarrow 0$. Alors

$$\begin{aligned} 0 &= \mathbb{P}_X(\emptyset) = \mathbb{P}_X\left(\bigcap_n]a, a + h_n]\right) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}_X(]a, a + h_n]) \\ &= \lim_{n \rightarrow \infty} F_X(a + h_n) - F_X(a) \end{aligned}$$

et

$$\begin{aligned} \mathbb{P}\{X = a\} &= \mathbb{P}_X\{a\} \\ &= \mathbb{P}_X\left(\bigcap_n]a - h_n, a]\right) \\ &= \lim_{n \rightarrow \infty} (F_X(a) - F_X(a - h_n)), \end{aligned}$$

c.à.d,

$$\lim_{n \rightarrow \infty} F_X(a - h_n) = F_X(a) - \mathbb{P}\{X = a\};$$

en particulier cette limite existe.

- (3) On trouve

$$\begin{aligned} \lim_{x \rightarrow -\infty} F_X(x) &= \lim_{n \rightarrow \infty} F_X(-n) = \mathbb{P}_X\left(\bigcap_n]-\infty, -n]\right) = \mathbb{P}_X(\emptyset) = 0; \\ \lim_{x \rightarrow +\infty} F_X(x) &= \lim_{n \rightarrow \infty} F_X(n) = \mathbb{P}_X\left(\bigcup_n]-\infty, n]\right) = \mathbb{P}_X(\mathbb{R}) = 1. \end{aligned}$$

□

REMARQUE 5.18. Soit X une variable aléatoire et F_X sa fonction de répartition.

- (1) Pour tout $x \in \mathbb{R}$ l'on a

$$F_X(x) - F_X(x-) = \mathbb{P}_X(\{x\}) = \mathbb{P}\{X = x\}.$$

En particulier, F_X est une fonction continue si et seulement si

$$\mathbb{P}\{X = x\} = 0, \quad \forall x \in \mathbb{R}.$$

- (2) F_X admet au plus un nombre dénombrable de points de discontinuité.

DÉMONSTRATION. (1) est clair.

Pour (2) on note D l'ensemble des points de discontinuité de F_X et

$$D_n = \left\{ x \in \mathbb{R} : F_X(x) - F_X(x-) \geq \frac{1}{n} \right\}$$

l'ensemble des points de discontinuité avec un saut d'amplitude $\geq 1/n$. Alors

$$\text{card}(D_n) \leq n,$$

puisque $0 \leq F_X \leq 1$. D'autre part, on a évidemment $D = \cup_n D_n$, ce qui montre que D est dénombrable. \square

DÉFINITION 5.19. Soit X une variable aléatoire. On note \mathbb{P}_X la loi et F_X la fonction de répartition de X , c.à.d.

$$\mathbb{P}_X(A) = \mathbb{P}\{X \in A\}, \quad A \in \mathcal{B}(\mathbb{R}),$$

$$F_X(a) = \mathbb{P}\{X \leq a\} = \mathbb{P}_X(]-\infty, a]), \quad a \in \mathbb{R}.$$

Une *densité de X* est une fonction $f_X: \mathbb{R} \rightarrow \mathbb{R}_+$ qui est intégrable au sens de Riemann et telle que

$$F_X(a) = \int_{-\infty}^a f_X(t) dt, \quad a \in \mathbb{R}.$$

En particulier, on a:

$$\int_{-\infty}^{\infty} f_X(t) dt = 1.$$

DÉFINITION 5.20. Soit X une variable aléatoire. On dit que X est une *variable aléatoire continue*, si X admet une densité, c.à.d. s'il existe une fonction $f_X: \mathbb{R} \rightarrow \mathbb{R}_+$ telle que

$$F_X(a) = \int_{-\infty}^a f_X(t) dt, \quad a \in \mathbb{R}.$$

REMARQUE 5.21. Une densité f_X n'est pas forcément une fonction continue. Mais, en cas que f_X est continue, on déduit de $F_X(a) = \int_{-\infty}^a f_X(t) dt$ la relation

$$f_X(t) = \left. \frac{d}{da} \right|_{a=t} F_X(a).$$

5.5. Quelques lois classiques

EXEMPLE 5.22. Soit X une variable aléatoire de densité f_X .

(1) *La loi uniforme* sur $[a, b]$ avec $a < b$.

Cette loi a pour densité

$$f_X(t) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq t \leq b \\ 0 & \text{sinon.} \end{cases}$$

La fonction de répartition est donnée par

$$F_X(x) = \int_{-\infty}^x f_X(t) dt = \begin{cases} 0 & \text{si } x \leq a \\ \frac{x-a}{b-a} & \text{si } a < x \leq b \\ 1 & \text{si } b < x. \end{cases}$$



Interprétation: Cette loi constitue “l’analogue continu” de la loi uniforme sur un ensemble d’entiers $\{p, p + 1, \dots, p + q\}$. Chaque point de $[a, b]$ est “également vraisemblable”.

(2) *Loi exponentielle* de paramètre $\alpha > 0$.
Cette loi a pour densité

$$f_X(t) = \begin{cases} \alpha e^{-\alpha t} & \text{si } t \geq 0 \\ 0 & \text{sinon,} \end{cases}$$

c.à.d.

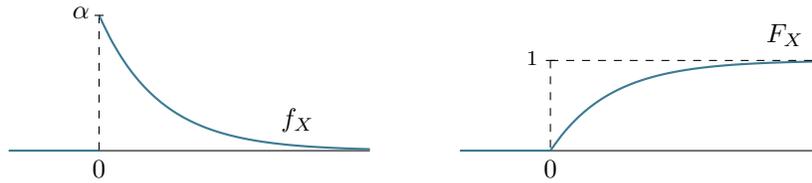
$$f_X(t) = \alpha e^{-\alpha t} \mathbb{1}_{\mathbb{R}_+}(t).$$

On vérifie que

$$\int_{-\infty}^{\infty} f_X(t) dt = \int_0^{\infty} \alpha e^{-\alpha t} dt = -e^{-\alpha t} \Big|_0^{\infty} = 1,$$

et de même façon

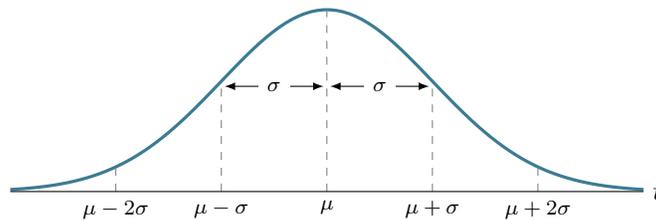
$$F_X(x) = \int_{-\infty}^x f_X(t) dt = \begin{cases} 1 - e^{-\alpha x} & \text{si } x \geq 0 \\ 0 & \text{sinon.} \end{cases}$$



Interprétation: Cette loi constitue “l’analogue continu” de la loi géométrique.

(3) *La loi normale* $N(\mu, \sigma^2)$ de paramètres μ et σ^2 ($\mu \in \mathbb{R}, \sigma > 0$). La densité est donnée par

$$f_{\mu, \sigma^2}(t) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}, \quad t \in \mathbb{R}.$$





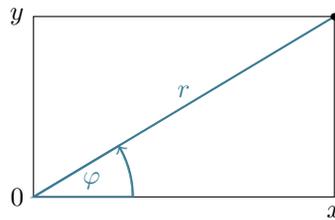
Pour la preuve que

$$\int_{-\infty}^{\infty} f_{\mu, \sigma^2}(t) dt = 1$$

on utilisera les coordonnées polaires.

LEMME 5.23 (Coordonnées polaires de \mathbb{R}^2). Pour $(x, y) \in \mathbb{R}^2$ on pose

$$r = \sqrt{x^2 + y^2}, \quad \tan \varphi = \frac{y}{x}.$$



Alors

$$dxdy = r d\varphi dr,$$

c.à.d. en particulier

$$\int_{\mathbb{R}^2} h(x^2 + y^2) dxdy = \int_0^{\infty} \left(\int_0^{2\pi} h(r^2) r d\varphi \right) dr = 2\pi \int_0^{\infty} h(r^2) r dr.$$

D'où on obtient

$$\begin{aligned} \left(\int_{-\infty}^{\infty} e^{-\frac{t^2}{2\sigma^2}} dt \right)^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} dxdy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2\sigma^2}} dxdy \\ &= 2\pi \int_0^{\infty} e^{-\frac{r^2}{2\sigma^2}} r dr = 2\pi \left[-\sigma^2 e^{-\frac{r^2}{2\sigma^2}} \right]_{r=0}^{r \rightarrow +\infty} = 2\pi\sigma^2. \end{aligned}$$

Autement dit,

$$\int_{-\infty}^{\infty} e^{-\frac{t^2}{2\sigma^2}} dt = \sqrt{2\pi}\sigma,$$

et donc,

$$\int_{-\infty}^{\infty} f_{\mu, \sigma^2}(t) dt = \int_{-\infty}^{\infty} f_{0, \sigma^2}(t) dt = 1.$$

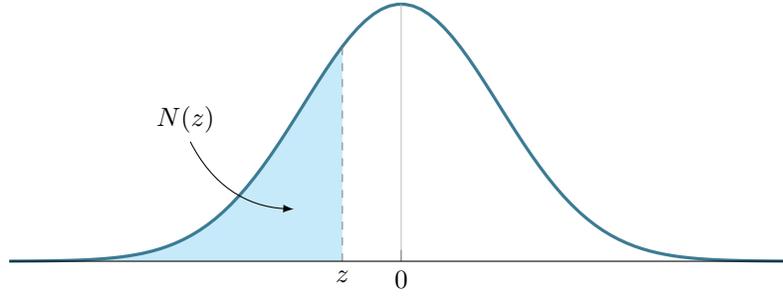
Ainsi:

$$\begin{aligned} F_{\mu, \sigma^2}(x) &= \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \\ &= \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \quad (\text{avec la substitution } z = \frac{t-\mu}{\sigma}) \\ &= N\left(\frac{x-\mu}{\sigma}\right) \end{aligned}$$

où

$$N(z) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}} dt, \quad z \in \mathbb{R},$$

est la fonction de répartition d'une variable normale centrée réduite.



PROPOSITION 5.24 (Caractérisation de la loi exponentielle). *Soit X une variable aléatoire non-négative avec une densité continue f_X . Alors les deux propriétés suivantes sont équivalentes:*

- (1) X suit une loi exponentielle, i.e. il existe $\alpha > 0$ tel que $f_X(t) = \alpha e^{-\alpha t} \mathbb{1}_{\mathbb{R}_+}(t)$.
- (2) X est sans mémoire, i.e. $\forall s, t \geq 0$,

$$\mathbb{P}\{X > s + t \mid X > t\} = \mathbb{P}\{X > s\}.$$

LEMME 5.25. *Soit $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$ une fonction continue telle que*

$$\phi(s + t) = \phi(s)\phi(t), \quad \forall s, t \geq 0.$$

Si $\phi \neq 0$, alors il existe $\lambda \in \mathbb{R}$ tel que

$$\phi(t) = e^{\lambda t}, \quad \forall t \geq 0.$$

DÉMONSTRATION. Pour $n \in \mathbb{N}^* = \{1, 2, \dots\}$, on obtient

$$\phi(n) = \phi((n-1) + 1) = \phi(n-1)\phi(1) = \dots = \phi(1)^n$$

et

$$\phi(1) = \phi\left(n \cdot \frac{1}{n}\right) = \phi\left(\underbrace{\frac{1}{n} + \dots + \frac{1}{n}}_{n \text{ fois}}\right) = \phi\left(\frac{1}{n}\right)^n.$$

Donc $\phi(1/n) = \phi(1)^{1/n}$. Plus généralement, soit $q \in \mathbb{Q}$, $q > 0$, tel que

$$q = \frac{n}{m}, \quad n, m \in \mathbb{N}^*.$$

Alors

$$\begin{aligned}\phi(q) &= \phi\left(\frac{n}{m}\right) = \phi\left(\underbrace{\frac{1}{m} + \dots + \frac{1}{m}}_{n \text{ fois}}\right) \\ &= \phi\left(\frac{1}{m}\right)^n = \left(\phi(1)^{1/m}\right)^n \\ &= \phi(1)^{n/m} = \phi(1)^q.\end{aligned}$$

On constate que $\phi(1) = \phi(1/2)^2 \geq 0$. En plus, on a $\phi(1) > 0$, car si $\phi(1) = 0$, alors $\phi(q) = \phi(1)^q = 0$ pour tout $0 < q \in \mathbb{Q}$, et donc $\phi \equiv 0$ par la continuité de ϕ . On pose

$$\lambda := \log \phi(1),$$

c.à.d., $\phi(1) = e^\lambda$. Ainsi:

$$\phi(t) = (e^\lambda)^t = e^{\lambda t}, \quad \forall t \in \mathbb{Q},$$

et en utilisant la continuité de ϕ ,

$$\phi(t) = e^{\lambda t}, \quad \forall t \geq 0. \quad \square$$

DÉMONSTRATION (de Proposition 5.24). Soit $X: \Omega \rightarrow \mathbb{R}_+$ la variable variable aléatoire et

$$F_X(x) = \mathbb{P}\{X \leq x\} = \int_0^x f_X(t) dt, \quad x \geq 0,$$

sa fonction de répartition.

(1) \Rightarrow (2): Supposons que la variable X suit une loi exponentielle de paramètre $\alpha > 0$, c.à.d.

$$f_X(t) = \begin{cases} \alpha e^{-\alpha t}, & t \geq 0 \\ 0, & t < 0 \end{cases} \quad \text{et} \quad F_X(x) = \begin{cases} 1 - \alpha e^{-\alpha x}, & x \geq 0 \\ 0, & x \leq 0. \end{cases}$$

Alors, l'on a pour tout $\forall s, t \geq 0$:

$$\begin{aligned}\mathbb{P}\{X > s+t \mid X > t\} &= \frac{\mathbb{P}\{X > s+t, X > t\}}{\mathbb{P}\{X > t\}} \\ &= \frac{\mathbb{P}\{X > s+t\}}{\mathbb{P}\{X > t\}} = \frac{1 - \mathbb{P}\{X \leq s+t\}}{1 - \mathbb{P}\{X \leq t\}} \\ &= \frac{1 - F_X(s+t)}{1 - F_X(t)} \\ &= \frac{e^{-\alpha(s+t)}}{e^{-\alpha t}} = e^{-\alpha s} \\ &= 1 - F_X(s) = \mathbb{P}\{X > s\}.\end{aligned}$$

(1) \Leftarrow (2): On sait que $\mathbb{P}\{X > s+t \mid X > t\} = \mathbb{P}\{X > s\}$ pour $s, t \geq 0$. Alors

$$\underbrace{\mathbb{P}\{X > s+t\}}_{=\phi(s+t)} = \underbrace{\mathbb{P}\{X > s\}}_{=\phi(s)} \underbrace{\mathbb{P}\{X > t\}}_{=\phi(t)}$$

où $\phi(x) = \mathbb{P}\{X > x\} = \int_x^\infty f_X(r) dr$.

En utilisant que $\phi \neq 0$, on obtient par Lemme 5.25,

$$\phi(x) = e^{\lambda x} \quad \text{avec } \lambda \in \mathbb{R}.$$

Car $\phi(x) \rightarrow 0$, quand $x \rightarrow \infty$, on constate que $\lambda < 0$. Posons donc $\alpha := -\lambda > 0$, on a

$$\phi(x) = \mathbb{P}\{X > x\} = e^{-\alpha x}, \quad x \geq 0$$

et

$$F_X(x) = \mathbb{P}\{X \leq x\} = 1 - e^{-\alpha x}, \quad x \geq 0.$$

Autrement dit, la variable X suit la loi exponentielle de paramètre $\alpha > 0$. \square

5.6. Loi et densité commune

DÉFINITION 5.26. On appelle *vecteur aléatoire* tout n -uple $X = (X_1, \dots, X_n)$ de variables aléatoires X_1, \dots, X_n , i.e., $X: \Omega \rightarrow \mathbb{R}^n$ variable aléatoire au sens que

$$\{X \in B\} \equiv X^{-1}(B) \in \mathcal{A}, \quad \text{pour tout } B \in \mathcal{B}(\mathbb{R}^n),$$

où $\mathcal{B}(\mathbb{R}^n)$ désigne la tribu borélienne de \mathbb{R}^n .

