

Probabilités et Statistiques

Licence 3ème année

Université d'Orléans

Nils Berglund

Version de Mars 2010

Table des matières

1	Probabilités discrètes	1
1.1	Espace probabilisé discret	1
1.1.1	Probabilités conditionnelles	4
1.1.2	Indépendance	6
1.2	Variables aléatoires	7
1.2.1	Loi d'une variable aléatoire	8
1.2.2	Espérance	9
1.2.3	Variance et covariance	11
1.2.4	Fonction génératrice	14
2	Probabilités continues	17
2.1	Variables aléatoires réelles à densité	17
2.1.1	Densité et fonction de répartition	17
2.1.2	Espérance et variance	21
2.1.3	Fonction caractéristique	22
2.1.4	Changement de variable	24
2.2	Vecteurs aléatoires à densité	24
2.2.1	Densité conjointe	24
2.2.2	Indépendance, convolution	28
2.2.3	Changement de variable	29
2.3	Mesures de probabilité et espaces probabilisés*	33
2.3.1	Mesures	33
2.3.2	Applications mesurables	34
2.3.3	Intégrale de Lebesgue	35
3	Théorèmes limite	39
3.1	La "loi des petits nombres"	39
3.1.1	Convergence en loi	40
3.1.2	Convergence en distance ℓ_1	41
3.2	La loi des grands nombres	43
3.2.1	Loi faible des grands nombres	43
3.2.2	Grandes déviations*	44
3.3	Le théorème de la limite centrale	45
3.3.1	Cas de la loi binomiale : formule de Moivre–Laplace	46
3.3.2	Cas général	48

4	Introduction à la statistique	51
4.1	Estimateurs	51
4.1.1	Estimateurs empiriques	52
4.1.2	Estimateur de maximum de vraisemblance	53
4.2	Intervalles de confiance	55
4.2.1	Intervalle de confiance pour une proportion	55
4.2.2	Intervalles de confiance pour les paramètres d'une loi gaussienne	56
4.3	Test d'hypothèses	59
4.3.1	Tests sur une loi gaussienne	60
4.3.2	Tests d'adéquation et d'indépendance du χ^2	61

Chapitre 1

Probabilités discrètes

La théorie des probabilités sert à modéliser des situations dont notre connaissance est imparfaite. Le manque d'informations est alors remplacé par une composante aléatoire.

Par exemple, lors du jet d'un dé, les lois de Newton devraient en principe nous permettre de calculer la trajectoire exacte du dé, connaissant sa position et sa vitesse initiales, et d'en déduire sur quelle face il va tomber. En pratique, non seulement ce calcul est extrêmement difficile, mais le résultat dépend aussi de manière très sensible des conditions initiales. Il est alors plus simple d'admettre que le dé peut tomber sur chacune de ses six faces avec la même *probabilité* de $1/6$ (si le dé est parfaitement symétrique – sinon, il peut être préférable d'associer des probabilités différentes aux différentes faces).

En théorie des probabilités, on suppose donnés un ensemble de résultats possibles de l'“expérience” considérée, et leurs probabilités respectives. On cherche alors à en déduire les probabilités d'événements plus compliqués, ou les résultats d'expériences plus complexes, comme par exemple le lancer d'un grand nombre de dés.

1.1 Espace probabilisé discret

Un espace probabilisé discret est caractérisé par trois ingrédients:

1. Un *univers* Ω : c'est l'ensemble des *événements élémentaires* de l'expérience, supposé ici discret (fini ou dénombrable).
2. Un ensemble d'*événements* (ou *événements composés*) \mathcal{F} : tout événement $A \in \mathcal{F}$ est un sous-ensemble de Ω ($A \subset \Omega$).
3. Une *distribution de probabilité* $p : \Omega \rightarrow [0, 1]$, satisfaisant

$$\sum_{\omega \in \Omega} p(\omega) = 1 . \tag{1.1.1}$$

Pour tout $\omega \in \Omega$, $p(\omega)$ est appelée la *probabilité de l'événement élémentaire* ω .

Exemple 1.1.1.

1. Pour un jet de dé (non pipé), on pourra prendre l'univers $\Omega = \{1, 2, 3, 4, 5, 6\}$, et comme distribution $p(\omega) = 1/6$ pour tout $\omega \in \Omega$ (distribution *uniforme*). Un exemple d'événement composé est $A = \{\omega : \omega \text{ est pair}\} = \{2, 4, 6\}$.
2. Pour un jet de deux pièces de monnaie, pouvant indiquer Pile (P) ou Face (F), on peut prendre $\Omega = \{PP, PF, FP, FF\}$, avec $p(\omega) = 1/4$ pour tout $\omega \in \Omega$.

3. Si l'on tire successivement trois boules d'un sac contenant exactement trois boules numérotées de 1 à 3, on pourra prendre $\Omega = \{(1, 2, 3), (1, 3, 2), \dots, (3, 2, 1)\}$, de nouveau avec la distribution uniforme $p(\omega) = 1/6$ pour tout $\omega \in \Omega$.

Remarque 1.1.2. Le choix de l'univers Ω n'est pas unique, en fait on peut choisir n'importe quel ensemble contenant au moins autant d'éléments qu'il y a d'événements considérés comme distinguables par l'expérience. Par exemple, dans le cas de deux pièces, on aurait pu considérer qu'on ne sait pas distinguer les pièces l'une de l'autre, et choisir $\Omega = \{PP, PF, FF\}$. Cette fois, PF désigne l'événement "les deux pièces ne sont pas tombées du même côté". On prendra alors $p(PP) = p(FF) = 1/4$, et $p(PF) = 1/2$.

Définition 1.1.3 (Espace probabilisé discret). *Un espace probabilisé discret (Ω, p) est donné par un ensemble dénombrable Ω et une application $p : \Omega \rightarrow [0, 1]$ telle que*

$$\sum_{\omega \in \Omega} p(\omega) = 1. \quad (1.1.2)$$

Remarque 1.1.4. Nous n'avons pas exclu la possibilité que l'univers Ω soit infini. Dans ce cas, la somme (1.1.2) doit être considérée comme la somme d'une série numérique. Comme tous les termes de la série sont non-négatifs, la somme est indépendante de leur ordre (ce qui n'est pas nécessairement le cas pour des séries à termes positifs et négatifs).

Exemple 1.1.5. On jette une pièce jusqu'à obtention du premier pile. On peut alors choisir $\Omega = \mathbb{N}^*$, où $\omega \in \Omega$ désigne le numéro du jet lors duquel on obtient le premier pile, avec $p(\omega) = 2^{-\omega}$. Il s'agit ici d'un exemple d'univers dénombrable mais infini. On a bien

$$\sum_{\omega \in \Omega} p(\omega) = \sum_{i=1}^{\infty} \frac{1}{2^i} = 1. \quad (1.1.3)$$

Définition 1.1.6 (Événements). *L'espace des événements (ou événements composés) d'un espace probabilisé discret (Ω, p) est l'ensemble des parties de Ω :*

$$\mathcal{F} = \mathcal{P}(\Omega) = \{A : A \subset \Omega\}. \quad (1.1.4)$$

La probabilité de l'événement A est

$$\mathbb{P}(A) = \sum_{\omega \in A} p(\omega). \quad (1.1.5)$$

L'ensemble vide \emptyset est l'événement impossible et $\mathbb{P}(\emptyset) = 0$ par définition.

L'univers entier Ω est l'événement certain.

Les opérations logiques élémentaires sur les événements correspondent à des opérations de théorie des ensembles, selon le tableau suivant:

Opération logique	Équivalent ensembliste
A et B	$A \cap B$
A ou B	$A \cup B$
non A	$A^c = \Omega \setminus A$
A, B incompatibles	$A \cap B = \emptyset$
A implique B	$A \subset B$

Proposition 1.1.7.

1. Pour tout $A \in \mathcal{F}$, $0 \leq \mathbb{P}(A) \leq 1$.
2. $\mathbb{P}(\Omega) = 1$ et $\mathbb{P}(\emptyset) = 0$.
3. Pour tout ensemble d'événements $\{A_i\}_{i \in \mathbb{N}}$, on a

$$\mathbb{P}\left(\bigcup_{i \in \mathbb{N}} A_i\right) \leq \sum_{i \in \mathbb{N}} \mathbb{P}(A_i). \quad (1.1.6)$$

4. Si les événements $\{A_i\}_{i \in \mathbb{N}}$ sont deux à deux incompatibles, c'est-à-dire $A_i \cap A_j = \emptyset$ pour $i \neq j$, alors

$$\mathbb{P}\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} \mathbb{P}(A_i). \quad (1.1.7)$$

5. Si $A \subset B$, alors $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$.
6. Si $A \subset B$, alors $\mathbb{P}(A) \leq \mathbb{P}(B)$.
7. On a $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

DÉMONSTRATION. En exercice. □

Remarque 1.1.8.

1. L'ensemble $\mathcal{F} = \mathcal{P}(\Omega)$ a la propriété que le complémentaire de tout élément de \mathcal{F} , toute intersection et toute réunion d'éléments de \mathcal{F} sont encore dans \mathcal{F} . On dit que \mathcal{F} forme une *tribu* ou *σ -algèbre*.
2. (Ω, p) est un espace probabilisé discret si et seulement si l'application $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ définie par (1.1.5) satisfait les deux *axiomes de Kolmogorov*:
 - (K1) $\mathbb{P}(\Omega) = 1$.
 - (K2) Si I est un ensemble dénombrable et $\{A_i\}_{i \in I}$ est une famille d'événements deux à deux incompatibles, alors

$$\mathbb{P}\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mathbb{P}(A_i). \quad (1.1.8)$$

En effet, la Proposition 1.1.7 montre que \mathbb{P} satisfait (K1) et (K2). Inversement, soit $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ une application satisfaisant (K1) et (K2). Alors le fait que $\Omega = \Omega \cup \emptyset$ implique $\mathbb{P}(\emptyset) = 0$. Si l'on définit p par $p(\omega) = \mathbb{P}(\{\omega\})$, il est clair que (1.1.5) est satisfaite, et en particulier $1 = \mathbb{P}(\Omega) = \sum_{\omega \in \Omega} p(\omega)$. (Ω, p) est donc bien un espace probabilisé discret.

On peut donc également définir un espace probabilisé par la donnée d'un triplet $(\Omega, \mathcal{F}, \mathbb{P})$, où $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ satisfait les axiomes de Kolmogorov (K1) et (K2). Cette définition a l'avantage d'être généralisable à des Ω non dénombrables.

Exemple 1.1.9.

1. Pour le lancer de deux dés équilibrés, on peut prendre $\Omega = \{(1, 1), (1, 2), \dots, (6, 6)\}$, dont le cardinal est $|\Omega| = 36$, et $p(\omega) = 1/36$ pour tout $\omega \in \Omega$.
L'événement "la somme des points vaut 8" est $A = \{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}$, et sa probabilité est $\mathbb{P}(A) = 5/36$.
2. On jette n fois une pièce de monnaie. Prenons alors $\Omega = \{(\omega_1, \dots, \omega_n) : \omega_i \in \{P, F\} \forall i\}$, qui est de cardinal 2^n .
L'événement A_k "on obtient k fois pile" est de cardinal

$$|A_k| = C_n^k \equiv \binom{n}{k} = \frac{n!}{k!(n-k)!}, \quad (1.1.9)$$

et on a donc $\mathbb{P}(A_k) = \binom{n}{k} 2^{-n}$.

3. Une urne contient r boules rouges et n boules noires. On tire une boule au hasard. On peut choisir $\Omega = \{1, \dots, r+n\}$ avec la distribution équiprobable. La probabilité de tirer une boule rouge est alors

$$\mathbb{P}(\{1, \dots, r\}) = \frac{r}{r+n}. \quad (1.1.10)$$

1.1.1 Probabilités conditionnelles

La notion de probabilité conditionnelle est une notion fondamentale. En effet, on a souvent accès à la probabilité qu'un certain événement A soit réalisé, sous la condition qu'un événement B ait eu lieu, ce qui revient à restreindre l'univers Ω à B .

Définition 1.1.10 (Probabilité conditionnelle). *Soit $B \subset \Omega$ un événement tel que $\mathbb{P}(B) > 0$. Pour tout $A \subset \Omega$, on appelle probabilité conditionnelle de A sachant B la quantité*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (1.1.11)$$

Remarquons que si l'on définit $q(\omega) = \mathbb{P}(\{\omega\}|B)$ pour tout $\omega \in B$, alors (B, q) est un espace probabilisé discret, puisque

$$\sum_{\omega \in B} q(\omega) = \sum_{\omega \in B} \frac{\mathbb{P}(\{\omega\} \cap B)}{\mathbb{P}(B)} = \frac{1}{\mathbb{P}(B)} \sum_{\omega \in B} \mathbb{P}(\{\omega\}) = 1. \quad (1.1.12)$$

Proposition 1.1.11. *Pour tous $A, B \subset \Omega$ avec $\mathbb{P}(B) > 0$,*

1. $\mathbb{P}(A|B) = \mathbb{P}(A \cap B|B)$.
2. $A \supset B$ implique $\mathbb{P}(A|B) = 1$.
3. $A \cap B = \emptyset$ implique $\mathbb{P}(A|B) = 0$.
4. Si $\{A_i\}_{i \in \mathbb{N}}$ est une famille d'événements deux à deux incompatibles, alors

$$\mathbb{P}\left(\bigcup_{i \in \mathbb{N}} A_i \mid B\right) = \sum_{i \in \mathbb{N}} \mathbb{P}(A_i|B). \quad (1.1.13)$$

5. On a $\mathbb{P}(A^c|B) = 1 - \mathbb{P}(A|B)$.

DÉMONSTRATION. En exercice. □

Exemple 1.1.12.

1. On lance deux dés. La probabilité qu'au moins l'un des dés indique 2, sachant que la somme des points vaut 6 est donnée par

$$\frac{\mathbb{P}(\{(2, 4), (4, 2)\})}{\mathbb{P}(\{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\})} = \frac{2}{5}. \quad (1.1.14)$$

2. On considère une ligne transmettant des signaux binaires. Le système est modélisé par l'univers

$$\Omega = \{(E, R) : E \in \{0, 1\}, R \in \{0, 1\}\}, \quad (1.1.15)$$

où E désigne le signal émis, et R le signal reçu. Dans une ligne parfaite, on aurait toujours $R = E$, mais nous allons supposer que la ligne altère le signal transmis avec une certaine probabilité.

Considérons les événements

- $E_i = \{(i, 0), (i, 1)\}$: le signal i est émis;
- $R_i = \{(0, i), (1, i)\}$: le signal i est reçu;
- $F = \{(0, 1), (1, 0)\}$: faute de transmission.

Supposons que l'on connaisse la probabilité qu'une faute de transmission ait lieu, en fonction du signal émis. C'est-à-dire que l'on connaît

- $f_0 = \mathbb{P}(F|E_0) = \mathbb{P}(\{0, 1\}|E_0) = \mathbb{P}(R_1|E_0)$,
- $f_1 = \mathbb{P}(F|E_1) = \mathbb{P}(\{1, 0\}|E_0) = \mathbb{P}(R_0|E_1)$.

Alors, la probabilité qu'une faute de transmission ait lieu est donnée par

$$\begin{aligned} \mathbb{P}(F) &= \mathbb{P}(F \cap E_0) + \mathbb{P}(F \cap E_1) \\ &= f_0\mathbb{P}(E_0) + f_1\mathbb{P}(E_1). \end{aligned} \quad (1.1.16)$$

Cette dernière relation est une application de la loi des probabilités totales, que nous allons énoncer de manière générale, ainsi que son corollaire, la *formule de Bayes*.

Théorème 1.1.13 (Loi de la probabilité totale et formule de Bayes). *Soient B_1, \dots, B_n des événements incompatibles deux à deux, tels que $\mathbb{P}(B_i) > 0$ pour tout i .*

1. Pour tout $A \subset \bigcup_j B_j$,

$$\mathbb{P}(A) = \sum_{j=1}^n \mathbb{P}(A|B_j)\mathbb{P}(B_j) \quad (1.1.17)$$

(loi de la probabilité totale).

2. Pour tout $A \subset \bigcup_j B_j$,

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{j=1}^n \mathbb{P}(A|B_j)\mathbb{P}(B_j)} \quad (1.1.18)$$

(formule de Bayes).

DÉMONSTRATION. La première relation se montre en écrivant

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_j (A \cap B_j)\right) = \sum_j \mathbb{P}(A \cap B_j) = \sum_j \mathbb{P}(A|B_j)\mathbb{P}(B_j). \quad (1.1.19)$$

La seconde s'obtient en notant que

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(B_i \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\mathbb{P}(A)}, \quad (1.1.20)$$

puis en appliquant (1.1.17). \square

Il est inutile d'apprendre la formule de Bayes par coeur, mieux vaut savoir la redériver! Cette formule permet d'"inverser une probabilité conditionnelle", un procédé qui est souvent source de confusion dans les applications.

Exemple 1.1.14. Revenons au problème de transmission de l'exemple précédent. Un problème important est de caractériser la fiabilité de la ligne. Par exemple, sachant que l'on a reçu le signal 1, quelle est la probabilité que ce soit effectivement le signal 1 qui a été émis? L'expression (1.1.18) nous donne

$$\mathbb{P}(E_1|R_1) = \frac{\mathbb{P}(R_1|E_1)\mathbb{P}(E_1)}{\sum_j \mathbb{P}(R_1|E_j)\mathbb{P}(E_j)} = \frac{(1 - f_1)\mathbb{P}(E_1)}{(1 - f_1)\mathbb{P}(E_1) + f_0\mathbb{P}(E_0)}. \quad (1.1.21)$$

Supposons par exemple que la ligne transmet les 0 avec une fiabilité de 95% et les 1 avec une fiabilité de 99%. On transmet un signal contenant un quart de 1. Avec $f_0 = 0.05$, $f_1 = 0.01$, $\mathbb{P}(E_0) = 0.75$ et $\mathbb{P}(E_1) = 0.25$, on obtient $\mathbb{P}(E_1|R_1) \simeq 0.87$, donc il y a tout de même une probabilité d'erreur de 13% pour la transmission des 1!

1.1.2 Indépendance

La notion de probabilité conditionnelle est intimement liée à la notion d'indépendance.

Définition 1.1.15 (Indépendance).

1. Deux événements A et B sont indépendants si

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \quad (1.1.22)$$

Dans ce cas, on a donc $\mathbb{P}(A|B) = \mathbb{P}(A)$ (pourvu que $\mathbb{P}(B) > 0$).

2. n événements A_1, \dots, A_n sont indépendants si

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \dots \mathbb{P}(A_{i_k}) \quad (1.1.23)$$

pour tout choix d'indices $\{i_1, \dots, i_k\} \subset \{1, \dots, n\}$.

Remarque 1.1.16. L'indépendance par paires n'implique pas l'indépendance. Un exemple (artificiel) est de prendre, pour $\Omega = \{1, 2, 3, 4\}$ et la distribution uniforme $p(\omega) = 1/4$ pour tout $\omega \in \Omega$, les événements $A = \{1, 2\}$, $B = \{2, 3\}$ et $C = \{1, 3\}$. Alors on a $\mathbb{P}(A \cap B) = 1/4 = \mathbb{P}(A)\mathbb{P}(B)$, donc A et B sont indépendants. De même, B et C sont indépendants, et A et C également. Néanmoins, $\mathbb{P}(A \cap B \cap C) = 0 \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$, et donc A , B et C ne sont pas indépendants.

Un exemple important d'événements indépendants apparaît lorsque l'on effectue plusieurs expériences, différentes ou non, qui ne s'influencent pas mutuellement. On peut alors construire un espace probabilisé unique pour l'ensemble des expériences, dont les événements liés à des expériences différentes seront naturellement indépendants.

Définition 1.1.17 (Espace produit). Soient $(\Omega_1, p_1), \dots, (\Omega_n, p_n)$ des espaces probabilisés discrets (identiques ou non). On appelle espace produit de ces espaces l'espace probabilisé discret (Ω, p) , où $\Omega = \Omega_1 \times \dots \times \Omega_n$ est le produit cartésien des Ω_i , et la distribution p est définie par

$$p(\omega = (\omega_1, \dots, \omega_n)) = p_1(\omega_1) \dots p_n(\omega_n). \quad (1.1.24)$$

Proposition 1.1.18. Soit (Ω, p) l'espace produit de $(\Omega_1, p_1), \dots, (\Omega_n, p_n)$. Pour chaque $i \in \{1, \dots, n\}$ on choisit un événement $A_i \subset \Omega_i$, et on définit

$$\widehat{A}_i = \{\omega \in \Omega : \omega_i \in A_i\} \subset \Omega. \quad (1.1.25)$$

Alors les événements $\widehat{A}_1, \dots, \widehat{A}_n$ sont indépendants.

DÉMONSTRATION. Commençons par le cas $n = 2$. Alors

$$\begin{aligned} \mathbb{P}(\widehat{A}_1 \cap \widehat{A}_2) &= \sum_{\omega \in \widehat{A}_1 \cap \widehat{A}_2} p(\omega) = \sum_{\omega_1 \in A_1} \sum_{\omega_2 \in A_2} p_1(\omega_1)p_2(\omega_2) \\ &= \sum_{\omega_1 \in A_1} p_1(\omega_1) \sum_{\omega_2 \in A_2} p_2(\omega_2) = \sum_{\omega \in \widehat{A}_1} p(\omega) \sum_{\omega \in \widehat{A}_2} p(\omega) \\ &= \mathbb{P}(\widehat{A}_1)\mathbb{P}(\widehat{A}_2). \end{aligned} \quad (1.1.26)$$

On procède ensuite par récurrence sur n . □

Exemple 1.1.19 (Expérience de Bernoulli). Considérons une expérience dont le résultat est soit un succès S , ayant lieu avec probabilité $q \in [0, 1]$, soit un échec E . Par exemple, dans le cas d'un lancer de dé pour lequel seule l'obtention d'un 6 est considéré comme succès, on aura $q = 1/6$. L'espace probabilisé associé est donc (Ω_1, p_1) avec $\Omega_1 = \{S, E\}$ et $p_1(S) = q$, $p_1(E) = 1 - q$. Pour n répétitions indépendantes de l'expérience, on pourra considérer l'espace produit (Ω, p) de n copies de (Ω_1, p_1) .

Soit B_k l'événement "la k -ème expérience est un succès". Alors les B_k sont indépendants, et pour tout choix de $k_1, \dots, k_m \in \{1, \dots, n\}$ distincts,

$$\mathbb{P}(B_{k_1} \cap \dots \cap B_{k_m}) = \mathbb{P}(B_{k_1}) \dots \mathbb{P}(B_{k_m}) = q^m. \quad (1.1.27)$$

Plus généralement, pour des indices $k_1, \dots, k_{m_1}, \ell_1, \dots, \ell_{m_2} \in \{1, \dots, n\}$, tous distincts, l'événement "les expériences k_1, \dots, k_{m_1} sont des succès, et les expériences $\ell_1, \dots, \ell_{m_2}$ sont des échecs" a la probabilité

$$\begin{aligned} \mathbb{P}(B_{k_1} \cap \dots \cap B_{k_{m_1}} \cap B_{\ell_1}^c \dots B_{\ell_{m_2}}^c) &= \mathbb{P}(B_{k_1}) \dots \mathbb{P}(B_{k_{m_1}}) \mathbb{P}(B_{\ell_1}^c) \dots \mathbb{P}(B_{\ell_{m_2}}^c) \\ &= q^{m_1} (1 - q)^{m_2}. \end{aligned} \quad (1.1.28)$$

Un autre événement important est l'événement S_k " k succès parmi n expériences". Comme il y a $\binom{n}{k} \equiv C_n^k$ manières de choisir les k expériences couronnées de succès, sa probabilité vaut

$$\mathbb{P}(S_k) = \binom{n}{k} q^k (1 - q)^{n-k} =: b(k; n, q). \quad (1.1.29)$$

1.2 Variables aléatoires

Définition 1.2.1 (Variable aléatoire). Soit (Ω, p) un espace probabilisé discret. Une variable aléatoire (discrète) est une application $X : \Omega \rightarrow \mathbb{R}$.

Notation 1.2.2.

- Nous noterons $X(\Omega) := \{X(\omega) : \omega \in \Omega\}$ l'image de X .
- Si $A \subset \mathbb{R}$, $X^{-1}(A) := \{\omega \in \Omega : X(\omega) \in A\}$ dénote l'événement " X appartient à A ", que nous abrègerons aussi $\{X \in A\}$. L'événement $X^{-1}(z)$ sera dénoté $\{X = z\}$.
- $\mathbb{P}(\{X \in A\})$ sera abrégé $\mathbb{P}\{X \in A\}$ ou $\mathbb{P}(X \in A)$, de même nous utiliserons les notations $\mathbb{P}\{X = z\}$, $\mathbb{P}\{X \leq z\}$, etc.
- $\mathbb{P}(\{X \in A\} \cap \{Y \in B\})$ sera abrégé $\mathbb{P}\{X \in A, Y \in B\}$.

Exemple 1.2.3.

