

Théorie des Probabilités

Licence MASS, 3ème année

Université du Sud Toulon–Var

Nils Berglund

Version de Novembre 2006

Table des matières

1	Probabilités discrètes	1
1.1	Espace probabilisé discret	1
1.2	Probabilités conditionnelles, indépendance	4
1.3	Variabes aléatoires	7
1.4	Loi des grands nombres	13
1.5	Théorème de la limite centrale	16
1.6	Loi de Poisson et “loi des petits nombres”	20
1.7	Fonction génératrice	24
2	Probabilités continues	27
2.1	Variabes aléatoires réelles à densité	27
2.2	Vecteurs aléatoires à densité	33
2.3	Mesures de probabilité et espaces probabilisés	39
3	Introduction aux processus stochastiques	45
3.1	La marche aléatoire unidimensionnelle symétrique	45
3.2	Le processus ponctuel de Poisson	51
3.3	Les chaînes de Markov	55
3.4	Chaînes de Markov absorbantes	58
3.5	Chaînes de Markov ergodiques	62

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 considéré 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

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

1.2 Probabilités conditionnelles, indépendance

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 1.2.1 (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

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (1.2.1)$$

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

$$\sum_{\omega \in B} q(\omega) = \sum_{\omega \in B} \frac{\mathbb{P}(\{\omega\} \cap B)}{\mathbb{P}(B)} = \frac{1}{\mathbb{P}(B)} \sum_{\omega \in B} \mathbb{P}(\{\omega\}) = 1. \quad (1.2.2)$$

Proposition 1.2.2. 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

$$\mathbb{P}\left(\bigcup_{i \in \mathbb{N}} A_i \mid B\right) = \sum_{i \in \mathbb{N}} \mathbb{P}(A_i|B). \quad (1.2.3)$$

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

DÉMONSTRATION. En exercice. □

Exemple 1.2.3.

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

$$\frac{\mathbb{P}(\{(2, 4), (4, 2)\})}{\mathbb{P}(\{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\})} = \frac{2}{5}. \quad (1.2.4)$$

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

$$\Omega = \{(E, R) : E \in \{0, 1\}, R \in \{0, 1\}\}, \quad (1.2.5)$$

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

$$\begin{aligned} \mathbb{P}(F) &= \mathbb{P}(F \cap E_0) + \mathbb{P}(F \cap E_1) \\ &= f_0\mathbb{P}(E_0) + f_1\mathbb{P}(E_1). \end{aligned} \quad (1.2.6)$$

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 1.2.4 (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 $A \subset \bigcup_j B_j$,*

$$\mathbb{P}(A) = \sum_{j=1}^n \mathbb{P}(A|B_j)\mathbb{P}(B_j) \quad (1.2.7)$$

(loi de la probabilité totale).

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

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{j=1}^n \mathbb{P}(A|B_j)\mathbb{P}(B_j)} \quad (1.2.8)$$

(formule de Bayes).

DÉMONSTRATION. La première relation se montre en écrivant

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_j (A \cap B_j)\right) = \sum_j \mathbb{P}(A \cap B_j) = \sum_j \mathbb{P}(A|B_j)\mathbb{P}(B_j). \quad (1.2.9)$$

La seconde s'obtient en notant que

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(B_i \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\mathbb{P}(A)}, \quad (1.2.10)$$

puis en appliquant (1.2.7). □

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 1.2.5. 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 (1.2.8) nous donne

$$\mathbb{P}(E_1|R_1) = \frac{\mathbb{P}(R_1|E_1)\mathbb{P}(E_1)}{\sum_j \mathbb{P}(R_1|E_j)\mathbb{P}(E_j)} = \frac{(1 - f_1)\mathbb{P}(E_1)}{(1 - f_1)\mathbb{P}(E_1) + f_0\mathbb{P}(E_0)}. \quad (1.2.11)$$

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!

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

Définition 1.2.6 (Indépendance).

1. Deux événements A et B sont indépendants si

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \quad (1.2.12)$$

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

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \dots \mathbb{P}(A_{i_k}) \quad (1.2.13)$$

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

Remarque 1.2.7. 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, $\mathbb{P}(A \cap B \cap C) = 0 \neq \mathbb{P}(A) \cap \mathbb{P}(B) \cap \mathbb{P}(C)$, 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 1.2.8 (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

$$p(\omega = (\omega_1, \dots, \omega_n)) = p_1(\omega_1) \dots p_n(\omega_n). \quad (1.2.14)$$

Proposition 1.2.9. 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

$$\hat{A}_i = \{\omega \in \Omega : \omega_i \in A_i\} \subset \Omega. \quad (1.2.15)$$

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

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

$$\begin{aligned} \mathbb{P}(\hat{A}_1 \cap \hat{A}_2) &= \sum_{\omega \in \hat{A}_1 \cap \hat{A}_2} p(\omega) = \sum_{\omega_1 \in A_1} \sum_{\omega_2 \in A_2} p_1(\omega_1)p_2(\omega_2) \\ &= \sum_{\omega_1 \in A_1} p_1(\omega_1) \sum_{\omega_2 \in A_2} p_2(\omega_2) = \sum_{\omega \in \hat{A}_1} p(\omega) \sum_{\omega \in \hat{A}_2} p(\omega) \\ &= \mathbb{P}(\hat{A}_1)\mathbb{P}(\hat{A}_2). \end{aligned} \quad (1.2.16)$$