(1) La loi \mathbb{P}_X de X est définie par:

$$\mathbb{P}_X(B) := \mathbb{P}\{X \in B\}, \quad B \in \mathcal{B}(\mathbb{R}^n).$$

(2) On appelle *fonction de répartition* de X la fonction:

$$F_X: \mathbb{R}^n \rightarrow [0, 1], \quad F_X(a_1, \dots, a_n) := \mathbb{P}\{X_1 \leq a_1, \dots, X_n \leq a_n\}.$$

(3) Soit $f_X: \mathbb{R}^n \rightarrow \mathbb{R}_+$ une fonction intégrable (au sens de Riemann) telle que

$$\int_{\mathbb{R}^n} f_X(x) dx = 1.$$

On dit que X admet f_X pour densité si, pour tout $(a_1, \dots, a_n) \in \mathbb{R}^n$:

$$\begin{aligned} F_X(a_1, \dots, a_n) &= \int_{]-\infty, a_1] \times \dots \times]-\infty, a_n]} f_X(x) dx \\ &= \int_{-\infty}^{a_1} \dots \int_{-\infty}^{a_n} f_X(x_1, \dots, x_n) dx_1 \dots dx_n. \end{aligned}$$

La fonction f_X est appelé *densité commune* de $X = (X_1, \dots, X_n)$.

REMARQUE 5.27. Soient X_1, \dots, X_n de variables aléatoire.

(1) Les conditions suivantes sont équivalentes:

- (a) la famille (X_1, \dots, X_n) est indépendante;
- (b) $\{X_1 \in B_1\}, \dots, \{X_n \in B_n\}$ sont indépendants, $\forall B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$;
- (c) $\mathbb{P}\{X_1 \in B_1, \dots, X_n \in B_n\} = \prod_{i=1}^n \mathbb{P}\{X_i \in B_i\}$, $\forall B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$;
- (d) $\mathbb{P}\{X_1 \leq a_1, \dots, X_n \leq a_n\} = \prod_{i=1}^n \mathbb{P}\{X_i \leq a_i\}$, $\forall a_1, \dots, a_n \in \mathbb{R}$;
- (e) $F_X(a_1, \dots, a_n) = F_{X_1}(a_1) \cdot \dots \cdot F_{X_n}(a_n)$, $\forall a_1, \dots, a_n \in \mathbb{R}$.

(2) Pour tout $i \in \{1, \dots, n\}$ soit f_{X_i} une densité de X_i . Les conditions suivantes sont équivalentes:

- (a) les variables X_1, \dots, X_n sont indépendantes;
- (b) la fonction

$$f: \mathbb{R}^n \rightarrow \mathbb{R}_+, \quad t \mapsto f(t) := f_{X_1}(t_1) \cdot \dots \cdot f_{X_n}(t_n)$$

est une densité commune pour $X = (X_1, \dots, X_n)$.

DÉMONSTRATION. Il suffit à démontrer (2).

(b) \Rightarrow (a): Soit

$$t \mapsto f_X(t) = f_{X_1}(t_1) \cdot \dots \cdot f_{X_n}(t_n), \quad t = (t_1, \dots, t_n) \in \mathbb{R}^n,$$

une densité commune. Alors, pour tout $(a_1, \dots, a_n) \in \mathbb{R}$,

$$\begin{aligned} F_X(a_1, \dots, a_n) &= \mathbb{P}\{X_1 \leq a_1, \dots, X_n \leq a_n\} \\ &= \int_{-\infty}^{a_1} \dots \int_{-\infty}^{a_n} f_X(t_1, \dots, t_n) dt_1 \dots dt_n \\ &= \int_{-\infty}^{a_1} \dots \int_{-\infty}^{a_n} f_{X_1}(t_1) \cdot \dots \cdot f_{X_n}(t_n) dt_1 \dots dt_n \\ &= \prod_{i=1}^n \int_{-\infty}^{a_i} f_{X_i}(t_i) dt_i = \prod_{i=1}^n F_{X_i}(a_i). \end{aligned}$$

Autrement dit, la famille (X_1, \dots, X_n) est indépendante.

(a) \Rightarrow (b): Soit maintenant la famille (X_1, \dots, X_n) indépendante. Alors

$$\begin{aligned} F_X(a_1, \dots, a_n) &= F_{X_1}(a_1) \cdot \dots \cdot F_{X_n}(a_n) \\ &= \prod_{i=1}^n \mathbb{P}\{X_i \leq a_i\} = \prod_{i=1}^n \int_{-\infty}^{a_i} f_{X_i}(t_i) dt_i \\ &= \int_{-\infty}^{a_1} \dots \int_{-\infty}^{a_n} \underbrace{f_{X_1}(t_1) \cdot \dots \cdot f_{X_n}(t_n)}_{=: f(t_1, \dots, t_n)} dt_1 \dots dt_n. \end{aligned}$$

La fonction $f := f_{X_1} \cdot \dots \cdot f_{X_n}$ est donc une densité commune. \square

L'espérance de variables aléatoires continues

6.1. L'espérance d'une variable continue

Problème Comment définir l'espérance $\mathbb{E}[X]$ d'une variable aléatoire continue X de densité f_X ?

REMARQUE 6.1 (Cas discret). Soit X une variable aléatoire non-négative à valeurs dans un ensemble dénombrable $\{x_1, x_2, \dots\} \subset [0, \infty[$. Alors:

$$\begin{aligned}\mathbb{E}[X] &= \sum_{i=1}^{\infty} x_i \mathbb{P}\{X = x_i\} = \sum_{i=1}^{\infty} x_i p_i; \\ F_X(t) &= \mathbb{P}\{X \leq t\} = \sum_{i: x_i \leq t} p_i, \quad \text{où } p_i := \mathbb{P}\{X = x_i\}.\end{aligned}$$

En particulier,

$$\begin{aligned}\mathbb{E}[X] &= \sum_{i=1}^{\infty} p_i \int_0^{x_i} 1 dt = \sum_{i=1}^{\infty} p_i \int_0^{\infty} \mathbb{1}_{[0, x_i[}(t) dt \\ &= \sum_{i=1}^{\infty} \int_0^{\infty} p_i \mathbb{1}_{[0, x_i[}(t) dt = \int_0^{\infty} \sum_{i: x_i > t} p_i dt = \int_0^{\infty} \mathbb{P}\{X > t\} dt.\end{aligned}$$

On remarque que la formule

$$\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}\{X > t\} dt$$

a sens pour une variable aléatoire $X \geq 0$ quelconque (en ce compris le cas où la variable X n'est pas discrète ni continue).

DÉFINITION 6.2 (Cas général). Soit $X: \Omega \rightarrow [0, \infty[$ une variable aléatoire non-négative. On pose:

$$\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}\{X > t\} dt = \int_0^{\infty} (1 - F_X(t)) dt \in [0, \infty].$$

En *cas général* d'une variable aléatoire

$$X = X_+ - X_-$$

on définit

$$\mathbb{E}[X] := \mathbb{E}[X_+] - \mathbb{E}[X_-], \quad \text{si } \mathbb{E}[X_+] < \infty \text{ ou } \mathbb{E}[X_-] < \infty.$$

PROPOSITION 6.3. Soit X une variable aléatoire de densité f_X . Alors

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} s f_X(s) ds \quad (\text{si l'intégrale est bien définie}).$$

DÉMONSTRATION. (1) Traitons d'abord le cas $X \geq 0$:

$$\begin{aligned}
 \mathbb{E}[X] &= \int_0^\infty \mathbb{P}\{X > t\} dt \\
 &= \int_0^\infty \left(\int_t^\infty f_X(s) ds \right) dt \\
 &= \int_0^\infty \left(\int_0^\infty \mathbb{1}_{\{s \geq t\}}(s) f_X(s) ds \right) dt \\
 &= \int_0^\infty \left(\int_0^\infty \mathbb{1}_{[0,s]}(t) f_X(s) dt \right) ds \\
 &= \int_0^\infty \left(\int_0^s f_X(s) dt \right) ds \\
 &= \int_0^\infty s f_X(s) ds.
 \end{aligned}$$

(2) Pour le cas général:

$$X = X_+ - X_-$$

on remarque d'abord que

$$\begin{aligned}
 \mathbb{E}[X_+] &= \int_0^\infty \mathbb{P}\{X_+ > t\} dt \\
 &= \int_0^\infty \mathbb{P}\{X > t\} dt \\
 &= \int_0^\infty s f_X(s) ds \quad (\text{avec le même calcul que ci-dessus})
 \end{aligned}$$

et

$$\begin{aligned}
 \mathbb{E}[X_-] &= \int_0^\infty \mathbb{P}\{X_- > t\} dt = \int_0^\infty \mathbb{P}\{X < -t\} dt \\
 &= \int_0^\infty \left(\int_{-\infty}^{-t} f_X(s) ds \right) dt \\
 &= \int_0^\infty \left(\int_{-\infty}^0 \mathbb{1}_{\{s \leq -t\}} f_X(s) ds \right) dt \\
 &= \int_{-\infty}^0 \left(\int_0^\infty \mathbb{1}_{\{t \leq -s\}} dt \right) f_X(s) ds \\
 &= \int_{-\infty}^0 \left(\int_0^\infty \mathbb{1}_{[0,-s]}(t) dt \right) f_X(s) ds \\
 &= \int_{-\infty}^0 (-s) f_X(s) ds = - \int_{-\infty}^0 s f_X(s) ds.
 \end{aligned}$$

Donc on a

$$\begin{aligned}
 \mathbb{E}[X_+] &= \int_0^\infty \mathbb{P}\{X_+ > t\} dt = \int_0^\infty s f_X(s) ds \\
 \mathbb{E}[X_-] &= \int_0^\infty \mathbb{P}\{X_- > t\} dt = - \int_{-\infty}^0 s f_X(s) ds,
 \end{aligned}$$

et puis

$$\mathbb{E}[X] = \mathbb{E}[X_+] - \mathbb{E}[X_-] = \int_{-\infty}^{\infty} s f_X(s) ds,$$

si l'intégrale est bien définie. \square

PROPOSITION 6.4. *Soit X une variable aléatoire de densité f_X et soit $h: \mathbb{R} \rightarrow \mathbb{R}$ une fonction non-négative ou bornée. Alors*

$$\mathbb{E}[h(X)] = \int_{-\infty}^{\infty} h(s) f_X(s) ds.$$

En particulier, pour $p = 1, 2, \dots$,

$$\mathbb{E}[X^p] = \int_{-\infty}^{\infty} s^p f_X(s) ds \quad (\text{si l'intégrale est bien définie}),$$

et

$$\begin{aligned} \text{var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{\infty} (s - \mathbb{E}[X])^2 f_X(s) ds \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \int_{-\infty}^{\infty} s^2 f_X(s) ds - \left(\int_{-\infty}^{\infty} s f_X(s) ds \right)^2, \\ \sigma(X) &= \sqrt{\text{var}(X)}. \end{aligned}$$

DÉMONSTRATION. Sans restrictions il suffit se ramener au cas que $h \geq 0$. Alors

$$\begin{aligned} \mathbb{E}[h(X)] &= \int_0^{\infty} \mathbb{P}\{h(X) > t\} dt \\ &= \int_0^{\infty} \left(\int_{\{s: h(s) > t\}} f_X(s) ds \right) dt \\ &= \int_0^{\infty} \left(\int_0^{\infty} \mathbb{1}_{\{s: h(s) > t\}}(s) f_X(s) ds \right) dt \\ &= \int_0^{\infty} \left(\int_0^{\infty} \mathbb{1}_{[0, h(s)[(t) dt} \right) f_X(s) ds \\ &= \int_0^{\infty} h(s) f_X(s) ds. \quad \square \end{aligned}$$

EXEMPLE 6.5 (Loi uniforme sur $[a, b]$). Pour les moments d'une variable aléatoire X suivant la loi uniforme sur $[a, b]$, on a

$$\mathbb{E}[X^p] = \int_{-\infty}^{\infty} x^p f_X(x) dx = \frac{1}{b-a} \int_a^b x^p dx = \frac{1}{p+1} \frac{b^{p+1} - a^{p+1}}{b-a}.$$

On utilise que

$$f_X(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x).$$

En particulier,

$$\mathbb{E}[X] = \frac{1}{2} \frac{b^2 - a^2}{b-a} = \frac{a+b}{2} \quad \text{et} \quad \mathbb{E}[X^2] = \frac{1}{3} \frac{b^3 - a^3}{b-a}.$$

On constate que

$$\text{var}(X) = \frac{(b-a)^2}{12}.$$

EXEMPLE 6.6 (Loi exponentielle de paramètre α). Pour les moments d'une variable aléatoire X suivant la loi exponentielle de paramètre α , on a

$$\mathbb{E}[X^p] = \int_{-\infty}^{\infty} x^p f_X(x) dx = \int_0^{\infty} x^p \alpha e^{-\alpha x} dx.$$

On utilise que

$$f_X(x) = \alpha e^{-\alpha x} \mathbf{1}_{\mathbb{R}_+}(x).$$

La formule d'intégration par parties:

$$\int_0^{\infty} h \cdot g' = h \cdot g \Big|_0^{\infty} - \int_0^{\infty} h' \cdot g,$$

avec $h(x) = x^p$ et $g(x) = -e^{-\alpha x}$, donne

$$\mathbb{E}[X^p] = p \int_0^{\infty} x^{p-1} e^{-\alpha x} dx = \frac{p}{\alpha} \mathbb{E}[X^{p-1}] = \dots = \frac{p!}{\alpha^p}.$$

En particulier,

$$\mathbb{E}[X] = \frac{1}{\alpha}, \quad \mathbb{E}[X^2] = \frac{2}{\alpha^2} \quad \text{et} \quad \text{var}(x) = \frac{1}{\alpha^2}.$$

EXEMPLE 6.7 (Loi normale de paramètres μ et σ^2).

Soit X une variable aléatoire de loi normale $N(\mu, \sigma^2)$, c.à.d.

$$f_X(x) = f_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Alors

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_{\mu, \sigma^2}(x) dx = \mu + \underbrace{\int_{-\infty}^{\infty} (x-\mu) f_{\mu, \sigma^2}(x) dx}_{=0} = \mu;$$

$$\begin{aligned} \text{var}(X) &= \int_{-\infty}^{\infty} (x-\mu)^2 f_{\mu, \sigma^2}(x) dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-(x-\mu)^2/2\sigma^2} dx \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-y^2/2} dy \quad (\text{avec } y = (x-\mu)/\sigma^2) \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \left[-y e^{-y^2/2} \right]_{-\infty}^{\infty} + \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy \quad (\text{intégration par parties}) \\ &= \sigma^2. \end{aligned}$$

6.2. La loi d'une somme de variables aléatoires indépendantes

PROPOSITION 6.8. Soient X_1 et X_2 deux variables aléatoires indépendantes admettant des densités f_{X_1} et f_{X_2} , i.e., pour tout $x \in \mathbb{R}$,

$$F_{X_i}(x) = \mathbb{P}\{X_i \leq x\} = \int_{-\infty}^x f_{X_i}(t) dt, \quad i = 1, 2.$$

Alors la variable aléatoire

$$X = X_1 + X_2$$

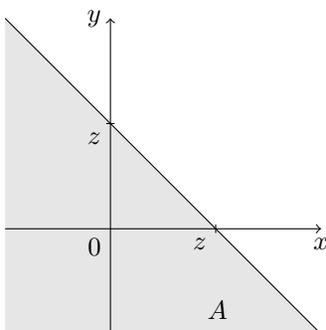
admet une densité f_X donnée par la convolution $f_{X_1} * f_{X_2}$ des fonctions f_{X_1} et f_{X_2} , c.a.d.

$$\begin{aligned} f_X(t) &= (f_{X_1} * f_{X_2})(t) \\ &= \int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(t-x) dx \\ &= \int_{-\infty}^{\infty} f_{X_1}(t-y) f_{X_2}(y) dy, \quad t \in \mathbb{R}. \end{aligned}$$

DÉMONSTRATION. D'abord on note que

$$\mathbb{P}\{X_1 + X_2 \leq z\} = \mathbb{P}\{(X_1, X_2) \in A\},$$

où $A = \{(x, y) \in \mathbb{R}^2 \mid x + y \leq z\}$.