1. Soit X la somme des points obtenus en jetant deux dés. Nous prendrons comme univers $\Omega = \{1, \dots, 6\} \times \{1, \dots, 6\}$, avec la distribution uniforme, et $X(\omega) = \omega_1 + \omega_2$. Nous aurons $X(\Omega) = \{2, \dots, 12\}$ et, par exemple

$$\begin{aligned} \mathbb{P}\{X = 4\} &= \mathbb{P}(\{(1, 3), (2, 2), (3, 1)\}) = \frac{3}{36} = \frac{1}{12}, \\ \mathbb{P}\{X \leq 3\} &= \mathbb{P}(\{(1, 1), (1, 2), (2, 1)\}) = \frac{3}{36} = \frac{1}{12}. \end{aligned} \quad (1.2.1)$$

2. Pour tout événement $A \subset \Omega$, on note 1_A la fonction indicatrice définie par

$$1_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon.} \end{cases} \quad (1.2.2)$$

C'est une variable aléatoire, d'image $1_A(\Omega) = \{0, 1\}$.

3. Soit X le nombre de succès dans une expérience de Bernoulli de longueur n , avec probabilité de succès q . On peut écrire X sous la forme

$$X(\omega) = \sum_{i=1}^n 1_S(\omega_i), \quad (1.2.3)$$

où $1_S(\omega_i)$ est la fonction indicatrice d'un succès lors de la i -ème expérience. On a par conséquent $X(\Omega) = \{0, 1, \dots, n\}$.

1.2.1 Loi d'une variable aléatoire

Définition 1.2.4 (Loi d'une variable aléatoire). *L'application*

$$\begin{aligned} f : X(\Omega) &\mapsto [0, 1] \\ z &\mapsto f(z) = \mathbb{P}\{X = z\} \equiv \mathbb{P}(X^{-1}\{z\}) \end{aligned} \quad (1.2.4)$$

est la loi (ou la distribution) de X .

Remarque 1.2.5. Le fait que

$$\Omega = \bigcup_{z \in X(\Omega)} X^{-1}(z) \equiv \bigcup_{z \in X(\Omega)} \{X = z\} \quad (1.2.5)$$

implique

$$\sum_{z \in X(\Omega)} f(z) = 1. \quad (1.2.6)$$

Par conséquent, $(X(\Omega), f)$ est un espace probabilisé discret.

Exemple 1.2.6.

1. Pour le jet de deux dés, la loi de la somme des points obtenus est donnée par

$$\begin{aligned} f(2) = f(12) &= \frac{1}{36}, & f(3) = f(11) &= \frac{2}{36}, & f(4) = f(10) &= \frac{3}{36}, \\ f(5) = f(9) &= \frac{4}{36}, & f(6) = f(8) &= \frac{5}{36}, & f(7) &= \frac{6}{36}. \end{aligned} \quad (1.2.7)$$

2. *Loi de Bernoulli*: C'est la loi d'une fonction indicatrice $X = 1_A$, c'est-à-dire telle que

$$\begin{aligned} \mathbb{P}\{X = 1\} &= q, \\ \mathbb{P}\{X = 0\} &= 1 - q. \end{aligned} \quad (1.2.8)$$

où $q = \mathbb{P}(A) \in [0, 1]$.

3. *Loi binomiale*: C'est la loi du nombre de succès dans l'expérience de Bernoulli d'ordre n et paramètre q :

$$\mathbb{P}\{X = k\} = \binom{n}{k} q^k (1 - q)^{n-k} =: b(k; n, q), \quad (1.2.9)$$

pour $k \in X(\Omega) = \{0, 1, \dots, n\}$.

4. *Loi de Poisson*: C'est par définition la loi donnée par

$$\mathbb{P}\{X = k\} = \pi_\lambda(k) := e^{-\lambda} \frac{\lambda^k}{k!} \quad \forall k \in X(\Omega) = \mathbb{N}, \quad (1.2.10)$$

où $\lambda > 0$ est un paramètre. Nous verrons qu'elle donne une bonne approximation de la loi binomiale lorsque $n \gg 1$ et $nq = \lambda$.

5. *Loi géométrique*: Soit X le nombre de fois que l'expérience de Bernoulli a été répétée jusqu'au premier succès (nous admettons que l'expérience se poursuit jusqu'au premier succès au moins). A strictement parler, nous n'avons pas défini jusqu'ici d'espace produit d'ordre infini. Nous pouvons contourner cette difficulté en travaillant avec l'univers $\Omega = \mathbb{N}^*$, où ω est le numéro de la première expérience couronnée de succès. Nous avons donc

$$p(\omega) = (1 - q)^{\omega-1}q, \quad (1.2.11)$$

puisque les $\omega - 1$ premières expériences sont des échecs, suivis d'un succès. On a bien

$$\sum_{\omega \in \Omega} p(\omega) = q \sum_{i=1}^{\infty} (1 - q)^{i-1} = q \frac{1}{1 - (1 - q)} = 1, \quad (1.2.12)$$

donc (Ω, p) est un espace probabilisé discret. La loi de X , donné par $X(\omega) = \omega$, est $\mathbb{P}\{X = n\} = q(1 - q)^{n-1}$ pour $n \in \mathbb{N}^*$ et s'appelle la loi géométrique.

La loi géométrique a une propriété remarquable: La probabilité d'un succès au $(n + 1)$ -ème essai, sachant que les n premiers essais ont échoué, ne dépend pas de n . Autrement dit, la probabilité de gagner dans une loterie au $(n + 1)$ -ème essai, sachant que l'on n'a jamais gagné auparavant, est la même que de gagner du premier coup. Plus généralement, la probabilité de gagner au $(n + k)$ -ème essai, sachant que l'on n'a pas gagné lors des n premiers essais, est la même que de gagner au k -ème essai.

Proposition 1.2.7. *Pour une variable aléatoire X de loi géométrique, et tout $k \geq 1$,*

$$\mathbb{P}\{X = n + k | X > n\} = \mathbb{P}\{X = k\} \quad \forall n. \quad (1.2.13)$$

DÉMONSTRATION. Par calcul direct. On a

$$\mathbb{P}\{X = n + k | X > n\} = \frac{\mathbb{P}\{X = n + k\}}{\mathbb{P}\{X > n\}} = \frac{q(1 - q)^{n-1+k}}{\sum_{m>n} q(1 - q)^{m-1}}, \quad (1.2.14)$$

et le dénominateur est égal à $(1 - q)^n$. \square

1.2.2 Espérance

Définition 1.2.8 (Espérance). *Soit X une variable aléatoire telle que*

$$\sum_{z \in X(\Omega)} |z| \mathbb{P}\{X = z\} < \infty. \quad (1.2.15)$$

Alors l'espérance de X est définie par

$$\mathbb{E}(X) = \sum_{z \in X(\Omega)} z \mathbb{P}\{X = z\} \equiv \sum_{\omega \in \Omega} X(\omega)p(\omega). \quad (1.2.16)$$

Remarquons que pour un univers infini, l'espérance existe si et seulement si la série de terme général $X(\omega)p(\omega)$ converge absolument, et alors l'espérance est égale à la somme de cette série.

On sera souvent amené à considérer la variable aléatoire $\varphi(X)$, où $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction donnée, par exemple $\varphi(x) = x^2$. Son espérance vaut alors

$$\mathbb{E}(\varphi(X)) = \sum_{z \in X(\Omega)} \varphi(z) \mathbb{P}\{X = z\} \equiv \sum_{\omega \in \Omega} \varphi(X(\omega))p(\omega). \quad (1.2.17)$$

Proposition 1.2.9 (Linéarité de l'espérance). *Soient X_1, \dots, X_n des variables aléatoires dont l'espérance existe, et soient a_1, \dots, a_n des nombres réels. Alors l'espérance de la combinaison linéaire $a_1X_1 + \dots + a_nX_n$ existe et vaut*

$$\mathbb{E}(a_1X_1 + \dots + a_nX_n) = a_1\mathbb{E}(X_1) + \dots + a_n\mathbb{E}(X_n). \quad (1.2.18)$$

DÉMONSTRATION. En exercice. \square

Exemple 1.2.10.

1. *Loi de Bernoulli*: L'espérance d'une fonction indicatrice est donnée par

$$\mathbb{E}(1_A) = 1 \cdot \mathbb{P}\{1_A = 1\} + 0 \cdot \mathbb{P}\{1_A = 0\} = \mathbb{P}\{1_A = 1\} = \mathbb{P}(A). \quad (1.2.19)$$

L'espérance d'une variable aléatoire suivant une loi de Bernoulli de paramètre q est donc égale à q .

2. *Loi binomiale*: L'espérance d'une variable aléatoire X de loi binomiale $b(k; n, q)$ est donnée par la somme

$$\mathbb{E}(X) = \sum_{k=0}^n kb(k; n, q) = \sum_{k=0}^n k \frac{n!}{k!(n-k)!} q^k (1-q)^{n-k}. \quad (1.2.20)$$

Cette somme peut en principe se calculer à l'aide de la formule du binôme. Il existe toutefois une manière bien plus simple de calculer l'espérance de X . Écrivons, comme dans l'exemple 1.2.3, $X = X_1 + \dots + X_n$, où $X_i = 1_S(\omega_i)$ est l'indicatrice d'un succès lors de la i -ème expérience. Comme $\mathbb{E}(X_i) = q$, il suit, par la Proposition 1.2.9, que

$$\mathbb{E}(X) = \sum_{i=1}^n \mathbb{E}(X_i) = nq. \quad (1.2.21)$$

3. *Loi de Poisson*: L'espérance d'une variable aléatoire X suivant une loi de Poisson de paramètre λ est donnée par la somme

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda. \quad (1.2.22)$$

4. *Loi géométrique*: L'espérance d'une variable aléatoire X de loi géométrique est donnée par la somme

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} kq(1-q)^{k-1}. \quad (1.2.23)$$

Pour calculer cette somme, introduisons la fonction

$$G(z) = \frac{z}{1-z} = \sum_{k=1}^{\infty} z^k. \quad (1.2.24)$$

La série converge absolument pour $|z| < 1$, et dans ce cas, nous avons le droit de d'intervertir somme et dérivée. Il suit

$$\frac{1}{(1-z)^2} = G'(z) = \sum_{k=1}^{\infty} kz^{k-1}. \quad (1.2.25)$$

L'espérance de X est donc donnée par

$$\mathbb{E}(X) = q \sum_{k=1}^{\infty} k(1-q)^{k-1} = qG'(1-q) = q \frac{1}{q^2} = \frac{1}{q}. \quad (1.2.26)$$

Par exemple, l'espérance du nombre de jets de dé nécessaires jusqu'à l'obtention du premier 6 est égale à 6.

5. Soit X une variable aléatoire telle que $\mathbb{P}\{X = k\} = 6/(k\pi)^2$ pour $k \in \mathbb{N}^*$ (le facteur $6/\pi^2$ a été choisi à cause du fait que la somme des $1/k^2$ vaut $\pi^2/6$). Alors

$$\sum_{k=1}^{\infty} |k| \mathbb{P}\{X = k\} = \frac{6}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{k} = \infty, \quad (1.2.27)$$

et nous dirons que l'espérance de X est infinie (ici la somme ne dépend pas de l'ordre des termes, car ceux-ci sont tous positifs).

1.2.3 Variance et covariance

Définition 1.2.11 (Variance). *Soit X une variable aléatoire dont l'espérance existe. On appelle variance de X la somme*

$$\text{Var}(X) = \mathbb{E}([X - \mathbb{E}(X)]^2) = \sum_{z \in X(\Omega)} [z - \mathbb{E}(X)]^2 \mathbb{P}\{X = z\}, \quad (1.2.28)$$

si la série converge. Si la série diverge, on dit que la variance de X est infinie. On appelle écart-type de X la quantité $\sigma(X) = \sqrt{\text{Var}(X)}$.

Proposition 1.2.12.

1. On a $\text{Var}(X) \geq 0$, et $\text{Var}(X) = 0$ si et seulement si $\mathbb{P}\{X = \mathbb{E}(X)\} = 1$.
2. On a $\text{Var}(X) < \infty \Leftrightarrow \mathbb{E}(X^2) < \infty$.
3. Si $\text{Var}(X) < \infty$, alors $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.
4. Pour $a, b \in \mathbb{R}$, on a $\text{Var}(a + bX) = b^2 \text{Var}(X)$.
5. Si $\text{Var}(X) < \infty$ et $\text{Var}(Y) < \infty$, alors $\text{Var}(X + Y) < \infty$.

DÉMONSTRATION. En exercice. □

Exemple 1.2.13.

1. *Loi de Bernoulli:* Soit X une variable aléatoire suivant une loi de Bernoulli. Alors $\mathbb{E}(X^2) = 1^2 \cdot q + 0^2 \cdot (1 - q) = q$, et donc

$$\text{Var}(X) = q - q^2 = q(1 - q). \quad (1.2.29)$$

2. *Loi de Poisson:* Pour une variable aléatoire X suivant une loi de Poisson, on obtient

$$\mathbb{E}(X(X - 1)) = \sum_{k=0}^{\infty} k(k - 1) e^{-\lambda} \frac{\lambda^k}{k!} = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k - 2)!} = \lambda^2, \quad (1.2.30)$$

d'où $\mathbb{E}(X^2) = \lambda^2 + \mathbb{E}(X) = \lambda^2 + \lambda$, et enfin

$$\text{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda. \quad (1.2.31)$$

3. *Loi géométrique*: Soit X une variable aléatoire suivant une loi géométrique. En dérivant la relation (1.2.25), on obtient, pour $|z| < 1$,

$$\sum_{k=1}^{\infty} k(k-1)z^{k-2} = G''(z) = \frac{2}{(1-z)^3}. \quad (1.2.32)$$

Ceci permet d'écrire, successivement,

$$\begin{aligned} \mathbb{E}(X(X-1)) &= \sum_{k=1}^{\infty} k(k-1)q(1-q)^{k-1} = q(1-q)\frac{2}{q^3} = 2\frac{1-q}{q^2}, \\ \mathbb{E}(X^2) &= \mathbb{E}(X(X-1)) + \mathbb{E}(X) = \frac{2-q}{q^2}, \\ \text{Var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{1-q}{q^2}. \end{aligned} \quad (1.2.33)$$

La variance d'une variable aléatoire de loi géométrique est donc égale à $(1-q)/q^2$.

En général, la variance de la somme $X + Y$ n'est pas égale à la somme des variances de X et de Y . En fait, nous avons

$$\begin{aligned} \text{Var}(X+Y) &= \mathbb{E}([X+Y - \mathbb{E}(X+Y)]^2) \\ &= \mathbb{E}([X - \mathbb{E}(X) + Y - \mathbb{E}(Y)]^2) \\ &= \mathbb{E}([X - \mathbb{E}(X)]^2) + \mathbb{E}([Y - \mathbb{E}(Y)]^2) + 2\mathbb{E}([X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]) \\ &= \text{Var}(X) + \text{Var}(Y) + 2\mathbb{E}([X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]). \end{aligned} \quad (1.2.34)$$

Ce calcul motive la définition suivante.

Définition 1.2.14 (Covariance). *On appelle covariance de deux variables aléatoires X et Y la quantité*

$$\begin{aligned} \text{cov}(X, Y) &= \mathbb{E}([X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]) \\ &= \mathbb{E}(XY - X\mathbb{E}(Y) - Y\mathbb{E}(X) + \mathbb{E}(X)\mathbb{E}(Y)) \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y). \end{aligned} \quad (1.2.35)$$

Par conséquent, on a

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{cov}(X, Y). \quad (1.2.36)$$

On dit que X et Y sont non corrélées si $\text{cov}(X, Y) = 0$. Dans ce cas, on a $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$.

Proposition 1.2.15.

1. On a $\text{cov}(X, Y) = \text{cov}(Y, X)$.
2. La covariance est une forme bilinéaire: Pour $a, b \in \mathbb{R}$,

$$\begin{aligned} \text{cov}(aX, bY) &= ab\text{cov}(X, Y), \\ \text{cov}(X_1 + X_2, Y) &= \text{cov}(X_1, Y) + \text{cov}(X_2, Y). \end{aligned} \quad (1.2.37)$$

3. Pour des variables aléatoires X_1, \dots, X_n , on a

$$\operatorname{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \operatorname{Var}(X_i) + \sum_{i \neq j} \operatorname{cov}(X_i, X_j). \quad (1.2.38)$$

DÉMONSTRATION. En exercice. \square

La variante suivante de l'inégalité de Cauchy–Schwarz permet de borner la covariance de deux variables aléatoires.

Proposition 1.2.16. *Si $\operatorname{Var}(X)$ et $\operatorname{Var}(Y)$ existent, alors $\operatorname{cov}(X, Y)$ existe et*

$$|\operatorname{cov}(X, Y)| \leq \sigma(X)\sigma(Y) = \sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}. \quad (1.2.39)$$

DÉMONSTRATION. L'existence suit du fait que $|\mathbb{E}(XY)| < \infty$ en vertu de l'inégalité $|X(\omega)Y(\omega)| \leq \frac{1}{2}(X(\omega)^2 + Y(\omega)^2)$. Considérons alors l'application

$$\begin{aligned} g: \quad \mathbb{R}^2 &\rightarrow \mathbb{R}_+ \\ (a, b) &\mapsto g(a, b) = \operatorname{Var}(aX + bY). \end{aligned} \quad (1.2.40)$$

Les résultats précédents impliquent que

$$\begin{aligned} g(a, b) &= a^2 \operatorname{Var}(X) + 2ab \operatorname{cov}(X, Y) + b^2 \operatorname{Var}(Y) \\ &= (a \quad b) \begin{pmatrix} \operatorname{Var}(X) & \operatorname{cov}(X, Y) \\ \operatorname{cov}(X, Y) & \operatorname{Var}(Y) \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}. \end{aligned} \quad (1.2.41)$$

Comme $g(a, b) \geq 0$ pour tout choix de a, b , la matrice dans cette dernière expression est semi-définie positive. Son déterminant est donc non-négatif, or celui-ci est précisément égal à $\operatorname{Var}(X)\operatorname{Var}(Y) - \operatorname{cov}(X, Y)^2$. Ainsi, $|\operatorname{cov}(X, Y)| \leq \sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}$. \square

Remarque 1.2.17. En statistique, on introduit le *coefficient de corrélation*

$$\rho_{X,Y} = \frac{\operatorname{cov}(X, Y)}{\sigma(X)\sigma(Y)} \in [-1, 1]. \quad (1.2.42)$$

La preuve de la proposition montre que $|\rho_{X,Y}| = 1$ si et seulement s'il existe a, b tels que $\operatorname{Var}(aX + bY) = 0$, donc si et seulement si $\mathbb{P}\{aX + bY = c\} = 1$ pour un certain c . C'est-à-dire que les variables X et Y sont en fait linéairement dépendantes.

Nous établissons maintenant un lien entre la non-corrélation et l'indépendance.

Définition 1.2.18 (Indépendance de variables aléatoires). *Les variables aléatoires X_1, \dots, X_n sont dites indépendantes si*

$$\mathbb{P}\{X_1 = z_1, \dots, X_n = z_n\} = \mathbb{P}\{X_1 = z_1\} \dots \mathbb{P}\{X_n = z_n\} \quad (1.2.43)$$

pour tout choix de $z_1 \in X_1(\Omega), \dots, z_n \in X_n(\Omega)$.

Cette définition est compatible avec la Définition 1.1.15 de l'indépendance d'événements. En effet, on a le résultat suivant.

Proposition 1.2.19. *Les propriétés suivantes sont équivalentes:*

1. X_1, \dots, X_n sont indépendantes.

2. Pour tout choix de $A_1, \dots, A_n \subset \mathbb{R}$,

$$\mathbb{P}\{X_1 \in A_1, \dots, X_n \in A_n\} = \mathbb{P}\{X_1 \in A_1\} \dots \mathbb{P}\{X_n \in A_n\}. \quad (1.2.44)$$

3. Pour tout choix de $A_1, \dots, A_n \subset \mathbb{R}$, les événements $\{X_1 \in A_1\}, \dots, \{X_n \in A_n\}$ sont indépendants.

4. Pour tout choix de $z_1 \in X_1(\Omega), \dots, z_n \in X_n(\Omega)$, les événements $\{X_1 = z_1\}, \dots, \{X_n = z_n\}$ sont indépendants.

DÉMONSTRATION. En exercice. □

Et voici le lien annoncé entre indépendance et absence de corrélation.

Proposition 1.2.20. *Deux variables aléatoire indépendantes, dont l'espérance existe, sont non corrélées.*

DÉMONSTRATION. Supposons d'abord que la covariance existe. Alors

$$\begin{aligned} \mathbb{E}(XY) &= \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} xy \mathbb{P}\{X = x, Y = y\} \\ &= \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} xy \mathbb{P}\{X = x\} \mathbb{P}\{Y = y\} = \mathbb{E}(X)\mathbb{E}(Y), \end{aligned} \quad (1.2.45)$$

donc $\text{cov}(X, Y) = 0$. Un calcul similaire montre que $\mathbb{E}(|XY|) = \mathbb{E}(|X|)\mathbb{E}(|Y|)$, ce qui justifie l'existence de la covariance. □

Il suit de ce résultat que si X_1, \dots, X_n sont indépendantes, alors la variance de leur somme est égale à la somme de leurs variances.

Exemple 1.2.21.

1. *Loi binomiale:* On a vu qu'une variable aléatoire X de loi binomiale pouvait s'écrire $X = X_1 + \dots + X_n$, où $X_i = 1_S(\omega_i)$ prend la valeur 1 avec probabilité q , et 0 avec probabilité $1 - q$. Ainsi

$$\text{Var}(X_i) = 0^2 \cdot (1 - q) + 1^2 \cdot q - \mathbb{E}(X_i)^2 = q(1 - q). \quad (1.2.46)$$

Comme les X_i sont indépendants, il suit $\text{Var}(X) = nq(1 - q)$.

2. Soit $\Omega = \{-1, 0, 1\}$ avec $p(\omega) = 1/3$ pour chaque $\omega \in \Omega$. Considérons les variables aléatoires $X(\omega) = \omega$ et $Y = |X|$. Alors $\mathbb{E}(X) = 0$, $\mathbb{E}(Y) = 2/3$ et $\mathbb{E}(XY) = 0$. Par conséquent, X et Y sont non corrélées. Cependant elles ne sont pas indépendantes: Par exemple les événements $\{X = 1\}$ et $\{Y = 0\}$ ne sont pas indépendants.

1.2.4 Fonction génératrice

Nous terminons cette première partie en introduisant la notion de fonction génératrice, qui est un outil permettant de simplifier le calcul d'espérances.

Définition 1.2.22 (Fonction génératrice). *Soit X une variable aléatoire à valeurs dans \mathbb{N} . On appelle fonction génératrice de X la fonction $G_X : \mathbb{C} \rightarrow \mathbb{C}$ donnée par la série entière*

$$G_X(z) = \sum_{k \geq 0} z^k \mathbb{P}\{X = k\} \equiv \mathbb{E}(z^X). \quad (1.2.47)$$

On remarque que

$$G_X(1) = \sum_{k \geq 0} \mathbb{P}\{X = k\} = 1, \quad (1.2.48)$$

donc le rayon de convergence R de la série entière est supérieur ou égal à 1. Si $R > 1$, on peut échanger somme et dérivée, ce qui donne

$$\begin{aligned} G'_X(z) &= \sum_{k \geq 0} k z^{k-1} \mathbb{P}\{X = k\} & \Rightarrow & \quad G'_X(1) = \sum_{k \geq 0} k \mathbb{P}\{X = k\} = \mathbb{E}(X), \\ G''_X(z) &= \sum_{k \geq 0} k(k-1) z^{k-2} \mathbb{P}\{X = k\} & \Rightarrow & \quad G''_X(1) = \sum_{k \geq 0} k(k-1) \mathbb{P}\{X = k\} \\ & & & = \mathbb{E}(X(X-1)). \end{aligned} \quad (1.2.49)$$

Nous avons donc le résultat suivant.

Proposition 1.2.23. *Soit X une variable aléatoire à valeurs dans \mathbb{N} dont la fonction génératrice $G_X(z)$ admet un rayon de convergence strictement supérieur à 1. Alors on a*

$$\begin{aligned} \mathbb{E}(X) &= G'_X(1), \\ \text{Var}(X) &= G''_X(1) + G'_X(1) - G'_X(1)^2. \end{aligned} \quad (1.2.50)$$

Exemple 1.2.24.

1. *Loi de Bernoulli:* La fonction génératrice se réduit à deux termes:

$$G_X(z) = z^0 \cdot (1-q) + z^1 \cdot q = qz + 1 - q. \quad (1.2.51)$$

2. *Loi géométrique:* $\mathbb{P}\{X = k\} = q(1-q)^{k-1}$ pour $k \geq 1$. La fonction génératrice est une série géométrique:

$$G_X(z) = \sum_{k \geq 1} z^k q (1-q)^{k-1} = \frac{q}{1-q} \sum_{k \geq 1} [z(1-q)]^k = \frac{qz}{1 - (1-q)z}. \quad (1.2.52)$$

Le rayon de convergence est $R = 1/(1-q)$.