On procède ensuite par récurrence sur n . □

Exemple 1.2.10 (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,

$$\mathbb{P}(B_{k_1} \cap \dots \cap B_{k_m}) = \mathbb{P}(B_{k_1}) \dots \mathbb{P}(B_{k_m}) = q^m. \quad (1.2.17)$$

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

$$\begin{aligned} \mathbb{P}(B_{k_1} \cap \dots \cap B_{k_{m_1}} \cap B_{\ell_1}^c \dots B_{\ell_{m_2}}^c) &= \mathbb{P}(B_{k_1}) \dots \mathbb{P}(B_{k_{m_1}}) \mathbb{P}(B_{\ell_1}^c) \dots \mathbb{P}(B_{\ell_{m_2}}^c) \\ &= q^{m_1} (1 - q)^{m_2}. \end{aligned} \quad (1.2.18)$$

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

$$\mathbb{P}(S_k) = \binom{n}{k} q^k (1 - q)^{n-k} =: b(k; n, q). \quad (1.2.19)$$

1.3 Variables aléatoires

Définition 1.3.1 (Variable aléatoire). Soit (Ω, p) un espace probabilisé discret. Une variable aléatoire (discrète) est une application $X : \Omega \rightarrow \mathbb{R}$.

Notation 1.3.2.

- 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 1.3.3.

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

$$\begin{aligned} \mathbb{P}\{X = 4\} &= \mathbb{P}(\{(1, 3), (2, 2), (3, 1)\}) = \frac{3}{36} = \frac{1}{12}, \\ \mathbb{P}\{X \leq 3\} &= \mathbb{P}(\{(1, 1), (1, 2), (2, 1)\}) = \frac{3}{36} = \frac{1}{12}. \end{aligned} \quad (1.3.1)$$

2. 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

$$X(\omega) = \sum_{i=1}^n 1_S(\omega_i), \quad (1.3.2)$$

où

$$1_S(\omega_i) = \begin{cases} 1 & \text{si } \omega_i = S \\ 0 & \text{si } \omega_i = E \end{cases} \quad (1.3.3)$$

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

Définition 1.3.4 (Loi d'une variable aléatoire). *L'application*

$$\begin{aligned} f : X(\Omega) &\mapsto [0, 1] \\ z &\mapsto f(z) = \mathbb{P}\{X = z\} \equiv \mathbb{P}(X^{-1}\{z\}) \end{aligned} \quad (1.3.4)$$

est la loi (ou la distribution) de X .

Remarque 1.3.5. Le fait que

$$\Omega = \bigcup_{z \in X(\Omega)} X^{-1}(z) \equiv \bigcup_{z \in X(\Omega)} \{X = z\} \quad (1.3.5)$$

implique

$$\sum_{z \in X(\Omega)} f(z) = 1. \quad (1.3.6)$$

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

Exemple 1.3.6.

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

$$\begin{aligned} f(2) = f(12) = \frac{1}{36}, & \quad f(3) = f(11) = \frac{2}{36}, & \quad f(4) = f(10) = \frac{3}{36}, \\ f(5) = f(9) = \frac{4}{36}, & \quad f(6) = f(8) = \frac{5}{36}, & \quad f(7) = \frac{6}{36}. \end{aligned} \quad (1.3.7)$$

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

$$\mathbb{P}\{X = k\} = \binom{n}{k} q^k (1 - q)^{n-k} =: b(k; n, q), \quad (1.3.8)$$

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

3. *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

$$p(\omega) = (1 - q)^{\omega-1} q, \quad (1.3.9)$$

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

$$\sum_{\omega \in \Omega} p(\omega) = q \sum_{i=1}^{\infty} (1 - q)^{i-1} = q \frac{1}{1 - (1 - q)} = 1, \quad (1.3.10)$$

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

$$\mathbb{P}\{X = n+k | X > n\} = \mathbb{P}\{X = k\} \quad \forall n. \quad (1.3.11)$$

DÉMONSTRATION. Par calcul direct. On a

$$\mathbb{P}\{X = n+k | X > n\} = \frac{\mathbb{P}\{X = n+k\}}{\mathbb{P}\{X > n\}} = \frac{q(1-q)^{n-1+k}}{\sum_{m>n} q(1-q)^{m-1}}, \quad (1.3.12)$$

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

Définition 1.3.8 (Espérance). *Soit X une variable aléatoire telle que*

$$\sum_{z \in X(\Omega)} |z| \mathbb{P}\{X = z\} < \infty. \quad (1.3.13)$$

Alors l'espérance de X est définie par

$$\mathbb{E}(X) = \sum_{z \in X(\Omega)} z \mathbb{P}\{X = z\} \equiv \sum_{\omega \in \Omega} X(\omega)p(\omega). \quad (1.3.14)$$

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.

Proposition 1.3.9 (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*

$$\mathbb{E}(a_1X_1 + \dots + a_nX_n) = a_1\mathbb{E}(X_1) + \dots + a_n\mathbb{E}(X_n). \quad (1.3.15)$$

DÉMONSTRATION. En exercice. \square

Exemple 1.3.10.