Donc

$$\begin{aligned} \mathbb{P}\{(X_1, X_2) \in A\} &= \iint_A f_{(X_1, X_2)}(x, y) dx dy \\ &= \iint_A f_{X_1}(x) f_{X_2}(y) dx dy \quad (\text{puisque } X_1, X_2 \text{ sont indépendantes}) \\ &= \int_{-\infty}^{\infty} f_{X_1}(x) \left(\int_{-\infty}^{z-x} f_{X_2}(y) dy \right) dx \\ &= \int_{-\infty}^{\infty} f_{X_1}(x) \left(\int_{-\infty}^z f_{X_2}(y-x) dy \right) dx \\ &= \int_{-\infty}^z \left(\int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(y-x) dx \right) dy \\ &= \int_{-\infty}^z (f_{X_1} * f_{X_2})(y) dy, \quad \text{pour tout } z \in \mathbb{R}. \end{aligned}$$

Autrement dit, $f_{X_1} * f_{X_2}$ est une densité de $X = X_1 + X_2$. \square

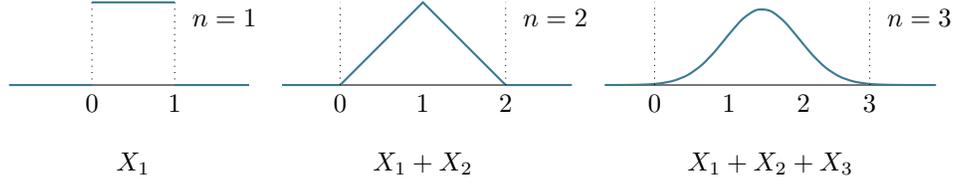
EXEMPLE 6.9. Soient X_1, X_2 indépendantes de loi uniforme sur $[0, 1]$. Alors $X_1 + X_2$ admet la densité:

$$f_{X_1+X_2}(t) = \begin{cases} 0, & t \leq 0 \\ t, & 0 \leq t \leq 1 \\ 2-t, & 1 \leq t \leq 2 \\ 0, & t \geq 2. \end{cases}$$

En effet, on a

$$\begin{aligned} f_{X_1+X_2}(t) &= \int_{-\infty}^{\infty} \mathbb{1}_{[0,1]}(x) \mathbb{1}_{[0,1]}(t-x) dx \\ &= \int_{-\infty}^{\infty} \mathbb{1}_{[0,1] \cap [t-1,t]}(x) dx = |[0,1] \cap [t-1,t]|. \end{aligned}$$

Plus généralement, soit $X = X_1 + X_2 + \dots + X_n$ avec X_i indépendantes et de loi uniforme sur $[0, 1]$. De la même manière on peut déterminer la densité de X .



REMARQUE 6.10 (Linéarité de l'espérance). Soient X, Y deux variables aléatoires et $f_{(X,Y)}$ une densité commune de (X, Y) . Alors on a pour la densité de X et Y , respectivement $X + Y$, les formules suivantes:

$$\begin{aligned} f_X(s) &= \int_{-\infty}^{+\infty} f_{(X,Y)}(s, t) dt, \\ f_Y(t) &= \int_{-\infty}^{+\infty} f_{(X,Y)}(s, t) ds, \\ f_{X+Y}(t) &= \int_{-\infty}^{+\infty} f_{(X,Y)}(s, t-s) ds. \end{aligned}$$

(En effet, pour tout $a \in \mathbb{R}$ on a

$$\begin{aligned} F_X(a) &= \mathbb{P}\{X \leq a, Y < +\infty\} \\ &= \int_{-\infty}^a \int_{-\infty}^{+\infty} f_{(X,Y)}(s, t) ds dt \\ &= \int_{-\infty}^a \left(\int_{-\infty}^{+\infty} f_{(X,Y)}(s, t) dt \right) ds = \int_{-\infty}^a f_X(s) ds \end{aligned}$$

ou

$$\begin{aligned} F_{X+Y}(a) &= \mathbb{P}\{X + Y \leq a\} \\ &= \mathbb{P}\{(X, Y) \in A\} \quad \text{où } A := \{(x, y) \in \mathbb{R}^2 \mid x + y \leq a\} \\ &= \iint_A f_{(X,Y)}(s, t) ds dt \\ &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{a-s} f_{(X,Y)}(s, t) dt \right) ds \\ &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^a f_{(X,Y)}(s, t-s) dt \right) ds \\ &= \int_{-\infty}^a \left(\int_{-\infty}^{+\infty} f_{(X,Y)}(s, t-s) ds \right) dt = \int_{-\infty}^a f_{X+Y}(t) dt \end{aligned}$$

etc).

En particulier, on trouve

$$\begin{aligned}
\mathbb{E}[X + Y] &= \int_{-\infty}^{+\infty} t f_{X+Y}(t) dt \\
&= \int_{-\infty}^{+\infty} t \left(\int_{-\infty}^{+\infty} f_{(X,Y)}(s, t-s) ds \right) dt \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} t f_{(X,Y)}(s, t-s) ds dt \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (s+t) f_{(X,Y)}(s, t) ds dt \\
&= \int_{-\infty}^{+\infty} s \left(\int_{-\infty}^{+\infty} f_{(X,Y)}(s, t) dt \right) ds + \int_{-\infty}^{+\infty} t \left(\int_{-\infty}^{+\infty} f_{(X,Y)}(s, t) ds \right) dt \\
&= \int_{-\infty}^{+\infty} s f_X(s) ds + \int_{-\infty}^{+\infty} t f_Y(t) dt = \mathbb{E}[X] + \mathbb{E}[Y].
\end{aligned}$$

6.3. La loi gamma et le paradoxe de l'autobus

PROPOSITION 6.11. Soient T_1, \dots, T_n des variables aléatoires indépendantes de loi exponentielle de paramètre α . Alors $S_n := T_1 + \dots + T_n$ admet la densité:

$$f_{S_n}(t) = \begin{cases} \frac{\alpha^n t^{n-1}}{(n-1)!} e^{-\alpha t}, & t \geq 0 \\ 0, & t < 0 \end{cases}.$$

DÉFINITION 6.12. Soit X une variable aléatoire de densité f_X . On dit que X suit une *loi gamma* de paramètres n et α , si

$$f_X(t) = \frac{\alpha^n t^{n-1}}{(n-1)!} e^{-\alpha t} \mathbf{1}_{\{t \geq 0\}}.$$

DÉMONSTRATION (de Proposition 6.11). Supposons la formule démontrée pour n . En utilisant que $S_{n+1} = S_n + T_{n+1}$ avec S_n et T_{n+1} indépendants, on obtient par récurrence:

$$\begin{aligned}
f_{S_{n+1}}(t) &= \int_0^\infty f_{S_n}(x) f_{T_{n+1}}(t-x) dx \\
&= \int_0^\infty f_{S_n}(x) \alpha e^{-\alpha(t-x)} \mathbf{1}_{\mathbb{R}_+}(t-x) dx \\
&= \alpha e^{-\alpha t} \frac{\alpha^n}{(n-1)!} \int_0^t x^{n-1} e^{-\alpha x} e^{\alpha x} dx \quad (\text{par hypothèse}) \\
&= \alpha e^{-\alpha t} \frac{\alpha^n}{(n-1)!} \frac{t^n}{n} \\
&= \frac{\alpha^{n+1} t^n}{n!} e^{-\alpha t}. \quad \square
\end{aligned}$$

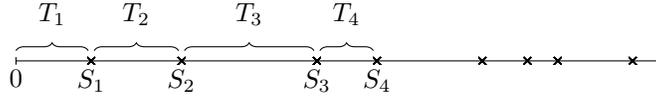
PROPOSITION 6.13 (Temps d'attente consécutifs). *On interprète les variables T_1, \dots, T_n comme les temps d'attente consécutifs jusqu'à l'instant auquel un certain événement se produit, et*

$$S_n = T_1 + \dots + T_n$$

comme temps total d'attente après lequel se produit le $n^{\text{ième}}$ coup. Pour $t \geq 0$, on note N_t le nombre d'événements qui se produisent avant le temps t , c.à.d.

$$N_t := \max\{n \mid S_n \leq t\}.$$

Alors, pour $t > 0$, la variable aléatoire N_t suit une loi de Poisson de paramètre αt .



DÉMONSTRATION. Soit $t \geq 0$ et $n \in \mathbb{N}$. Alors on a:

$$\begin{aligned} N_t = n &\iff S_1 \leq t, \dots, S_n \leq t \text{ et } S_{n+1} > t \\ &\iff n = \max\{k \mid S_k \leq t\} = \sum_{k \geq 1} \mathbb{1}_{\{S_k \leq t\}}. \end{aligned}$$

En particulier,

$$\begin{aligned} \mathbb{P}\{N_t = n\} &= \mathbb{P}\{S_n \leq t\} - \mathbb{P}\{S_{n+1} \leq t\} \\ &= F_{S_n}(t) - F_{S_{n+1}}(t), \end{aligned}$$

où

$$F_{S_n}(t) = \mathbb{P}\{S_n \leq t\} = \int_0^t f_{S_n}(x) dx.$$

Nous démontrons que

$$F_{S_n}(t) = 1 - \sum_{k=0}^{n-1} \frac{\alpha^k t^k}{k!} e^{-\alpha t}, \quad t \geq 0.$$

En écrivant

$$f_n(t) := f_{S_n}(t) \quad \text{et} \quad F_n(t) := 1 - \sum_{k=0}^{n-1} \frac{\alpha^k t^k}{k!} e^{-\alpha t}$$

il suffit à démontrer que

$$(*) \quad F'_n(t) = f_n(t), \quad t \geq 0.$$

(En effet, alors on a $F'_{S_n}(t) - F'_n(t) = 0$, et donc $F_{S_n} - F_n = C$ pour une constante C ; car $\lim_{t \rightarrow \infty} F_n(t) = \lim_{t \rightarrow \infty} F_{S_n}(t) = 1$, on conclut que $C = 0$.)

Démontrons maintenant (*). En effet,

$$\begin{aligned} F'_n(t) &= - \sum_{k=1}^{n-1} \frac{\alpha^k}{k!} k t^{k-1} e^{-\alpha t} + \sum_{k=0}^{n-1} \frac{\alpha^k t^k}{k!} \alpha e^{-\alpha t} \\ &= - \sum_{k=1}^{n-1} \frac{\alpha^k}{(k-1)!} t^{k-1} e^{-\alpha t} + \sum_{k=1}^n \frac{\alpha^k t^{k-1}}{(k-1)!} e^{-\alpha t} \\ &= \frac{\alpha^n}{(n-1)!} t^{n-1} e^{-\alpha t} = f_n(t). \end{aligned}$$

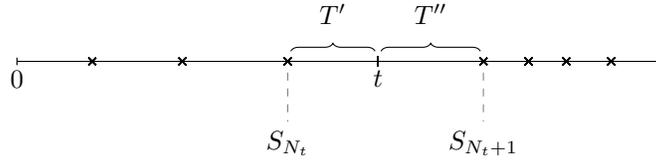
Il en résulte que

$$\mathbb{P}\{N_t = n\} = F_{S_n}(t) - F_{S_{n+1}}(t) = \frac{(\alpha t)^n}{n!} e^{-\alpha t},$$

ce qui démontre que N_t suit une loi de Poisson de paramètre αt . \square

EXEMPLE 6.14 (Paradoxe de l'autobus). On fixe un instant t et on se demande combien de temps doit on attendre en moyenne jusqu'au prochain événement. Plus précieusement, on veut déterminer la loi commune de (T', T'') où

$$T' = t - S_{N_t} \quad \text{et} \quad T'' = S_{N_t+1} - t.$$



Rappelons que

$$S_n = T_1 + \dots + T_n,$$

où les T_i sont indépendantes et suivent une loi exponentielle de paramètre α . En particulier, on a $\mathbb{E}[T_i] = 1/\alpha$ pour tout i . La variable

$$N_t = \max\{n \mid S_n \leq t\}$$

est alors Poisson de paramètre αt .

Soient maintenant $0 \leq x \leq t$ et $y \geq 0$.

$$\begin{aligned} T' = t - S_n \geq x & \quad T'' = S_{n+1} - t \geq y \\ \iff S_n \leq t - x & \quad \iff S_{n+1} \geq y + t \end{aligned}$$

Alors on a:

$$\begin{aligned} \mathbb{P}\{T' \geq x, T'' \geq y\} &= \sum_{n \geq 0} \mathbb{P}\{S_n \leq t - x, S_{n+1} \geq t + y\} \\ &= \mathbb{P}\{S_1 \geq t + y\} + \sum_{n \geq 1} \mathbb{P}\{S_n \leq t - x, S_{n+1} \geq t + y\} \\ &= \mathbb{P}\{T_1 \geq t + y\} + \sum_{n \geq 1} \mathbb{P}\{S_n \leq t - x, T_{n+1} \geq t - S_n + y\} \\ &= \mathbb{P}\{T_1 \geq t + y\} + \sum_{n \geq 1} \mathbb{P}\{(S_n, T_{n+1}) \in A\}. \end{aligned}$$

où

$$A := \{(s, r) \in \mathbb{R}^2 : s \leq t - x, r \geq t - s + y\}.$$

Par l'indépendance de S_n et T_{n+1} , la densité commune de (S_n, T_{n+1}) s'écrit

$$f_{(S_n, T_{n+1})}(s, r) = f_{S_n}(s) f_{T_{n+1}}(r).$$

En utilisant la loi exponentielle de T_i on trouve

$$\begin{aligned}
\mathbb{P}\{T' \geq x, T'' \geq y\} &= e^{-\alpha(t+y)} + \sum_{n \geq 1} \int_{s \leq t-x} \int_{r \geq t-s+y} f_{S_n}(s) f_{T_{n+1}}(r) ds dr \\
&= e^{-\alpha(t+y)} + \sum_{n \geq 1} \int_{s \leq t-x} \int_{r \geq t-s+y} f_{S_n}(s) \alpha e^{-\alpha r} ds dr \\
&= e^{-\alpha(t+y)} + \sum_{n \geq 1} \int_{s \leq t-x} f_{S_n}(s) e^{-\alpha(t-s+y)} ds \\
&= e^{-\alpha(t+y)} \left[1 + \int_0^{t-x} \left(\sum_{n \geq 1} f_{S_n}(s) \right) e^{\alpha s} ds \right]
\end{aligned}$$

Mais

$$\sum_{n \geq 1} f_{S_n}(s) = \sum_{n \geq 1} \frac{\alpha^n s^{n-1}}{(n-1)!} e^{-\alpha s} = \alpha \sum_{n \geq 0} \frac{\alpha^n s^n}{n!} e^{-\alpha s} = \alpha,$$

et donc

$$\begin{aligned}
\mathbb{P}\{T' \geq x, T'' \geq y\} &= e^{-\alpha(t+y)} \left[1 + \int_0^{t-x} \alpha e^{\alpha s} ds \right] \\
&= e^{-\alpha(t+y)} \left[1 + (e^{\alpha(t-x)} - 1) \right] \\
&= e^{-\alpha x} e^{-\alpha y}.
\end{aligned}$$

Ainsi, pour $0 \leq x \leq t$ et $y \geq 0$, l'on a

$$\mathbb{P}\{T' \geq x, T'' \geq y\} = e^{-\alpha x} e^{-\alpha y}.$$

REMARQUE 6.15 (Paradoxe d'autobus; conséquences).

(1) (la loi de T'') On trouve que

$$\mathbb{P}\{T'' \geq y\} = \mathbb{P}\{T' \geq 0, T'' \geq y\} = e^{-\alpha y}, \quad y \geq 0.$$

La variable T'' est donc *exponentielle de paramètre α* ; en particulier,

$$\mathbb{E}[T''] = \frac{1}{\alpha}.$$

Interprétation. Pour le prochain "bus" il faut attendre aussi longtemps que si l'attente démarrait à nouveau à l'instant t .