3. *Loi binomiale:* $\mathbb{P}\{X = k\} = \binom{n}{k} q^k (1-q)^{n-k}$ pour $k = 0, \dots, n$. La fonction génératrice se calcule à l'aide de la formule du binôme:

$$G_X(z) = \sum_{k=0}^n \binom{n}{k} (qz)^k (1-q)^{n-k} = (qz + 1 - q)^n. \quad (1.2.53)$$

C'est un polynôme de degré n , donc le rayon de convergence est bien sûr infini.

4. *Loi de Poisson:* $\mathbb{P}\{X = k\} = e^{-\lambda} \lambda^k / k!$ pour $k \geq 0$. La fonction génératrice est une exponentielle:

$$G_X(z) = e^{-\lambda} \sum_{k \geq 0} \frac{(\lambda z)^k}{k!} = e^{\lambda(z-1)}. \quad (1.2.54)$$

Le rayon de convergence est infini.

On vérifiera que les relations (1.2.50) donnent bien les espérances et les variances calculées précédemment. Nous les résumons dans le tableau suivant.

Loi	$\mathbb{P}\{X = k\}$	$G_X(z)$	$\mathbb{E}(X)$	$\text{Var}(X)$
Bernoulli	$\begin{cases} \mathbb{P}\{X = 0\} = 1 - q \\ \mathbb{P}\{X = 1\} = q \end{cases}$	$qz + 1 - q$	q	$q(1 - q)$
Binomiale $b(k; n, q)$	$\binom{n}{k} q^k (1 - q)^{n-k}$	$(qz + 1 - q)^n$	nq	$nq(1 - q)$
Géométrique	$q(1 - q)^{k-1}$	$\frac{qz}{1 - (1 - q)z}$	$\frac{1}{q}$	$\frac{1 - q}{q^2}$
Poisson $\pi_\lambda(k)$	$e^{-\lambda} \frac{\lambda^k}{k!}$	$e^{\lambda(z-1)}$	λ	λ

Finalement, le résultat suivant permet d'obtenir des résultats sur la loi de la somme de certaines variables aléatoires indépendantes.

Proposition 1.2.25. *Si X et Y sont indépendantes, à valeurs dans \mathbb{N} , avec fonctions génératrices G_X et G_Y , alors $G_{X+Y} = G_X G_Y$.*

DÉMONSTRATION. On a

$$\begin{aligned}
 G_{X+Y}(z) &= \sum_{k \geq 0} z^k \mathbb{P}\{X + Y = k\} \\
 &= \sum_{k \geq 0} \sum_{\ell=0}^k z^k \mathbb{P}\{X = \ell, Y = k - \ell\} \\
 &= \sum_{k \geq 0} \sum_{\ell=0}^k z^k \mathbb{P}\{X = \ell\} \mathbb{P}\{Y = k - \ell\} \\
 &= \sum_{\ell \geq 0} \sum_{n \geq 0} z^{\ell+n} \mathbb{P}\{X = \ell\} \mathbb{P}\{Y = n\} \\
 &= G_X(z) G_Y(z). \tag{1.2.55}
 \end{aligned}$$

□

On remarquera en particulier que la somme de deux variables aléatoires de loi binomiale suit encore une loi binomiale, et que la somme de deux variables aléatoires de loi de Poisson suit encore une loi de Poisson:

Corollaire 1.2.26. *Si X et Y sont indépendantes et suivent des lois de Poisson, de paramètres respectifs λ et μ , alors $X + Y$ suit une loi de Poisson de paramètre $\lambda + \mu$.*

DÉMONSTRATION. Il suffit de remarquer que $G_{X+Y}(z) = G_X(z) G_Y(z) = e^{(\lambda+\mu)(z-1)}$. □

Chapitre 2

Probabilités continues

Dans ce chapitre, nous abordons l'étude d'espaces probabilisés non dénombrables, en particulier le cas de variables aléatoires continues. Dans un premier temps, nous allons considérer la situation de variables aléatoires admettant une densité, qui est plus simple à traiter que le cas général. Dans tout ce chapitre, les intégrales sont définies au sens de Lebesgue, mais dans le cas de variables aléatoires à densité, l'intégrale de Lebesgue sera synonyme de l'intégrale de Riemann. Nous introduirons l'intégrale de Lebesgue dans la Section 2.3.

2.1 Variables aléatoires réelles à densité

2.1.1 Densité et fonction de répartition

Exemple 2.1.1. On laisse tomber une aiguille à tricoter. La probabilité qu'elle pointe exactement vers le nord est nulle. Par contre, la probabilité qu'elle pointe dans une direction α , comprise entre le nord et l'est, est de $1/4$. Plus généralement, si $\varphi_1 < \varphi_2 < \varphi_1 + 2\pi$, on aura

$$\mathbb{P}\{\varphi_1 \leq \alpha \leq \varphi_2\} = \frac{\varphi_2 - \varphi_1}{2\pi} = \int_{\varphi_1}^{\varphi_2} \frac{1}{2\pi} d\varphi. \quad (2.1.1)$$

On dit que α suit la *loi uniforme* sur $[0, 2\pi]$.

D'une manière générale, si la probabilité qu'une variable aléatoire X appartienne à un intervalle peut s'écrire comme l'intégrale d'une fonction f sur cet intervalle, on dira que cette variable aléatoire admet la densité f .

Définition 2.1.2 (Densité). Une fonction $f : \mathbb{R} \rightarrow [0, \infty[$, intégrable selon Lebesgue, s'appelle une densité si

$$\int_{-\infty}^{\infty} f(x) dx = 1. \quad (2.1.2)$$

Exemple 2.1.3.

1. Densité de la *loi uniforme* sur $[a, b]$:

$$\frac{1}{b-a} 1_{[a,b]}(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b], \\ 0 & \text{sinon.} \end{cases} \quad (2.1.3)$$

2. Densité de la *loi normale standard* (ou loi de Gauss, Figure 2.1a):

$$\varphi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} . \quad (2.1.4)$$

Nous vérifierons un peu plus loin (cf. Exemple 2.2.13) que $\varphi(x)$ est effectivement une densité de probabilité, c'est-à-dire qu'on a

$$\int_{-\infty}^{\infty} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = 1 . \quad (2.1.5)$$

3. Densité de la *loi normale de moyenne μ , variance σ^2* :

$$\varphi(x; \mu, \sigma^2) = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi} \sigma} . \quad (2.1.6)$$

Le changement de variable $y = (x - \mu)/\sigma$ donne

$$\int_{-\infty}^{\infty} \varphi(x; \mu, \sigma^2) dx = \int_{-\infty}^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy = 1 . \quad (2.1.7)$$

4. Densité de la *loi exponentielle de paramètre $\lambda > 0$* (Figure 2.1b):

$$f(x) = 1_{\{x \geq 0\}} \lambda e^{-\lambda x} = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 , \\ 0 & \text{si } x < 0 . \end{cases} \quad (2.1.8)$$

5. Densité de la *loi de Cauchy de paramètre $c > 0$* (Figure 2.1c):

$$f(x) = \frac{c}{\pi} \frac{1}{x^2 + c^2} . \quad (2.1.9)$$

Définition 2.1.4 (Fonction de répartition). *Une fonction $F : \mathbb{R} \rightarrow [0, 1]$ est une fonction de répartition si*

- F est croissante: $x \leq y \Rightarrow F(x) \leq F(y)$.
- F est continue à droite: $\lim_{y \rightarrow x+} F(y) = F(x) \forall x$.
- $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$.

Une fonction de répartition F est dite absolument continue de densité f si

$$F(x) = \int_{-\infty}^x f(y) dy . \quad (2.1.10)$$

On a les propriétés suivantes:

1. Si f est une densité, la fonction F définie par (2.1.10) est une fonction de répartition, qui de plus est continue.
2. Si F est une fonction de répartition continue, elle n'admet pas nécessairement une densité (on peut construire des contre-exemples, faisant intervenir, par exemple, des objets fractals).

Le lien entre la notion de fonction de répartition et les variables aléatoires vient du fait que pour toute variable aléatoire réelle, $\mathbb{P}\{X \leq t\}$ est une fonction de répartition. En effet,

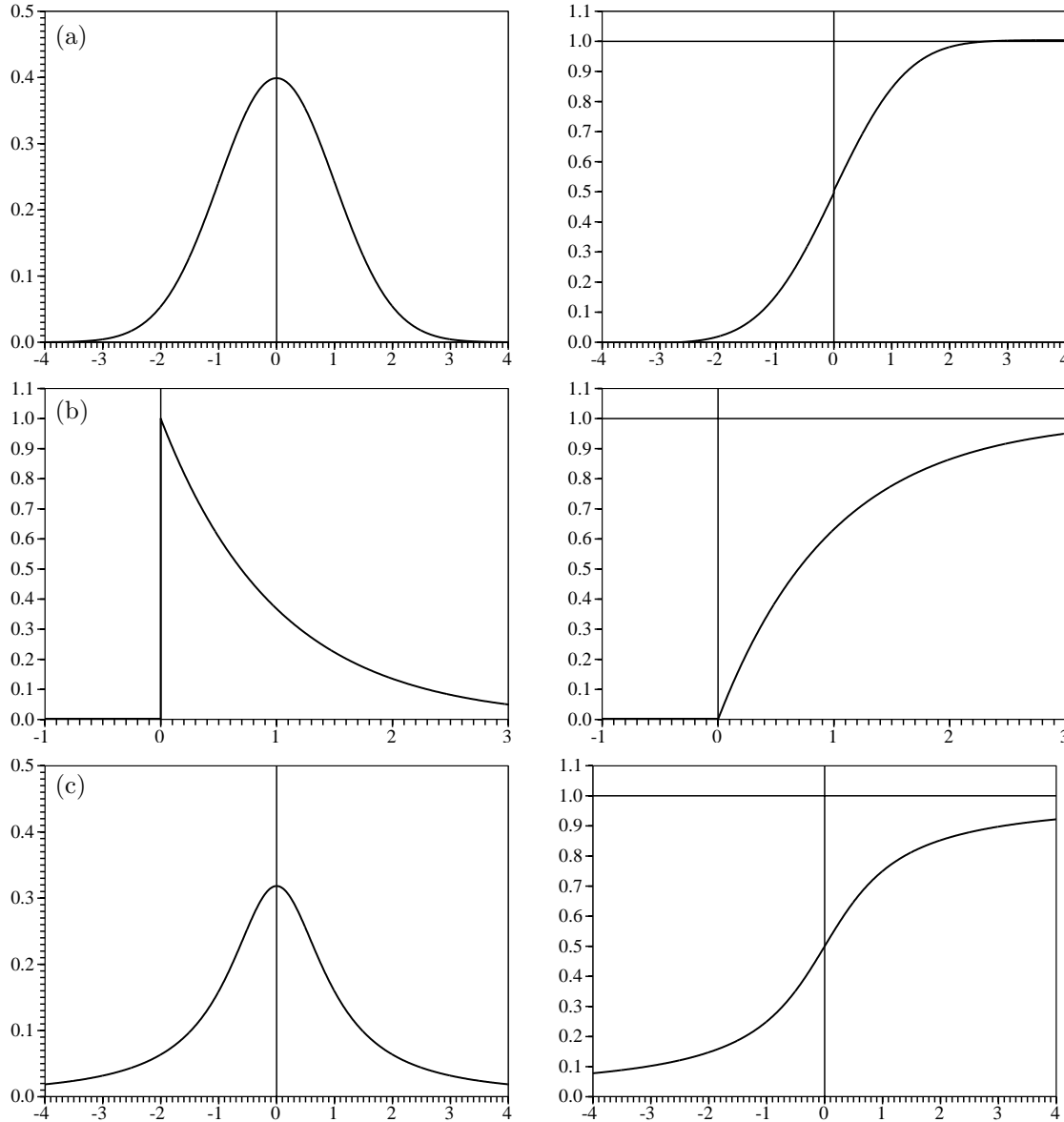


FIGURE 2.1. Densités et fonctions de répartition correspondantes: (a) loi normale standard $\mathcal{N}(0,1)$, (b) loi exponentielle $\text{Exp}(1)$, (c) loi de Cauchy de paramètre 1.

- si $s \leq t$, alors $\{X \leq s\} \subset \{X \leq t\}$, et donc $\mathbb{P}\{X \leq s\} \leq \mathbb{P}\{X \leq t\}$;
- $\lim_{s \rightarrow t+} \mathbb{P}\{X \leq s\} - \mathbb{P}\{X \leq t\} = \lim_{s \rightarrow t+} \mathbb{P}\{t < X \leq s\} = 0$;
- $\lim_{t \rightarrow -\infty} \mathbb{P}\{X \leq t\} = 0$ et $\lim_{t \rightarrow +\infty} \mathbb{P}\{X \leq t\} = 1$.

Ceci motive la définition suivante.

Définition 2.1.5 (Variable aléatoire à densité). *Si X est une variable aléatoire,*

$$F_X(t) = \mathbb{P}\{X \leq t\} \quad (2.1.11)$$

est appelée fonction de répartition de X . Si F_X est absolument continue de densité f , on

dit que X admet la densité f et on a les relations

$$\mathbb{P}\{X \leq t\} = \int_{-\infty}^t f(s) \, ds, \quad (2.1.12)$$

$$\mathbb{P}\{a < X \leq b\} = \mathbb{P}\{X \leq b\} - \mathbb{P}\{X \leq a\} = \int_a^b f(s) \, ds. \quad (2.1.13)$$

Dans ce cas, on peut remplacer $<$ par \leq et inversement.

Exemple 2.1.6.

1. Soit (Ω, p) un espace probabilisé discret. La fonction de répartition d'une variable aléatoire $X : \Omega \rightarrow X(\Omega) \subset \mathbb{R}$ est

$$F_X(t) = \sum_{z \in X(\Omega) : z \leq t} \mathbb{P}\{X = z\} = \sum_{\omega \in \Omega : X(\omega) \leq t} p(\omega) = \sum_{\omega \in \Omega} 1_{[X(\omega), \infty[}(t) p(\omega). \quad (2.1.14)$$

$F_X(t)$ est donc constante partout sauf aux points $z \in X(\Omega)$, où elle effectue un saut de $\mathbb{P}\{X = z\}$. La variable aléatoire X n'admet pas de densité.

2. La fonction de répartition associée à la densité de la loi uniforme sur $[a, b]$ est donnée par

$$F(t) = \int_{-\infty}^t 1_{[a,b]}(s) \, ds = \begin{cases} 0 & \text{si } t \leq a, \\ \frac{t-a}{b-a} & \text{si } a < t \leq b, \\ 1 & \text{si } t > b. \end{cases} \quad (2.1.15)$$

Si X admet cette fonction de répartition, on notera $X \sim \mathcal{U}([a, b])$.

3. On notera $\Phi(t)$ l'intégrale

$$\Phi(t) = \int_{-\infty}^t \varphi(s) \, ds, \quad (2.1.16)$$

où $\varphi(s) = e^{-s^2/2} / \sqrt{2\pi}$ est la densité de la loi normale standard. Comme φ est une densité, Φ est une fonction de répartition. Si X admet la fonction de répartition Φ , on dira qu'elle *suit la loi normale standard*, et on notera $X \sim \mathcal{N}(0, 1)$. Nous dirons que X suit la *loi normale de moyenne μ et variance σ^2* et nous noterons $X \sim \mathcal{N}(\mu, \sigma^2)$ si elle admet la densité $\varphi(x; \mu, \sigma^2)$ introduite en (2.1.6).

4. Supposons que X soit une variable aléatoire satisfaisant

$$\mathbb{P}\{X > t\} = \begin{cases} 1 & \text{si } t < 0, \\ e^{-\lambda t} & \text{si } t \geq 0. \end{cases} \quad (2.1.17)$$

Alors la fonction de répartition de X est donnée par

$$F_X(t) = 1 - \mathbb{P}\{X > t\} = \begin{cases} 0 & \text{si } t < 0, \\ 1 - e^{-\lambda t} & \text{si } t \geq 0. \end{cases} \quad (2.1.18)$$

On s'aperçoit que

$$F'_X(t) = \begin{cases} 0 & \text{si } t < 0, \\ \lambda e^{-\lambda t} & \text{si } t > 0 \end{cases} \quad (2.1.19)$$

est la densité de la loi exponentielle (sauf en $t = 0$, où la fonction F_X n'est pas dérivable). La discontinuité en 0 de cette densité n'ayant pas d'influence sur l'intégrale,

X est absolument continue de densité exponentielle, donnée par (2.1.8). On dira que X suit la loi exponentielle, et on notera $X \sim \text{Exp}(\lambda)$.

Remarquons que si $t > s > 0$, on a

$$\mathbb{P}\{X > t | X > s\} = \frac{\mathbb{P}\{X > t\}}{\mathbb{P}\{X > s\}} = e^{-\lambda(t-s)} = \mathbb{P}\{X > t - s\}. \quad (2.1.20)$$

La loi exponentielle est donc un analogue continu de la loi géométrique, cf. la Proposition 1.2.7.

2.1.2 Espérance et variance

Pour des variables aléatoires admettant une densité, l'espérance et la variance se définissent de manière analogue au cas discret, en remplaçant les sommes par des intégrales.

Définition 2.1.7 (Espérance et variance). *Soit X une variable aléatoire admettant la densité f .*

1. Si $x \mapsto |x|f(x)$ est intégrable selon Lebesgue, on appelle espérance de X la grandeur

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x) dx. \quad (2.1.21)$$

2. Si de plus $x \mapsto (x - \mathbb{E}(X))^2 f(x)$ est intégrable selon Lebesgue, on appelle variance de X la grandeur

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - \mathbb{E}(X))^2 f(x) dx = \mathbb{E}(X^2) - \mathbb{E}(X)^2. \quad (2.1.22)$$

Exemple 2.1.8.

1. *Loi uniforme*: Si $X \sim \mathcal{U}([a, b])$, alors

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \frac{1}{b-a} 1_{[a,b]}(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2} \quad (2.1.23)$$

et comme on a

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 \frac{1}{b-a} 1_{[a,b]}(x) dx = \int_a^b \frac{x^2}{b-a} dx = \frac{a^2 + ab + b^2}{3}, \quad (2.1.24)$$

la variance est donnée par

$$\text{Var}(X) = \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(a-b)^2}{12}. \quad (2.1.25)$$

2. *Loi normale standard*: On a

$$\lim_{L \rightarrow \infty} \int_{-L}^L |x| \varphi(x) dx = \lim_{L \rightarrow \infty} 2 \int_0^L x \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \lim_{L \rightarrow \infty} \frac{2}{\sqrt{2\pi}} \left[-e^{-x^2/2}\right]_0^L = \frac{2}{\sqrt{2\pi}}, \quad (2.1.26)$$

donc $x \mapsto |x|\varphi(x)$ est intégrable. Par conséquent, si $X \sim \mathcal{N}(0, 1)$, alors $\mathbb{E}(X)$ existe. En fait, $\mathbb{E}(X) = 0$ puisque $x \mapsto x\varphi(x)$ est impaire.

Quant à la variance, une intégration par parties donne

$$\text{Var}(X) = \int_{-\infty}^{\infty} x^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \left[-x \frac{e^{-x^2/2}}{\sqrt{2\pi}}\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = 1. \quad (2.1.27)$$

Ce résultat explique la notation $X \sim \mathcal{N}(0, 1)$: les arguments 0 et 1 se réfèrent à l'espérance et la variance de X .

3. *Loi normale de moyenne μ et variance σ^2* : Le changement de variable $y = (x - \mu)/\sigma$ donne

$$\begin{aligned}\mathbb{E}(X) &= \int_{-\infty}^{\infty} x \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma y + \mu) e^{-y^2/2} dy = \mu, \\ \text{Var}(X) &= \int_{-\infty}^{\infty} (x - \mu)^2 \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma} dx = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-y^2/2} dy = \sigma^2.\end{aligned}\quad (2.1.28)$$

4. *Loi exponentielle de paramètre λ* :

$$\begin{aligned}\mathbb{E}(X) &= \int_0^{\infty} x \lambda e^{-\lambda x} dx = \left[-x e^{-\lambda x}\right]_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx = \frac{1}{\lambda}, \\ \text{Var}(X) &= \int_0^{\infty} \left(x - \frac{1}{\lambda}\right)^2 \lambda e^{-\lambda x} dx = \left[-\left(x - \frac{1}{\lambda}\right)^2 e^{-\lambda x}\right]_0^{\infty} + \int_0^{\infty} 2\left(x - \frac{1}{\lambda}\right) e^{-\lambda x} dx \\ &= \frac{1}{\lambda^2} + \frac{2}{\lambda} \mathbb{E}(X) - \frac{2}{\lambda} \int_0^{\infty} e^{-\lambda x} dx = \frac{1}{\lambda^2}.\end{aligned}\quad (2.1.29)$$

5. *Loi de Cauchy de paramètre c* : Comme

$$\lim_{L \rightarrow \infty} \int_{-L}^L |x| \frac{c}{\pi x^2 + c^2} dx = \frac{2c}{\pi} \lim_{L \rightarrow \infty} \int_0^L \frac{x}{x^2 + c^2} dx = \frac{c}{\pi} \lim_{L \rightarrow \infty} \left[\log(x^2 + c^2)\right]_0^L = +\infty,\quad (2.1.30)$$

l'espérance de la loi de Cauchy n'existe pas (bien que sa densité soit paire, et qu'on pourrait être tenté de considérer son espérance comme nulle).

2.1.3 Fonction caractéristique

Définition 2.1.9 (Fonction caractéristique). *Soit X une variable aléatoire admettant la densité f . On appelle fonction caractéristique de X la fonction*

$$\phi_X(z) = \int_{-\infty}^{\infty} e^{izx} f(x) dx \equiv \mathbb{E}(e^{izX}).\quad (2.1.31)$$

On remarquera que $\phi_X(z)$ n'est autre que la transformée de Fourier de la densité de X (à un facteur près, selon la convention adoptée pour la transformée de Fourier). On a $\phi_X(0) = 1$, et pour z réel,

$$|\phi_X(z)| \leq \int_{-\infty}^{\infty} |e^{izx}| f(x) dx = 1,\quad (2.1.32)$$

donc l'intégrale existe sur l'axe réel. Si ϕ_X est analytique au voisinage de $z = 0$, on aura

$$\phi'_X(z) = i \int_{-\infty}^{\infty} x e^{izx} f(x) dx,\quad (2.1.33)$$

et par conséquent

$$\phi'_X(0) = i\mathbb{E}(X), \quad \phi''_X(0) = -\mathbb{E}(X^2).\quad (2.1.34)$$

La fonction caractéristique permet donc de calculer l'espérance et la variance de X (et plus généralement l'espérance de n'importe quel polynôme en X).

Exemple 2.1.10.

1. *Loi exponentielle de paramètre λ* : Si $X \sim \text{Exp}(\lambda)$, alors

$$\phi_X(z) = \lambda \int_0^\infty e^{izx - \lambda x} dx = \left[\frac{\lambda}{iz - \lambda} e^{(iz - \lambda)x} \right]_0^\infty = \frac{\lambda}{\lambda - iz}. \quad (2.1.35)$$

On remarque que pour $|z| < \lambda$, l'on peut développer $\phi_X(z)$ en série entière

$$\phi_X(z) = \frac{1}{1 - iz/\lambda} = \sum_{k=0}^{\infty} \left(\frac{iz}{\lambda} \right)^k, \quad (2.1.36)$$

et comme par ailleurs

$$\phi_X(z) = \mathbb{E}(e^{izX}) = \sum_{k=0}^{\infty} \frac{(iz)^k}{k!} \mathbb{E}(X^k), \quad (2.1.37)$$

on obtient par identification

$$\mathbb{E}(X^k) = \frac{k!}{\lambda^k}. \quad (2.1.38)$$

2. *Loi normale standard*: Comme $x^2 - 2izx = (x - iz)^2 + z^2$, le théorème de Cauchy permet d'écrire

$$\phi_X(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{izx - x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-iz)^2/2} dx e^{-z^2/2} = e^{-z^2/2}. \quad (2.1.39)$$

En développant l'exponentielle, on obtient

$$\phi_X(z) = \sum_{\ell=0}^{\infty} \frac{(-1)^\ell z^{2\ell}}{\ell! 2^\ell}, \quad (2.1.40)$$

ce qui donne

$$\mathbb{E}(X^k) = \begin{cases} 0 & \text{si } k \text{ est impair,} \\ \frac{(2\ell)!}{2^\ell \ell!} & \text{si } k = 2\ell \text{ est pair.} \end{cases} \quad (2.1.41)$$

En séparant les termes pairs et impairs de $(2\ell)!$, la dernière expression peut être réécrite sous la forme

$$\begin{aligned} \frac{(2\ell)!}{2^\ell \ell!} &= \frac{2\ell(2\ell-2)\dots 4 \cdot 2}{2^\ell \ell!} (2\ell-1)(2\ell-3)\dots 5 \cdot 3 \cdot 1 \\ &= (2\ell-1)(2\ell-3)\dots 5 \cdot 3 \cdot 1, \end{aligned} \quad (2.1.42)$$

cette dernière quantité étant souvent dénotée $(2\ell-1)!!$.