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

$$\mathbb{E}(X) = \sum_{k=0}^n kb(k; n, q) = \sum_{k=0}^n k \frac{n!}{k!(n-k)!} q^k (1-q)^{n-k}. \quad (1.3.16)$$

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 1.3.3, $X = X_1 + \dots + X_n$, où $X_i = 1_{\mathcal{S}}(\omega_i)$ est l'indicatrice d'un succès lors de la i -ème expérience. Nous avons

$$\mathbb{E}(X_i) = 0 \cdot \mathbb{P}\{X_i = 0\} + 1 \cdot \mathbb{P}\{X_i = 1\} = q, \quad (1.3.17)$$

et donc, par la Proposition 1.3.9,

$$\mathbb{E}(X) = \sum_{i=1}^n \mathbb{E}(X_i) = nq . \quad (1.3.18)$$

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

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} kq(1-q)^{k-1} . \quad (1.3.19)$$

Pour calculer cette somme, introduisons la fonction

$$G(z) = \frac{z}{1-z} = \sum_{k=1}^{\infty} z^k . \quad (1.3.20)$$

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

$$\frac{1}{(1-z)^2} = G'(z) = \sum_{k=1}^{\infty} kz^{k-1} . \quad (1.3.21)$$

L'espérance de X est donc donnée par

$$\mathbb{E}(X) = q \sum_{k=1}^{\infty} k(1-q)^{k-1} = qG'(1-q) = q \frac{1}{q^2} = \frac{1}{q} . \quad (1.3.22)$$

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

3. 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

$$\sum_{k=1}^{\infty} |k| \mathbb{P}\{X = k\} = \frac{6}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{k} = \infty , \quad (1.3.23)$$

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).

Définition 1.3.11 (Variance). Soit X une variable aléatoire dont l'espérance existe. On appelle variance de X la somme

$$V(X) = \mathbb{E}([X - \mathbb{E}(X)]^2) = \sum_{z \in X(\Omega)} [z - \mathbb{E}(X)]^2 \mathbb{P}\{X = z\} , \quad (1.3.24)$$

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 $\sigma(X) = \sqrt{V(X)}$.

Proposition 1.3.12.

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

5. Si $V(X) < \infty$ et $V(Y) < \infty$, alors $V(X + Y) < \infty$.

DÉMONSTRATION. En exercice. \square

Exemple 1.3.13 (Loi géométrique). Soit X une variable aléatoire suivant une loi géométrique. En dérivant la relation (1.3.21), on obtient, pour $|z| < 1$,

$$\sum_{k=1}^{\infty} k(k-1)z^{k-2} = G''(z) = \frac{2}{(1-z)^3}. \quad (1.3.25)$$

Ceci permet d'écrire, successivement,

$$\begin{aligned} \mathbb{E}(X(X-1)) &= \sum_{k=1}^{\infty} k(k-1)q(1-q)^{k-1} = q(1-q)\frac{2}{q^3} = 2\frac{1-q}{q^2}, \\ \mathbb{E}(X^2) &= \mathbb{E}(X(X-1)) + \mathbb{E}(X) = \frac{2-q}{q^2}, \\ V(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{1-q}{q^2}. \end{aligned} \quad (1.3.26)$$

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

$$\begin{aligned} V(X+Y) &= \mathbb{E}([X+Y - \mathbb{E}(X+Y)]^2) \\ &= \mathbb{E}([X - \mathbb{E}(X)] + [Y - \mathbb{E}(Y)]^2) \\ &= \mathbb{E}([X - \mathbb{E}(X)]^2) + \mathbb{E}([Y - \mathbb{E}(Y)]^2) + 2\mathbb{E}([X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]) \\ &= V(X) + V(Y) + 2\mathbb{E}([X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]). \end{aligned} \quad (1.3.27)$$

Ce calcul motive la définition suivante.

Définition 1.3.14 (Covariance). On appelle covariance de deux variables aléatoires X et Y la quantité

$$\begin{aligned} \text{cov}(X, Y) &= \mathbb{E}([X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]) \\ &= \mathbb{E}(XY - X\mathbb{E}(Y) - Y\mathbb{E}(X) + \mathbb{E}(X)\mathbb{E}(Y)) \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y). \end{aligned} \quad (1.3.28)$$

Par conséquent, on a

$$V(X+Y) = V(X) + V(Y) + 2\text{cov}(X, Y). \quad (1.3.29)$$

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

Proposition 1.3.15.

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

$$\begin{aligned} \text{cov}(aX, bY) &= ab\text{cov}(X, Y), \\ \text{cov}(X_1 + X_2, Y) &= \text{cov}(X_1, Y) + \text{cov}(X_2, Y). \end{aligned} \quad (1.3.30)$$

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

$$V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i) + \sum_{i \neq j} \text{cov}(X_i, X_j). \quad (1.3.31)$$

DÉMONSTRATION. En exercice. \square

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

Proposition 1.3.16. *Si $V(X)$ et $V(Y)$ existent, alors $\text{cov}(X, Y)$ existe et*

$$|\text{cov}(X, Y)| \leq \sigma(X)\sigma(Y) = \sqrt{V(X)V(Y)}. \quad (1.3.32)$$

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

$$\begin{aligned} g: \quad \mathbb{R}^2 &\rightarrow \mathbb{R}_+ \\ (a, b) &\mapsto g(a, b) = V(aX + bY). \end{aligned} \quad (1.3.33)$$