(2) (l'indépendance de T' et T'') Pour $0 \leq x \leq t$ et $y \geq 0$ on trouve que

$$\begin{aligned}
\mathbb{P}\{T' \leq x, T'' \leq y\} &= 1 - \mathbb{P}\{T' \geq x \text{ ou } T'' \geq y\} \\
&= 1 - (\mathbb{P}\{T' \geq x\} + \mathbb{P}\{T'' \geq y\} - \mathbb{P}\{T' \geq x, T'' \geq y\}) \\
&= 1 - (e^{-\alpha x} + e^{-\alpha y} - e^{-\alpha x} e^{-\alpha y}) \\
&= (1 - e^{-\alpha x})(1 - e^{-\alpha y}) \\
&= \mathbb{P}\{T' \leq x\} \mathbb{P}\{T'' \leq y\};
\end{aligned}$$

en particulier, les variables T' et T'' sont indépendantes.

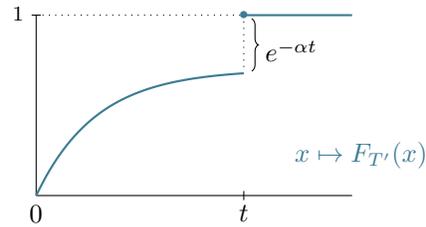
(3) (la loi de T') On a

$$\mathbb{P}\{T' \geq x\} = e^{-\alpha x} \text{ pour } 0 \leq x \leq t;$$

$$\mathbb{P}\{T' \geq t\} = e^{-\alpha t} \text{ et } \mathbb{P}\{T' > t\} = 0,$$

et donc $\mathbb{P}\{T' \geq t\} = e^{-\alpha t} = \mathbb{P}\{T' = t\}$. Par conséquent,

$$F_{T'}(x) = \mathbb{P}\{T' \leq x\} = \begin{cases} 1 - e^{-\alpha x}, & x \leq t \\ 1 & , x > t. \end{cases}$$



En particulier,

$$\begin{aligned} \mathbb{E}[T'] &= \int_0^\infty \mathbb{P}\{T' \geq x\} dx \\ &= \int_0^t e^{-\alpha x} dx \\ &= -\frac{1}{\alpha} e^{-\alpha x} \Big|_0^t = \frac{1}{\alpha} (1 - e^{-\alpha t}), \end{aligned}$$

et puis

$$\mathbb{E}[T' + T''] = \frac{1 - e^{-\alpha t}}{\alpha} + \frac{1}{\alpha} = \frac{2}{\alpha} - \frac{e^{-\alpha t}}{\alpha} \rightarrow \frac{2}{\alpha}, \text{ quand } t \rightarrow \infty.$$

Interprétation (paradoxe de l'autobus). Pour t grand, en moyenne, la longueur d'intervalle aléatoire d'attente qui couvre le point t est $\approx 2/\alpha$; même si un intervalle typique a en moyenne une longueur de $1/\alpha$.

La loi du tout ou rien et la loi forte des grands nombres

7.1. Lemme de Borel-Cantelli

LEMME 7.1 (Lemme de Borel-Cantelli). *Soit $A_1, A_2, \dots \in \mathcal{A}$ une suite d'événements et soit*

$$\begin{aligned} A_\infty &:= \bigcap_n \bigcup_{k \geq n} A_k = \text{“une infinité d'événements } A_n \text{ se réalisent”} \\ &= \text{“} A_n \text{ a lieu une infinité de fois”} \\ &= \text{“} A_n \text{ infiniment souvent”} \equiv \text{“} A_n \text{ i.s.”} \end{aligned}$$

- (1) Si $\sum_k \mathbb{P}(A_k) < \infty$, alors $\mathbb{P}(A_\infty) = 0$;
 (2) Si $\sum_k \mathbb{P}(A_k) = \infty$ et A_1, A_2, \dots indépendants, alors $\mathbb{P}(A_\infty) = 1$.

DÉMONSTRATION. (1) Comme $B_n := \bigcup_{k \geq n} A_k \downarrow$ et $A_\infty = \bigcap_n B_n$, on a

$$\mathbb{P}(A_\infty) = \lim_{n \rightarrow \infty} \mathbb{P}(B_n) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcup_{k \geq n} A_k\right) \leq \lim_{n \rightarrow \infty} \sum_{k \geq n} \mathbb{P}(A_k) = 0.$$

(2) On remarque d'abord que

$$\mathbb{P}(A_\infty^c) = \mathbb{P}\left(\bigcup_n \bigcap_{k \geq n} A_k^c\right) \leq \sum_n \mathbb{P}\left(\bigcap_{k \geq n} A_k^c\right).$$

Mais, pour tout n , on a:

$$\begin{aligned} \mathbb{P}\left(\bigcap_{k \geq n} A_k^c\right) &= \lim_{m \rightarrow \infty} \mathbb{P}\left(\bigcap_{n \leq k \leq m} A_k^c\right) \\ &= \lim_{m \rightarrow \infty} \prod_{n \leq k \leq m} \mathbb{P}(A_k^c) \quad (\text{les } A_k \text{ sont indépendants}) \\ &= \lim_{m \rightarrow \infty} \prod_{n \leq k \leq m} (1 - \mathbb{P}(A_k)) \\ &\leq \lim_{m \rightarrow \infty} \prod_{n \leq k \leq m} e^{-\mathbb{P}(A_k)} \quad (\text{car } e^{-x} \geq 1 - x, \text{ pour } 0 \leq x \leq 1) \\ &\leq \lim_{m \rightarrow \infty} e^{-\sum_{k=n}^m \mathbb{P}(A_k)} = 0. \end{aligned}$$

(pour la dernière égalité on utilise que $\sum_k \mathbb{P}(A_k) = +\infty$. □

COROLLAIRE 7.2. Soit A_1, A_2, \dots une suite indépendante d'événements. Alors on a

$$\mathbb{P}(A_\infty) \in \{0, 1\},$$

et ceci selon la série $\sum_n \mathbb{P}(A_n)$ est convergente ou divergente.

EXEMPLE 7.3 (Marche aléatoire sur \mathbb{Z}). Soit $(X_n)_{n \geq 1}$ i.i.d. telle que

$$\mathbb{P}\{X_n = 1\} = 1 - \mathbb{P}\{X_n = -1\}.$$

On pose $p := \mathbb{P}\{X_n = 1\}$ et $S_n = X_1 + \dots + X_n$. Pour

$$A_n = \{S_n = 0\} \quad \text{et} \quad A_\infty = \{S_n = 0 \text{ i.s.}\}$$

on trouve

$$\mathbb{P}(A_\infty) = \mathbb{P}\{S_n = 0 \text{ i.s.}\} = \begin{cases} 1 & p = \frac{1}{2}, \\ 0 & p \neq \frac{1}{2}. \end{cases}$$

DÉMONSTRATION. Cas $p \neq \frac{1}{2}$: On a

$$\mathbb{P}(A_{2n}) = \mathbb{P}\{S_{2n} = 0\} = \binom{2n}{n} p^n (1-p)^n.$$

Il suffit démontrer que $\sum_n \mathbb{P}(A_{2n}) < \infty$; alors $\mathbb{P}\{S_n = 0 \text{ i.s.}\} = 0$ par le lemme de Borel-Cantelli (1).

Rappel (Critère du quotient) Si a_n est une suite de réels non-négatifs telle que la suite de quotients a_{n+1}/a_n converge vers un réel $c < 1$, alors la série $\sum_n a_n$ est convergente.

En effet,

$$\begin{aligned} \frac{\mathbb{P}(A_{2n+2})}{\mathbb{P}(A_{2n})} &= \frac{\binom{2n+2}{n+1} p^{n+1} (1-p)^{n+1}}{\binom{2n}{n} p^n (1-p)^n} \\ &= p(1-p) \frac{(2n+2)!}{(n+1)!(n+1)!} \frac{n!n!}{(2n)!} \\ &= \underbrace{p(1-p)}_{<1/4} \underbrace{\frac{(2n+2)(2n+1)}{(n+1)(n+1)}}_{\rightarrow 4} \rightarrow c < 1. \end{aligned}$$

Cas $p = \frac{1}{2}$: Soit

$$p_N := \mathbb{P}\{\exists n \geq 1 : S_n = N\}, \quad N \in \mathbb{N}.$$

Alors $p_1 = p_{-1}$ par symétrie, et

$$\begin{aligned} p_0 &= \mathbb{P}\{\exists n \geq 1 : S_1 \neq 0, \dots, S_{n-1} \neq 0, S_n = -1, S_{n+1} = 0\} + \\ &\quad \mathbb{P}\{\exists n \geq 1 : S_1 \neq 0, \dots, S_{n-1} \neq 0, S_n = 1, S_{n+1} = 0\} \\ &= \mathbb{P}\{\exists n \geq 1 : S_1 \neq 0, \dots, S_{n-1} \neq 0, S_n = -1, X_{n+1} = 1\} + \\ &\quad \mathbb{P}\{\exists n \geq 1 : S_1 \neq 0, \dots, S_{n-1} \neq 0, S_n = 1, X_{n+1} = -1\} \\ &= \frac{1}{2} p_{-1} + \frac{1}{2} p_1 = \frac{1}{2} p_1 + \frac{1}{2} p_1 = p_1. \end{aligned}$$

Evidémet,

$$p_2 = p_1^2.$$

En plus,

$$\begin{aligned}
 p_1 &= \mathbb{P}\{\exists n \geq 1 : S_n = 1\} \\
 &= \mathbb{P}\{S_1 = 1\} + \mathbb{P}\{\exists n \geq 1 : S_n = 1, S_1 = -1\} \\
 &= \frac{1}{2} + \mathbb{P}\{\exists n \geq 1 : S_n = 1 \mid S_1 = -1\} \mathbb{P}\{S_1 = -1\} \\
 &= \frac{1}{2} + \frac{1}{2} p_2 = \frac{1}{2} + \frac{1}{2} p_1^2.
 \end{aligned}$$

Mais

$$p_1^2 - 2p_1 + 1 = 0 \iff (p_1 - 1)^2 = 0 \iff p_1 = 1.$$

Donc,

$$p_0 = 1, p_1 = 1, p_2 = 1, \dots$$

Autrement dit,

$$p_n = 1, \quad \forall n \in \mathbb{Z}.$$

À partir d'ici, l'assertion est évidente. Soit, par exemple,

$$\tau = \inf\{n \geq 1 : S_n = 0\}.$$

Alors $(S_{\tau+n})_{n \geq 0}$ est aussi un marche aléatoire sur \mathbb{Z} de même loi que S_n . D'où $\mathbb{P}\{S_n = 0 \text{ i.s.}\} = 1$. \square

EXEMPLE 7.4 (Le singe savant). Un singe est placé devant une machine à écrire. Il tape au hasard un caractère après l'autre, chaque caractère ayant même probabilité d'être tapé. On note SH le texte (ou la concaténation) de toutes les œuvres de Shakespeare, c.à.d.

SH = Roméo et Juliette + Macbeth + Hamlet + Othello + Le Roi Lear + ...

Alors la probabilité que le singe tape SH une infinité de fois est 1, c.à.d.

$$P\{\text{"le singe tape SH i.s."}\} = 1.$$

DÉMONSTRATION. Soit \mathbb{A} l'alphabet avec la loi uniforme. Alors

$$\Omega = \{\omega = (\omega_1, \omega_2, \dots) \mid \omega_i \in \mathbb{A}, i = 1, 2, \dots\}.$$

On note

$$X_i(\omega) = \omega_i, \quad i = 1, 2, \dots$$

Alors la suite (X_i) est i.i.d. avec

$$\mathbb{P}\{X_i = a\} = \frac{1}{|\mathbb{A}|}, \quad a \in \mathbb{A}.$$

Soit SH = (a_1, \dots, a_ℓ) l'œuvre de Shakespeare. Pour $n = 0, 1, 2, \dots$ on pose:

$$A_n = \{\omega \in \Omega \mid \omega_{n+1} = a_1, \omega_{n+2} = a_2, \dots, \omega_{n+\ell} = a_\ell\}.$$

On obtient

$$\mathbb{P}(A_n) = \left(\frac{1}{|\mathbb{A}|}\right)^\ell > 0.$$

Considérons maintenant les événements

$$B_n = A_{\ell(n-1)}, \quad n = 1, 2, \dots$$

Alors $(B_n)_{n \geq 1}$ est une famille indépendante, et pour tout n , on a

$$\mathbb{P}(B_n) = \left(\frac{1}{|\mathbb{A}|}\right)^\ell =: c > 0.$$

En particulier,

$$\sum_n \mathbb{P}(B_n) = \sum_n c = +\infty,$$

et par le lemme de Borel-Cantelli on déduit

$$\mathbb{P}\{B_n \text{ i.s.}\} = 1,$$

c'est-à-dire avec la probabilité = 1, le singe tape SH infiniment souvent. \square

7.2. Convergence de variables aléatoires

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires, définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Il existe plusieurs notions de convergence de X_n vers une variable aléatoire X .

(1) *Convergence en probabilité*

$$\lim_{n \rightarrow \infty} \mathbb{P}\{|X_n - X| \geq \varepsilon\} = 0 \quad \forall \varepsilon > 0 \quad \text{“} X_n \rightarrow X \text{ en probabilité”}$$

(2) *Convergence presque sûrement*

$$\mathbb{P}\left\{\omega \in \Omega \mid \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right\} = 1 \quad \text{“} X_n \rightarrow X \text{ presque sûrement”}$$

(3) *Convergence dans L^p , $p \geq 1$*

$$\lim_{n \rightarrow \infty} \mathbb{E}[|X_n - X|^p] = 0 \quad \text{“} X_n \rightarrow X \text{ dans } L^p\text{”}$$

REMARQUE 7.5. On constate que (3) \Rightarrow (1) et (2) \Rightarrow (1).

DÉMONSTRATION. (3) \Rightarrow (1) est claire: pour tout $\varepsilon > 0$ on a

$$\begin{aligned} \mathbb{P}\{|X_n - X| \geq \varepsilon\} &= \mathbb{P}\{|X_n - X|^p \geq \varepsilon^p\} \\ &\leq \frac{\mathbb{E}[|X_n - X|^p]}{\varepsilon^p} \rightarrow 0, \quad \text{quand } n \rightarrow \infty. \end{aligned}$$

(2) \Rightarrow (1):

D'abord on remarque que

$$\begin{aligned} \mathbb{P}\{|X_n - X| \geq \varepsilon \text{ i.s.}\} &= \mathbb{P}\left(\bigcap_{m \geq 1} \bigcup_{n \geq m} \{|X_n - X| \geq \varepsilon\}\right) \\ &= \lim_{m \rightarrow \infty} \mathbb{P}\left(\underbrace{\bigcup_{n \geq m} \{|X_n - X| \geq \varepsilon\}}_{\geq \mathbb{P}\{|X_m - X| \geq \varepsilon\}}\right). \end{aligned}$$

Pour conclure il suffit à démontrer que $\mathbb{P}\{|X_n - X| \geq \varepsilon \text{ i.s.}\} = 0$ sous la condition que “ $X_n \rightarrow X$ presque sûrement”. Pour ceci on utilise le lemme suivant. \square

LEMME 7.6. *On a*

$$X_n \rightarrow X \text{ presque sûrement} \iff \forall \varepsilon > 0, \mathbb{P}\{|X_n - X| \geq \varepsilon \text{ i.s.}\} = 0.$$

DÉMONSTRATION. Pour $\{X_n \not\rightarrow X\} \equiv \{\omega : X_n(\omega) \not\rightarrow X(\omega)\}$ on peut écrire

$$\{X_n \not\rightarrow X\} = \bigcup_{k \geq 1} \bigcap_{m \geq 1} \bigcup_{n \geq m} \left\{|X_n - X| \geq \frac{1}{k}\right\}.$$

Donc:

$$\begin{aligned} \underbrace{\mathbb{P}\{|X_n - X| \not\rightarrow 0\}}_{\mathbb{P}\{|X_n - X| \rightarrow 0\} = 1} = 0 &\iff \mathbb{P}\left(\underbrace{\bigcap_{m \geq 1} \bigcup_{n \geq m} \left\{|X_n - X| \geq \frac{1}{k}\right\}}_{\mathbb{P}\{|X_n - X| \geq 1/k \text{ i.s.}\}}\right) = 0. \quad \square \\ &\iff \mathbb{P}\{|X_n - X| \rightarrow 0\} = 1 \end{aligned}$$

REMARQUE 7.7. Par le lemme de Borel-Cantelli, la condition

$$\sum_n \mathbb{P}\{|X_n - X| \geq \varepsilon\} < \infty \quad \forall \varepsilon > 0 \quad \text{“convergence rapide en probabilité”}$$

implique la convergence $X_n \rightarrow X$ presque sûrement.