3. *Loi de Cauchy*: En écrivant

$$\frac{c}{\pi} \frac{1}{x^2 + c^2} = \frac{1}{2\pi i} \left[\frac{1}{x - ic} - \frac{1}{x + ic} \right], \quad (2.1.43)$$

la fonction caractéristique peut être calculée par la méthode des résidus, et donne

$$\phi_X(z) = e^{-c|z|} \quad (2.1.44)$$

pour z réel. Cette fonction n'est pas dérivable en zéro, reflétant le fait que l'espérance n'existe pas.

2.1.4 Changement de variable

Il est souvent utile de connaître la densité d'une variable aléatoire Y , fonction de la variable aléatoire X , en fonction de la densité de X .

Proposition 2.1.11 (Formule de changement de variable). *Soit X une variable aléatoire admettant la densité f . Soit $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continûment différentiable et strictement monotone (donc g' ne s'annule jamais). Si Y est une variable aléatoire telle que $X = g(Y)$, alors elle admet la densité*

$$f(g(t))|g'(t)|. \quad (2.1.45)$$

DÉMONSTRATION. Considérons le cas où g est croissante. Alors la fonction de répartition de Y peut s'écrire

$$F_Y(t) = \mathbb{P}\{Y \leq t\} = \mathbb{P}\{g(Y) \leq g(t)\} = \mathbb{P}\{X \leq g(t)\} = \int_{-\infty}^{g(t)} f(x) dx. \quad (2.1.46)$$

Le résultat suit alors en prenant la dérivée par rapport à t (par le théorème sur la dérivée d'une intégrale dont les bornes dépendent d'un paramètre). Dans le cas où g est décroissante, le raisonnement est similaire, sauf que le sens d'une inégalité change. \square

Exemple 2.1.12. Soit X une variable de loi uniforme $\mathcal{U}([0, 1])$. Soit $Y = -\log(X)$, c'est-à-dire $X = e^{-Y}$. Alors la formule (2.1.46) avec $f(x) = 1_{[0,1]}(x)$ et $g(t) = e^{-t}$ montre que la densité de Y est donnée par

$$1_{[0,1]}(e^{-t})|-e^{-t}| = e^{-t} 1_{\{t \geq 0\}}, \quad (2.1.47)$$

c'est-à-dire que Y suit la loi exponentielle $\text{Exp}(1)$.

2.2 Vecteurs aléatoires à densité

2.2.1 Densité conjointe

Définition 2.2.1 (Vecteur aléatoire). *On appelle vecteur aléatoire de dimension n un n -uplet $X = (X_1, \dots, X_n)$ où chaque X_i est une variable aléatoire réelle.*

Nous aimerions associer, quand c'est possible, une densité $f(x_1, \dots, x_n)$ à un tel vecteur aléatoire. C'est-à-dire que la probabilité que X appartienne à un sous-ensemble $B \subset \mathbb{R}^n$ devrait pouvoir s'écrire comme une intégrale multiple de f sur B .

Considérons d'abord le cas $n = 2$. Supposons que pour une fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}_+$, l'intégrale

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2 \quad (2.2.1)$$

existe pour chaque $x_1 \in \mathbb{R}$. Supposons de plus que la fonction f_1 est intégrable. Le théorème de Fubini affirme que

$$\int_{-\infty}^{\infty} f_1(x_1) dx_1 = \int_{-\infty}^{\infty} f_2(x_2) dx_2 \quad (2.2.2)$$

où l'on a posé

$$f_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_1. \quad (2.2.3)$$

On note alors

$$\int_{\mathbb{R}^2} f(x_1, x_2) \, dx_1 \, dx_2 := \int_{-\infty}^{\infty} f_1(x_1) \, dx_1 \equiv \int_{-\infty}^{\infty} f_2(x_2) \, dx_2 . \quad (2.2.4)$$

Si $B \subset \mathbb{R}^n$, on introduit de plus

$$\int_B f(x_1, x_2) \, dx_1 \, dx_2 = \int_{\mathbb{R}^2} 1_B(x_1, x_2) f(x_1, x_2) \, dx_1 \, dx_2 , \quad (2.2.5)$$

pour autant que cette dernière intégrale existe, où

$$1_B(x_1, x_2) = \begin{cases} 1 & \text{si } (x_1, x_2) \in B , \\ 0 & \text{sinon} \end{cases} \quad (2.2.6)$$

désigne la fonction indicatrice de $x \in B$. Ces notations se généralisent à des fonctions $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

Définition 2.2.2 (Densité d'un vecteur aléatoire). *Une fonction intégrable $f : \mathbb{R}^n \rightarrow [0, \infty[$ est une densité si elle satisfait*

$$\int_{\mathbb{R}^n} f(x_1, \dots, x_n) \, dx_1 \dots dx_n = 1 . \quad (2.2.7)$$

On dit qu'un vecteur aléatoire $X = (X_1, \dots, X_n)$ admet la densité f si

$$\mathbb{P}\{X_1 \leq a_1, \dots, X_n \leq a_n\} = \int_{]-\infty, a_1] \times \dots \times]-\infty, a_n]} f(x_1, \dots, x_n) \, dx_1 \dots dx_n \quad (2.2.8)$$

pour tout choix de $a_1, \dots, a_n \in \mathbb{R}$. On dit aussi que f est la densité conjointe des variables aléatoires X_1, \dots, X_n .

Le théorème suivant, sur lequel nous reviendrons dans la section suivante, permet d'étendre la relation (2.2.8) à des événements plus généraux. Pour l'instant il suffira de savoir que les *Boréliens* sont une classe de sous-ensembles de \mathbb{R}^n , comprenant en particulier les ouverts et les fermés de \mathbb{R}^n , ainsi que toutes leurs réunions et intersections.

Théorème 2.2.3. *Soit X un vecteur aléatoire admettant la densité f . Alors pour tout Borélien $B \subset \mathbb{R}^n$, on a*

$$\mathbb{P}\{X \in B\} = \int_B f(x_1, \dots, x_n) \, dx_1 \dots dx_n . \quad (2.2.9)$$

Exemple 2.2.4.

1. *L'aiguille de Buffon*: On laisse tomber une aiguille de longueur $a < 1$ sur une feuille recouverte de lignes parallèles, distantes de 1 (Figure 2.2). Avec quelle probabilité l'aiguille coupe-t-elle une des lignes?

On peut admettre que la position $y \in [0, 1[$ du milieu de l'aiguille, mesurée par rapport à la ligne située immédiatement en-dessous de ce milieu, et l'angle $\varphi \in [0, 2\pi[$ que l'aiguille fait avec les lignes, sont distribués uniformément:

$$f(y, \varphi) = \frac{1}{2\pi} 1_{[0, 1[\times [0, 2\pi[}(y, \varphi) . \quad (2.2.10)$$

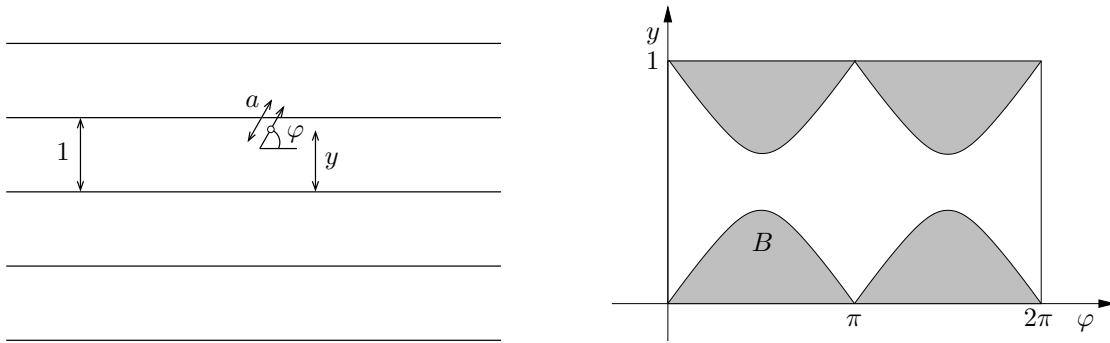


FIGURE 2.2. Expérience de l'aiguille de Buffon. Le domaine B correspond aux valeurs de φ et y pour lesquelles il y a une intersection entre l'aiguille et une ligne horizontale.

On remarque alors que l'aiguille coupe une ligne si et seulement si (y, φ) appartient à l'ensemble

$$B = \left\{ (y, \varphi) \in [0, 1] \times [0, 2\pi[: y - \frac{a}{2} |\sin \varphi| < 0 \text{ ou } y + \frac{a}{2} |\sin \varphi| > 1 \right\}. \quad (2.2.11)$$

La probabilité d'une intersection est donc donnée par

$$\begin{aligned} \mathbb{P}\{(y, \varphi) \in B\} &= \int_B f(y, \varphi) \, dy \, d\varphi = 4 \int_{\{(y, \varphi) : 0 \leq \varphi \leq \pi, 0 \leq y \leq \frac{a}{2} \sin \varphi\}} \frac{1}{2\pi} \, dy \, d\varphi \\ &= \frac{4}{2\pi} \int_0^\pi \int_0^{\frac{a}{2} \sin \varphi} \, dy \, d\varphi = \frac{2}{\pi} \int_0^\pi \frac{a}{2} \sin \varphi \, d\varphi = \frac{2a}{\pi}. \end{aligned} \quad (2.2.12)$$

Cette expérience permet donc d'estimer la valeur de π (en lançant l'aiguille un grand nombre de fois, et en invoquant la loi des grands nombres, que nous verrons au chapitre suivant).

2. *Densités marginales*: Dans le cas $n = 2$, appliquons le Théorème 2.2.3 pour l'ensemble $B =]-\infty, a] \times \mathbb{R}$ (Figure 2.3a). On a $(x_1, x_2) \in B$ si et seulement si $x_1 \leq a$, et donc

$$\mathbb{P}\{X_1 \leq a\} = \mathbb{P}\{X \in B\} = \int_{-\infty}^a \underbrace{\int_{-\infty}^{\infty} f(x_1, x_2) \, dx_2}_{f_1(x_1)} \, dx_1. \quad (2.2.13)$$

Ainsi, $f_1(x_1)$ est la densité de la variable aléatoire X_1 . On l'appelle *première densité marginale de f* . On définit de manière analogue la seconde densité marginale $f_2(x_2)$. Plus généralement, pour une densité de dimension $n > 2$, la k -ème densité marginale est obtenue par intégration sur tous les x_i avec $i \neq k$.

3. *Densité d'une somme*: Soient X_1 et X_2 deux variables aléatoires. Alors le changement

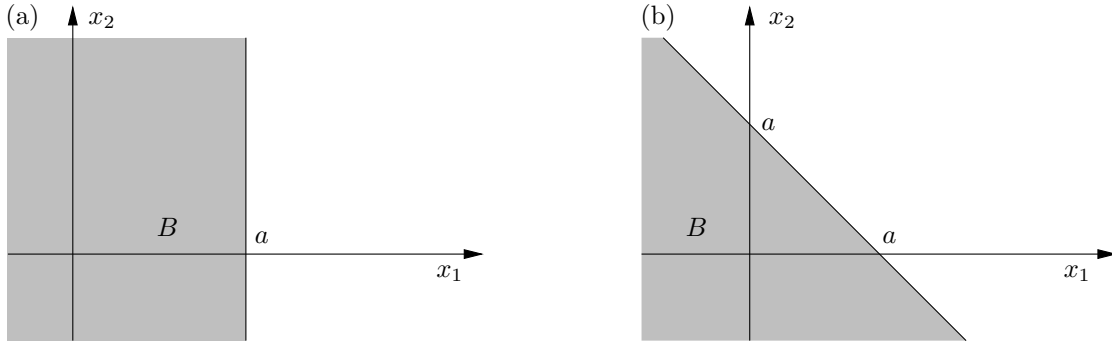


FIGURE 2.3. (a) Domaine d'intégration déterminant la densité marginale d'une densité conjointe. (b) Domaine d'intégration déterminant la densité de la somme $X_1 + X_2$.

de variable $x_1 = z - x_2$ nous permet d'écrire (c.f. Figure 2.3b)

$$\begin{aligned}
 \mathbb{P}\{X_1 + X_2 \leq a\} &= \int_{\{(x_1, x_2): x_1 \leq a - x_2\}} f(x_1, x_2) \, dx_1 \, dx_2 \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{a - x_2} f(x_1, x_2) \, dx_1 \, dx_2 \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^a f(z - x_2, x_2) \, dz \, dx_2 \\
 &= \int_{-\infty}^a \underbrace{\left[\int_{-\infty}^{\infty} f(z - x_2, x_2) \, dx_2 \right]}_{=: g(z)} \, dz . \tag{2.2.14}
 \end{aligned}$$

La densité de la somme $Z = X_1 + X_2$ est donc donnée par

$$g(z) = \int_{-\infty}^{\infty} f(z - x_2, x_2) \, dx_2 . \tag{2.2.15}$$

Proposition 2.2.5. *Si X_1, X_2 admettent la densité conjointe f , et si leurs espérances existent, alors l'espérance de $X_1 + X_2$ existe et vaut*

$$\mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2) . \tag{2.2.16}$$

DÉMONSTRATION. En utilisant (2.2.16) et le changement de variable $z = x_1 + x_2$, on a

$$\begin{aligned}
 \mathbb{E}(X_1 + X_2) &= \int_{-\infty}^{\infty} z g(z) \, dz = \int_{-\infty}^{\infty} z \int_{-\infty}^{\infty} f(z - x_2, x_2) \, dx_2 \, dz \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 + x_2) f(x_1, x_2) \, dx_1 \, dx_2 \\
 &= \int_{-\infty}^{\infty} x_1 \underbrace{\int_{-\infty}^{\infty} f(x_1, x_2) \, dx_2}_{f_1(x_1)} \, dx_1 + \int_{-\infty}^{\infty} x_2 \underbrace{\int_{-\infty}^{\infty} f(x_1, x_2) \, dx_1}_{f_2(x_2)} \, dx_2 \\
 &= \mathbb{E}(X_1) + \mathbb{E}(X_2) . \tag{2.2.17}
 \end{aligned}$$

□

2.2.2 Indépendance, convolution

Définition 2.2.6 (Indépendance). *Les variables aléatoires réelles X_1, \dots, X_n sont dites indépendantes si*

$$\mathbb{P}\{X_1 \leq a_1, \dots, X_n \leq a_n\} = \mathbb{P}\{X_1 \leq a_1\} \dots \mathbb{P}\{X_n \leq a_n\} \quad (2.2.18)$$

pour tout choix de $a_1, \dots, a_n \in \mathbb{R}$.

Théorème 2.2.7. *Soient X_1, \dots, X_n des variables aléatoires telles que X_i admette une densité f_i pour chaque i . Alors elles sont indépendantes si et seulement si*

$$f(x_1, \dots, x_n) := f_1(x_1) \dots f_n(x_n) \quad (2.2.19)$$

est la densité conjointe de X_1, \dots, X_n .

DÉMONSTRATION.

\Leftarrow Si f est la densité de X , alors

$$\begin{aligned} \mathbb{P}\{X_1 \leq a_1, \dots, X_n \leq a_n\} &= \int_{-\infty}^{a_1} \dots \int_{-\infty}^{a_n} f_1(x_1) \dots f_n(x_n) \, dx_1 \dots dx_n \\ &= \int_{-\infty}^{a_1} f_1(x_1) \, dx_1 \dots \int_{-\infty}^{a_n} f_n(x_n) \, dx_n \\ &= \mathbb{P}\{X_1 \leq a_1\} \dots \mathbb{P}\{X_n \leq a_n\}. \end{aligned} \quad (2.2.20)$$

\Rightarrow Si X_1, \dots, X_n sont indépendantes, alors

$$\begin{aligned} \mathbb{P}\{X_1 \leq a_1, \dots, X_n \leq a_n\} &= \mathbb{P}\{X_1 \leq a_1\} \dots \mathbb{P}\{X_n \leq a_n\} \\ &= \int_{-\infty}^{a_1} f_1(x_1) \, dx_1 \dots \int_{-\infty}^{a_n} f_n(x_n) \, dx_n \\ &= \int_{-\infty}^{a_1} \dots \int_{-\infty}^{a_n} f(x_1, \dots, x_n) \, dx_1 \dots dx_n, \end{aligned} \quad (2.2.21)$$

donc $f(x_1, \dots, x_n)$ est la densité conjointe de (X_1, \dots, X_n) . \square

Remarquons que les $f_i(x_i)$ sont évidemment les densités marginales de $f(x_1, \dots, x_n)$. La relation (2.2.14) implique immédiatement le résultat suivant.

Corollaire 2.2.8. *Si X_1 et X_2 sont des variables aléatoires indépendantes, de densités respectives f_1 et f_2 , alors $X_1 + X_2$ admet la densité*

$$g(z) = \int_{-\infty}^{\infty} f_1(z - x_2) f_2(x_2) \, dx_2. \quad (2.2.22)$$

Ceci motive la définition suivante.

Définition 2.2.9 (Convolution). *Si f_1 et f_2 sont deux densités, la fonction*

$$f_1 \star f_2(z) = \int_{-\infty}^{\infty} f_1(z - x_2) f_2(x_2) \, dx_2 \quad (2.2.23)$$

s'appelle la convolution de f_1 et f_2 .

Nous appliquons maintenant ces résultats au cas particulier important de variables aléatoires gaussiennes, c'est-à-dire suivant des lois normales.

Théorème 2.2.10. *Soient X_1, \dots, X_n des variables aléatoires indépendantes, suivant chacune une loi normale, de moyenne μ_i et de variance σ_i^2 . Alors la somme $X_1 + \dots + X_n$ suit une loi normale de moyenne $\mu_1 + \dots + \mu_n$ et de variance $\sigma_1^2 + \dots + \sigma_n^2$.*

DÉMONSTRATION. Nous allons considérer le cas $n = 2$, le cas général s'en déduit par récurrence. Il est plus commode de travailler avec $Y_1 = X_1 - \mu_1$ et $Y_2 = X_2 - \mu_2$, qui suivent des lois normales centrées (c'est-à-dire de moyenne nulle). Soit

$$u = u(y_2) = \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sigma_1 \sigma_2} y_2 - \frac{\sigma_2}{\sigma_1 \sqrt{\sigma_1^2 + \sigma_2^2}} z. \quad (2.2.24)$$

Nous allons utiliser le fait que

$$u^2 + \frac{z^2}{\sigma_1^2 + \sigma_2^2} = \frac{(z - y_2)^2}{\sigma_1^2} + \frac{y_2^2}{\sigma_2^2}, \quad (2.2.25)$$

qui se vérifie par un calcul élémentaire. La densité g de $Y_1 + Y_2$ est alors donnée par la convolution

$$\begin{aligned} g(z) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \sigma_1} e^{-(z-y_2)^2/2\sigma_1^2} \frac{1}{\sqrt{2\pi} \sigma_2} e^{-y_2^2/2\sigma_2^2} dy_2 \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left(\frac{(z-y_2)^2}{\sigma_1^2} + \frac{y_2^2}{\sigma_2^2}\right)\right\} dy_2 \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} e^{-u^2/2} e^{-z^2/2(\sigma_1^2+\sigma_2^2)} \frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}} du \\ &= \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}} e^{-z^2/2(\sigma_1^2+\sigma_2^2)} = \varphi(z; 0, \sigma_1^2 + \sigma_2^2), \end{aligned} \quad (2.2.26)$$

où nous avons utilisé le changement de variable $y_2 \mapsto u(y_2)$. Ceci montre que $Y_1 + Y_2$ suit la loi $\mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$, et par conséquent $X_1 + X_2$ suit la loi $\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. \square

2.2.3 Changement de variable

Considérons, comme dans le cas réel, la relation entre la densité f d'un vecteur aléatoire X et la densité d'une fonction de X . Dans le cas réel, nous avons dérivé une formule dans le cas où la fonction était continûment différentiable et strictement monotone. Dans le cas vectoriel, nous aurons besoin de supposer que la fonction est un difféomorphisme.

Définition 2.2.11 (Difféomorphisme, Jacobien). *Soient A, B des ouverts de \mathbb{R}^n . Une fonction $g : A \rightarrow B$ est un difféomorphisme si elle est continûment différentiable, et admet une réciproque g^{-1} qui est également continûment différentiable. On appelle Jacobien de g le déterminant*

$$\frac{\partial g}{\partial x}(x) = \det \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(x) & \frac{\partial g_1}{\partial x_2}(x) & \dots & \frac{\partial g_1}{\partial x_n}(x) \\ \frac{\partial g_2}{\partial x_1}(x) & \frac{\partial g_2}{\partial x_2}(x) & \dots & \frac{\partial g_2}{\partial x_n}(x) \\ \vdots & \vdots & & \vdots \\ \frac{\partial g_n}{\partial x_1}(x) & \frac{\partial g_n}{\partial x_2}(x) & \dots & \frac{\partial g_n}{\partial x_n}(x) \end{pmatrix}. \quad (2.2.27)$$

Rappelons le théorème fondamental du changement de variables dans une intégrale multiple.

Théorème 2.2.12 (Théorème fondamental du changement de variables). *Soit $g : B \rightarrow A$ un difféomorphisme. Alors pour toute fonction intégrable $f : A \rightarrow \mathbb{R}^n$ et tout $D \subset B$, on a*

$$\int_{g(D)} f(x) \, dx_1 \dots dx_n = \int_D f(g(y)) \left| \frac{\partial g}{\partial y}(y) \right| \, dy_1 \dots dy_n. \quad (2.2.28)$$

Exemple 2.2.13 (Coordonnées polaires). Considérons l'application $g : (r, \varphi) \mapsto (x_1, x_2)$ définie par

$$\begin{aligned} x_1 &= r \cos(\varphi), \\ x_2 &= r \sin(\varphi). \end{aligned} \quad (2.2.29)$$

Elle définit un difféomorphisme de $]0, \infty[\times]0, 2\pi[$ vers $\mathbb{R}^2 \setminus \{(0, 0)\}$. En effet, g est bien continûment différentiable, et sa réciproque est définie par

$$\begin{aligned} r &= \sqrt{x_1^2 + x_2^2}, \\ \varphi &= \arg(x_1 + ix_2), \end{aligned} \quad (2.2.30)$$

où l'argument $\arg(z)$ d'un nombre complexe z est l'unique nombre $\varphi \in [0, 2\pi[$ tel que $|z| e^{i\varphi} = z$ (on peut l'exprimer à l'aide de $\arctan(x_2/x_1)$). Le Jacobien de g est donné par

$$\frac{\partial(x_1, x_2)}{\partial(r, \varphi)} = \det \begin{pmatrix} \frac{\partial(r \cos \varphi)}{\partial r} & \frac{\partial(r \cos \varphi)}{\partial \varphi} \\ \frac{\partial(r \sin \varphi)}{\partial r} & \frac{\partial(r \sin \varphi)}{\partial \varphi} \end{pmatrix} = \begin{pmatrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & r \cos \varphi \end{pmatrix} = r. \quad (2.2.31)$$

La formule de changement de variable donne donc, pour une fonction intégrable et continue f ,

$$\int_{\mathbb{R}^2} f(x_1, x_2) \, dx_1 \, dx_2 = \int_0^{2\pi} \int_0^\infty f(r \cos \varphi, r \sin \varphi) r \, dr \, d\varphi \quad (2.2.32)$$

(on notera que les intégrales sur \mathbb{R}^2 et sur $\mathbb{R}^2 \setminus \{(0, 0)\}$ sont les mêmes pour une fonction continue). Appliquons cette égalité à la fonction

$$f(x_1, x_2) = \frac{1}{2\pi} e^{-(x_1^2 + x_2^2)/2} = \left(\frac{1}{\sqrt{2\pi}} e^{-x_1^2/2} \right) \left(\frac{1}{\sqrt{2\pi}} e^{-x_2^2/2} \right). \quad (2.2.33)$$

On obtient

$$\begin{aligned} \left(\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx \right)^2 &= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-x_1^2/2} \, dx_1 \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-x_2^2/2} \, dx_2 \\ &= \int_{\mathbb{R}^2} \frac{1}{2\pi} e^{-(x_1^2 + x_2^2)/2} \, dx_1 \, dx_2 \\ &= \int_0^{2\pi} \int_0^\infty \frac{e^{-r^2/2}}{2\pi} r \, dr \, d\varphi \\ &= \int_0^{2\pi} \left[\frac{-e^{-r^2/2}}{2\pi} \right]_0^\infty d\varphi = \int_0^{2\pi} \frac{1}{2\pi} d\varphi = 1, \end{aligned} \quad (2.2.34)$$

ce qui démontre (2.1.5), c'est-à-dire le fait que $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ est bien une densité.