Les résultats précédents impliquent que

$$\begin{aligned} g(a, b) &= a^2 V(X) + 2ab \text{cov}(X, Y) + b^2 V(Y) \\ &= (a \ b) \begin{pmatrix} V(X) & \text{cov}(X, Y) \\ \text{cov}(X, Y) & V(Y) \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}. \end{aligned} \quad (1.3.34)$$

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 $V(X)V(Y) - \text{cov}(X, Y)^2$. Ainsi, $|\text{cov}(X, Y)| \leq \sqrt{V(X)V(Y)}$. \square

Remarque 1.3.17. En statistique, on introduit le *coefficient de corrélation*

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)} \in [-1, 1]. \quad (1.3.35)$$

La preuve de la proposition montre que $|\rho_{X,Y}| = 1$ si et seulement s'il existe a, b tels que $V(aX + bY) = 0$, 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 1.3.18 (Indépendance de variables aléatoires). *Les variables aléatoires X_1, \dots, X_n sont dites indépendantes si*

$$\mathbb{P}\{X_1 = z_1, \dots, X_n = z_n\} = \mathbb{P}\{X_1 = z_1\} \dots \mathbb{P}\{X_n = z_n\} \quad (1.3.36)$$

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

Cette définition est compatible avec la Définition 1.2.6 de l'indépendance d'événements. En effet, on a le résultat suivant.

Proposition 1.3.19. *Les propriétés suivantes sont équivalentes:*

1. X_1, \dots, X_n sont indépendantes.

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

$$\mathbb{P}\{X_1 \in A_1, \dots, X_n \in A_n\} = \mathbb{P}\{X_1 \in A_1\} \dots \mathbb{P}\{X_n \in A_n\}. \quad (1.3.37)$$

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

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

$$\begin{aligned} \mathbb{E}(XY) &= \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} xy \mathbb{P}\{X = x, Y = y\} \\ &= \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} xy \mathbb{P}\{X = x\} \mathbb{P}\{Y = y\} = \mathbb{E}(X)\mathbb{E}(Y), \end{aligned} \quad (1.3.38)$$

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 1.3.21.

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

$$V(X_i) = 0^2 \cdot (1 - q) + 1^2 \cdot q - \mathbb{E}(X_i)^2 = q(1 - q). \quad (1.3.39)$$

Comme les X_i sont indépendants, il suit $V(X) = nq(1 - q)$.

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.

1.4 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 1.3.9 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 faible des grands nombres. Pour démontrer cette loi, nous devons d'abord prouver l'inégalité de Bienaymé–Chebychev.

Lemme 1.4.1. *Soit X une variable aléatoire dont l'espérance existe. Pour tout $a > 0$,*

$$\mathbb{P}\{|X| \geq a\} \leq \frac{\mathbb{E}(|X|)}{a}. \quad (1.4.1)$$

DÉMONSTRATION. On a

$$\begin{aligned} \mathbb{P}\{|X| \geq a\} &= \sum_{x \in X(\Omega), |x| \geq a} \mathbb{P}\{X = x\} \leq \sum_{x \in X(\Omega), |x| \geq a} \frac{|x|}{a} \mathbb{P}\{X = x\} \\ &\leq \frac{1}{a} \sum_{x \in X(\Omega)} |x| \mathbb{P}\{X = x\} = \frac{1}{a} \mathbb{E}(|X|). \end{aligned} \quad (1.4.2)$$

□

Corollaire 1.4.2 (Inégalité de Bienaymé–Chebychev). *Soit X une variable aléatoire dont l'espérance et la variance existent. Alors, pour tout $a > 0$,*

$$\mathbb{P}\{|X - \mathbb{E}(X)| \geq a\} \leq \frac{1}{a^2} \mathbb{V}(X). \quad (1.4.3)$$

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

$$\mathbb{P}\{|X - \mathbb{E}(X)| \geq a\} = \mathbb{P}\{Y \geq a^2\} \leq \frac{1}{a^2} \mathbb{E}(Y) = \frac{1}{a^2} \mathbb{V}(X). \quad (1.4.4)$$

□

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 $\sigma(X)$. On dit que X est “concentrée dans un intervalle d'ordre de grandeur $\sigma(X)$ autour de son espérance”.

Théorème 1.4.3 (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*

$$S_n = \sum_{i=1}^n X_i. \quad (1.4.5)$$

Alors, pour tout $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left\{\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right\} = 0. \quad (1.4.6)$$

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

$$\mathbb{P}\left\{\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right\} \leq \frac{1}{\varepsilon^2} \mathbb{V}\left(\frac{S_n}{n}\right) = \frac{1}{\varepsilon^2 n^2} \mathbb{V}(S_n) = \frac{1}{\varepsilon^2 n^2} n \sigma^2 = \frac{\sigma^2}{n \varepsilon^2}, \quad (1.4.7)$$

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

Remarque 1.4.4. Il existe également une loi forte des grands nombres. Celle-ci affirme que sous certaines conditions, on a

$$\mathbb{P}\left\{\lim_{n \rightarrow \infty} \frac{S_n}{n} = \mu\right\} = 1. \quad (1.4.8)$$