7.3. La loi forte des grands nombres

D’abord, rappelons la loi faible des grands nombres.

REMARQUE 7.8 (Loi faible des grands nombres). Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires, deux à deux non-corrélées telle que $\mathbb{E}[X_i^2] \leq c < \infty$ pour tout $i \in \mathbb{N}$. On pose $S_n = X_1 + \dots + X_n$. Alors

$$\mathbb{P}\left(\left|\frac{S_n - \mathbb{E}[S_n]}{n}\right| > \varepsilon\right) \rightarrow 0, \quad \text{quand } n \rightarrow \infty,$$

c’est-à-dire,

$$\frac{S_n - \mathbb{E}[S_n]}{n} \rightarrow 0 \quad \text{en probabilité.}$$

Si de plus $\mathbb{E}[X_i] = \mathbb{E}[X_1]$ pour tout i , alors

$$\frac{S_n}{n} \rightarrow \mathbb{E}[X_1] \quad \text{en probabilité.}$$

THÉORÈME 7.9 (Loi forte des grands nombres). Soient $(X_i)_{i \geq 1}$ une suite de variables aléatoires, deux à deux non-corrélées telles que $\mathbb{E}[X_i^2] \leq c < \infty$ pour tout $i \in \mathbb{N}$. On pose $S_n = X_1 + \dots + X_n$. Alors

$$\frac{S_n - \mathbb{E}[S_n]}{n} \rightarrow 0, \quad \text{presque sûrement.}$$

En particulier, si de plus $\mathbb{E}[X_i] = \mathbb{E}[X_1]$ pour tout i , alors

$$\frac{S_n}{n} \rightarrow \mathbb{E}[X_1], \quad \text{presque sûrement.}$$

DÉMONSTRATION. D’abord on note que $\mathbb{E}|X_i| < \infty$ à cause de $\mathbb{E}[X_i^2] < \infty$ (voir Exemple 3.14). Sans perte de généralités, on peut supposer que $\mathbb{E}[X_i] = 0$ pour tous les i . Alors, on a par hypothèse,

$$\mathbb{E}[S_n^2] = \sum_{i=1}^n \mathbb{E}[X_i^2] + \underbrace{\sum_{\substack{i,j=1 \\ i \neq j}}^n \mathbb{E}[X_i X_j]}_{=0} \leq nc.$$

Utilisant l’inégalité de Tchebychev, on obtient que

$$\mathbb{P}\{|S_n| \geq n\varepsilon\} \leq \frac{nc}{n^2\varepsilon^2} = \frac{c}{n\varepsilon^2},$$

et par conséquent,

$$\sum_n \mathbb{P}\{|S_n| \geq n^2\varepsilon\} \leq \sum_n \frac{c}{n^2\varepsilon^2} < \infty.$$

Par le lemme de Borel-Cantelli on obtient

$$\mathbb{P} \{ |S_{n^2}| \geq n^2 \varepsilon \text{ i.s.} \} = 0,$$

et donc avec le Lemme 7.6,

$$\frac{S_{n^2}}{n^2} \rightarrow 0 \text{ presque sûrement, quand } n \rightarrow \infty.$$

Il reste à estimer les déviations de S_k pour $n^2 < k < (n+1)^2$ par rapport au S_{n^2} .

On pose

$$D_n := \max_{n^2 \leq k < (n+1)^2} |S_k - S_{n^2}|, \quad n \geq 1.$$

Alors

$$\begin{aligned} \mathbb{E}[D_n^2] &\leq \sum_{k=n^2}^{(n+1)^2-1} \mathbb{E} \left[|S_k - S_{n^2}|^2 \right] \\ &= \sum_{k=n^2}^{(n+1)^2-1} \mathbb{E} \left[\left| \sum_{i=n^2+1}^k X_i \right|^2 \right] \\ &= \sum_{k=n^2}^{(n+1)^2-1} \sum_{i=n^2+1}^k \mathbb{E}[X_i^2] \\ &\leq 2n \sum_{k=n^2+1}^{(n+1)^2-1} \mathbb{E}[X_i^2] \leq 4n^2 c. \end{aligned}$$

On obtient

$$\mathbb{P} \{ D_n \geq n^2 \varepsilon \} \leq \frac{4n^2 c}{n^4 \varepsilon^2} = \frac{4c}{\varepsilon^2} \frac{1}{n^2},$$

et d'où comme avant

$$\frac{D_n}{n^2} \rightarrow 0 \text{ presque sûrement, quand } n \rightarrow \infty.$$

Finalement, on a pour $n^2 \leq k < (n+1)^2$:

$$\frac{|S_k|}{k} \leq \frac{|S_{n^2}| + D_n}{n^2} \rightarrow 0 \text{ presque sûrement, quand } n \rightarrow \infty. \quad \square$$

REMARQUE 7.10. Il existe des versions renforcées de la loi forte des grands nombres, par exemple celle de Kolmogorov:

THÉORÈME 7.11 (Kolmogorov 1930). *Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires i.i.d. telle que $\mathbb{E}|X_i| < \infty$. Alors:*

$$\frac{S_n}{n} \rightarrow \mathbb{E}[X_1], \quad \text{presque sûrement.}$$

EXEMPLE 7.12. Soit $(X_i)_{i \geq 1}$ i.i.d. et $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction mesurable telle que $\mathbb{E}|f(X_i)| < \infty$. Alors la famille $(f(X_i))_{i \geq 1}$ est aussi i.i.d. et par la loi forte des grands nombres:

$$\frac{1}{n} \sum_{i=1}^n f(X_i) \rightarrow \mathbb{E}[f(X_1)], \quad \text{quand } n \rightarrow \infty.$$

En particulier:

- (1) Soient X_i de la loi uniforme sur $[a, b]$ et $f : [a, b] \rightarrow \mathbb{R}$ une fonction continue. Alors, on a

$$\frac{1}{n} \sum_{i=1}^n f(X_i) \rightarrow \frac{1}{b-a} \int_a^b f(x) dx \quad \text{presque sûrement}$$

(méthode de Monte-Carlo pour le calcul d'intégrales).

- (2) Soit $f = \mathbb{1}_B$ avec $B \subset \mathbb{R}$ mesurable (e.g. $B = [a, b]$), alors on a pour “la fréquence relative d'événements $\{X_i \in B\}$ ”

$$\frac{1}{n} \sum_{i=1}^n f(X_i) = \frac{|\{i \in \{1, \dots, n\} : X_i \in B\}|}{n} \rightarrow \mathbb{E}[\mathbb{1}_B(X_i)] = \mathbb{P}\{X_i \in B\}.$$

- (3) Soit $f = \mathbb{1}_{]-\infty, t]}$ et

$$F_n(\omega, t) := \frac{|\{i \in \{1, \dots, n\} : X_i(\omega) \leq t\}|}{n}, \quad t \in \mathbb{R}, \omega \in \Omega,$$

la “fonction de répartition empirique” de X_i . Alors

$$F_n(\cdot, t) \rightarrow F_{X_1}(t) \quad \text{presque sûrement.}$$

EXEMPLE 7.13 (Nombres normaux). Fixons $2 \leq d \in \mathbb{N}$. Soit $\omega \in [0, 1[$ et

$$\omega = \frac{\omega_1}{d} + \frac{\omega_2}{d^2} + \frac{\omega_3}{d^3} + \dots = \sum_{n=1}^{\infty} \frac{\omega_n}{d^n}, \quad \omega_n \in \{0, 1, \dots, d-1\},$$

le développement d -adique de ω . On appelle ω un nombre d -normal si

$$\forall k \in \{0, 1, \dots, d-1\}, \quad \frac{|\{i \leq n : \omega_i = k\}|}{n} \rightarrow \frac{1}{d}, \quad \text{quand } n \rightarrow \infty.$$

REMARQUE 7.14. Soit $d = 10$. Alors $0, 1234567890123 \dots$ est d -normal.

Question: Sont $\pi - 3$, $e - 2$, $\sqrt{2} - 1$ des nombres d -normaux? La réponse est inconnue!

THÉORÈME 7.15 (Émile Borel 1909). Soit $2 \leq d \in \mathbb{N}$. Sauf un ensemble de mesure de Lebesgue 0, tout nombre ω dans $[0, 1[$ est d -normal.

DÉMONSTRATION. Soit $\Omega = [0, 1[$, $\mathcal{A} = \mathcal{B}([0, 1[)$ et \mathbb{P} la mesure de Lebesgue sur $[0, 1[$. Notons

$$X_n : \Omega \rightarrow \{0, 1, \dots, d-1\}, \quad X_n(\omega) := \omega_n.$$

Alors

$$\begin{aligned} \mathbb{P}\{X_1 = n_1, \dots, X_k = n_k\} &= \mathbb{P}\{\omega \in \Omega \mid X_1(\omega) = n_1, \dots, X_k(\omega) = n_k\} \\ &= \mathbb{P}\left\{\omega \in \Omega \mid \frac{n_1}{d} + \frac{n_2}{d^2} + \dots + \frac{n_k}{d^k} \leq \omega < \frac{n_1}{d} + \frac{n_2}{d^2} + \dots + \frac{n_k + 1}{d^k}\right\} \\ &= \mathbb{P}\left[\left[\frac{n_1}{d} + \frac{n_2}{d^2} + \dots + \frac{n_k}{d^k}, \frac{n_1}{d} + \frac{n_2}{d^2} + \dots + \frac{n_k + 1}{d^k}\right]\right] = \frac{1}{d^k} \\ &= \mathbb{P}\{X_1 = n_1\} \cdot \dots \cdot \mathbb{P}\{X_k = n_k\}, \end{aligned}$$

c.à.d. $(X_n)_{n \geq 1}$ est i.i.d. et $\mathbb{P}\{X_n = k\} = 1/d$ pour tout $k \in \{0, 1, \dots, d-1\}$. En particulier, la suite

$$(\mathbb{1}_{\{X_n = k\}})_{n \geq 1} \quad \text{est i.i.d. pour tout } k \in \{0, 1, \dots, d-1\}.$$

En conséquence, par la loi forte des grands nombres, on obtient donc tout $k \in \{0, 1, \dots, d-1\}$,

$$\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_n=k\}} \xrightarrow{n \rightarrow \infty} \mathbb{E}[\mathbb{1}_{\{X_1=k\}}] = \mathbb{P}\{X_1 = k\} = \frac{1}{d} \quad \text{p.s.}$$

Il reste à noter que

$$\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_n=k\}}(\omega) = \frac{|\{i \leq n : \omega_i = k\}|}{n}. \quad \square$$

Le théorème de la limite centrale

8.1. Fonctions caractéristiques

DÉFINITION 8.1. Soit X une variable aléatoire réelle sur $(\Omega, \mathcal{A}, \mathbb{P})$. La fonction

$$\varphi_X : \mathbb{R} \rightarrow \mathbb{C}, \quad \varphi_X(t) := \mathbb{E}[e^{itX}] \equiv \mathbb{E}[\cos(tX)] + i \mathbb{E}[\sin(tX)],$$

est appelée *fonction caractéristique* de X .

REMARQUE 8.2. La fonction caractéristique a les propriétés suivantes:

- (1) $\varphi_X(0) = 1$
- (2) $\varphi_X(-t) = \overline{\varphi_X(t)}$, en effet

$$\begin{aligned} \varphi_X(-t) &= \mathbb{E}[\cos(-tX)] + i \mathbb{E}[\sin(-tX)] \\ &= \mathbb{E}[\cos(tX)] + (-i) \mathbb{E}[\sin(tX)] = \overline{\varphi_X(t)} \end{aligned}$$

- (3) $|\varphi_X(t)| \leq \mathbb{E}[|e^{itX}|] = 1$
- (4) $\varphi_{aX+b}(t) = \mathbb{E}[e^{it(aX+b)}] = \mathbb{E}[e^{itaX}] \mathbb{E}[e^{itb}] = e^{itb} \varphi_X(at)$
- (5) Si X, Y sont indépendantes, alors

$$\varphi_{X+Y}(t) = \mathbb{E}[e^{it(X+Y)}] = \mathbb{E}[e^{itX} e^{itY}] = \mathbb{E}[e^{itX}] \mathbb{E}[e^{itY}] = \varphi_X(t) \varphi_Y(t).$$

Plus généralement, soient X_1, \dots, X_n indépendantes, alors

$$\varphi_{X_1+\dots+X_n}(t) = \prod_{k=1}^n \varphi_{X_k}(t).$$

EXEMPLE 8.3. Soit X une variable aléatoire réelle sur $(\Omega, \mathcal{A}, \mathbb{P})$.

- (1) Soit X Bernoulli, c'est-à-dire $\mathbb{P}\{X = 1\} = p$ et $\mathbb{P}\{X = 0\} = 1 - p = q$. Alors

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = pe^{it} + q.$$

- (2) Soit X binômiale de paramètres n et p . Sans perte de généralités,

$$X = X_1 + \dots + X_n$$

avec X_1, \dots, X_n indépendantes de loi de Bernoulli telles que $\mathbb{P}\{X = 1\} = p$. Alors, en utilisant Remarque 8.2 (5),

$$\varphi_X(t) = \prod_{k=1}^n \varphi_{X_k}(t) = (q + pe^{it})^n.$$

- (3) Soit $\mathbb{P}\{X = 1\} = \mathbb{P}\{X = -1\} = \frac{1}{2}$. Alors

$$\begin{aligned} \varphi_X(t) &= \mathbb{E}[e^{itX}] = e^{it} \mathbb{P}\{X = 1\} + e^{-it} \mathbb{P}\{X = -1\} \\ &= \frac{e^{it} + e^{-it}}{2} = \cos t. \end{aligned}$$

(4) Soit X Poisson de paramètre $\lambda > 0$. Alors

$$\begin{aligned}\varphi_X(t) &= \sum_{k=0}^{\infty} e^{itk} \mathbb{P}\{X = k\} \\ &= \sum_{k=0}^{\infty} e^{itk} \frac{e^{-\lambda} \lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{it})^k}{k!} \\ &= e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda(e^{it}-1)}.\end{aligned}$$

(5) Soit X une variable normale $N(0, 1)$. Alors

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} f_{0,1}(x) dx \quad \text{avec } f_{0,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Donc

$$\begin{aligned}\varphi_X(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\cos(tx) + i \sin(tx)) e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos(tx) e^{-x^2/2} dx,\end{aligned}$$

car la fonction $x \mapsto \sin(tx) e^{-x^2/2}$ est impaire. D'autre part, on a

$$\varphi'_X(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (-x) \sin(tx) e^{-x^2/2} dx.$$

En utilisant intégration par parties

$$\int_{-\infty}^{\infty} hg' = - \int_{-\infty}^{\infty} h'g \quad \text{avec } h(x) = \sin(tx) \text{ et } g(x) = e^{-x^2/2},$$

on déduit que

$$\begin{aligned}\varphi'_X(t) &= -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t \cos(tx) e^{-x^2/2} dx \\ &= -t\varphi_X(t).\end{aligned}$$

Autrement dit, pour tout $t \in \mathbb{R}$,

$$\varphi'_X(t) = -t\varphi_X(t),$$

et puis

$$\begin{aligned}\frac{d}{dt} (\varphi_X(t) e^{t^2/2}) &= \varphi'_X(t) e^{t^2/2} + \varphi_X(t) t e^{t^2/2} \\ &= \{\varphi'_X(t) + t\varphi_X(t)\} e^{t^2/2} = 0.\end{aligned}$$

Donc

$$\varphi_X(t) e^{t^2/2} = \varphi_X(0) = 1, \quad \text{pour } t \in \mathbb{R},$$

c.à.d.

$$\varphi_X(t) = e^{-t^2/2}.$$

(6) Soit $X \equiv X_{\mu, \sigma^2}$ normale $N(\mu, \sigma^2)$. Alors

$$X_{\mu, \sigma^2} = \sigma X_{0,1} + \mu$$

avec $X_{0,1}$ une variable $N(0, 1)$, et puis

$$\varphi_{X_{\mu, \sigma^2}}(t) = \varphi_{\sigma X_{0,1} + \mu}(t) = \varphi_{X_{0,1}}(\sigma t) e^{i\mu t} = e^{i\mu t - t^2 \sigma^2 / 2}.$$

PROPOSITION 8.4.