Corollaire 2.2.14 (Formule de transformation d'une densité). *Soit X un vecteur aléatoire, admettant la densité f . Soit g un homéomorphisme, et soit Y un vecteur aléatoire tel que $X = g(Y)$. Alors Y admet la densité*

$$f(g(y)) \left| \frac{\partial g}{\partial y}(y) \right|. \quad (2.2.35)$$

DÉMONSTRATION. Soit $B \subset \mathbb{R}^n$ et $A = g(B)$. Alors on a

$$\mathbb{P}\{Y \in B\} = \mathbb{P}\{g(Y) \in g(B)\} = \mathbb{P}\{X \in A\} = \int_A f(x) \, dx = \int_B f(g(y)) \left| \frac{\partial g}{\partial y}(y) \right| \, dy, \quad (2.2.36)$$

en vertu du théorème précédent. \square

Exemple 2.2.15 (Algorithme de Box–Müller). On se donne deux variables aléatoires U et V , indépendantes et de loi uniforme $\mathcal{U}([0, 1])$. On cherche la densité du couple

$$(X, Y) = (\sqrt{-2 \log(1 - U)} \cos(2\pi V), \sqrt{-2 \log(1 - U)} \sin(2\pi V)). \quad (2.2.37)$$

Remarquons pour commencer que la fonction $u \mapsto -2 \log(1 - u)$ envoie l'intervalle $[0, 1[$ sur \mathbb{R}_+ . Le domaine de (X, Y) est donc \mathbb{R}^2 . Le Jacobien de la transformation $(U, V) \mapsto (X, Y)$ est donné par $-2\pi/(1 - U)$. Le Jacobien de la réciproque est égal à l'inverse de cette quantité, à savoir $-(1 - U)/(2\pi) = e^{-(X^2 + Y^2)/2} / (2\pi)$. Le couple (U, V) admet la densité $1_{[0,1]}(u)1_{[0,1]}(v)$. En appliquant la formule de changement de variable, on obtient que (X, Y) admet la densité

$$\frac{e^{-(x^2 + y^2)/2}}{2\pi} = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \frac{e^{-y^2/2}}{\sqrt{2\pi}}. \quad (2.2.38)$$

Cela signifie que X et Y sont indépendantes, et suivent les deux une loi normale standard $\mathcal{N}(0, 1)$.

Exemple 2.2.16 (Vecteur gaussien). La densité conjointe des n variables gaussiennes indépendantes est donnée par

$$f(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2} \sigma_1 \dots \sigma_n} \exp \left\{ -\frac{1}{2} \left[\frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \dots + \frac{(x_n - \mu_n)^2}{\sigma_n^2} \right] \right\}, \quad (2.2.39)$$

on l'appelle une densité gaussienne à n dimensions. Considérons le cas où les X_i suivent la loi normale standard, c'est-à-dire

$$f(x) = f(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}} e^{-\|x\|^2/2}. \quad (2.2.40)$$

Par l'indépendance, nous aurons $\text{cov}(X_i, X_j) = 0$ pour $i \neq j$, alors que $\text{Var}(X_i) = 1$, et par conséquent $\mathbb{E}(X_i X_j)$ vaut 1 si $i = j$, 0 sinon. Soit alors S une matrice non singulière, c'est-à-dire telle que $\det S \neq 0$, et $T = S^{-1}$. Considérons la variable aléatoire

$$Y = TX + \mu, \quad (2.2.41)$$

dont l'espérance vaut μ . Quelle est sa densité? Par la formule de changement de variable, la densité de Y est donnée par

$$\tilde{f}(y) = \frac{|\det S|}{(2\pi)^{n/2}} e^{-\|S(y - \mu)\|^2/2}. \quad (2.2.42)$$

Remarquons que l'on peut récrire

$$\begin{aligned} \|S(y - \mu)\|^2 &= \langle S(y - \mu) | S(y - \mu) \rangle = \langle y - \mu | A(y - \mu) \rangle \\ &= \sum_{i,j=1}^n a_{ij}(y_i - \mu_i)(y_j - \mu_j), \end{aligned} \quad (2.2.43)$$

où A est la matrice symétrique définie positive $A = S^T S$, d'éléments a_{ij} , qui satisfait $\det A = (\det S)^2$. Nous avons donc

$$\tilde{f}(y_1, \dots, y_n) = \frac{\sqrt{\det A}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2} \sum_{i,j=1}^n a_{ij}(y_i - \mu_i)(y_j - \mu_j)\right\}, \quad (2.2.44)$$

qui est la forme générale d'une densité gaussienne. Nous avons déjà vu que son espérance vaut μ . Quant à sa covariance, elle se calcule en observant que

$$\begin{aligned} \mathbb{E}(Y_i Y_j) &= \mathbb{E}\left(\left[\sum_{k=1}^n T_{ik} X_k + \mu_i\right] \left[\sum_{\ell=1}^n T_{j\ell} X_\ell + \mu_j\right]\right) \\ &= \sum_{k,\ell=1}^n T_{ik} T_{j\ell} \mathbb{E}(X_k X_\ell) + \mu_i \mu_j \\ &= \sum_{k=1}^n T_{ik} T_{jk} + \mu_i \mu_j = (TT^T)_{ij} + \mu_i \mu_j, \end{aligned} \quad (2.2.45)$$

d'où, comme $TT^T = A^{-1}$,

$$\text{cov}(Y_i, Y_j) = (A^{-1})_{ij}. \quad (2.2.46)$$

La matrice $C = A^{-1}$ s'appelle la *matrice de covariance* de la densité gaussienne (2.2.44). Nous voyons en particulier que les variables Y_1, \dots, Y_n sont non corrélées si et seulement si A est diagonale, ce qui est le cas si et seulement si les variables sont indépendantes.

Définition 2.2.17 (Densité gaussienne). *Soit C une matrice $n \times n$ symétrique, définie positive, et $\mu \in \mathbb{R}^n$. La densité*

$$\varphi(x; \mu, C) = \frac{1}{\sqrt{(2\pi)^n \det C}} \exp\left\{-\frac{1}{2} \langle x - \mu | C^{-1}(x - \mu) \rangle\right\} \quad (2.2.47)$$

s'appelle densité gaussienne de moyenne μ et matrice de covariance C .

Voici pour terminer une autre propriété remarquable des densités gaussiennes.

Proposition 2.2.18. *Les marginales d'une densité gaussienne sont des densités gaussiennes.*

DÉMONSTRATION. Il suffit à nouveau de considérer le cas $n = 2$, $\mu = 0$, les cas de dimension plus élevée pouvant être traités par récurrence. Or l'identité

$$a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 = a_{11}\left(x_1 + \frac{a_{12}}{a_{11}}x_2\right)^2 + \frac{a_{11}a_{22} - a_{12}^2}{a_{11}}x_2^2 \quad (2.2.48)$$

permet d'effectuer l'intégrale de la densité sur x_1 , à l'aide du changement de variable $u = x_1 + (a_{12}/a_{11})x_2$. La densité marginale est donc proportionnelle à $e^{-x_2^2/2\sigma_2^2}$, avec σ_2 donné par $\sigma_2^2 = a_{11}/\det A$. \square

2.3 Mesures de probabilité et espaces probabilisés*

Nous donnons dans cette section une introduction à la théorie générale des espaces probabilisés, telle que fondée par Kolmogorov. Cette théorie permet de traiter de manière unifiée les espaces probabilisés discrets, les variables aléatoires à densité, et bien d'autres situations. La construction comporte trois étapes principales:

1. l'espace probabilisé est défini à l'aide des notions de tribu et de mesure de probabilité;
2. les variables aléatoires sont définies comme des fonctions mesurables;
3. les espérances sont définies comme des intégrales par rapport à la mesure de probabilité.

2.3.1 Mesures

Définition 2.3.1 (Tribu). *Soit Ω un ensemble. Une famille non vide \mathcal{F} de sous-ensembles de Ω s'appelle une tribu (ou σ -algèbre) si*

- $\Omega \in \mathcal{F}$;
- pour tout $A \in \mathcal{F}$, on a $A^c \in \mathcal{F}$;
- pour toute famille $\{A_n\}_{n \in \mathbb{N}}$ d'éléments de \mathcal{F} , on a $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Définition 2.3.2 (Mesure de probabilité). *Une mesure sur une tribu \mathcal{F} est une application $\mu : \mathcal{F} \rightarrow [0, \infty]$ telle que*

- $\mu(\emptyset) = 0$;
- σ -additivité: pour toute famille $\{A_n\}_{n \in \mathbb{N}}$ d'éléments de \mathcal{F} , disjoints deux à deux,

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n). \quad (2.3.1)$$

Si de plus $\mu(\Omega) = 1$, μ est une mesure de probabilité, et on la notera \mathbb{P} .

Définition 2.3.3 (Espace probabilisé). *Si \mathcal{F} est une tribu sur Ω , (Ω, \mathcal{F}) est appelé un espace mesurable, et les éléments de \mathcal{F} sont dits mesurables. Si μ est une mesure, $(\Omega, \mathcal{F}, \mu)$ est appelé un espace mesuré. Si \mathbb{P} est une mesure de probabilité, $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé.*

Exemple 2.3.4.

1. $\mathcal{F} = \{\Omega, \emptyset\}$ est une σ -algèbre, appelée la σ -algèbre *triviale*. La seule mesure de probabilité sur \mathcal{F} est donnée par $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\emptyset) = 0$.
2. Soit Ω un ensemble fini ou dénombrable, et $\mathcal{F} = \mathcal{P}(\Omega) = \{A : A \subset \Omega\}$ l'ensemble de toutes les parties de Ω . Alors \mathcal{F} est une tribu, et (Ω, \mathcal{F}) est un espace mesurable. Si $p : \Omega \rightarrow [0, 1]$ satisfait $\sum_{\omega \in \Omega} p(\omega) = 1$, alors l'application $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ définie par

$$\mathbb{P}(A) = \sum_{\omega \in A} p(\omega) \quad (2.3.2)$$

est une mesure de probabilité. On la note

$$\mathbb{P} = \sum_{\omega \in \Omega} p(\omega) \delta_{\omega}, \quad (2.3.3)$$

où $\delta_{\omega}(A)$ vaut 1 si $\omega \in A$, 0 sinon. L'espace $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé discret.

3. Si $\Omega = \mathbb{R}$, la tribu des Boréliens \mathcal{B} est par définition la plus petite tribu contenant tous les intervalles de la forme $]a, b] \subset \mathbb{R}$. On dit que c'est la tribu engendrée par ces intervalles.

L'application λ définie par

$$\lambda(]a, b]) = b - a, \quad (2.3.4)$$

et étendue à \mathcal{B} par σ -additivité, est une mesure, et s'appelle la *mesure de Lebesgue*. Le mesure de Lebesgue d'un intervalle est donc simplement sa longueur. Cette mesure n'est pas une mesure de probabilité car $\lambda(\mathbb{R}) = \infty$.

Si $F : \mathbb{R} \rightarrow [0, 1]$ est une fonction de répartition, l'application $\mathbb{P} : \mathcal{B} \rightarrow [0, 1]$ définie par

$$\mathbb{P}(]a, b]) = F(b) - F(a), \quad (2.3.5)$$

et étendue à \mathcal{B} par σ -additivité, est une mesure de probabilité. Ainsi $(\mathbb{R}, \mathcal{B}, \mathbb{P})$ est un espace probabilisé.

4. Si $\Omega = \mathbb{R}^n$, on définit de manière analogue la tribu des Boréliens \mathcal{B}_n comme celle engendrée par les hypercubes $]a_1, b_1] \times \cdots \times]a_n, b_n]$. La mesure de Lebesgue sur \mathbb{R}^n est définie par

$$\lambda(]a_1, b_1] \times \cdots \times]a_n, b_n]) = (b_1 - a_1) \cdots (b_n - a_n), \quad (2.3.6)$$

et étendue à \mathcal{B}_n par σ -additivité.

5. Soient $(\Omega_i, \mathcal{F}_i)$, $i \in I$, des espaces mesurables. Pour $A \in \mathcal{F}_k$, notons

$$A^{(k)} = \left\{ \omega \in \prod_i \Omega_i : \omega_k \in A \right\}. \quad (2.3.7)$$

L'ensemble de tous les $A^{(k)}$ engendre une tribu sur $\prod_i \Omega_i$, appelée *tribu produit* des \mathcal{F}_i , et notée $\otimes_i \mathcal{F}_i$. L'espace mesurable $(\prod_i \Omega_i, \otimes_i \mathcal{F}_i)$ est *l'espace produit* des $(\Omega_i, \mathcal{F}_i)$. Par exemple, dans une expérience de Pile ou Face, prenons $\Omega_i = \{P, F\}$ pour chaque jet, avec la tribu

$$\mathcal{F}_i = \{\emptyset, \{P\}, \{F\}, \{P, F\}\}. \quad (2.3.8)$$

Une mesure de probabilité sur l'espace produit est définie par

$$\mathbb{P}(\{\omega \in \Omega : (\omega_1, \dots, \omega_m) \in A \subset \{F, P\}^m\}) = \frac{|A|}{2^m} \quad (2.3.9)$$

pour tout m .

2.3.2 Applications mesurables

Définition 2.3.5 (Application mesurable).

- Si (Ω, \mathcal{F}) et (Ω', \mathcal{F}') sont des espaces mesurables, une application $f : \Omega \rightarrow \Omega'$ est dite \mathcal{F} - \mathcal{F}' -mesurable si elle satisfait

$$f^{-1}(\mathcal{F}') := \{f^{-1}(A) : A \in \mathcal{F}'\} \subset \mathcal{F}, \quad (2.3.10)$$

c'est-à-dire si la préimage de tout élément de \mathcal{F}' est un élément de \mathcal{F} .

- Une application $f : \Omega \rightarrow \mathbb{R}$ est dite \mathcal{F} -mesurable, ou simplement mesurable, si elle est \mathcal{F} - \mathcal{B} -mesurable. En particulier, toute fonction continue $f : \mathbb{R} \rightarrow \mathbb{R}$ est \mathcal{B} -mesurable.

Définition 2.3.6 (Variable aléatoire).

- Si $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé et $X : \Omega \rightarrow \mathbb{R}$ est \mathcal{F} -mesurable, on dit que c'est une variable aléatoire. On vérifie que

$$\begin{aligned} \mathbb{P}X^{-1} : \mathcal{B} &\rightarrow [0, 1] \\ A &\mapsto \mathbb{P}X^{-1}(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}\{X \in A\} \end{aligned} \quad (2.3.11)$$

est une mesure de probabilité sur $(\mathbb{R}, \mathcal{B})$, appelée la loi de X .

- L'application

$$\begin{aligned} F_X : \mathbb{R} &\rightarrow [0, 1] \\ t &\mapsto F_X(t) = \mathbb{P}X^{-1}(-\infty, t] = \mathbb{P}\{X \leq t\} \end{aligned} \quad (2.3.12)$$

est appelée la fonction de répartition de X .

Les fonctions mesurables d'un espace probabilisé sont donc simplement celles pour lesquelles on peut définir $\mathbb{P}\{X \in A\}$ pour tout Borélien $A \in \mathcal{B}$. La situation est symbolisée dans le diagramme suivant (notons que X , étant mesurable, peut être identifiée à une application de \mathcal{F} dans \mathcal{B}):

$$\begin{array}{ccc} & \mathbb{P} & \\ (\Omega, \mathcal{F}) & \longrightarrow & [0, 1] \\ & X \searrow & \nearrow \mathbb{P}X^{-1} \\ & (\mathbb{R}, \mathcal{B}) & \end{array} \quad (2.3.13)$$

2.3.3 Intégrale de Lebesgue

Finalement, il nous faut définir la notion d'intégrale sur un espace mesuré ou probabilisé. Nous nous bornerons ici à donner les définitions et le résultat principal.

Définition 2.3.7 (Intégrale d'une fonction simple).

Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré.

- On appelle fonction simple toute fonction $e : \Omega \rightarrow \mathbb{R}$ de la forme

$$e(\omega) = \sum_{i=1}^k a_i 1_{A_i}(\omega), \quad (2.3.14)$$

où les a_i sont dans \mathbb{R} , les A_i sont dans \mathcal{F} , et $1_{A_i}(\omega)$ est la fonction indicatrice de $\omega \in A_i$.

- L'intégrale d'une telle fonction simple est définie par

$$\int_{\Omega} e \, d\mu = \sum_{i=1}^k a_i \mu(A_i). \quad (2.3.15)$$

Théorème 2.3.8. Toute fonction mesurable $f : \Omega \rightarrow [0, \infty]$ peut s'écrire comme

$$f = \lim_{n \rightarrow \infty} e_n, \quad (2.3.16)$$

pour une suite croissante de fonctions simples e_n (c'est-à-dire telle que $e_n(\omega) \leq e_{n+1}(\omega)$ pour tout $\omega \in \Omega$). De plus

$$\lim_{n \rightarrow \infty} \int_{\Omega} e_n \, d\mu \in [0, \infty] \quad (2.3.17)$$

existe et ne dépend pas de la suite de e_n convergeant vers f .

Ce résultat justifie la définition suivante.

Définition 2.3.9 (Intégrale d'une fonction mesurable).

- Soit $f : \Omega \rightarrow [0, \infty]$ une fonction mesurable. Alors son intégrale est définie par

$$\int_{\Omega} f \, d\mu = \lim_{n \rightarrow \infty} \int_{\Omega} e_n \, d\mu, \quad (2.3.18)$$

où e_n est une suite croissante de fonctions simples convergeant vers f .

- Soit $f : \Omega \rightarrow [-\infty, \infty]$ une fonction mesurable. Alors son intégrale est définie par

$$\int_{\Omega} f \, d\mu = \int_{\Omega} f_+ \, d\mu - \int_{\Omega} f_- \, d\mu, \quad (2.3.19)$$

où $f_+(\omega) = \max\{f(\omega), 0\}$ et $f_-(\omega) = \max\{-f(\omega), 0\}$ désignent la partie positive et la partie négative de f .

- Pour tout $A \in \mathcal{F}$, on pose

$$\int_A f \, d\mu = \int_{\Omega} 1_A f \, d\mu, \quad (2.3.20)$$

- On note $\mathcal{L}_1(\Omega, \mathcal{F}, \mu)$ l'espace des fonctions mesurables $X : \Omega \rightarrow \mathbb{R}$ telles que l'intégrale de $|X|$ soit finie, et $\mathcal{L}_p(\Omega, \mathcal{F}, \mu)$ l'espace des fonctions mesurables $X : \Omega \rightarrow \mathbb{R}$ telles que l'intégrale de $|X|^p$ soit finie.

Dans le cas particulier $(\Omega, \mathcal{F}, \mu) = (\mathbb{R}, \mathcal{B}, \lambda)$, l'intégrale introduite ci-dessus s'appelle l'intégrale de Lebesgue. Par définition, l'intégrale de Lebesgue d'une fonction simple est égale à la surface comprise sous le graphe de cette fonction, donc à son intégrale de Riemann. Par conséquent, si une fonction est intégrable selon Riemann, son intégrale de Lebesgue est égale à son intégrale de Riemann, et on écrit

$$\int_{\mathbb{R}} f \, d\lambda = \int_{\mathbb{R}} f(x) \, dx = \int_{-\infty}^{\infty} f(x) \, dx. \quad (2.3.21)$$

Revenons maintenant au cas particulier d'un espace probabilisé.

Définition 2.3.10 (Espérance). Soit X une variable aléatoire sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ telle qu'on ait soit $X(\omega) \geq 0$ pour tout ω , soit $X \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$. Alors on appelle espérance de X l'intégrale

$$\mathbb{E}(X) = \int_{\Omega} X \, d\mathbb{P}. \quad (2.3.22)$$

Plus généralement, si $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ est \mathcal{B} -mesurable, on pose

$$\mathbb{E}(\varphi(X)) = \int_{\Omega} \varphi(X) \, d\mathbb{P}. \quad (2.3.23)$$

On notera que les notations suivantes sont toutes équivalentes:

$$\begin{aligned} \mathbb{P}\{X \leq t\} &= \int_{\{X \leq t\}} d\mathbb{P} = \int_{\Omega} 1_{\{X \leq t\}} \, d\mathbb{P} = \mathbb{E}(1_{\{X \leq t\}}) \\ &= \mathbb{P}X^{-1}(]-\infty, t]) = \int_{\mathbb{R}} 1_{]-\infty, t]} \, d(\mathbb{P}X^{-1}) = \int_{-\infty}^t d(\mathbb{P}X^{-1}). \end{aligned} \quad (2.3.24)$$

Exemple 2.3.11.

- Dans le cas d'un espace probablisé discret, c'est-à-dire

$$\mathbb{P} = \sum_{\omega \in \Omega} p(\omega) \delta_{\omega}, \quad (2.3.25)$$

toute variable aléatoire $X : \Omega \rightarrow \mathbb{R}$ est une fonction simple, car elle peut s'écrire

$$X = \sum_{\omega \in \Omega} X(\omega) 1_{\{\omega\}}. \quad (2.3.26)$$

Par conséquent, son espérance est donnée par

$$\mathbb{E}(X) = \int_{\Omega} X \, d\mathbb{P} = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\}) = \sum_{\omega \in \Omega} X(\omega) p(\omega). \quad (2.3.27)$$

- Si X est une variable aléatoire dont la fonction de répartition admet une densité f , on a

$$\mathbb{P}\{X \in A\} = \int_A f(x) \, dx. \quad (2.3.28)$$

Comme par ailleurs

$$\mathbb{P}\{X \in A\} = \mathbb{P}X^{-1}(A) = \int_A d(\mathbb{P}X^{-1}), \quad (2.3.29)$$

on s'aperçoit que la loi de X est donnée par $d(\mathbb{P}X^{-1}) = f(x) \, dx$. On retrouve donc

$$\mathbb{E}(X) = \int_{\Omega} X \, d\mathbb{P} = \int_{\mathbb{R}} x \, d(\mathbb{P}X^{-1}) = \int_{\mathbb{R}} x f(x) \, dx. \quad (2.3.30)$$

Chapitre 3

Théorèmes limite

Dans certains cas il s'avère qu'une suite X_0, X_1, X_2, \dots de variables aléatoires converge vers une variable aléatoire X . Ce genre de résultat, appelé théorème limite, est utile car la limite X fournit en général une bonne approximation de X_n pour n suffisamment grand, et est souvent plus simple à étudier que les X_n de la suite.

Il existe plusieurs notions de convergence pour une suite de variables aléatoires : convergence en loi, en probabilité, en moyenne d'ordre p , presque sûre. Notre but ici n'est pas de développer la théorie générale de ces différentes notions de convergence, mais plutôt d'étudier quelques-uns des théorèmes limite les plus importants, dont la loi des grands nombres, qui nous permettra d'interpréter la notion d'espérance, et le théorème de la limite centrale, qui fournit une interprétation de la variance.

3.1 La “loi des petits nombres”

En 1898, Ladislaus Bortkiewicz nota que la loi de Poisson décrivait bien les événements rares dans de grandes populations. Il cite notamment comme exemple le nombre de décès annuels par suite de ruade de cheval dans la cavalerie prussienne, qui suit de près une loi de Poisson de paramètre $\lambda = 0.61$:

Nombre k de décès	0	1	2	3	4
Nombre de corps de cavalerie avec k décès	109	65	22	3	1
$200 \cdot \pi_{0.61}(k)$	108.67	66.28	20.21	4.11	0.63

(la valeur $\lambda = 0.61$ correspond au nombre moyen de décès par corps).

En fait, il est naturel de modéliser la situation en supposant que chaque soldat a une probabilité très faible q d'être tué par une ruade. En supposant les différents accidents indépendants, le nombre de décès serait alors décrit par une loi binomiale de paramètres n et q , où n est le nombre de soldats. Or il s'avère effectivement que pour n grand et q petit, avec nq de l'ordre de l'unité, la loi binomiale de paramètres n et q est bien approchée par une loi de Poisson de paramètre $\lambda = nq$.

Cette observation peut être utile en pratique. Bien que nous ayons une expression explicite pour la loi binomiale, les probabilités ne sont pas toujours aisées à calculer si le nombre n est grand. Supposons par exemple que l'on joue $n = 1000$ fois à un jeu de hasard dans lequel la probabilité de gagner est de $q = 1/1000$. Le calcul de la probabilité de gagner 20 fois parmi ces 1000 jeux nécessite la détermination du coefficient binomial

$\binom{1000}{20}$, donc des nombres $20!$ et $1000!/980!$. En particulier le second nombre est difficile à calculer précisément sur une calculatrice, à moins de multiplier explicitement les entiers de 981 jusqu'à 1000. Par contre il est facile de calculer $\pi_1(20) = e^{-1}/20!$.