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 1.4.5. 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

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \left| \frac{S_n}{n} - q \right| \geq \varepsilon \right\} = \lim_{n \rightarrow \infty} \sum_{k: |k-nq| \geq n\varepsilon} b(k; n, q) = 0 \quad (1.4.9)$$

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

$$\mathbb{P} \left\{ \left| \frac{S_n}{n} - \frac{1}{2} \right| \geq \varepsilon \right\} \leq \frac{1}{\varepsilon^2} V \left(\frac{S_n}{n} \right) = \frac{1}{4n\varepsilon^2}. \quad (1.4.10)$$

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

$$\mathbb{P} \left\{ \left| \frac{S_{1000}}{1000} - \frac{1}{2} \right| \geq \varepsilon \right\} = \mathbb{P} \left\{ S_{1000} \notin [400, 600] \right\} \leq \frac{1}{40} = 0.025. \quad (1.4.11)$$

Digression 1.4.6 (Grandes déviations). En fait, la majoration (1.4.10) 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

$$\mathbb{P} \left\{ S_n - \frac{n}{2} \geq n\varepsilon \right\} = \mathbb{P} \left\{ e^{\lambda(S_n - n/2)} \geq e^{\lambda n\varepsilon} \right\} \leq \frac{1}{e^{\lambda n\varepsilon}} \mathbb{E} \left(e^{\lambda(S_n - n/2)} \right). \quad (1.4.12)$$

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

$$\mathbb{E}(Z_i) = e^{-\lambda/2} \mathbb{P}\{Z_i = 0\} + e^{\lambda/2} \mathbb{P}\{Z_i = 1\} = \cosh(\lambda/2). \quad (1.4.13)$$

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

$$\mathbb{E} \left(e^{\lambda(S_n - n/2)} \right) = \mathbb{E} \left(\prod_{i=1}^n Z_i \right) = \prod_{i=1}^n \mathbb{E}(Z_i) = [\cosh(\lambda/2)]^n. \quad (1.4.14)$$

Nous pouvons écrire cette dernière expression comme $\exp[n \log \cosh(\lambda/2)]$. L'inégalité (1.4.12) devient donc

$$\mathbb{P} \left\{ S_n - \frac{n}{2} \geq n\varepsilon \right\} \leq \exp \left\{ -n [\lambda\varepsilon - \log \cosh(\lambda/2)] \right\}. \quad (1.4.15)$$

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

$$\begin{aligned} I(\varepsilon) &= f \left(\log \frac{1 + 2\varepsilon}{1 - 2\varepsilon} \right) \\ &= \varepsilon \log \frac{1 + 2\varepsilon}{1 - 2\varepsilon} - \log \left[\sqrt{(1 - 2\varepsilon)(1 + 2\varepsilon)} \right] \\ &= \left(\frac{1}{2} + \varepsilon \right) \log(1 + 2\varepsilon) + \left(\frac{1}{2} - \varepsilon \right) \log(1 - 2\varepsilon). \end{aligned} \quad (1.4.16)$$

En utilisant (1.4.15) pour cette valeur de λ , et la même estimation pour $\mathbb{P}\{S_n - \frac{n}{2} \leq -n\varepsilon\}$, nous obtenons

$$\mathbb{P}\left\{\left|\frac{S_n}{n} - \frac{1}{2}\right| \geq \varepsilon\right\} \leq 2e^{-nI(\varepsilon)}. \quad (1.4.17)$$

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

$$\mathbb{P}\left\{S_{1000} \notin [400, 600]\right\} \leq 2e^{-1000I(0.1)} \simeq 3.6 \cdot 10^{-9}. \quad (1.4.18)$$

1.5 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 1.4.3, nous avons vu que

$$\mathbb{P}\left\{\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right\} \leq \frac{\sigma^2}{n\varepsilon^2}. \quad (1.5.1)$$

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 rééchelonnée

$$\widehat{S}_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}. \quad (1.5.2)$$

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

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

$$\lim_{n \rightarrow \infty} \mathbb{P}\{a \leq \widehat{S}_n \leq b\} = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx. \quad (1.5.3)$$

Commençons par vérifier que le membre de droite définit bien une probabilité.

Lemme 1.5.1.

$$\int_{-\infty}^{\infty} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = 1. \quad (1.5.4)$$

DÉMONSTRATION. L'astuce consiste à calculer le carré de l'intégrale, qui vaut

$$\left(\int_{-\infty}^{\infty} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx\right) \left(\int_{-\infty}^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy\right) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)/2} dx dy. \quad (1.5.5)$$

En passant en coordonnées polaires (r, φ) , et tenant compte du Jacobien de la transformation $dx dy = r dr d\varphi$, cette expression devient

$$\frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-r^2/2} r dr d\varphi = \left[-e^{-r^2/2}\right]_0^{\infty} = 1. \quad (1.5.6)$$

□

On a donc bien

$$\lim_{n \rightarrow \infty} \mathbb{P}\left\{-\infty \leq \widehat{S}_n \leq \infty\right\} = 1. \quad (1.5.7)$$

Pour démontrer le théorème de la limite centrale dans le cas particulier de la loi binomiale $b(k; n, q)$, il est utile de connaître la *formule de Stirling*:

Lemme 1.5.2 (Formule de Stirling).

$$\lim_{n \rightarrow \infty} \frac{n!}{n^n e^{-n} \sqrt{2\pi n}} = 1. \quad (1.5.8)$$

DÉMONSTRATION. Considérons la *fonction Gamma d'Euler*

$$\Gamma(n) = \int_0^{\infty} t^{n-1} e^{-t} dt. \quad (1.5.9)$$

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

$$\Gamma(n+1) = \int_0^{\infty} t^n e^{-t} dt = \left[-t^n e^{-t}\right]_0^{\infty} + n \int_0^{\infty} t^{n-1} e^{-t} dt = n\Gamma(n), \quad (1.5.10)$$

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

$$\begin{aligned} n! = \Gamma(n+1) &= \int_0^{\infty} t^n e^{-t} dt \\ &= \int_{-\sqrt{n}}^{\infty} (n + \sqrt{n}s)^n e^{-n - \sqrt{n}s} \sqrt{n} ds \\ &= n^n e^{-n} \sqrt{n} \int_{-\sqrt{n}}^{\infty} \left(1 + \frac{s}{\sqrt{n}}\right)^n e^{-\sqrt{n}s} ds \\ &= n^n e^{-n} \sqrt{2\pi n} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-h_n(s)} ds, \end{aligned} \quad (1.5.11)$$

où nous avons posé

$$h_n(s) = \begin{cases} \sqrt{n}s - n \log\left(1 + \frac{s}{\sqrt{n}}\right) & \text{pour } s > -\sqrt{n}, \\ +\infty & \text{autrement.} \end{cases} \quad (1.5.12)$$

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

$$h_n(s) = \frac{s^2}{2} + \mathcal{O}\left(\frac{s^3}{\sqrt{n}}\right) \quad \text{pour } s > -\sqrt{n}, \quad (1.5.13)$$

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

$$h_n(s) \geq g(s) = \begin{cases} \frac{s^2}{6} & \text{pour } s \leq 1, \\ \frac{s-1}{2} & \text{pour } s \geq 1, \end{cases} \quad (1.5.14)$$

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

$$\lim_{n \rightarrow \infty} \frac{n!}{n^n e^{-n} \sqrt{2\pi n}} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \lim_{n \rightarrow \infty} e^{-h_n(s)} ds = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-s^2/2} ds, \quad (1.5.15)$$

et cette dernière intégrale vaut 1 par le lemme précédent. \square

Nous pouvons maintenant prouver le théorème de la limite centrale pour la loi binomiale par un calcul relativement direct. Rappelons que cette loi a espérance nq et variance $nq(1 - q)$.

Théorème 1.5.3 (Moivre–Laplace). *Si S_n suit la loi binomiale de paramètre q , alors $\widehat{S}_n = (S_n - nq)/\sqrt{nq(1 - q)}$ satisfait*

$$\lim_{n \rightarrow \infty} \mathbb{P}\left\{a \leq \widehat{S}_n \leq b\right\} = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \quad (1.5.16)$$

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

DÉMONSTRATION. 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

$$\begin{aligned} \mathbb{P}\left\{a \leq \widehat{S}_n \leq b\right\} &= \mathbb{P}\left\{\frac{n}{2} + \frac{\sqrt{n}}{2}a \leq S_n \leq \frac{n}{2} + \frac{\sqrt{n}}{2}b\right\} \\ &= \sum_{k \in A} b(k; n, 1/2). \end{aligned} \quad (1.5.17)$$

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

$$\lim_{n \rightarrow \infty} \sup_{k \in A} \left| \frac{x(k, n)}{y(k, n)} - 1 \right| = 0. \quad (1.5.18)$$

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

$$\begin{aligned} b(k; n, 1/2) &= 2^{-n} \frac{n!}{k!(n - k)!} \\ &\sim 2^{-n} \frac{n^n e^{-n} \sqrt{2\pi n}}{k^k e^{-k} \sqrt{2\pi k} (n - k)^{n-k} e^{-(n-k)} \sqrt{2\pi(n - k)}} \\ &= \frac{2^{-n}}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n - k)}} \frac{n^n}{k^k (n - k)^{n-k}} \\ &= \frac{1}{\sqrt{2\pi n}} \frac{1}{\sqrt{(k/n)(1 - k/n)}} \frac{2^{-n}}{(k/n)^k (1 - k/n)^{n-k}}. \end{aligned} \quad (1.5.19)$$

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

$$\frac{k}{n} = \frac{1}{2} \left(1 + \frac{x}{\sqrt{n}}\right), \quad 1 - \frac{k}{n} = \frac{1}{2} \left(1 - \frac{x}{\sqrt{n}}\right), \quad (1.5.20)$$

et par conséquent

$$b(k; n, 1/2) \sim \frac{1}{\sqrt{2\pi n}} \frac{2}{\sqrt{1 - x^2/n}} \psi(x, n), \quad (1.5.21)$$

où nous avons introduit

$$\psi(x, n) = \left(1 + \frac{x}{\sqrt{n}}\right)^{-\left(1+x/\sqrt{n}\right)n/2} \left(1 - \frac{x}{\sqrt{n}}\right)^{-\left(1-x/\sqrt{n}\right)n/2}. \quad (1.5.22)$$