(1) Soit X une variable aléatoire discrète à valeurs dans \mathbb{Z} . Alors

$$\varphi_X(t) = \sum_{k=-\infty}^{\infty} \mathbb{P}\{X = k\} e^{itk} \quad (\text{série de Fourier}).$$

Les “coefficients de Fourier” $\mathbb{P}\{X = k\}$ sont déterminés par la relation:

$$\mathbb{P}\{X = k\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi_X(t) e^{-itk} dt.$$

(2) Soit X une variable aléatoire avec densité f_X . Alors

$$\varphi_X(t) = \int_{-\infty}^{\infty} f_X(x) e^{itx} dx =: \hat{f}(t) \quad (\text{transformée de Fourier}), \quad \text{et}$$

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi_X(t) e^{-itx} dt \quad (\text{transformée inverse de Fourier}).$$

En particulier,

$$\mathbb{P}\{a < X \leq b\} = \frac{1}{2\pi} \int_a^b \left(\int_{-\infty}^{\infty} \varphi_X(t) e^{-itx} dt \right) dx.$$

REMARQUE 8.5 (Théorème d’unicité). La fonction caractéristique φ_X caractérise la loi de X , c.à.d. si deux variables aléatoires admettant la même fonction caractéristique, elles ont même loi.

EXEMPLE 8.6.

(1) Soient X et Y deux variables aléatoires indépendantes telles que

- la variable X est Poisson de paramètre $\lambda > 0$, et
- la variable Y est Poisson de paramètre $\mu > 0$.

Alors $X + Y$ est Poisson de paramètre $\lambda + \mu$.

(2) Soient X_1, \dots, X_n indépendantes, X_ℓ normale $N(\mu_\ell, \sigma_\ell^2)$ pour $\ell = 1, \dots, n$.

Alors $X_1 + \dots + X_n$ est normale $N(\mu, \sigma^2)$ où

$$\begin{aligned} \mu &= \mu_1 + \dots + \mu_n \quad \text{et} \\ \sigma^2 &= \sigma_1^2 + \dots + \sigma_n^2. \end{aligned}$$

DÉMONSTRATION. (1) En effet, par l’indépendance de X et Y , on obtient

$$\varphi_{X+Y}(t) = \varphi_X(t) \varphi_Y(t) = e^{\lambda(e^{it} - 1)} e^{\mu(e^{it} - 1)} = e^{(\lambda + \mu)(e^{it} - 1)}.$$

(2) Par l’indépendance de X_1, \dots, X_n , on a

$$\varphi_{S_n}(t) = \prod_{\ell=1}^n \varphi_{X_\ell}(t) = \prod_{\ell=1}^n e^{i\mu_\ell - \sigma_\ell^2 t^2 / 2} = e^{i\mu - \sigma^2 t^2 / 2}. \quad \square$$

REMARQUE 8.7. Soit $n \in \{1, 2, \dots\}$. Supposons que X est une variable aléatoire telle que

$$\mathbb{E}[|X|^n] < \infty.$$

Alors, φ_X est n fois dérivable, et

$$\varphi_X^{(k)}(t) = i^k \mathbb{E}[X^k e^{itX}], \quad k = 1, \dots, n.$$

En particulier,

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

8.2. Le Théorème de Moivre-Laplace

Pour une variable aléatoire X on note $\varphi_X(t) = \mathbb{E}[e^{itX}]$ sa fonction caractéristique. Rappelons que si $\varphi_X = \varphi_Y$, alors

$$\mathbb{P}\{a < X \leq b\} = \mathbb{P}\{a < Y \leq b\}, \quad \text{pour tout } a, b \in \mathbb{R}.$$

THÉORÈME 8.8 (de Moivre-Laplace). Soit $(X_k)_{k \geq 1}$ une suite i.i.d. de variables aléatoires avec $0 < \sigma^2 := \text{var}(X_k) < \infty$. Soit

$$S_n = X_1 + \dots + X_n$$

et

$$S_n^* := \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\text{var}(S_n)}} = \frac{S_n - n\mathbb{E}[X_1]}{\sqrt{n}\sigma}.$$

Alors, les variables S_n^* convergent en loi vers une variable aléatoire $N(0, 1)$, c.à.d. pour tout $a, b \in \mathbb{R}$ avec $a < b$,

$$\mathbb{P}\{a < S_n^* \leq b\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx = N(b) - N(a), \quad \text{quand } n \rightarrow \infty,$$

où

$$N(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-x^2/2} dx$$

est la fonction de répartition de la loi normale $N(0, 1)$.

DÉMONSTRATION. On utilise que S_n^* convergent vers $N(0, 1)$ en loi si et seulement si

$$\varphi_{S_n^*}(t) \rightarrow \varphi_{N(0,1)}(t) = e^{-t^2/2} \quad \forall t \in \mathbb{R}.$$

Nous démontrons que

$$\varphi_{S_n^*}(t) \rightarrow e^{-t^2/2}, \quad \text{quand } n \rightarrow \infty,$$

pour tout t .

Fait utilisé dans la preuve: Si $z_n \rightarrow z$ dans \mathbb{C} , alors $(1 + \frac{z_n}{n})^n \rightarrow e^z$.

Sans restrictions soit (X_k) une suite centrée, c.à.d. $\mathbb{E}[X_k] = 0$ pour tout k . Puisque $\sigma^2 = \mathbb{E}[X_k^2] < \infty$, la fonction $\varphi := \varphi_{X_k}$ est de classe C^2 et en plus

$$\begin{aligned} \varphi'(0) &= i \mathbb{E}[X_k] = 0, \quad \text{et} \\ \varphi''(0) &= i^2 \mathbb{E}[X_k^2] = -\sigma^2. \end{aligned}$$

Le développement de Taylor en $t = 0$ donne

$$\varphi(t) = 1 - \frac{\sigma^2}{2} t^2 + r(t), \quad \text{où } \frac{|r(t)|}{t^2} \rightarrow 0, \quad \text{quand } t \rightarrow 0.$$

Comme

$$\varphi_{S_n^*}(t) = \prod_{i=1}^n \varphi_{X_i}(t) = \varphi_{X_1}(t)^n,$$

on obtient

$$\begin{aligned} \varphi_{S_n^*}(t) &= \varphi_{S_n} \left(\frac{t}{\sqrt{n}\sigma} \right) = \varphi_{X_1} \left(\frac{t}{\sqrt{n}\sigma} \right)^n = \varphi \left(\frac{t}{\sqrt{n}\sigma} \right)^n \\ &= \left(1 - \frac{\sigma^2}{2} \frac{t^2}{n\sigma^2} + r \left(\frac{t}{\sqrt{n}\sigma} \right) \right)^n \rightarrow e^{-t^2/2}, \quad n \rightarrow \infty. \end{aligned}$$

La convergence est uniforme sur les intervalles compacts. On a utilisé que

$$nr \left(\frac{t}{\sqrt{n}\sigma} \right) \rightarrow 0, \quad n \rightarrow \infty. \quad \square$$

8.3. Applications

EXEMPLE 8.9. On jette un dé $n = 12000$ fois. Soit P la probabilité que le nombre de “6” est dans $]1800, 2100]$. La formule exacte est

$$P = \sum_{1800 < k \leq 2100} \binom{12000}{k} \left(\frac{1}{6} \right)^k \left(\frac{5}{6} \right)^{n-k}.$$

On note

$$X_k = \begin{cases} 1, & \text{si “6” au } k^{\text{ième}} \text{ coup} \\ 0, & \text{sinon.} \end{cases}$$

Alors, les variables (X_k) sont i.i.d. avec $\mathbb{P}\{X_k = 1\} = \frac{1}{6}$. En particulier,

$$\mathbb{E}[X_k] = \frac{1}{6} = \mathbb{E}[X_k^2] \quad \text{et} \quad \text{var}(X_k) = \mathbb{E}[X_k^2] - \mathbb{E}[X_k]^2 = \frac{1}{6} - \frac{1}{36} = \frac{5}{36} =: \sigma^2.$$

Soit

$$S_n = X_1 + \dots + X_n.$$

Pour $a < b$ on a:

$$\begin{aligned} \mathbb{P}\{a < S_n \leq b\} &= \mathbb{P} \left\{ \frac{a - n\mathbb{E}[X_1]}{\sqrt{n}\sigma} < \underbrace{\frac{S_n - n\mathbb{E}[X_1]}{\sqrt{n}\sigma}}_{=S_n^*} \leq \frac{b - n\mathbb{E}[X_1]}{\sqrt{n}\sigma} \right\} \\ &\approx N \left(\frac{b - n\mathbb{E}[X_1]}{\sqrt{n}\sigma} \right) - N \left(\frac{a - n\mathbb{E}[X_1]}{\sqrt{n}\sigma} \right), \end{aligned}$$

c.à.d.

$$P \approx N \left(\frac{2100 - 2000}{\sqrt{12000} \frac{5}{36}} \right) - N \left(\frac{1800 - 2000}{\sqrt{12000} \frac{5}{36}} \right) = N(\sqrt{6}) - N(-2\sqrt{6}) \approx 0,992.$$

EXEMPLE 8.10 (Roulette américaine).

On a

- 36 poches numérotées de 1 à 36 (18 rouges et 18 noires)
- 2 poches numérotées 0 et 00 qui sont vertes.

Le joueur mise 1 \$ sur “rouge” (ou “noir”). On pose

$$X_k = \begin{cases} 1 & \text{si “rouge” au } k^{\text{ième}} \text{ coup;} \\ -1 & \text{sinon.} \end{cases}$$

Alors on a

$$\mathbb{P}\{X_k = 1\} = \frac{18}{38} \quad \text{et} \quad \mathbb{P}\{X_k = -1\} = \frac{20}{38},$$

par conséquent

$$\mathbb{E}[X_k] = \frac{18}{38} - \frac{20}{38} = \frac{-1}{19} \quad \text{et} \quad \text{var}(X_k) = \mathbb{E}[X_k^2] - \mathbb{E}[X_k]^2 = 1 - \left(\frac{1}{19}\right)^2 \approx 0,9972.$$

Considérons

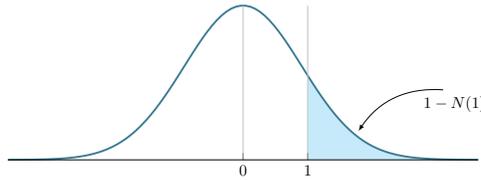
$$\mathbb{P}\{S_n \geq 0\} = \mathbb{P}\left\{ \underbrace{\frac{S_n - n\mathbb{E}[X_1]}{\sqrt{n\sigma}}}_{=S_n^*} \geq \frac{-n\mathbb{E}[X_1]}{\sqrt{n\sigma}} \right\}.$$

Par exemple, soit $n = 361 = 19^2$, et pour simplicité, $\sigma=1$. Alors

$$-\frac{n\mathbb{E}[X_1]}{\sqrt{n\sigma}} = 1,$$

c'est-à-dire,

$$\mathbb{P}\{S_n \geq 0\} \approx 1 - N(1) = 1 - 0,8413 = 0,1587.$$



Conclusion Après 361 coups, le joueur a perdu en moyenne 19 \$; seulement avec une probabilité d'environ 16 % son bilan est encore positif.

Statistiques: l'estimation et tests d'hypothèses

9.1. Modèles statistiques et estimateurs

DÉFINITION 9.1. On appelle *modèle statistique* la donnée de:

- un espace Ω (= l'ensemble des résultats possibles de l'expérience), muni d'une tribu \mathcal{A} des événements;
- une famille $(\mathbb{P}_\vartheta)_{\vartheta \in \Theta}$ de probabilités sur (Ω, \mathcal{A}) .

- (1) Si on veut déterminer la valeur de ϑ , il s'agit *d'estimation*.
- (2) Si on veut savoir si ϑ se trouve dans une partie $\Theta_0 \subset \Theta$, il s'agit d'un *test*.

EXEMPLES 9.2.

- *Fabrication des pièces sur une machine*

$$\begin{aligned} \vartheta &= \text{la probabilité (inconnue) d'être défectueuse} \\ \Omega &= \{0, 1\}, \quad \mathbb{P}_\vartheta\{1\} = \vartheta = 1 - \mathbb{P}_\vartheta\{0\}, \quad \vartheta \in \Theta = [0, 1]. \end{aligned}$$

Nous appelons ce modèle un *modèle de Bernoulli*.

- *Taille des enfants de six ans en France*

$$(\mathbb{P}_\vartheta)_{\vartheta \in \Theta} = \{N(\mu, \sigma^2): \mu \in \mathbb{R}, \sigma^2 > 0\}, \quad \text{modèle gaussien.}$$

On a $\vartheta = (\mu, \sigma^2) \in \mathbb{R} \times]0, \infty[$ si les deux paramètres sont inconnus ou $\vartheta = \mu \in \mathbb{R}$ si la variance σ^2 est connue.

DÉFINITION 9.3. Un *n-échantillon* est une suite (X_1, \dots, X_n) de n variables aléatoires indépendantes de même loi \mathbb{P}_ϑ .

EXEMPLE 9.4. Dans le modèle de Bernoulli on note

$$X_i = \begin{cases} 1 & \text{si la } i^{\text{ème}} \text{ pièce est défectueuse,} \\ 0 & \text{sinon.} \end{cases}$$

et puis,

$$\begin{aligned} \Omega^n &= \{0, 1\}^n, \quad \mathbb{P}_\vartheta^n = \mathbb{P}_\vartheta \otimes \dots \otimes \mathbb{P}_\vartheta, \\ \forall \omega &= (i_1, \dots, i_n) \in \Omega^n, \quad X(\omega) := X_1(\omega) + \dots + X_n(\omega) = i_1 + \dots + i_n, \end{aligned}$$

et

$$\mathbb{P}_\vartheta^n\{\omega\} = \vartheta^{X(\omega)}(1 - \vartheta)^{n - X(\omega)}.$$

DÉFINITION 9.5.

- (1) On appelle *estimateur* du paramètre ϑ toute variable aléatoire (“statistique”) de la forme

$$T = f(X_1, \dots, X_n)$$

à valeurs dans Θ , c.à.d. toute variable aléatoire construite à partir de l'échantillon (X_1, \dots, X_n) .

- (2) Un estimateur T de ϑ est dit *sans biais* si

$$\forall \vartheta \in \Theta, \quad \mathbb{E}_\vartheta[T] = \vartheta,$$

(où \mathbb{E}_ϑ désigne l'espérance par rapport à la probabilité $\mathbb{P}_\vartheta^n = \mathbb{P}_\vartheta \otimes \dots \otimes \mathbb{P}_\vartheta$).

EXEMPLE 9.6. Supposons qu'on observe un n -échantillon de la loi $N(\vartheta, \sigma^2)$. Un estimateur de ϑ sans biais est la *moyenne empirique*

$$T = \bar{X} = \frac{1}{n}(X_1 + \dots + X_n).$$

En fait,

$$\mathbb{E}_\vartheta[T] = \frac{1}{n}(\mathbb{E}_\vartheta[X_1] + \dots + \mathbb{E}_\vartheta[X_n]) = \frac{1}{n}n\vartheta = \vartheta.$$

DÉFINITION 9.7. (1) Le *risque quadratique* de l'estimateur T de ϑ est

$$R_T(\vartheta) := \mathbb{E}_\vartheta[(T - \vartheta)^2] = \text{var}_\vartheta(T),$$

où on a utilisé que $\vartheta = \mathbb{E}_\vartheta[T]$.

- (2) Étant donné deux estimateurs T et T' de ϑ , on dit que T est *meilleur* que T' (au sens du risque quadratique) si

$$R_T(\vartheta) \leq R_{T'}(\vartheta), \quad \forall \vartheta \in \Theta,$$

resp. *strictement meilleur* s'il est meilleur, et si de plus $R_T(\vartheta) < R_{T'}(\vartheta)$ pour au moins une valeur de ϑ .