3.1.1 Convergence en loi

Nous commençons par établir la convergence en loi (c'est-à-dire des probabilités à k fixé) de la loi binomiale vers la loi de Poisson. Rappelons que ces deux lois sont données par

$$\begin{aligned} b(k; n, q) &= \frac{n!}{(n-k)!k!} q^k (1-q)^{n-k} & k \in \{0, \dots, n\}, \\ \pi_\lambda(k) &= e^{-\lambda} \frac{\lambda^k}{k!} & k \in \mathbb{N}. \end{aligned} \quad (3.1.1)$$

Proposition 3.1.1. *Soit $\{q_n\}_{n \geq 0}$ une suite telle que $\lim_{n \rightarrow \infty} nq_n = \lambda > 0$. Alors, pour tout $k \in \mathbb{N}$,*

$$\lim_{n \rightarrow \infty} b(k; n, q_n) = \pi_\lambda(k). \quad (3.1.2)$$

DÉMONSTRATION. Soit $\lambda_n = nq_n$. Alors

$$\begin{aligned} b(k; n, q_n) &= b\left(k; n, \frac{\lambda_n}{n}\right) \\ &= \frac{n(n-1)\dots(n-k+1)}{k!} \frac{\lambda_n^k}{n^k} \left(1 - \frac{\lambda_n}{n}\right)^{n-k} \\ &= \frac{\lambda_n^k}{k!} \frac{1}{(1 - \lambda_n/n)^k} \left(1 - \frac{1}{n}\right) \dots \left(1 - \frac{k-1}{n}\right) \left(1 - \frac{\lambda_n}{n}\right)^n. \end{aligned} \quad (3.1.3)$$

Lorsque $n \rightarrow \infty$, on a $\lambda_n \rightarrow \lambda$, $\lambda_n/n \rightarrow 0$ et $j/n \rightarrow 0$ pour $j = 0, \dots, k-1$, donc

$$\lim_{n \rightarrow \infty} b(k; n, q_n) = \frac{\lambda^k}{k!} \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda_n}{n}\right)^n = \frac{\lambda^k}{k!} e^{-\lambda}, \quad (3.1.4)$$

la dernière égalité pouvant se montrer par un développement limité de $\log[(1 - \lambda_n/n)^n]$. \square

Exemple 3.1.2. La probabilité de gagner au tiercé est de $1/1000$. Quelle est la probabilité de gagner k fois en jouant 2000 fois?

Nous modélisons la situation par une expérience de Bernoulli de longueur $n = 2000$ et paramètre $q = 1/1000$. La probabilité de gagner k fois sera donnée par

$$b(k; 2000, 1/1000) \simeq \pi_2(k) = e^{-2} \frac{2^k}{k!}. \quad (3.1.5)$$

Le tableau suivant compare quelques valeurs des deux lois.

k	0	1	2	3	4	5
$b(k; 2000, 1/1000)$	0.13520	0.27067	0.27081	0.18053	0.09022	0.03605
$\pi_2(k)$	0.13534	0.27067	0.27067	0.18045	0.09022	0.03609

3.1.2 Convergence en distance ℓ_1

On peut en fait faire beaucoup mieux que de montrer la convergence simple de la loi binomiale vers la loi de Poisson, et borner la “distance ℓ_1 ” entre les deux lois. Pour cela, nous commençons par établir un résultat dit de couplage.

Lemme 3.1.3. *Soient X et Y deux variables aléatoires à valeurs dans \mathbb{N} . Alors*

$$\sum_{k=0}^{\infty} |\mathbb{P}\{X = k\} - \mathbb{P}\{Y = k\}| \leq 2\mathbb{P}\{X \neq Y\}. \quad (3.1.6)$$

DÉMONSTRATION. Posons, pour abréger l’écriture, $f(k) = \mathbb{P}\{X = k\}$, $g(k) = \mathbb{P}\{Y = k\}$ et $A = \{k : f(k) > g(k)\}$. Alors

$$\begin{aligned} \sum_{k=0}^{\infty} |f(k) - g(k)| &= \sum_{k \in A} (f(k) - g(k)) - \sum_{k \notin A} (f(k) - g(k)) \\ &= 2 \sum_{k \in A} (f(k) - g(k)) - \underbrace{\sum_{k \in \mathbb{N}} (f(k) - g(k))}_{=1-1=0}. \end{aligned} \quad (3.1.7)$$

Or nous pouvons écrire

$$\begin{aligned} \sum_{k \in A} (f(k) - g(k)) &= \mathbb{P}\{X \in A\} - \mathbb{P}\{Y \in A\} \\ &= \mathbb{P}\{X \in A, Y \in A\} + \mathbb{P}\{X \in A, Y \notin A\} - \mathbb{P}\{Y \in A\} \\ &\leq \mathbb{P}\{X \in A, Y \in A\} + \mathbb{P}\{X \neq Y\} - \mathbb{P}\{Y \in A\} \\ &\leq \mathbb{P}\{X \neq Y\}, \end{aligned} \quad (3.1.8)$$

ce qui conclut la démonstration. \square

Théorème 3.1.4. *On a*

$$\sum_{k=0}^{\infty} |b(k; n, q) - \pi_{nq}(k)| \leq nq^2. \quad (3.1.9)$$

DÉMONSTRATION. La démonstration que nous allons donner est une petite merveille de la théorie des probabilités. Elle n’utilise pratiquement pas d’analyse, mais un grand nombre de concepts de probabilités discrètes vus au chapitre 1.

Nous commençons par introduire des espaces probabilisés (Ω_i, p_i) , pour $i = 1, \dots, n$, donnés par $\Omega_i = \{-1, 0, 1, 2, \dots\}$ et

$$p_i(k) = \begin{cases} e^{-q} - (1 - q) & \text{si } k = -1, \\ 1 - q & \text{si } k = 0, \\ e^{-q} \frac{q^k}{k!} & \text{si } k \geq 1. \end{cases} \quad (3.1.10)$$

On vérifiera que les p_i définissent bien une distribution de probabilité. Sur chaque Ω_i , nous introduisons les deux variables aléatoires

$$X_i(\omega_i) = \begin{cases} 0 & \text{si } \omega_i = 0, \\ 1 & \text{sinon,} \end{cases} \quad Y_i(\omega_i) = \begin{cases} \omega_i & \text{si } \omega_i \geq 1, \\ 0 & \text{sinon.} \end{cases} \quad (3.1.11)$$

De cette manière, on a $\mathbb{P}\{X_i = 0\} = 1 - q$, $\mathbb{P}\{X_i = 1\} = q$, et $\mathbb{P}\{Y_i = k\} = \pi_q(k)$ pour tout $k \geq 0$. De plus,

$$\begin{aligned} \mathbb{P}\{X_i = Y_i\} &= \mathbb{P}\{X_i = 0, Y_i = 0\} + \mathbb{P}\{X_i = 1, Y_i = 1\} \\ &= p_i(0) + p_i(1) = 1 - q + qe^{-q}, \end{aligned} \quad (3.1.12)$$

donc

$$\mathbb{P}\{X_i \neq Y_i\} = q(1 - e^{-q}) \leq q^2. \quad (3.1.13)$$

Soit (Ω, p) l'espace produit des (Ω_i, p_i) . Alors

- $X = X_1 + \dots + X_n$ suit la loi binomiale $\mathbb{P}\{X = k\} = b(k; n, q)$;
- $Y = Y_1 + \dots + Y_n$ suit la loi de Poisson $\mathbb{P}\{Y = k\} = \pi_{nq}(k)$, en vertu du Corollaire 1.2.26.

Comme $X \neq Y$ implique que $X_i \neq Y_i$ pour un i au moins, il suit de (3.1.13) que

$$\mathbb{P}\{X \neq Y\} \leq \sum_{i=1}^n \mathbb{P}\{X_i \neq Y_i\} \leq nq^2. \quad (3.1.14)$$

Le résultat suit alors du Lemme 3.1.3. \square

Le corollaire suivant montre l'utilité de la relation (3.1.9).

Corollaire 3.1.5. *Soit $f : \mathbb{N} \rightarrow \mathbb{R}$ une fonction telle que $|f(k)| \leq M$ pour tout k . Soit X une variable aléatoire de loi binomiale $b(k; n, q)$ et Y une variable aléatoire de loi de Poisson $\pi_{nq}(k)$. Alors*

$$|\mathbb{E}(f(X)) - \mathbb{E}(f(Y))| \leq Mnq^2. \quad (3.1.15)$$

En particulier, pour tout $A \subset \mathbb{N}$, on a

$$|\mathbb{P}\{X \in A\} - \mathbb{P}\{Y \in A\}| \leq nq^2. \quad (3.1.16)$$

DÉMONSTRATION.

$$|\mathbb{E}(f(X)) - \mathbb{E}(f(Y))| \leq \sum_{k=0}^{\infty} |f(k)| |b(k; n, q) - \pi_{nq}(k)| \leq Mnq^2. \quad (3.1.17)$$

La relation (3.1.16) est un cas particulier de (3.1.15), avec $f(X) = 1_A(X)$ égale à la fonction indicatrice de $X \in A$. En effet, on a $\mathbb{E}(1_A(X)) = \sum_{k \in A} \mathbb{P}\{X = k\} = \mathbb{P}\{X \in A\}$ et $|1_A(k)| \leq 1$ pour tout $k \in A$. \square

Exemple 3.1.6. Revenons à l'exemple 3.1.2. Avec $n = 2000$ et $q = 1/1000$, la relation (3.1.16) devient

$$|\mathbb{P}\{X \in A\} - \mathbb{P}\{Y \in A\}| \leq 0.002. \quad (3.1.18)$$

En particulier, la différence entre $b(k; 2000, 1/1000)$ et $\pi_2(k)$ est au plus de 0.002. En fait, elle beaucoup plus petite, puisque c'est la somme de ces différences qui est inférieure à 0.002.

3.2 La loi des grands nombres

Dans cette section, nous allons donner une interprétation de l'espérance d'une variable aléatoire. Supposons que X_1, \dots, X_n sont des variables aléatoires obtenues en répétant n fois la même expérience. Si $\mathbb{E}(X_i) = \mu$ pour chaque i , nous savons par la Proposition 1.2.9 que l'espérance de la moyenne $(X_1 + \dots + X_n)/n$ vaut également μ . Mais pouvons-nous affirmer plus, à savoir que la moyenne des X_i a une grande probabilité d'être proche de μ ? C'est ce que fait la loi des grands nombres.

3.2.1 Loi faible des grands nombres

Nous allons montrer ici la version dite faible de la loi des grands nombres. Pour faire cela, nous devons d'abord prouver l'inégalité de Markov, ainsi que son corollaire, l'inégalité de Bienaymé–Chebychev.

Lemme 3.2.1 (Inégalité de Markov). *Soit X une variable aléatoire dont l'espérance existe. Pour tout $a > 0$,*

$$\mathbb{P}\{|X| \geq a\} \leq \frac{\mathbb{E}(|X|)}{a}. \quad (3.2.1)$$

DÉMONSTRATION. On a

$$\mathbb{P}\{|X| \geq a\} = \mathbb{E}(1_{\{|X| \geq a\}}) \leq \mathbb{E}\left(\frac{|X|}{a} 1_{\{|X| \geq a\}}\right) \leq \frac{1}{a} \mathbb{E}(|X|). \quad (3.2.2)$$

□

Corollaire 3.2.2 (Inégalité de Bienaymé–Chebychev). *Soit X une variable aléatoire dont l'espérance et la variance existent. Alors, pour tout $a > 0$,*

$$\mathbb{P}\{|X - \mathbb{E}(X)| \geq a\} \leq \frac{1}{a^2} \text{Var}(X). \quad (3.2.3)$$

DÉMONSTRATION. Soit $Y = [X - \mathbb{E}(X)]^2$. Il suffit alors d'écrire

$$\mathbb{P}\{|X - \mathbb{E}(X)| \geq a\} = \mathbb{P}\{Y \geq a^2\} \leq \frac{1}{a^2} \mathbb{E}(Y) = \frac{1}{a^2} \text{Var}(X). \quad (3.2.4)$$

□

Nous remarquons que la probabilité $\mathbb{P}\{|X - \mathbb{E}(X)| \geq a\}$ devient faible dès que a^2 est sensiblement plus grand que la variance de X , donc dès que a est sensiblement plus grand que l'écart-type $\sigma(X) = \sqrt{\text{Var}(X)}$. On dit que X est “concentrée dans un intervalle d'ordre de grandeur $\sigma(X)$ autour de son espérance”.

Théorème 3.2.3 (Loi faible des grands nombres). *Pour tout entier positif n , on se donne des variables aléatoires X_1, \dots, X_n , non corrélées, pouvant dépendre de n . On suppose que chaque X_i a la même espérance μ et la même variance σ^2 . Soit*

$$S_n = \sum_{i=1}^n X_i. \quad (3.2.5)$$

Alors, S_n/n converge vers μ en probabilité, c'est-à-dire que pour tout $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left\{\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right\} = 0. \quad (3.2.6)$$

DÉMONSTRATION. Par l'inégalité de Bienaymé–Chebychev, on a

$$\mathbb{P}\left\{\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right\} \leq \frac{1}{\varepsilon^2} \operatorname{Var}\left(\frac{S_n}{n}\right) = \frac{1}{\varepsilon^2 n^2} \operatorname{Var}(S_n) = \frac{1}{\varepsilon^2 n^2} n\sigma^2 = \frac{\sigma^2}{n\varepsilon^2}, \quad (3.2.7)$$

qui tend vers 0 lorsque $n \rightarrow \infty$. \square

Remarque 3.2.4. Il existe également une loi forte des grands nombres. Celle-ci affirme que sous certaines conditions, on a

$$\mathbb{P}\left\{\lim_{n \rightarrow \infty} \frac{S_n}{n} = \mu\right\} = 1. \quad (3.2.8)$$

À ce stade, cependant, nous ne pouvons pas donner de sens mathématique précis à cette équation, car nous n'avons pas défini d'espace probabilisé commun à tous les S_n .

Exemple 3.2.5. Soit $S_n = \sum_{i=1}^n 1_S(\omega_i)$ le nombre de succès dans une expérience de Bernoulli de longueur n , et probabilité de succès q . La loi faible des grands nombres affirme que

$$\lim_{n \rightarrow \infty} \mathbb{P}\left\{\left|\frac{S_n}{n} - q\right| \geq \varepsilon\right\} = \lim_{n \rightarrow \infty} \sum_{k: |k-nq| \geq n\varepsilon} b(k; n, q) = 0 \quad (3.2.9)$$

pour tout $\varepsilon > 0$. Cela signifie que $S_n \in [nq - n\varepsilon, nq + n\varepsilon]$ avec grande probabilité pour n suffisamment grand (mais pas que $S_n = nq$ avec grande probabilité).

Dans le cas d'une pièce de monnaie équilibrée, S_n désigne le nombre de Pile parmi n jets, et l'on prendra $q = 1/2$. Alors S_n/n sera proche de $1/2$ pour n grand. En fait, l'inégalité de Bienaymé–Chebychev nous donne

$$\mathbb{P}\left\{\left|\frac{S_n}{n} - \frac{1}{2}\right| \geq \varepsilon\right\} \leq \frac{1}{\varepsilon^2} \operatorname{Var}\left(\frac{S_n}{n}\right) = \frac{1}{4n\varepsilon^2}. \quad (3.2.10)$$

Par exemple, si on jette 1000 fois la pièce, et $\varepsilon = 0.1$,

$$\mathbb{P}\left\{\left|\frac{S_{1000}}{1000} - \frac{1}{2}\right| \geq \varepsilon\right\} = \mathbb{P}\{S_{1000} \notin [400, 600]\} \leq \frac{1}{40} = 0.025. \quad (3.2.11)$$

3.2.2 Grandes déviations*

En fait, la majoration (3.2.10) est trop pessimiste, et la probabilité en question est beaucoup plus faible. On peut améliorer l'inégalité de Bienaymé–Chebychev de la manière suivante. Pour tout $\lambda > 0$, on a

$$\mathbb{P}\left\{S_n - \frac{n}{2} \geq n\varepsilon\right\} = \mathbb{P}\left\{e^{\lambda(S_n - n/2)} \geq e^{\lambda n\varepsilon}\right\} \leq \frac{1}{e^{\lambda n\varepsilon}} \mathbb{E}(e^{\lambda(S_n - n/2)}). \quad (3.2.12)$$

Posons $X_i = 1_S(\omega_i)$ et $Z_i = e^{\lambda(X_i - 1/2)}$. On a donc

$$\mathbb{E}(Z_i) = e^{-\lambda/2} \mathbb{P}\{X_i = 0\} + e^{\lambda/2} \mathbb{P}\{X_i = 1\} = \cosh(\lambda/2). \quad (3.2.13)$$

Comme $S_n = \sum_{i=1}^n X_i$, on a $e^{\lambda(S_n - n/2)} = \prod_{i=1}^n Z_i$, et l'indépendance des Z_i implique

$$\mathbb{E}(e^{\lambda(S_n - n/2)}) = \mathbb{E}\left(\prod_{i=1}^n Z_i\right) = \prod_{i=1}^n \mathbb{E}(Z_i) = [\cosh(\lambda/2)]^n. \quad (3.2.14)$$

Nous pouvons écrire cette dernière expression comme $\exp[n \log \cosh(\lambda/2)]$. L'inégalité (3.2.12) devient donc

$$\mathbb{P}\left\{S_n - \frac{n}{2} \geq n\varepsilon\right\} \leq \exp\left\{-n[\lambda\varepsilon - \log \cosh(\lambda/2)]\right\}. \quad (3.2.15)$$

Considérons la fonction $f(\lambda) = \lambda\varepsilon - \log \cosh(\lambda/2)$. Un calcul montre qu'elle prend sa valeur minimale lorsque $\tanh(\lambda/2) = 2\varepsilon$, donc pour $\lambda = \log[(1 + 2\varepsilon)/(1 - 2\varepsilon)]$. Posons alors

$$\begin{aligned} I(\varepsilon) &= f\left(\log \frac{1 + 2\varepsilon}{1 - 2\varepsilon}\right) \\ &= \varepsilon \log \frac{1 + 2\varepsilon}{1 - 2\varepsilon} - \log[\sqrt{(1 - 2\varepsilon)(1 + 2\varepsilon)}] \\ &= \left(\frac{1}{2} + \varepsilon\right) \log(1 + 2\varepsilon) + \left(\frac{1}{2} - \varepsilon\right) \log(1 - 2\varepsilon). \end{aligned} \quad (3.2.16)$$

En utilisant (3.2.15) pour cette valeur de λ , et la même estimation pour $\mathbb{P}\{S_n - \frac{n}{2} \leq -n\varepsilon\}$, nous obtenons

$$\mathbb{P}\left\{\left|\frac{S_n}{n} - \frac{1}{2}\right| \geq \varepsilon\right\} \leq 2e^{-nI(\varepsilon)}. \quad (3.2.17)$$

Ce résultat s'appelle une *estimation de grande déviation*, et $I(\varepsilon)$ s'appelle la *fonction taux*. Comme $I(0.1) \simeq 0.02$, nous avons

$$\mathbb{P}\{S_{1000} \notin [400, 600]\} \leq 2e^{-1000I(0.1)} \simeq 3.6 \cdot 10^{-9}. \quad (3.2.18)$$

3.3 Le théorème de la limite centrale

Dans cette section, nous allons donner une interprétation plus précise de la variance d'une variable aléatoire. Examinons à nouveau la somme $S_n = X_1 + \dots + X_n$ de n variables aléatoires de même espérance μ et même variance σ^2 . Nous supposons les X_i *indépendants*. Nous avons vu dans la section précédente que la moyenne S_n/n a une grande probabilité d'être proche de μ , mais comment se comporte l'écart $S_n/n - \mu$? Dans la preuve du Théorème 3.2.3, nous avons vu que

$$\mathbb{P}\left\{\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right\} \leq \frac{\sigma^2}{n\varepsilon^2}. \quad (3.3.1)$$

Le membre de droite tend vers zéro lorsque $n \rightarrow \infty$, non seulement pour ε constant, mais également si $\varepsilon = \varepsilon(n)$ satisfait $\lim_{n \rightarrow \infty} n\varepsilon(n)^2 = \infty$. Par contre, il reste constant si $\varepsilon(n) = \sigma/\sqrt{n}$. On dit que la moyenne S_n/n "se concentre" dans un intervalle $[\mu - \sigma/\sqrt{n}, \mu + \sigma/\sqrt{n}]$. Pour étudier plus précisément la déviation entre la moyenne et l'espérance, on introduit la variable dite pivotale

$$\widehat{S}_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}. \quad (3.3.2)$$

Remarquons que $\mathbb{E}(\widehat{S}_n) = 0$ et $\text{Var}(\widehat{S}_n) = 1$.

Le *théorème de la limite centrale* affirme que pour tout intervalle $[a, b] \subset \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \mathbb{P}\{a \leq \widehat{S}_n \leq b\} = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \Phi(b) - \Phi(a), \quad (3.3.3)$$

où

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-x^2/2} dx \quad (3.3.4)$$

est la fonction de répartition de la loi normale standard. On dit que \widehat{S}_n converge en loi vers une variable normale standard.

Historiquement, le théorème de la limite centrale a d'abord été obtenu dans le cas particulier de sommes de variables de Bernoulli, c'est-à-dire lorsque S_n suit la loi binomiale. Dans ce cas, on l'appelle aussi la formule de Moivre–Laplace. La démonstration est un calcul relativement direct, qui n'apporte peut-être pas beaucoup à la compréhension du cas général, mais nous le donnerons tout de même dans la section suivante. Nous traiterons ensuite le cas général.

3.3.1 Cas de la loi binomiale : formule de Moivre–Laplace

Rappelons que la loi binomiale a espérance nq et variance $nq(1 - q)$. Le théorème de la limite centrale prend alors la forme suivante :

Théorème 3.3.1 (Moivre–Laplace). *Si S_n suit la loi binomiale de paramètre q , alors $\widehat{S}_n = (S_n - nq)/\sqrt{nq(1 - q)}$ satisfait*

$$\lim_{n \rightarrow \infty} \mathbb{P}\{a \leq \widehat{S}_n \leq b\} = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \quad (3.3.5)$$

pour tout intervalle $[a, b] \subset \mathbb{R}$.