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

$$\log \psi(x, n) = -\frac{n}{2} \left[\left(1 + \frac{x}{\sqrt{n}}\right) \log \left(1 + \frac{x}{\sqrt{n}}\right) + \left(1 - \frac{x}{\sqrt{n}}\right) \log \left(1 - \frac{x}{\sqrt{n}}\right) \right], \quad (1.5.23)$$

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

$$\log \psi(x, n) = -\frac{x^2}{2} + \mathcal{O}\left(\frac{x^3}{\sqrt{n}}\right). \quad (1.5.24)$$

Il suit donc de (1.5.21) que

$$b(k; n, 1/2) \sim \frac{2}{\sqrt{2\pi n}} e^{-x^2/2} \left[1 + \mathcal{O}\left(\frac{x}{\sqrt{n}}\right)\right] \sim \frac{2}{\sqrt{2\pi n}} e^{-x^2/2}, \quad (1.5.25)$$

puisque $x \in [a, b]$. En remplaçant dans (1.5.17), il vient

$$\mathbb{P}\left\{a \leq \widehat{S}_n \leq b\right\} \sim \frac{1}{\sqrt{2\pi}} \sum_{x \in [a, b]: (n + \sqrt{n}x)/2 \in \{0, \dots, n\}} \frac{2}{\sqrt{n}} e^{-x^2/2}. \quad (1.5.26)$$

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

Nous donnons maintenant, sans démonstration, l'énoncé général du théorème de la limite centrale.

Théorème 1.5.4 (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 . Alors la variable aléatoire $S_n = \sum_{i=1}^n X_i$ satisfait*

$$\lim_{n \rightarrow \infty} \mathbb{P}\left\{a \leq \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq b\right\} = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx. \quad (1.5.27)$$

Définition 1.5.5 (Loi normale). *La fonction*

$$\Phi(x) = \int_{-\infty}^x \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy \quad (1.5.28)$$

s'appelle la fonction de répartition de la loi normale. Ses valeurs sont tabulées. La fonction

$$\varphi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \quad (1.5.29)$$

est appelée la densité de la loi normale.

Le Théorème 1.5.4 montre que pour n grand, on a approximativement

$$\mathbb{P}\left\{\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \in [a, b]\right\} \simeq \Phi(b) - \Phi(a), \quad (1.5.30)$$

ou encore

$$\mathbb{P}\{S_n \in [\hat{a}, \hat{b}]\} \simeq \Phi\left(\frac{\hat{b} - n\mu}{\sqrt{n\sigma^2}}\right) - \Phi\left(\frac{\hat{a} - n\mu}{\sqrt{n\sigma^2}}\right). \quad (1.5.31)$$

Exemple 1.5.6. 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

$$\mathbb{P}\{S_n > 50\} = \mathbb{P}\{S_n \in [50, \infty[\} \simeq \Phi(\infty) - \Phi\left(\frac{50 - 40}{\sqrt{36}}\right) = 1 - \Phi(1.66\dots) \simeq 0.0485. \quad (1.5.32)$$

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

1.6 Loi de Poisson et “loi des petits nombres”

La loi normale donne une bonne approximation de la loi binomiale pour q fixé et n tendant vers l’infini. La loi de Poisson, quant à elle, décrit bien la loi binomiale pour n tendant vers l’infini et q tendant vers zéro, avec le produit nq tendant vers une constante. Elle modélise donc les expériences de Bernoulli avec une très faible probabilité de succès, mais avec un grand nombre d’essais, du même ordre de grandeur que l’inverse de la probabilité de succès.

Définition 1.6.1 (Loi de Poisson). La loi de Poisson de paramètre $\lambda > 0$ est définie par

$$\mathbb{P}\{X = k\} = \pi_\lambda(k) := e^{-\lambda} \frac{\lambda^k}{k!} \quad \forall k \in \mathbb{N}. \quad (1.6.1)$$

Remarquons que π_λ définit bien une loi de probabilité, puisque $e^{-\lambda} > \lambda^k/k!$ pour tout k et donc $\pi_\lambda(k) \in [0, 1]$, et

$$\sum_{k=0}^{\infty} \pi_\lambda(k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^\lambda = 1. \quad (1.6.2)$$

Proposition 1.6.2.

1. L’espérance et la variance de la loi de Poisson sont égales à λ .
2. 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.

1. Si X suit une loi de Poisson de paramètre λ , alors

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} k\pi_\lambda(k) = e^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda. \quad (1.6.3)$$

De manière similaire,

$$\mathbb{E}(X(X-1)) = \sum_{k=0}^{\infty} k(k-1)\pi_\lambda(k) = e^{-\lambda} \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} = \lambda^2, \quad (1.6.4)$$

d’où $\mathbb{E}(X^2) = \lambda^2 + \lambda$ et $V(X) = \lambda$.