EXEMPLE 9.8. Dans le *modèle gaussien* $N(\vartheta, \sigma^2)$ avec $\vartheta \in \mathbb{R}$ et $\sigma^2 > 0$, les variables

$$X_1 \quad \text{et} \quad \bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$$

sont des estimateurs sans biais de ϑ .

En fait, on a

$$X_1 \sim N(\vartheta, \sigma^2), \quad \bar{X} \sim N\left(\vartheta, \frac{\sigma^2}{n}\right),$$

car

$$\text{var}(\bar{X}) = \frac{1}{n^2} \underbrace{\sum_{i=1}^n \text{var}(X_i)}_{= n\sigma^2} = \frac{\sigma^2}{n}.$$

On constate donc que

$$R_{X_1}(\vartheta) = \sigma^2, \quad \text{mais} \quad R_{\bar{X}}(\vartheta) = \frac{\sigma^2}{n}.$$

Autrement dit, \bar{X} est *strictement meilleur* que X_1 dès que $n \geq 2$.

EXEMPLE 9.9 (Meilleur estimateur linéaire sans biais).

On observe un n -échantillon X_1, \dots, X_n de loi \mathbb{P}_ϑ sur \mathbb{R} (c.à.d. X_1, \dots, X_n i.i.d. et $X_i \sim \mathbb{P}_\vartheta$). Notons μ_ϑ et σ_ϑ^2 l'espérance et la variance de la loi \mathbb{P}_ϑ ,

$$\text{c.à.d. } \mathbb{E}_\vartheta[X_i] = \mu_\vartheta, \quad \text{var}_\vartheta(X_i) = \sigma_\vartheta^2.$$

On dit qu'un estimateur est *linéaire* s'il est de la forme

$$T = b + \sum_{i=1}^n a_i X_i \quad \text{avec } b, a_1, \dots, a_n \in \mathbb{R}.$$

On cherche à estimer μ_ϑ (supposons que $\vartheta \mapsto \mu_\vartheta$ prend au moins deux valeurs). Soit T un estimateur linéaire de μ_ϑ . Alors

$$\mathbb{E}_\vartheta[T] = b + \sum_{i=1}^n a_i \underbrace{\mathbb{E}_\vartheta[X_i]}_{=\mu_\vartheta} = b + \left(\sum_{i=1}^n a_i \right) \mu_\vartheta.$$

D'où

$$T \text{ sans biais} \iff \mathbb{E}_\vartheta[T] = \mu_\vartheta \iff b = 0 \text{ et } \sum_{i=1}^n a_i = 1.$$

Dans ce cas, le risque quadratique est

$$R_T(\vartheta) = \text{var}_\vartheta(T) = \sum_{i=1}^n a_i^2 \underbrace{\text{var}_\vartheta(X_i)}_{=\sigma_\vartheta^2} = \left(\sum_{i=1}^n a_i^2 \right) \sigma_\vartheta^2.$$

Comme $\sum_{i=1}^n a_i = 1$, on a toujours

$$\sum_{i=1}^n a_i^2 \geq \frac{1}{n},$$

avec égalité si et seulement si $a_i = \frac{1}{n}$ pour tout n , auquel cas $T = \bar{X}$ est la *moyenne empirique*.

COROLLAIRE 9.10. *Pour un n -échantillon admettant une variance, et pour estimer l'espérance, la moyenne empirique \bar{X} est le meilleur estimateur parmi tous les estimateurs linéaires sans biais.*

REMARQUE 9.11 (Estimateurs sans biais de la variance σ_ϑ^2).

(1) Si l'espérance $\mu_\vartheta = \mu$ est connue, alors

$$\Sigma^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$$

est un estimateur sans biais de σ_ϑ^2 . En fait,

$$\mathbb{E}_\vartheta[\Sigma^2] = \frac{1}{n} \sum_{i=1}^n \underbrace{\mathbb{E}_\vartheta[(X_i - \mu)^2]}_{=\sigma_\vartheta^2} = \sigma_\vartheta^2.$$

(2) Si μ est inconnu, alors un candidat naturel est

$$\Sigma^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

On trouve que

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 &= \frac{1}{n} \sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2) \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - 2\bar{X}^2 + \bar{X}^2 \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \frac{1}{n^2} \sum_{i,j=1}^n X_i X_j \\ &= \underbrace{\left(\frac{1}{n} - \frac{1}{n^2} \right)}_{\frac{n-1}{n^2}} \sum_{i=1}^n X_i^2 - \frac{1}{n^2} \sum_{\substack{i,j=1 \\ i \neq j}}^n X_i X_j. \end{aligned}$$

Donc

$$\begin{aligned} \mathbb{E}_\vartheta[\Sigma^2] &= \frac{1}{n} \frac{n-1}{n} \sum_{i=1}^n \mathbb{E}_\vartheta[X_i^2] - \frac{1}{n^2} \sum_{\substack{i,j=1 \\ i \neq j}}^n \underbrace{\mathbb{E}_\vartheta[X_i]}_{=\mu_\vartheta} \underbrace{\mathbb{E}_\vartheta[X_j]}_{=\mu_\vartheta} \\ &= \frac{1}{n} \frac{n-1}{n} (n\sigma_\vartheta^2 + n\mu_\vartheta^2) - \frac{1}{n^2} (n^2 - n)\mu_\vartheta^2 \\ &= \frac{n-1}{n} \sigma_\vartheta^2, \end{aligned}$$

où pour la deuxième égalité, nous avons utilisé que

$$\mathbb{E}[X_i^2] - \underbrace{\mathbb{E}[X_i]^2}_{=\mu_\vartheta^2} = \sigma_\vartheta^2.$$

Conclusion. La statistique

$$\Sigma^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

est biaisée, mais

$$S^2 := \frac{n}{n-1} \Sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (\text{variance empirique})$$

est un estimateur de la variance σ_ϑ^2 sans biais.

9.2. La loi du chi-carré et de Student

LEMME 9.12. Soit X_1, \dots, X_n un n -échantillon de la loi normal $N(0, 1)$, c.à.d. les X_1, \dots, X_n sont indépendantes et chaque X_i suit la loi normale de moyenne $\mu = 0$ et de variance $\sigma^2 = 1$. Soit $A \in \text{Matr}(n \times n; \mathbb{R})$ une matrice orthogonale, c.à.d. $A \cdot A^t = \mathbf{1}_n$. Alors les composantes Y_1, \dots, Y_n du vecteur aléatoire

$$Y = A \cdot X, \quad \text{où } X = (X_1, \dots, X_n),$$

forment aussi un n -échantillon de la loi normal $N(0, 1)$.

DÉMONSTRATION. On a

$$Y_i = \sum_{k=1}^n A_{ik} X_k,$$

donc chaque Y_i suit une loi normale. on obtient:

$$\begin{aligned} \text{cov}(Y_i, Y_j) &= \text{cov}\left(\sum_k A_{ik} X_k, \sum_\ell A_{j\ell} X_\ell\right) \\ &= \sum_{k, \ell} A_{ik} A_{j\ell} \underbrace{\text{cov}(X_k, X_\ell)}_{=\delta_{k, \ell}} \\ &= \sum_k A_{ik} A_{kj}^t = (AA^t)_{ij} = \delta_{i, j}. \end{aligned}$$

Autrement dit, les Y_1, \dots, Y_n suivent la loi $N(0, 1)$ et sont non-corelées. Pour conclure, on utilise le Lemme suivant. \square

LEMME 9.13. Soit $X = (X_1, \dots, X_n)$ un vecteur gaussien au sens que chaque composant X_i suit une loi normale. Alors les composantes X_1, \dots, X_n sont indépendantes

si et seulement si X_1, \dots, X_n sont non-corelées.

DÉMONSTRATION. Voir le Cours “Complements de Probabilités et Statistiques”. \square

DÉFINITION 9.14. Soit X_1, \dots, X_n un n -échantillon de la loi $N(0, 1)$. La loi de

$$U := X_1^2 + \dots + X_n^2$$

s'appelle loi du *chi-carré* à n -degrés de liberté. On la note χ_n^2 .

PROPOSITION 9.15. Soit X_1, \dots, X_n un n -échantillon de la loi $N(\mu, \sigma^2)$. Alors

- (1) \bar{X} et S^2 sont indépendantes;
- (2) \bar{X} suit la loi $N(\mu, \frac{\sigma^2}{n})$;
- (3) $(n-1) \frac{S^2}{\sigma^2}$ suit la loi χ_{n-1}^2 .

DÉMONSTRATION. En faisant le changement de variable

$$Z_i := \frac{X_i - \mu}{\sigma},$$

on remplace \bar{X} par $(\bar{X} - \mu)/\sigma$ et S^2 par S^2/σ^2 ; donc sans restrictions $\mu = 0$ et $\sigma^2 = 1$.

Choisissons une matrice orthogonale $A \in \text{Matr}(n \times n; \mathbb{R})$ dont les éléments de la dernière ligne sont tous égaux à $\frac{1}{\sqrt{n}}$. Alors, par le Lemme 9.12, les composantes (Y_1, \dots, Y_n) du vecteur

$$Y = AX, \quad \text{où } X = (X_1, \dots, X_n)$$

forment aussi un n -échantillon de $N(0, 1)$. On remarque que

$$Y_n = \frac{1}{\sqrt{n}}(X_1 + \dots + X_n) = \sqrt{n}\bar{X},$$

et

$$\sum_{i=1}^n Y_i^2 = \sum_{i=1}^n X_i^2.$$

En effet,

$$\begin{aligned} \sum_{i=1}^n X_i^2 &= \sum_{i,\ell,k} A_{i\ell} X_\ell A_{ik} X_k \\ &= \sum_{\ell,k} \left(\underbrace{\sum_i A_{i\ell}^t A_{ik}}_{=\delta_{\ell k}} \right) X_\ell X_k = \sum_{\ell=1}^n X_\ell^2. \end{aligned}$$

Donc, on obtient

$$\begin{aligned} (n-1)S^2 &= \sum_{i=1}^n (X_i - \bar{X})^2 \\ &= \sum_{i=1}^n X_i^2 - n\bar{X}^2 \\ &= \sum_{i=1}^n Y_i^2 - Y_n^2 = \sum_{i=1}^{n-1} Y_i^2. \end{aligned}$$

Comme les Y_1, \dots, Y_n sont indépendantes et de loi $N(0, 1)$, on trouve bien que S^2 et Y_n (donc aussi S^2 et \bar{X}) sont indépendantes, et que la variable $(n-1)S^2$ suit la loi χ_{n-1}^2 . \square

DÉFINITION 9.16. Soient X et Y deux variables aléatoires indépendantes avec X de loi $N(0, 1)$ et Y de loi χ_n^2 . Alors la loi de

$$T_n := \frac{X\sqrt{n}}{\sqrt{Y}}$$

s'appelle *loi de Student à n degrés de liberté*, et on la note t_n .

PROPOSITION 9.17. Soit X_1, \dots, X_n un n -échantillon de $N(\mu, \sigma^2)$. Alors

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S}$$

suit la loi de Student t_{n-1} à $n-1$ degrés de liberté.

DÉMONSTRATION. Les variables

$$X := \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad \text{et} \quad Y := (n-1) \frac{S^2}{\sigma^2}$$

sont indépendantes, et suivent les lois $N(0, 1)$, resp. χ_{n-1}^2 . Donc

$$T_{n-1} := \frac{X\sqrt{n-1}}{\sqrt{Y}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sqrt{n-1} \frac{\sigma}{\sqrt{n-1}S} = \sqrt{n} \frac{\bar{X} - \mu}{S}$$

suit la loi de Student t_{n-1} à $n-1$ degrés de liberté. \square

9.3. Intervalles de confiance

DÉFINITION 9.18. Soit T un estimateur de ϑ et $\alpha \in]0, 1[$ (on fixe a priori ce nombre proche de 1; typiquement on prend $\alpha = 0,95$ ou $\alpha = 0,99$).

On appelle *intervalle de confiance de niveau α* un intervalle aléatoire de la forme

$$[T(\omega) - A(\omega), T(\omega) + A(\omega)],$$

dans lequel le paramètre ϑ se trouve avec une probabilité au moins égale à α , c.à.d.

$$\mathbb{P}_\vartheta \{ |T - \vartheta| \leq A \} \geq \alpha, \quad \forall \vartheta \in \Theta.$$

REMARQUE 9.19. Par définition, A est une variable aléatoire positive, qu'on choisit souvent égale à une constante.

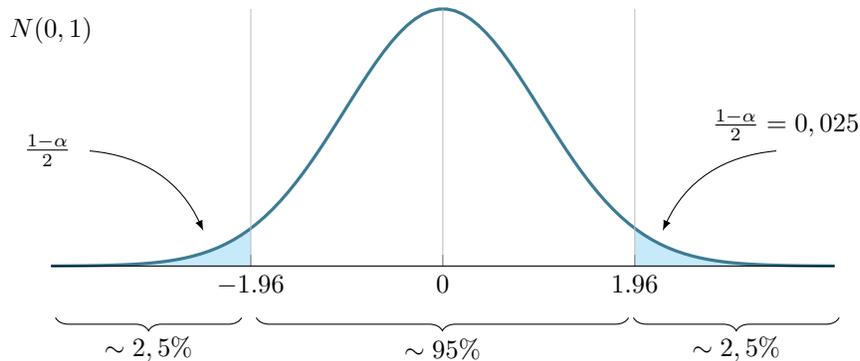
EXEMPLES 9.20.

(a) Considérons un n -échantillon de $N(\vartheta, \sigma^2)$ avec σ^2 connu. Rappelons que la moyenne empirique \bar{X} est un estimateur sans biais pour la moyenne ϑ . La loi de $\bar{X} - \vartheta$ sous \mathbb{P}_ϑ est $N(0, \frac{\sigma^2}{n})$ et par conséquent la variable

$$Z := \frac{\bar{X} - \vartheta}{\sigma/\sqrt{n}} = \sqrt{n} \frac{\bar{X} - \vartheta}{\sigma}$$

suit la loi $N(0, 1)$.

Fixons maintenant un niveau α , par exemple $\alpha = 0,95$:



On lit dans la table de la loi normale que

$$\mathbb{P}_\vartheta \left\{ \left| \sqrt{n} \frac{\bar{X} - \vartheta}{\sigma} \right| > 1,96 \right\} = 0,05.$$

En conséquence,

$$\left[\bar{X} - 1,96 \frac{\sigma}{\sqrt{n}}, \bar{X} + 1,96 \frac{\sigma}{\sqrt{n}} \right]$$

est un intervalle de niveau 0,95 pour la moyenne ϑ .

(b) Soit (X_1, \dots, X_n) un n -échantillon de $N(\vartheta, \sigma^2)$ avec σ^2 *inconnu*; on cherche un intervalle de confiance de niveau $\alpha = 0,95$ pour ϑ , lorsque $n = 25$.

Rappelons que

$$\frac{\sqrt{n}(\bar{X} - \vartheta)}{S} \sim t_{24}.$$

Pour une variable T qui suit la loi de Student t_{24} , on lit dans la table

$$\mathbb{P}\{|T| > 2,06\} \sim 0,05.$$

L'intervalle cherché est donc

$$(*) \quad \left[\bar{X} - 2,06 \frac{S}{\sqrt{n}}, \bar{X} + 2,06 \frac{S}{\sqrt{n}} \right] = [\bar{X} - 0,41 S, \bar{X} + 0,41 S].$$

REMARQUE 9.21. Si σ^2 est *connu*, l'intervalle est

$$(**) \quad \left[\bar{X} - 1,96 \frac{\sigma}{\sqrt{n}}, \bar{X} + 1,96 \frac{\sigma}{\sqrt{n}} \right].$$

Si on avait pris l'intervalle $(**)$ en remplaçant σ^2 par S^2 (resp. σ par S), on aurait obtenu un intervalle "trop optimiste" (i.e. trop petit); la différence entre 1,96 et 2,06 n'est pas très grande: plus n est grande, plus la différence entre $(*)$ et $(**)$ est faible.