La démonstration par calcul direct est basée sur la *formule de Stirling*:

Lemme 3.3.2 (Formule de Stirling).

$$\lim_{n \rightarrow \infty} \frac{n!}{n^n e^{-n} \sqrt{2\pi n}} = 1. \quad (3.3.6)$$

DÉMONSTRATION. Considérons la *fonction Gamma d'Euler*

$$\Gamma(n) = \int_0^{\infty} t^{n-1} e^{-t} dt. \quad (3.3.7)$$

Il est clair que $\Gamma(1) = 1$, et une intégration par parties montre que

$$\Gamma(n+1) = \int_0^{\infty} t^n e^{-t} dt = \left[-t^n e^{-t}\right]_0^{\infty} + n \int_0^{\infty} t^{n-1} e^{-t} dt = n\Gamma(n), \quad (3.3.8)$$

d'où $\Gamma(n+1) = n!$. Écrivons alors, à l'aide du changement de variables $t = n + \sqrt{n}s$,

$$\begin{aligned} n! = \Gamma(n+1) &= \int_0^{\infty} t^n e^{-t} dt \\ &= \int_{-\sqrt{n}}^{\infty} (n + \sqrt{n}s)^n e^{-n - \sqrt{n}s} \sqrt{n} ds \\ &= n^n e^{-n} \sqrt{n} \int_{-\sqrt{n}}^{\infty} \left(1 + \frac{s}{\sqrt{n}}\right)^n e^{-\sqrt{n}s} ds \\ &= n^n e^{-n} \sqrt{2\pi n} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-h_n(s)} ds, \end{aligned} \quad (3.3.9)$$

où nous avons posé

$$h_n(s) = \begin{cases} \sqrt{n}s - n \log\left(1 + \frac{s}{\sqrt{n}}\right) & \text{pour } s > -\sqrt{n}, \\ +\infty & \text{autrement.} \end{cases} \quad (3.3.10)$$

Un développement limité en $1/\sqrt{n}$ montre que

$$h_n(s) = \frac{s^2}{2} + \mathcal{O}\left(\frac{s^3}{\sqrt{n}}\right) \quad \text{pour } s > -\sqrt{n}, \quad (3.3.11)$$

et donc $h_n(s)$ converge simplement vers $-s^2/2$. D'autre part on vérifie la minoration

$$h_n(s) \geq g(s) = \begin{cases} \frac{s^2}{6} & \text{pour } s \leq 1, \\ \frac{s-1}{2} & \text{pour } s \geq 1, \end{cases} \quad (3.3.12)$$

valable pour tout $n \geq 1$. On vérifie facilement que la fonction $e^{-g(s)}$ est intégrable. Par conséquent, le théorème de la convergence dominée montre que l'on peut échanger limite $n \rightarrow \infty$ et intégrale, pour obtenir

$$\lim_{n \rightarrow \infty} \frac{n!}{n^n e^{-n} \sqrt{2\pi n}} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \lim_{n \rightarrow \infty} e^{-h_n(s)} ds = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-s^2/2} ds = 1. \quad (3.3.13)$$

□

DÉMONSTRATION DU THÉORÈME 3.3.1. Nous allons considérer le cas $q = 1/2$ afin de simplifier les notations, mais le cas d'un q quelconque se montre de manière totalement analogue. Il s'agit de calculer

$$\begin{aligned} \mathbb{P}\left\{a \leq \widehat{S}_n \leq b\right\} &= \mathbb{P}\left\{\frac{n}{2} + \frac{\sqrt{n}}{2}a \leq S_n \leq \frac{n}{2} + \frac{\sqrt{n}}{2}b\right\} \\ &= \sum_{k \in A} b(k; n, 1/2). \end{aligned} \quad (3.3.14)$$

où $A = \{0, \dots, n\} \cap [n/2 + \sqrt{n}a/2, n/2 + \sqrt{n}b/2]$. Dans la suite, nous dirons que deux fonctions $x(k, n)$ et $y(k, n)$ sont équivalentes, et nous noterons $x(k, n) \sim y(k, n)$, si

$$\lim_{n \rightarrow \infty} \sup_{k \in A} \left| \frac{x(k, n)}{y(k, n)} - 1 \right| = 0. \quad (3.3.15)$$

En particulier, nous avons $k \sim n/2$ et $(n-k) \sim n/2$. La formule de Stirling nous donne alors

$$\begin{aligned} b(k; n, 1/2) &= 2^{-n} \frac{n!}{k!(n-k)!} \\ &\sim 2^{-n} \frac{n^n e^{-n} \sqrt{2\pi n}}{k^k e^{-k} \sqrt{2\pi k} (n-k)^{n-k} e^{-(n-k)} \sqrt{2\pi(n-k)}} \\ &= \frac{2^{-n}}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} \frac{n^n}{k^k (n-k)^{n-k}} \\ &= \frac{1}{\sqrt{2\pi n}} \frac{1}{\sqrt{(k/n)(1-k/n)}} \frac{2^{-n}}{(k/n)^k (1-k/n)^{n-k}}. \end{aligned} \quad (3.3.16)$$

Écrivons k sous la forme $k = n/2 + x\sqrt{n}/2$, avec $x \in [a, b]$. On aura donc

$$\frac{k}{n} = \frac{1}{2} \left(1 + \frac{x}{\sqrt{n}} \right), \quad 1 - \frac{k}{n} = \frac{1}{2} \left(1 - \frac{x}{\sqrt{n}} \right), \quad (3.3.17)$$

et par conséquent

$$b(k; n, 1/2) \sim \frac{1}{\sqrt{2\pi n}} \frac{2}{\sqrt{1 - x^2/n}} \psi(x, n), \quad (3.3.18)$$

où nous avons introduit

$$\psi(x, n) = \left(1 + \frac{x}{\sqrt{n}} \right)^{-(1+x/\sqrt{n})n/2} \left(1 - \frac{x}{\sqrt{n}} \right)^{-(1-x/\sqrt{n})n/2}. \quad (3.3.19)$$

Observons que le logarithme de $\psi(x, n)$ s'écrit

$$\log \psi(x, n) = -\frac{n}{2} \left[\left(1 + \frac{x}{\sqrt{n}} \right) \log \left(1 + \frac{x}{\sqrt{n}} \right) + \left(1 - \frac{x}{\sqrt{n}} \right) \log \left(1 - \frac{x}{\sqrt{n}} \right) \right], \quad (3.3.20)$$

et qu'un développement limité en $1/\sqrt{n}$ donne

$$\log \psi(x, n) = -\frac{x^2}{2} + \mathcal{O}\left(\frac{x^3}{\sqrt{n}}\right). \quad (3.3.21)$$

Il suit donc de (3.3.18) que

$$b(k; n, 1/2) \sim \frac{2}{\sqrt{2\pi n}} e^{-x^2/2} \left[1 + \mathcal{O}\left(\frac{x}{\sqrt{n}}\right) \right] \sim \frac{2}{\sqrt{2\pi n}} e^{-x^2/2}, \quad (3.3.22)$$

puisque $x \in [a, b]$. En remplaçant dans (3.3.14), il vient

$$\mathbb{P}\left\{ a \leq \widehat{S}_n \leq b \right\} \sim \frac{1}{\sqrt{2\pi}} \sum_{x \in [a, b]: (n+\sqrt{n}x)/2 \in \{0, \dots, n\}} \frac{2}{\sqrt{n}} e^{-x^2/2}. \quad (3.3.23)$$

La somme s'effectue sur des x régulièrement espacés de $2/\sqrt{n}$. C'est donc une somme de Riemann, qui converge vers $\int_a^b e^{-x^2/2} dx$. \square

3.3.2 Cas général

Nous donnons maintenant une démonstration du théorème de la limite centrale dans le cas général, basée sur les fonctions caractéristiques. Celle-ci s'appuie sur le théorème de continuité de Lévy, que nous admettrons :

Théorème 3.3.3 (Théorème de continuité de Lévy). *Soit X_1, X_2, \dots une suite de variables aléatoires, de fonctions caractéristiques respectives $\phi_1(z), \phi_2(z), \dots$. Supposons que la suite des $\phi_n(z)$ converge simplement vers une fonction continue $\phi(z)$. Alors $\phi(z)$ est la fonction caractéristique d'une variable aléatoire X , et X_n converge en loi vers X , c'est-à-dire*

$$\lim_{n \rightarrow \infty} \mathbb{P}\left\{ a \leq X_n \leq b \right\} = \mathbb{P}\left\{ a \leq X \leq b \right\}, \quad (3.3.24)$$

pour tout intervalle $[a, b] \subset \mathbb{R}$.

Théorème 3.3.4 (Théorème de la limite centrale). *Soit X_1, X_2, \dots une suite de variables aléatoires indépendantes, identiquement distribuées (abrégé i.i.d.), d'espérance μ et de variance σ^2 . Soit $S_n = \sum_{i=1}^n X_i$. Alors la variable pivotale $\widehat{S}_n = (S_n - n\mu)/\sqrt{n\sigma^2}$ converge en loi vers une variable normale standard lorsque $n \rightarrow \infty$, c'est-à-dire*

$$\lim_{n \rightarrow \infty} \mathbb{P}\left\{a \leq \widehat{S}_n \leq b\right\} = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \Phi(b) - \Phi(a). \quad (3.3.25)$$

DÉMONSTRATION. Nous commençons par introduire des variables centrées et réduites

$$Y_i = \frac{X_i - \mu}{\sigma}. \quad (3.3.26)$$

Les Y_i sont également i.i.d. avec, par linéarité et bilinéarité de l'espérance et de la variance,

$$\mathbb{E}(Y_i) = 0, \quad \text{Var}(Y_i) = 1, \quad \mathbb{E}(Y_i^2) = 1. \quad (3.3.27)$$

De plus, nous avons

$$S_n = \sum_{i=1}^n (\mu + \sigma Y_i) = n\mu + \sigma \sum_{i=1}^n Y_i \quad \Rightarrow \quad \widehat{S}_n = \frac{S_n - n\mu}{\sqrt{n\sigma^2}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i. \quad (3.3.28)$$

Nous allons montrer que la fonction caractéristique de \widehat{S}_n converge, lorsque $n \rightarrow \infty$, vers la fonction caractéristique de la loi normale centrée réduite, à savoir $e^{-z^2/2}$:

$$\lim_{n \rightarrow \infty} \mathbb{E}(e^{iz\widehat{S}_n}) = e^{-z^2/2}. \quad (3.3.29)$$

Le Théorème 3.3.3 implique alors (3.3.25).

Les Y_i étant i.i.d., ils ont tous la même fonction caractéristique. Notons-la $\phi(z)$. Il suit de (3.3.27), des propriétés de base des fonctions caractéristiques et de la formule de Taylor que

$$\phi(z) = 1 + iz\mathbb{E}(X) - z^2\mathbb{E}(X^2) + r(z) = 1 - \frac{1}{2}z^2 + r(z), \quad (3.3.30)$$

où $r(z)$ est un reste satisfaisant

$$\lim_{z \rightarrow 0} \frac{r(z)}{z^2} = 0. \quad (3.3.31)$$

De plus, pour deux variables aléatoires indépendantes Z_1 et Z_2 , on a

$$\phi_{Z_1+Z_2}(z) = \mathbb{E}(e^{iz(Z_1+Z_2)}) = \mathbb{E}(e^{izZ_1})\mathbb{E}(e^{izZ_2}) = \phi_{Z_1}(z)\phi_{Z_2}(z). \quad (3.3.32)$$

Par conséquent, la fonction caractéristique de $\sqrt{n}\widehat{S}_n = \sum_{i=1}^n Y_i$ vaut

$$\phi_{\sqrt{n}\widehat{S}_n}(z) = \phi(z)^n, \quad (3.3.33)$$

et donc celle de \widehat{S}_n vaut

$$\phi_{\widehat{S}_n}(z) = \mathbb{E}(e^{iz\widehat{S}_n}) = \mathbb{E}(e^{i(z/\sqrt{n})\sqrt{n}\widehat{S}_n}) = \phi_{\sqrt{n}\widehat{S}_n}\left(\frac{z}{\sqrt{n}}\right) = \phi\left(\frac{z}{\sqrt{n}}\right)^n. \quad (3.3.34)$$

En effectuant un développement limité à partir de (3.3.30), on obtient

$$\log \phi\left(\frac{z}{\sqrt{n}}\right) = -\frac{z^2}{2n} + r_1\left(\frac{z}{\sqrt{n}}\right), \quad (3.3.35)$$

avec un reste satisfaisant

$$\lim_{z \rightarrow 0} \frac{r_1(z)}{z^2} = 0. \quad (3.3.36)$$

Il suit de (3.3.34) que

$$\log \phi_{\hat{S}_n}(z) = n \log \phi\left(\frac{z}{\sqrt{n}}\right) = -\frac{z^2}{2} + nr_1\left(\frac{z}{\sqrt{n}}\right). \quad (3.3.37)$$

Comme

$$\lim_{n \rightarrow \infty} nr_1\left(\frac{z}{\sqrt{n}}\right) = \lim_{n \rightarrow \infty} z^2 \frac{1}{(z/\sqrt{n})^2} r_1\left(\frac{z}{\sqrt{n}}\right) = 0, \quad (3.3.38)$$

on obtient

$$\lim_{n \rightarrow \infty} \log \phi_{\hat{S}_n}(z) = -\frac{z^2}{2}, \quad (3.3.39)$$

ce qui implique (3.3.29). \square

Le Théorème 3.3.4 montre que pour n grand, on a approximativement

$$\mathbb{P}\left\{\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \in [a, b]\right\} \simeq \Phi(b) - \Phi(a), \quad (3.3.40)$$

ou encore

$$\mathbb{P}\{S_n \in [\hat{a}, \hat{b}]\} \simeq \Phi\left(\frac{\hat{b} - n\mu}{\sqrt{n\sigma^2}}\right) - \Phi\left(\frac{\hat{a} - n\mu}{\sqrt{n\sigma^2}}\right). \quad (3.3.41)$$

Exemple 3.3.5. Une chaîne de montage produit des pièces défectueuses avec une probabilité de 10%. Quelle est la probabilité d'obtenir plus de 50 pièces défectueuses parmi 400?

Nous modélisons la situation par une expérience de Bernoulli de paramètre $q = 0.1$. Avec $n = 400$, $n\mu = nq = 40$ et $n\sigma^2 = nq(1 - q) = 36$, on obtient

$$\mathbb{P}\{S_n > 50\} = \mathbb{P}\{S_n \in [50, \infty[] \simeq \Phi(\infty) - \Phi\left(\frac{50 - 40}{\sqrt{36}}\right) = 1 - \Phi(1.66\dots) \simeq 0.0485. \quad (3.3.42)$$

Il y a donc un peu moins de 5% de chances d'obtenir plus de 50 pièces défectueuses.

Chapitre 4

Introduction à la statistique

Nous considérons dans ce dernier chapitre les bases de la statistique dite *inférentielle*, dont le but est de déduire des propriétés du système étudié d'observations incomplètes, par opposition à la statistique *descriptive*, dont le but est de résumer des données observées par des méthodes analytiques et graphiques. Voici deux exemples typiques de problèmes de statistique inférentielle :

Exemple 4.0.1. On effectue un sondage dans une population de N individus, afin de déterminer leur opinion sur une question politique. Admettons que chaque individu ait soit l'opinion A, soit l'opinion B. On demande leur opinion à n individus, avec $1 \ll n \ll N$. Quelle est la proportion d'individus de la population totale ayant l'opinion A? On ne peut répondre avec certitude à cette question, mais on peut estimer cette proportion. C'est un *problème d'estimation*. Plus précisément, on cherche un *intervalle de confiance*, c'est-à-dire un intervalle dans lequel la proportion inconnue se trouve avec une probabilité prescrite, par exemple 95%.

Exemple 4.0.2. On voudrait comparer l'efficacité de deux médicaments, par exemple des somnifères. On modélise la durée de sommeil procurée par chaque somnifère par deux variables aléatoires X et Y , de loi inconnue. On administre le premier somnifère à un groupe de n personnes, le second à un groupe de m personnes, et on mesure les durées de sommeil (c'est-à-dire qu'on observe n variables aléatoires de même loi que X , et m variables de même loi que Y). Au vu des résultats, peut-on penser que le premier somnifère est plus efficace que le second? C'est un problème de test d'hypothèse.

4.1 Estimateurs

On suppose que le résultat d'une expérience est donné par une variable aléatoire X , dont la loi n'est pas connue, mais supposée appartenir à une certaine famille $\{\mathbb{P}_\theta : \theta \in \Theta\}$. Le but est d'estimer le paramètre inconnu θ , ou plus généralement une grandeur $\lambda = g(\theta)$.

Exemple 4.1.1. L'expérience consiste à jeter une pièce de monnaie, dont on ne sait pas si elle est équilibrée. Soit $X = 1_{\text{Pile}}$ la variable aléatoire valant 1 si la pièce tombe sur Pile, 0 sinon. On peut supposer que X suit une loi de Bernoulli, de paramètre inconnu $\theta \in \Theta = [0, 1]$ (c'est-à-dire que θ est la probabilité que la pièce tombe sur Pile).

Pour estimer θ , on lance la pièce un grand nombre n de fois, c'est-à-dire que l'on génère des variables aléatoires X_1, \dots, X_n , indépendantes, et telles que $\mathbb{P}\{X_i = 1\} = \theta = 1 - \mathbb{P}\{X_i = 0\}$ pour tout i . Le but est d'estimer θ à partir de X_1, \dots, X_n . En d'autres

termes, on cherche une fonction $f : \mathbb{R}^n \rightarrow \Theta$ telle que $\hat{\theta} = f(X_1, \dots, X_n)$ donne une estimation du paramètre θ .

Définition 4.1.2. Soient X_1, \dots, X_n des variables aléatoires i.i.d., de loi \mathbb{P}_θ .

1. On dit que $\hat{\lambda} = f(X_1, \dots, X_n)$ est un estimateur sans biais de $\lambda = g(\theta)$ si on a

$$\mathbb{E}_\theta[\hat{\lambda}] = \lambda \quad (4.1.1)$$

pour tout $\theta \in \Theta$.

2. Le risque quadratique d'un estimateur $\hat{\lambda}$ de $g(\theta)$ est la quantité

$$R_\theta(\hat{\lambda}) = \mathbb{E}_\theta[(\hat{\lambda} - g(\theta))^2]. \quad (4.1.2)$$

Si $\hat{\lambda}$ est sans biais, son risque quadratique est donc égal à sa variance.

3. On dit qu'une suite $(\hat{\lambda}_n)$ d'estimateurs est consistante si elle converge vers $\lambda = g(\theta)$ en probabilité, c'est-à-dire si

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta\{|\hat{\lambda} - g(\theta)| > \varepsilon\} = 0 \quad (4.1.3)$$

pour tout $\varepsilon > 0$ et tout $\theta \in \Theta$.

4.1.1 Estimateurs empiriques

Supposons que le paramètre à estimer θ soit l'espérance de X : $\mathbb{E}_\theta(X) = \theta$. C'est le cas dans l'exemple 4.1.1.

Définition 4.1.3. On appelle moyenne empirique l'estimateur

$$\hat{\lambda} = \bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n). \quad (4.1.4)$$

Proposition 4.1.4. Supposons X de variance finie. Alors la moyenne empirique est un estimateur sans biais et consistant de l'espérance.

DÉMONSTRATION.

- Sans biais : $\mathbb{E}_\theta(\bar{X}_n) = \frac{1}{n}(\mathbb{E}_\theta(X_1) + \dots + \mathbb{E}_\theta(X_n)) = \frac{1}{n}n\mathbb{E}_\theta(X) = \theta$.
- Consistant : C'est exactement ce qu'affirme la loi des grands nombres.

Remarquons que le risque quadratique vaut $R_\theta(\bar{X}_n) = \frac{1}{n^2}n \text{Var}_\theta(X) = \text{Var}_\theta(X)/n$, et tend donc vers 0 lorsque n tend vers l'infini. \square

Dans l'exemple 4.1.1 de la pièce de monnaie, soit p le nombre de Pile obtenu en n jets. La moyenne empirique est simplement la proportion p/n de Pile parmi n jets, ce qui est assez naturel.

On peut construire un estimateur similaire pour la variance :

Définition 4.1.5. On appelle variance empirique l'estimateur

$$\hat{\lambda} = S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2. \quad (4.1.5)$$

Proposition 4.1.6. Supposons $\mathbb{E}_\theta(X^4) < \infty$ pour tout θ . Alors la variance empirique est un estimateur sans biais et consistant de la variance de X .

DÉMONSTRATION. Notons $\mu = \mathbb{E}_\theta(X)$, $\sigma^2 = \text{Var}_\theta(X)$ et

$$Q = \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \sum_{i=1}^n X_i^2 - n\bar{X}_n^2$$

(la dernière égalité étant obtenue en développant le carré). On a $\mathbb{E}_\theta(X_i^2) = \sigma^2 + \mu^2$ et

$$\begin{aligned} \mathbb{E}_\theta(\bar{X}_n^2) &= \frac{1}{n^2} \sum_{i,j=1}^n \mathbb{E}_\theta(X_i X_j) \\ &= \frac{1}{n^2} \left[\sum_{i \neq j} \mathbb{E}_\theta(X_i X_j) + \sum_{i=1}^n \mathbb{E}_\theta(X_i^2) \right] \\ &= \frac{1}{n^2} \left[n(n-1)\mu^2 + n(\sigma^2 + \mu^2) \right] \\ &= \mu^2 + \frac{\sigma^2}{n} \end{aligned}$$

d'où

$$\mathbb{E}_\theta(Q) = n(\sigma^2 + \mu^2) - n\mu^2 - \sigma^2 = (n-1)\sigma^2,$$

ce qui montre que l'estimateur S_n^2 est sans biais.

Pour montrer la consistance, on écrit

$$S_n^2 = \frac{n}{n-1} \left[\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2 \right].$$

Le premier terme entre crochets tend en probabilité vers $\mathbb{E}(X^2)$ par la loi des grands nombres. Le second tend vers $\mathbb{E}(X)^2$ en vertu d'une propriété affirmant que si Y_n tend en probabilité vers une constante a et que g est une fonction continue, alors $g(Y_n)$ tend en probabilité vers $g(a)$. \square

Dans l'exemple 4.1.1 de la pièce, on aura

$$S_n^2 = \frac{1}{n-1} \left[p \left(1 - \frac{p}{n}\right)^2 + (n-p) \left(\frac{p}{n}\right)^2 \right] = \frac{p(n-p)}{n(n-1)}. \quad (4.1.6)$$

4.1.2 Estimateur de maximum de vraisemblance

Une manière plus générale d'obtenir un estimateur consiste à maximiser la vraisemblance. C'est-à-dire qu'étant donné les valeurs observées de X_1, \dots, X_n , on cherche la valeur θ telle que la probabilité d'observer ces valeurs sous la loi \mathbb{P}_θ soit maximale.

Définition 4.1.7. Soient X_1, \dots, X_n des variables aléatoires i.i.d. de loi \mathbb{P}_θ et notons $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. On appelle vraisemblance la fonction $L : \mathbb{R}^n \times \Theta \rightarrow \mathbb{R}_+$ définie comme suit :

1. Pour une loi discrète

$$L_x(\theta) = \prod_{i=1}^n \mathbb{P}_\theta(X_i = x_i). \quad (4.1.7)$$

2. Pour une loi à densité f_θ

$$L_x(\theta) = \prod_{i=1}^n f_\theta(x_i) . \quad (4.1.8)$$

L'estimateur du maximum de vraisemblance est

$$\hat{\theta} = \arg \max_{\theta \in \Theta} L_x(\theta) . \quad (4.1.9)$$

La notation $\arg \max$ signifie que $\hat{\theta}$ est la valeur de θ pour laquelle $L_x(\theta)$ est maximale. En pratique, il est souvent plus facile de calculer la *log-vraisemblance* $\log(L_x(\theta))$, puis la valeur de θ qui annule sa dérivée.

Exemple 4.1.8. Si X suit une loi de Bernoulli de paramètre θ , on peut écrire

$$L_x(\theta) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} = \theta^{s_n} (1 - \theta)^{n-s_n} , \quad (4.1.10)$$

où $s_n = \sum_{i=1}^n x_i$. On a donc

$$\log L_x(\theta) = s_n \log \theta + (n - s_n) \log(1 - \theta) , \quad (4.1.11)$$

et

$$\frac{d}{d\theta} \log L_x(\theta) = \frac{s_n}{\theta} - \frac{n - s_n}{1 - \theta} , \quad (4.1.12)$$

qui s'annule pour

$$\theta = \hat{\theta} = \frac{S_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i . \quad (4.1.13)$$

On retrouve donc comme estimateur la moyenne empirique.

Les estimateurs de maximum de vraisemblance sont en général aisés à déterminer, mais peuvent être biaisés.

Exemple 4.1.9. Supposons que X_1, \dots, X_n suivent une loi normale d'espérance μ et de variance σ^2 inconnues. On a

$$L_x(\mu, \sigma) = \prod_{i=1}^n \frac{e^{-(x_i - \mu)^2 / 2\sigma^2}}{\sqrt{2\pi} \sigma} = \frac{e^{-\sum_{i=1}^n (x_i - \mu)^2 / 2\sigma^2}}{(2\pi)^{n/2} \sigma^n} , \quad (4.1.14)$$

donc

$$\log L_x(\mu, \sigma) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 - n \log \sigma - \frac{n}{2} \log(2\pi) , \quad (4.1.15)$$

et

$$\begin{aligned} \frac{\partial}{\partial \mu} \log L_x(\mu, \sigma) &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \\ \frac{\partial}{\partial \sigma} \log L_x(\mu, \sigma) &= \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - \mu)^2 - \frac{n}{\sigma} . \end{aligned} \quad (4.1.16)$$

On en déduit les estimateurs

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n , \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 . \quad (4.1.17)$$

L'estimateur $\hat{\mu}$ est non biaisé, mais la proposition 4.1.6 montre que $\mathbb{E}_{\mu, \sigma^2}[\hat{\sigma}^2] = \frac{n-1}{n} \sigma^2$.

4.2 Intervalles de confiance

Pour être plus précis, on cherche souvent à donner, plutôt qu'une valeur ponctuelle de λ , un intervalle dans lequel λ se trouve avec grande probabilité. On se donne donc un *coefficient de sécurité* $\beta \in (0, 1)$, proche de 1. Un *intervalle de confiance* est un intervalle I , dépendant des observations X_1, \dots, X_n , tel que

$$\mathbb{P}\{\lambda \in I\} \geq \beta. \quad (4.2.1)$$

On cherche donc deux fonctions $f_1, f_2 : \mathbb{R}^n \rightarrow \mathbb{R}$ telles que

$$\mathbb{P}_\theta\{f_1(X_1, \dots, X_n) \leq \lambda \leq f_2(X_1, \dots, X_n)\} \geq \beta. \quad (4.2.2)$$

Ces fonctions ne sont en général pas uniques, mais parfois des raisons de symétrie rendent un choix plus naturel que les autres.

4.2.1 Intervalle de confiance pour une proportion

Considérons le cas d'une population de N individus, divisés en deux catégories A et B. Le but est d'estimer la proportion p d'individus de catégorie A, en ne testant qu'un nombre n d'individus, $1 \ll n \ll N$.

Admettons que l'on effectue n tirages indépendants avec remise. Soit X_i la variable aléatoire valant 1 si le i^{e} individu est de catégorie A, 0 sinon. Chaque X_i suit une loi de Bernoulli de paramètre p , on a donc $\mu = \mathbb{E}(X_i) = p$ et $\sigma^2 = \text{Var}(X_i) = p(1-p)$ pour tout i .