9.4. Tests d'hypothèses

Soit $(\Omega, \mathcal{A}, (\mathbb{P}_\vartheta)_{\vartheta \in \Omega})$ un modèle statistique où Θ est divisé en une partie Θ_0 et son complémentaire $\Theta_1 := \Theta \setminus \Theta_0$.

L'*objectif*: On veut tester l'hypothèse

$$H_0: \vartheta \in \Theta_0 \quad (\text{"hypothèse nulle"}),$$

contre l'alternative

$$H_1: \vartheta \in \Theta_1 \quad (\text{"hypothèse alternative"}).$$

On considère $X = (X_1, \dots, X_n)$ un échantillon de variables aléatoires i.i.d. dans le modèle $(\mathbb{P}_\vartheta)_{\vartheta \in \Theta}$.

DÉFINITION 9.22. On appelle *test d'hypothèse* une règle

$$e: \mathbb{R}^n \rightarrow \{0, 1\},$$

telle que au vu de l'observation $x_1 = X_1(\omega), \dots, x_n = X_n(\omega)$, ou bien on rejette H_0 (c.à.d. $e(x_1, \dots, x_n) = 1$) ou bien on garde H_0 (c.à.d. $e(x_1, \dots, x_n) = 0$).

REMARQUE 9.23. Le test est déterminé par sa *région critique*

$$K = \{\omega \in \Omega \mid e(X_1(\omega), \dots, X_n(\omega)) = 1\},$$

resp. la partie correspondante dans \mathbb{R}^n ,

$$\{(x_1, \dots, x_n) \mid e(x_1, \dots, x_n) = 1\} \subset \mathbb{R}^n.$$

On considère la fonction

$$p: \Theta \rightarrow [0, 1], \quad \vartheta \mapsto \mathbb{P}_\vartheta(K).$$

Problème: On veut construire le test tel que

- p soit petite sur Θ_0 ;
- p soit grande sur $\Theta \setminus \Theta_0 = \Theta_1$.

NOTATION 9.24. La fonction $p|_{\Theta_1}$, c.à.d.

$$\Theta_1 \rightarrow [0, 1], \quad \vartheta \mapsto \mathbb{P}_\vartheta(K),$$

s'appelle *fonction puissance* du test.

Types d'erreurs

- (1) Si $\vartheta \in \Theta_0$, la probabilité $\mathbb{P}_\vartheta(K)$ s'appelle *risque d'erreur de première espèce*.
- (2) Si $\vartheta \in \Theta_1$, la probabilité $\mathbb{P}_\vartheta(K^c) = 1 - \mathbb{P}_\vartheta(K)$ s'appelle *risque d'erreur de seconde espèce*.

décision \ réalité	H_0 vraie	H_0 fausse
rejet de H_0	<i>rejet à tort</i> (erreur de 1 ^{ère} espèce)	correct
non-rejet de H_0	correct	<i>manque de puissance</i> (erreur de 2 ^{nde} espèce)

Méthodologie pratique

- (1) On minimise en priorité le risque de 1^{ère} espèce: soit

$$\alpha := \sup_{\vartheta \in \Theta_0} \mathbb{P}_\vartheta(K) \text{ "niveau du test" (ou niveau de la région critique } K).$$

Dans ce but on choisit une borne supérieure pour α ; les valeurs les plus courantes pour α sont $\alpha = 0,05$ ou $\alpha = 0,01$.

- (2) Parmi tous les tests de niveau α (ou de niveau $\leq \alpha$) on cherche à maximiser la fonction puissance

$$\Theta_1 \ni \vartheta \mapsto \mathbb{P}_\vartheta(K).$$

EXEMPLE 9.25. Dans le modèle

$$(\mathbb{P}_\vartheta)_{\vartheta \in \Theta} = \{N(\mu, \sigma^2) \mid \mu = \mu_0, \mu_1\}$$

avec σ^2 connu et $\mu_1 > \mu_0$, on souhaite tester

$$H_0: \mu = \mu_0 \quad \text{contre} \quad H_1: \mu = \mu_1.$$

On va rejeter H_0 si

$$\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$$

est suffisamment grande! Dans ce but on prend K de la forme $K = \{\bar{X} \geq a\}$.

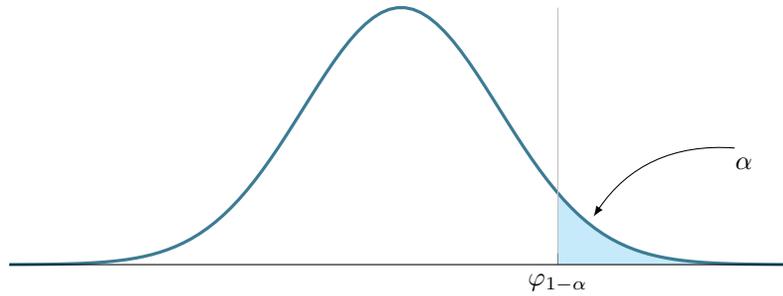
Erreur de 1^{ère} espèce.

Sous $H_0 : \bar{X} \sim N(\mu_0, \frac{\sigma^2}{n})$, et donc

$$Z := \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1).$$

On veut que

$$\mathbb{P}_{\mu_0}(K) = \mathbb{P}_{\mu_0}\{\bar{X} \geq a\} = \mathbb{P}_{\mu_0}\left\{Z \geq \sqrt{n} \frac{a - \mu_0}{\sigma}\right\} \stackrel{!}{=} \alpha.$$



$$\varphi_{1-\alpha} = \text{quantile d'ordre } 1 - \alpha$$

Par conséquent,

$$a := \mu_0 + \sigma \frac{\varphi_{1-\alpha}}{\sqrt{n}}$$

assure que le niveau du test est α .

Erreur de 2^{nde} espèce.

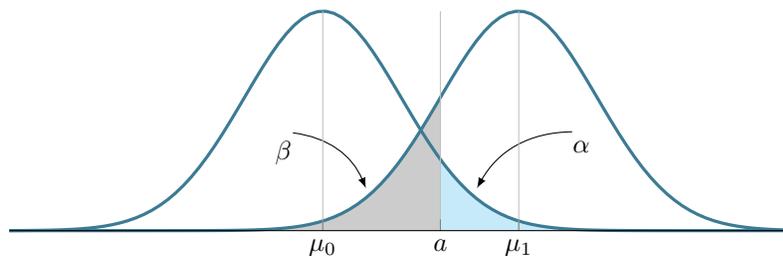
Pour ce choix, la probabilité de ne pas rejeter H_0 à tort est

$$\begin{aligned} \beta &:= \mathbb{P}_{\mu_1}(K^c) \\ &= \mathbb{P}_{\mu_1}\{\bar{X} < a\} \\ &= \mathbb{P}\left\{\mu_1 + \frac{\sigma Z}{\sqrt{n}} < \mu_0 + \sigma \frac{\varphi_{1-\alpha}}{\sqrt{n}}\right\} \\ &= \mathbb{P}\left\{Z < \varphi_{1-\alpha} + \sqrt{n} \frac{\mu_0 - \mu_1}{\sigma}\right\}. \end{aligned}$$

Pour μ_1 assez proche de μ_0 , la probabilité d'une erreur de 2^{nde} espèce peut être arbitrairement proche de

$$\mathbb{P}\{Z < \varphi_{1-\alpha}\} = 1 - \alpha$$

(donc le risque de 2^{nde} espèce est beaucoup moins bien contrôlé que le risque de 1^{ère} espèce).



On appelle

$$p\text{-valeur du test} := \mathbb{P}_{\mu_0}\{\bar{X} \geq \bar{x}^{\text{obs}}\} = \mathbb{P}\left\{Z \geq \sqrt{n} \frac{\bar{x}^{\text{obs}} - \mu_0}{\sigma}\right\} \quad \text{où } Z \sim N(0, 1).$$

On constate que

$$\begin{aligned} p\text{-valeur} \leq \alpha &\iff \bar{x}^{\text{obs}} \geq \mu_0 + \sigma \frac{\varphi_{1-\alpha}}{\sqrt{n}} \\ &\iff \bar{x}^{\text{obs}} \text{ est dans la région critique.} \end{aligned}$$

EXEMPLE 9.26. Soit maintenant

$$(\mathbb{P}_{\vartheta})_{\vartheta \in \Theta} = \{N(\mu, \sigma^2) : \mu \geq \mu_0, \sigma^2 > 0\} \quad \text{où } \mu_0 \in \mathbb{R}.$$

Test pour la moyenne μ . On veut tester

$$H_0: \mu = \mu_0 \quad \text{contre} \quad H_1: \mu > \mu_0$$

au niveau α . Considérons

$$\begin{aligned} \bar{X}_n &= \frac{1}{n} \sum_{i=1}^n X_i, \\ S_n^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2. \end{aligned}$$

Alors sous H_0 ,

$$T_n := \sqrt{n} \frac{\bar{X}_n - \mu_0}{S_n} \sim \text{la loi de Student } t_{n-1}.$$

Sous H_1 , on a par la loi forte des grands nombres:

$$\begin{aligned} \bar{X}_n - \mu_0 &\rightarrow \mu - \mu_0 > 0, \quad \text{p.s.} \\ S_n^2 &\rightarrow \sigma^2, \quad \text{p.s.} \end{aligned}$$

et puis $T_n \rightarrow +\infty$ p.s. lorsque $n \rightarrow \infty$.

On choisit donc $K = \{T_n \geq a\}$. Si $a \geq t_{n-1, 1-\alpha}$ (= quantile d'ordre $1 - \alpha$ pour t_{n-1}), alors

$$\begin{aligned} \forall \sigma^2 > 0, \quad \mathbb{P}_{(\mu_0, \sigma^2)}(K) &= \mathbb{P}_{(\mu_0, \sigma^2)}\{T_n \geq a\} \\ &\leq \mathbb{P}_{(\mu_0, \sigma^2)}\{T_n \geq t_{n-1, 1-\alpha}\} = \alpha. \end{aligned}$$

Comme on souhaite minimiser le risque de 2nde espèce:

$$\mathbb{P}_{(\mu, \sigma^2)}\{T_n < a\}, \quad \mu > \mu_0,$$

on choisit $a = t_{n-1, 1-\alpha}$. On remarque que la p -valeur est égale à $\mathbb{P}\{T \geq T_n^{\text{obs}}\}$, où $T \sim t_{n-1}$.

EXEMPLE 9.27 (Changement climatique). Considérons les données de la température moyenne au mois d'août à Paris. On sait que les données suivent la loi $N(20, \sigma^2)$ sur l'ensemble du 19^{ième} et 20^{ième} siècle.

Sur les 10 dernières années, on a observé les températures moyennes suivantes

$$x^{\text{obs}} = (x_1, \dots, x_{10}) = (22, 19, 21, 23, 20, 22, 24, 18, 20, 25).$$

On trouve

$$\bar{x}_{10} = 21,4 \quad \text{et} \quad S_{10} = 2,22$$

et par conséquent

$$T_{10}^{\text{obs}} \approx 1,99, \quad t_{9,1-\alpha} \approx 1,83 \quad (\text{avec } \alpha = 0,05).$$

Donc on rejette $H_0: \mu = 20$ au niveau $\alpha = 0,05$.

Index

- écart type, 31
- échantillon, 95
- égalité
 - de Bienaymé, 35
- événement, 2
- événements
 - indépendants, 15
- arrangement, 9
- axiomes de Kolmogonov, 56
- Bernstein
 - inégalité, 47
- biais, 96
- Bienaymé
 - égalité, 35
- Borel
 - théorème de Borel, 87
- Cauchy
 - inégalité, 32
- coefficient
 - de corrélation, 31
- combinaison, 9
- convergence
 - dans L^p , 84
 - en probabilité, 84
 - presque sûrement, 84
 - en loi, 92
- convolution
 - des fonctions, 72
- coordonnées polaires, 63
- corrélé négativement, 32
- corrélé positivement, 32
- couplage
 - d'expériences, 22
- couplage indépendant, 22
- covariance, 31
- de Moivre-Laplace, 92
- densité
 - d'une variable aléatoire, 61
- densité commune, 66
- distribution
 - de probabilité, 1
 - distribution de probabilité, 1
- erreur
 - de première espèce, 103
 - de seconde espèce, 103
- espérance, 26
 - d'une variable aléatoire, 69
 - de loi exponentielle, 72
 - de loi normale, 72
 - de loi uniforme, 71
- espace
 - de probabilité, 2
- espace probabilisé, 2
 - axiomes, 3
 - produit, 22
- estimateur, 96
 - linéaire, 97
 - meilleur, 96
 - sans biais, 96
 - strictement meilleur, 96
- estimation, 95
- expérience
 - de Bernoulli, 57
- expérience aléatoire, 1
- expériences répétées, 57
- fonction
 - caractéristique, 89
 - de répartition, 60
- fonction caractéristique, 89
- fonction de répartition, 59
 - propriétés, 60
- formule
 - de Bayes, 18
 - de Poincaré, 7
- Hoeffding
 - inégalité, 49
- inégalité
 - de Bernstein, 47
 - de Cauchy, 32
 - de Hoeffding, 49
 - de Jensen, 31
 - de Markov, 47

- de Tchebychev-Markov, 47
- indépendance, 15
- Intervalle de confiance, 101
 - niveau, 101
- Kolmogorov
 - axiomes, 56
 - loi forte, 86
- lemme
 - de Borel-Cantelli, 81
- loi
 - d'une somme indépendante, 72
 - d'une variable aléatoire, 59
 - de Student, 100
 - du chi-carré, 99
 - du tout ou rien, 81
 - exponentielle, 62
 - gamma, 75
 - normale, 62
 - uniforme, 61
 - binomiale, 12
 - hypergéométrique, 12
- loi de Student, 100
 - degrés de liberté, 100
- loi du chi-carré, 99
 - degrés de liberté, 99
- loi exponentielle
 - caractérisation, 64
- loi forte, 86
- Markov
 - inégalité, 47
- mesure
 - de Lebesgue, 56
- modèle statistique, 95
- moment, 31
 - centré, 31
 - d'ordre n , 31
- Montmort, 8
- moyenne
 - empirique, 97
- moyenne empirique, 96
- nombre normaux, 87
- non-corrélé, 32
- paradoxe
 - de Méré, 5
- probabilité, 2
 - conditionnelle, 15
- probabilité des causes, 18
- problème
 - du vestiaire, 8
- produit, 22
- puissance
 - d'un test, 103
- répartition, 59
- risque
 - quadratique, 96
- roulette
 - américaine, 94
- singe savant, 83
- Tchebychev-Markov
 - inégalité, 47
- test, 95
- théorème
 - de Borel, 87
 - de Moivre-Laplace, 92
- tirage
 - avec remise, 9
 - sans remise, 9
- transformation
 - orthogonal, 99
- tribu
 - engendrée, 56
- tribu borélienne, 58
- variable aléatoire, 58
 - continue, 61
 - sans mémoire, 64
- variables aléatoires
 - corrélées négativement, 32
 - corrélées positivement, 32
 - non-corrélées, 32
- variance, 31
 - empirique, 98
- vecteur
 - aléatoire, 66

Notations

$(\Omega, \mathcal{A}, \mathbb{P})$ (espace probabilisé), 2, 56
 A_∞ (A_n infiniment souvent), 81
 F_X (fonction de répartition de X), 59
 $N(\mu, \sigma^2)$ (loi normale), 62
 $\mathbb{E}[X]$ (espérance de X), 26, 69
 \mathbb{P} (probabilité), 2, 56
 $\mathbb{P}(A|B)$ (probabilité conditionnelle), 15
 \mathbb{P}_X (loi de X), 61
 \mathcal{A} (tribu), 56
 $\mathcal{B}(k, p)$ (loi binomiale), 12
 $\mathcal{H}(k; m, n)$ (loi hypergéométrique), 12
 χ_n^2 (loi du chi-carré), 99
 $\text{cov}(X, Y)$ (variance de X et Y), 31
 $\text{var}(X)$ (variance de X), 31
 $\rho(X, Y)$ (coefficient de corrélation), 31
 $\sigma(\mathcal{A}_0)$ (tribu engendrée par \mathcal{A}_0), 56
 φ_X (fonction caractéristique de X), 89
 f_X (densité de X), 61
 t_n (loi de Student), 100