Nous savons qu'un estimateur de p est la moyenne empirique

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i. \quad (4.2.3)$$

Le théorème de la limite centrale montre par ailleurs que la variable dite *pivotale*

$$\frac{n\bar{X}_n - n\mu}{\sqrt{n}\sigma} = \sqrt{n} \frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \quad (4.2.4)$$

suit une loi proche d'une loi normale standard pour n grand. Comme par la loi des grands nombres, p est proche de \bar{X}_n pour n grand, on peut en déduire que la variable aléatoire

$$T = \sqrt{n} \frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1 - \bar{X}_n)}} \quad (4.2.5)$$

suit également une loi proche d'une loi normale standard pour n grand. Notons alors c_β le nombre tel que

$$\Phi(c_\beta) - \Phi(-c_\beta) = \beta. \quad (4.2.6)$$

Comme $\Phi(-x) = 1 - \Phi(x)$, on a

$$c_\beta = \Phi^{-1}\left(\frac{1 + \beta}{2}\right). \quad (4.2.7)$$

Un exemple important correspond à $\beta = 0.95 = 95\%$. Comme $\Phi^{-1}(0.975) \simeq 1.96$, on a

$$c_{95\%} \simeq 1.96. \quad (4.2.8)$$

Pour n grand, nous aurons $\mathbb{P}\{T \in [-c_\beta, c_\beta]\} \simeq \beta$. En exprimant p en fonction de T à l'aide de (4.2.5), on obtient comme intervalle de confiance approximatif pour p

$$I_\beta = \left[\bar{X}_n - c_\beta \sqrt{\frac{\bar{X}_n(1 - \bar{X}_n)}{n}}, \bar{X}_n + c_\beta \sqrt{\frac{\bar{X}_n(1 - \bar{X}_n)}{n}} \right]. \quad (4.2.9)$$

Exemple 4.2.1. Afin de déterminer l'opinion de la population française sur une certaine question, on effectue un sondage sur $n = 1000$ personnes. Il s'avère que 450 personnes ont l'opinion A, alors que les 550 autres personnes ont l'opinion B. On estime donc la proportion d'opinions A dans la population à $\bar{X}_n = 45\%$. Pour simplifier nous modélisons le sondage par un tirage avec remise (le tirage sans remise serait modélisé par une loi hypergéométrique). Le calcul ci-dessus montre que l'intervalle de confiance pour un coefficient de sécurité de 95% est approximativement

$$I_{95\%} = [0.419, 0.481]. \quad (4.2.10)$$

On peut donc affirmer, avec 95% de confiance, que la proportion d'opinions A dans la population totale est comprise entre 41.9% et 48.1%. On peut améliorer (diminuer) cet intervalle en augmentant le nombre n .

4.2.2 Intervalles de confiance pour les paramètres d'une loi gaussienne

Nous considérons maintenant le cas où X suit une loi normale $\mathcal{N}(\mu, \sigma^2)$, de moyenne μ et variance σ^2 inconnues. Les estimateurs empiriques de μ et σ^2 sont

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2. \quad (4.2.11)$$

Si nous voulons déterminer des intervalles de confiance, il nous faut d'abord caractériser les lois de ces estimateurs. Un rôle important est joué par la loi dite du *khi-carré* (ou *khi-deux*) :

Définition 4.2.2. Soient X_1, \dots, X_n des variables aléatoires i.i.d. de loi normale standard. On appelle loi du khi-carré à n degrés de liberté, χ_n^2 , la loi de

$$Q = \sum_{i=1}^n X_i^2. \quad (4.2.12)$$

Proposition 4.2.3. La loi χ_n^2 admet la densité

$$f_{\chi_n^2} = \frac{1}{N(n)} x^{n/2-1} e^{-x/2} 1_{\{x \geq 0\}}, \quad (4.2.13)$$

où la normalisation est donnée par

$$N(n) = 2^{n/2} \Gamma(n/2) = \begin{cases} 2^{n/2} (\frac{n}{2} - 1)! & \text{si } n \text{ est pair,} \\ \sqrt{2\pi} (n-2)!! & \text{si } n \text{ est impair.} \end{cases} \quad (4.2.14)$$

DÉMONSTRATION.

- $n = 1$: Si $X \sim \mathcal{N}(0, 1)$ et $Y = X^2$, on a pour $y > 0$

$$\mathbb{P}\{Y < y\} = \mathbb{P}\{|X| < \sqrt{y}\} = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}),$$

où $\Phi(y) = (2\pi)^{-1/2} \int_{-\infty}^y e^{-t^2/2} dt$ est la fonction de répartition de la loi normale. En dérivant, on obtient

$$f_{\chi_1^2}(y) = \frac{d}{dy} \mathbb{P}\{Y < y\} = \frac{1}{\sqrt{2\pi y}} e^{-y/2} 1_{\{y \geq 0\}}.$$

- $n = 2$: La loi de $X_1^2 + X_2^2$ s'obtient par convolution :

$$f_{\chi_2^2}(x) = \int_{-\infty}^{\infty} f_{\chi_1^2}(t) f_{\chi_1^2}(x-t) dt = \frac{e^{-x/2}}{2\pi} \int_0^x \frac{dt}{\sqrt{t(x-t)}} = \frac{e^{-x/2}}{2} 1_{\{x \geq 0\}}.$$

- $n \geq 3$: On calcule la convolution des lois de $X_1^2 + X_2^2$ et de $X_3^2 + \dots + X_n^2$:

$$f_{\chi_{n+2}^2}(x) = \int_{-\infty}^{\infty} f_{\chi_n^2}(t) f_{\chi_2^2}(x-t) dt = \frac{e^{-x/2}}{2N(n)} \int_0^x t^{n/2-1} dt = \frac{x^{n/2} e^{-x/2}}{nN(n)} 1_{\{x \geq 0\}},$$

d'où la relation de récurrence $N(n+2) = nN(n)$. Les valeurs initiales $N(1) = \sqrt{2\pi}$ et $N(2) = 2$ permettent d'obtenir l'expression générale par récurrence. \square

Nous pouvons maintenant déterminer les lois des estimateurs (4.2.11).

Théorème 4.2.4 (Fisher). Soient \bar{X}_n et S_n^2 les variables définies en (4.2.11).

1. \bar{X}_n suit une loi normale $\mathcal{N}(\mu, \sigma^2/n)$.
2. $Q_n = (n-1)S_n^2$ suit la loi χ_{n-1}^2 .
3. \bar{X}_n et S_n^2 sont indépendantes.
4. La variable

$$T = \sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \tag{4.2.15}$$

suit la loi dite de Student à $n-1$ degrés de liberté, la loi de Student à ν degrés de liberté ayant densité

$$f_{T_\nu}(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi} \Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}. \tag{4.2.16}$$

DÉMONSTRATION.

1. Suit du fait qu'une somme de Gaussiennes est une Gaussienne (Théorème 2.2.10).
2. Notons $Y_i = (X_i - \bar{X}_n)$. Les Y_i ne sont pas indépendantes, en particulier la somme des Y_i est nulle. En remplaçant \bar{X}_n par sa définition on obtient (par un calcul similaire à celui de la preuve de la Proposition 4.1.6)

$$\mathbb{E}(Y_i Y_j) = \begin{cases} (1 - \frac{1}{n})\sigma^2 & \text{si } i = j, \\ -\frac{1}{n}\sigma^2 & \text{si } i \neq j. \end{cases}$$

Les Y_i étant centrées, la matrice de covariance des Y_i est donc

$$\text{cov}(Y) = \sigma^2 \left[\mathbb{1} - \frac{1}{n} J \right],$$

où $\mathbb{1}$ est la matrice identité, est J est la matrice dont tous les coefficients sont 1. C'est une matrice de rang 1, donc elle admet 0 comme valeur propre de multiplicité $n - 1$. La n ième valeur propre est n , le vecteur propre associé étant $(1, 1, \dots, 1)^T$. Il suit que $\text{cov}(Y)$ admet σ^2 comme valeur propre de multiplicité $n - 1$, et 0 comme valeur propre simple.

Il existe donc une matrice unitaire U telle que $U \text{cov}(Y) U^* = \sigma^2 \text{diag}(1, \dots, 1, 0)$. Cette matrice est la matrice de covariance du vecteur $Z = UY$ (Exemple 2.2.16). Autrement dit, Z_1, \dots, Z_{n-1} sont des variables aléatoires normales centrées indépendantes, de variance σ^2 , et $Z_n = 0$.

La transformation $Y \mapsto Z$ étant une isométrie, on a

$$Q_n = \sum_{i=1}^n Y_i^2 = \sum_{i=1}^n Z_i^2 = \sum_{i=1}^{n-1} Z_i^2 .$$

La proposition 4.2.3 montre que cette somme suit la loi χ_{n-1}^2 .

3. On a

$$\begin{aligned} \sigma^2 = \text{Var}(X_i) &= \underbrace{\text{Var}(Y_i)}_{=(1-\frac{1}{n})\sigma^2} + \underbrace{\text{Var}(\bar{X}_n)}_{=\frac{1}{n}\sigma^2} + 2 \text{cov}(Y_i, \bar{X}_n) \end{aligned}$$

d'où $\text{cov}(Y_i, \bar{X}_n) = 0$. Les variables \bar{X}_n et Y_i sont donc non corrélés, et, étant gaussiennes, elles sont indépendantes. \bar{X}_n est donc aussi indépendante de S_n .

4. Soit

$$Z = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} .$$

Il est immédiat que Z suit une loi normale standard, et qu'elle est indépendante des Y_i . Nous avons $T = \sigma Z / S_n$. La loi conjointe de Z et S_n^2 est connue, et le résultat suit d'un calcul de changement de variables dont nous omettons les détails. \square

Ce théorème nous permet de déterminer facilement des intervalles de confiance, de coefficient de sécurité β donné, pour les paramètres μ et σ^2 :

- Comme $\mu = \bar{X} - T S_n / \sqrt{n}$, on cherche le nombre t_β tel que

$$\mathbb{P}\{|T_{n-1}| \leq t_\beta\} = \frac{\Gamma(n/2)}{\sqrt{(n-1)\pi} \Gamma((n-1)/2)} \int_{-t_\beta}^{t_\beta} \left(1 + \frac{x^2}{n-1}\right)^{-n/2} dx = \beta , \quad (4.2.17)$$

et on prend comme intervalle de confiance pour la moyenne μ

$$I_\beta = \left[\bar{X} - t_\beta \frac{S_n}{\sqrt{n}}, \bar{X} + t_\beta \frac{S_n}{\sqrt{n}} \right] . \quad (4.2.18)$$

- Comme Q_n / σ^2 suit la loi χ_{n-1}^2 , on cherche deux nombres $u_\beta, u'_\beta \geq 0$ tels que

$$\mathbb{P}\{u_\beta \leq \chi_{n-1}^2 \leq u'_\beta\} = \frac{1}{2^{(n-1)/2} \Gamma((n-1)/2)} \int_{u_\beta}^{u'_\beta} x^{(n-3)/2} e^{-x/2} dx = \beta , \quad (4.2.19)$$

et on prend comme intervalle de confiance pour la variance σ^2

$$I_\beta = \left[\frac{Q_n}{u'_\beta}, \frac{Q_n}{u_\beta} \right] . \quad (4.2.20)$$

Le choix de l'intervalle n'est pas unique, il n'y a pas de choix canonique parce que la loi n'est pas symétrique.

4.3 Test d'hypothèses

Un point de vue légèrement différent est fourni par le problème du test d'hypothèses. Dans ce cas, on se donne une région $\Theta_0 \subset \Theta$ de l'espace des paramètres, et on aimerait tester l'hypothèse nulle $H_0 : \theta \in \Theta_0$ contre l'hypothèse alternative $H_1 : \theta \notin \Theta_0$. Pour ce faire, on dispose de n observations X_1, \dots, X_n , de loi inconnue \mathbb{P}_θ . Le problème consiste donc à construire une région $R \subset \Omega$, appelée *région de rejet*, telle qu'on considère l'hypothèse nulle comme peu probable si $(X_1, \dots, X_n) \in R$.

Cette procédure de test conduit à deux types d'erreur :

- Une *erreur de première espèce* se produit si on rejette l'hypothèse nulle alors que $\theta \in \Theta_0$. Elle est quantifiée par le *niveau de risque*

$$\alpha = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(R). \quad (4.3.1)$$

Plus α est petit, plus la probabilité d'une erreur de première espèce est faible. On appelle aussi *niveau de confiance* la valeur $1 - \alpha$.

- Une *erreur de seconde espèce* se produit si on on accepte l'hypothèse nulle alors que $\theta \notin \Theta_0$. Elle est quantifiée par la *puissance du test*

$$1 - \beta = \inf_{\theta \in \Theta \setminus \Theta_0} \mathbb{P}_\theta(R). \quad (4.3.2)$$

Plus β est petit, plus la probabilité d'une erreur de seconde espèce est faible.

En pratique, on cherche avant tout à trouver un test dont le niveau de risque est inférieur à un seuil donné, par exemple 5% ou 1%. Si l'on dispose de plusieurs tests de même niveau de risque α , le plus performant sera celui admettant la plus petite valeur de β .

Exemple 4.3.1. On aimerait décider si une pièce de monnaie est équilibrée en observant les résultats de n jets. La loi \mathbb{P}_θ est une loi de Bernoulli d'espérance θ , et l'hypothèse nulle est $\theta = 1/2$, c'est-à-dire que $\Theta_0 = \{1/2\}$. Un choix naturel pour la région de rejet est

$$R = \left\{ \left| \bar{X}_n - \frac{1}{2} \right| > \varepsilon \right\}, \quad (4.3.3)$$

pour un ε qui reste à déterminer. Pour ce faire, on observe que le niveau de risque est donné par

$$\alpha = \mathbb{P}_{1/2}(R) = \mathbb{P}_{1/2} \left\{ \left| \bar{X}_n - \frac{1}{2} \right| > \varepsilon \right\} \leq \frac{1}{\varepsilon^2} \text{Var}_{1/2}(\bar{X}_n) = \frac{1}{4n\varepsilon^2} \quad (4.3.4)$$

en vertu de l'inégalité de Bienaymé-Chebychev. Pour obtenir un niveau de confiance de $1 - \alpha = 95\%$, il suffit de choisir un ε de $1/\sqrt{4n \cdot 0.05} = 1/\sqrt{0.2n}$. Par exemple, si $n = 1000$, on rejette l'hypothèse que la pièce est équilibrée, avec un seuil de confiance de 95%, si $\bar{X}_n \notin [0.43, 0.57]$. Pour $n = 10\,000$ jets, la condition de rejet se réduit à $\bar{X}_n \notin [0.478, 0.522]$. Dans chaque cas, si \bar{X}_n appartient à l'intervalle indiqué, on dit que les valeurs observées ne permettent pas de rejeter l'hypothèse nulle (ce qui ne veut pas dire que l'hypothèse nulle est vraie avec 95% de confiance).

4.3.1 Tests sur une loi gaussienne

On suppose que X_1, \dots, X_n suivent une loi gaussienne $\mathcal{N}(\mu, \sigma^2)$, de paramètres inconnus μ et σ^2 . Voici quatre exemples classiques de tests :

1. **Test de $H_0 : \mu = \mu_0$ contre $H_1 : \mu \neq \mu_0$.**

Sous l'hypothèse H_0 , la variable pivotale

$$T_{n-1} = \sqrt{n} \frac{\bar{X}_n - \mu_0}{S_n} \quad (4.3.5)$$

suit la loi de Student à $(n - 1)$ degrés de liberté. On choisit donc une région de rejet de la forme

$$R = \left\{ \frac{|\bar{X}_n - \mu_0|}{S_n} > c_\alpha \right\}. \quad (4.3.6)$$

Comme

$$\mathbb{P}_{\mu_0}(R) = \mathbb{P}\{|T_{n-1}| > \sqrt{n}c_\alpha\}, \quad (4.3.7)$$

le paramètre c_α est déterminé en fonction du niveau de confiance $1 - \alpha$ par la relation

$$\mathbb{P}\{|T_{n-1}| \leq \sqrt{n}c_\alpha\} = \int_{-\sqrt{n}c_\alpha}^{\sqrt{n}c_\alpha} f_{T_{n-1}}(x) dx = 1 - \alpha, \quad (4.3.8)$$

avec $f_{T_{n-1}}$ la densité de la loi de Student donnée dans (4.2.16).

2. **Test de $H_0 : \mu \leq \mu_0$ contre $H_1 : \mu > \mu_0$.**

On choisit une région de rejet de la forme

$$R = \left\{ \frac{\bar{X}_n - \mu_0}{S_n} > c_\alpha \right\}. \quad (4.3.9)$$

Dans ce cas, on trouve, pour tout $\mu \leq \mu_0$,

$$\mathbb{P}_\mu(R) = \mathbb{P}_\mu \left\{ \sqrt{n} \frac{\bar{X}_n - \mu_0}{S_n} > \sqrt{n}c_\alpha \right\} \leq \mathbb{P}\{T_{n-1} > \sqrt{n}c_\alpha\}, \quad (4.3.10)$$

la borne étant atteinte pour $\mu = \mu_0$. On choisit donc c_α tel que

$$\mathbb{P}\{T_{n-1} > \sqrt{n}c_\alpha\} = \int_{\sqrt{n}c_\alpha}^{\infty} f_{T_{n-1}}(x) dx = \alpha. \quad (4.3.11)$$

3. **Test de $H_0 : \sigma = \sigma_0$ contre $H_1 : \sigma \neq \sigma_0$.**

Lorsque $\sigma = \sigma_0$, la variable

$$Q_{n-1} = \frac{1}{\sigma_0^2} \sum_{i=1}^n X_i^2 \quad (4.3.12)$$

suit la loi χ^2 à $n - 1$ degrés de liberté. On choisit donc une région de rejet de la forme

$$R = \{Q_{n-1} \notin [q_\alpha, q'_\alpha]\}, \quad (4.3.13)$$

avec q_α et q'_α choisis de telle manière que

$$\mathbb{P}\{Q_{n-1} \in [q_\alpha, q'_\alpha]\} = \int_{q_\alpha}^{q'_\alpha} f_{\chi_{n-1}^2}(x) dx = 1 - \alpha, \quad (4.3.14)$$

où $f_{\chi_{n-1}^2}(x)$ est la densité définie dans (4.2.13).

4. Test de $H_0 : \sigma \leq \sigma_0$ contre $H_1 : \sigma > \sigma_0$.

De manière similaire, on choisit

$$R = \{Q_{n-1} > q_\alpha\}, \quad (4.3.15)$$

avec q tel que

$$\mathbb{P}\{Q_{n-1} > q_\alpha\} = \int_{q_\alpha}^{\infty} f_{\chi_{n-1}^2}(x) dx = \alpha. \quad (4.3.16)$$

Un autre exemple de test sur des lois gaussiennes est le suivant. On dispose de deux échantillons indépendants : X_1, \dots, X_n sont distribuées selon une loi $\mathcal{N}(\mu_1, \sigma_1^2)$, alors que Y_1, \dots, Y_m sont distribuées selon une loi $\mathcal{N}(\mu_2, \sigma_2^2)$. Les paramètres des deux lois sont inconnus. On se demande si les deux échantillons admettent la même moyenne, c'est-à-dire que l'on voudrait tester l'hypothèse $H_0 : \mu_1 = \mu_2$ contre $H_1 : \mu_1 \neq \mu_2$. Ce genre de test intervient par exemple si l'on veut comparer l'efficacité de deux médicaments, le premier étant administré à n personnes, le second à m personnes différentes. On peut en effet admettre que pour un grand nombre de personnes, les valeurs observées (par exemple la durée du sommeil) ont des fluctuations normales autour d'une valeur moyenne qui mesure l'efficacité réelle.

Sous l'hypothèse H_0 , on obtient que la variable pivotale

$$T_{n+m-2} = \frac{\bar{Y}_m - \bar{X}_n}{\sqrt{\frac{\sum(X_i - \bar{X}_n)^2 + \sum(Y_i - \bar{Y}_m)^2}{n+m-2}}} \frac{1}{\sqrt{\frac{1}{n} + \frac{1}{m}}} \quad (4.3.17)$$

suit une loi de Student de paramètre $n+m-2$. On choisit donc une région de rejet de la forme

$$R = \{|T_{n+m-2}| > c_\alpha\}, \quad (4.3.18)$$

où c_α est tel que $\mathbb{P}\{|T_{n+m-2}| \leq c_\alpha\} = 1 - \alpha$.

4.3.2 Tests d'adéquation et d'indépendance du χ^2

Jusqu'ici, nous avons considéré des problèmes de statistique dite *paramétrique*, c'est-à-dire que la loi suivie par les valeurs observées est supposée appartenir à une famille donnée (par exemple les lois normales), dont seuls un ou plusieurs paramètres sont à déterminer. En statistique *non paramétrique*, au contraire, la loi elle-même est inconnue, et doit également être déterminée à partir des observations.

La situation la plus simple est celle d'une variable aléatoire discrète, prenant un nombre fini de valeurs x_1, \dots, x_k . A partir d'un échantillon de variables aléatoires i.i.d. X_1, \dots, X_n distribuées selon la loi inconnue, on cherche à estimer les probabilités $p_j = \mathbb{P}\{X = x_j\}$. Soit

$$N_n(j) = \sum_{i=1}^n 1_{\{X_i = x_j\}}, \quad j = 1, \dots, k, \quad (4.3.19)$$

le nombre de fois que la valeur x_j a été observée. Un estimateur de p_j est la fréquence empirique $N_n(j)/n$.

Pour un choix donné de $q_1, \dots, q_k \in [0, 1]$, avec $q_1 + \dots + q_k = 1$, on aimerait tester l'hypothèse nulle $H_0 : p_j = q_j \forall j$ contre $H_1 : \exists j, p_j \neq q_j$. C'est ce qu'on appelle un *test d'adéquation* de la loi inconnue p à la loi donnée q .

Sous H_0 , le vecteur des $N_n(j)$ suit une loi multinomiale, c'est-à-dire

$$\mathbb{P}\{N_n(1) = n_1, \dots, N_n(k) = n_k\} = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k}. \quad (4.3.20)$$

La quantité

$$Z_n = \sum_{j=1}^k \frac{(N_n(j) - nq_j)^2}{nq_j} = n \sum_{j=1}^k \frac{\left(\frac{N_n(j)}{n} - q_j\right)^2}{q_j} \quad (4.3.21)$$

mesure l'écart, proprement normalisé, entre les fréquences empiriques et la loi q . Karl Pearson a montré que dans la limite des grands n , la variable aléatoire

$$\sum_{j=1}^k \frac{(N_n(j) - np_j)^2}{np_j} \quad (4.3.22)$$

suit une loi χ^2 à $k-1$ degrés de liberté (parmi les $N_n(j)$, il n'y en a que $k-1$ indépendants, car leur somme vaut n). On prend donc une région de rejet de la forme

$$R = \{Z_n > z_\alpha\}, \quad (4.3.23)$$

où z_α est choisi en fonction du niveau de confiance $1 - \alpha$ de telle manière que

$$\int_{z_\alpha}^{\infty} f_{\chi_{k-1}^2}(x) dx = \alpha. \quad (4.3.24)$$

Il est important de se rappeler que ce test est basé sur un résultat asymptotique en n . En pratique, on considère que n doit être tel que les effectifs théoriques nq_j soient supérieurs à 5.

Une application importante de cette méthode est le *test d'indépendance du χ^2* . Soit X une variable aléatoire prenant des valeurs discrètes x_1, \dots, x_k , et soit Y une variable aléatoire prenant les valeurs y_1, \dots, y_l . On aimerait tester l'hypothèse nulle H_0 : les variables X et Y sont indépendantes contre H_1 : elles ne le sont pas.

La loi conjointe du couple (X, Y) est donnée par les probabilités

$$p_{ij} = \mathbb{P}\{X = x_i, Y = y_j\}, \quad i = 1, \dots, k, \quad j = 1, \dots, l. \quad (4.3.25)$$

Les lois marginales sont définies par

$$p_{i\bullet} = \sum_{j=1}^l p_{ij}, \quad p_{\bullet j} = \sum_{i=1}^k p_{ij}. \quad (4.3.26)$$

Si les variables X et Y sont indépendantes, alors on a

$$p_{ij} = p_{i\bullet} p_{\bullet j}, \quad i = 1, \dots, k, \quad j = 1, \dots, l. \quad (4.3.27)$$

Les p_{ij} peuvent être estimés par les fréquences empiriques

$$\hat{p}_{ij} = \frac{1}{n} N_{ij} = \frac{1}{n} \sum_{m=1}^n 1_{\{X_m = x_i, Y_m = y_j\}}. \quad (4.3.28)$$

Contrairement au cas précédent, on ne connaît pas la loi q de (X, Y) sous l'hypothèse H_0 . Il faut donc l'estimer aussi, ce que l'on fait à l'aide des fréquences marginales :

$$\hat{q}_{ij} = \frac{N_{i\bullet}}{n} \frac{N_{\bullet j}}{n}, \quad (4.3.29)$$

où

$$N_{i\bullet} = \sum_{j=1}^l N_{ij}, \quad N_{\bullet j} = \sum_{i=1}^k N_{ij}. \quad (4.3.30)$$

On montre alors que dans la limite $n \rightarrow \infty$, la variable

$$Z_n = \sum_{i=1}^k \sum_{j=1}^l \frac{(N_{ij} - n\hat{q}_{ij})^2}{n\hat{q}_{ij}} \quad (4.3.31)$$

converge en loi vers une variable de loi χ^2 à $(k-1)(l-1)$ degrés de liberté (le nombre $(k-1)(l-1)$ est le nombre de fréquences empiriques indépendantes). On rejette donc l'hypothèse d'indépendance si Z_n est supérieur à la valeur z_α telle que

$$\int_{z_\alpha}^{\infty} f_{\chi_{(k-1)(l-1)}^2}(x) dx = \alpha. \quad (4.3.32)$$