2. On a $X + Y = n$ si et seulement s'il existe un $k \in \{0, \dots, n\}$ tels que $X = k$ et $Y = n - k$. Par conséquent

$$\begin{aligned} \mathbb{P}\{X + Y = n\} &= \sum_{k=0}^n \mathbb{P}\{X = k, Y = n - k\} = \sum_{k=0}^n \mathbb{P}\{K = k\} \mathbb{P}\{Y = n - k\} \\ &= \sum_{k=0}^n e^{-\lambda} e^{-\mu} \frac{\lambda^k}{k!} \frac{\mu^{n-k}}{(n-k)!} = \frac{1}{n!} \underbrace{\sum_{k=0}^n \frac{n!}{k!(n-k)!} \lambda^k \mu^{n-k}}_{(\lambda+\mu)^n} e^{-(\lambda+\mu)} \\ &= \pi_{\lambda+\mu}(n). \end{aligned} \quad (1.6.5)$$

□

La proposition suivante établit la convergence (simple, c'est-à-dire à k fixé) de la loi binomiale vers la loi de Poisson.

Proposition 1.6.3. Soit $\{q_n\}_{n \geq 0}$ une suite telle que $\lim_{n \rightarrow \infty} nq_n = \lambda > 0$. Alors, pour tout $k \in \mathbb{N}$,

$$\lim_{n \rightarrow \infty} b(k; n, q_n) = \pi_\lambda(k). \quad (1.6.6)$$

DÉMONSTRATION. Soit $\lambda_n = nq_n$. Alors

$$\begin{aligned} b(k; n, q_n) &= b\left(k; n, \frac{\lambda_n}{n}\right) \\ &= \frac{n(n-1)\dots(n-k+1)}{k!} \frac{\lambda_n^k}{n^k} \left(1 - \frac{\lambda_n}{n}\right)^{n-k} \\ &= \frac{\lambda_n^k}{k!} \frac{1}{(1 - \lambda_n/n)^k} \left(1 - \frac{1}{n}\right) \dots \left(1 - \frac{k-1}{n}\right) \left(1 - \frac{\lambda_n}{n}\right)^n. \end{aligned} \quad (1.6.7)$$

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

$$\lim_{n \rightarrow \infty} b(k; n, q_n) = \frac{\lambda^k}{k!} \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda_n}{n}\right)^n = \frac{\lambda^k}{k!} e^{-\lambda}, \quad (1.6.8)$$

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

Exemple 1.6.4. 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

$$b(k; 2000, 1/1000) \simeq \pi_2(k) = e^{-2} \frac{2^k}{k!}. \quad (1.6.9)$$

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

L'avantage de la loi de Poisson est qu'elle est bien plus rapide à calculer, puisqu'elle ne fait pas intervenir de coefficients binomiaux.

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.

Théorème 1.6.5. *On a*

$$\sum_{k=0}^{\infty} |b(k; n, q) - \pi_{nq}(k)| \leq nq^2. \quad (1.6.10)$$

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 jusqu'ici.

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

$$p_i(k) = \begin{cases} e^{-q} - (1 - q) & \text{si } k = -1, \\ 1 - q & \text{si } k = 0, \\ e^{-q} \frac{q^k}{k!} & \text{si } k \geq 1. \end{cases} \quad (1.6.11)$$

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

$$X_i(\omega_i) = \begin{cases} 0 & \text{si } \omega_i = 0, \\ 1 & \text{sinon,} \end{cases} \quad Y_i(\omega_i) = \begin{cases} \omega_i & \text{si } \omega_i \geq 1, \\ 0 & \text{sinon.} \end{cases} \quad (1.6.12)$$

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,

$$\begin{aligned} \mathbb{P}\{X_i = Y_i\} &= \mathbb{P}\{X_i = 0, Y_i = 0\} + \mathbb{P}\{X_i = 1, Y_i = 1\} \\ &= p_i(0) + p_i(1) = 1 - q + qe^{-q}, \end{aligned} \quad (1.6.13)$$

donc

$$\mathbb{P}\{X_i \neq Y_i\} = q(1 - e^{-q}) \leq q^2. \quad (1.6.14)$$

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 de la Proposition 1.6.2.

Comme $X \neq Y$ implique que $X_i \neq Y_i$ pour un i au moins, il suit de (1.6.14) que

$$\mathbb{P}\{X \neq Y\} \leq \sum_{i=1}^n \mathbb{P}\{X_i \neq Y_i\} \leq nq^2. \quad (1.6.15)$$

Il reste donc à montrer que le membre de gauche de (1.6.10) est majoré par $2\mathbb{P}\{X \neq Y\}$. Un tel procédé s'appelle un argument de couplage. Nous posons, pour abrégé l'écriture,

$f(k) = \mathbb{P}\{X = k\}$, $g(k) = \mathbb{P}\{Y = k\}$ et $A = \{k : f(k) > g(k)\}$. Alors

$$\begin{aligned}
 \sum_{k=0}^{\infty} |b(k; n, q) - \pi_{nq}(k)| &= \sum_{k=0}^{\infty} |f(k) - g(k)| \\
 &= \sum_{k \in A} (f(k) - g(k)) - \sum_{k \notin A} (f(k) - g(k)) \\
 &= 2 \sum_{k \in A} (f(k) - g(k)) - \underbrace{\sum_{k \in \mathbb{N}} (f(k) - g(k))}_{=1-1=0}. \tag{1.6.16}
 \end{aligned}$$

Or nous pouvons écrire

$$\begin{aligned}
 \sum_{k \in A} (f(k) - g(k)) &= \mathbb{P}\{X \in A\} - \mathbb{P}\{Y \in A\} \\
 &= \mathbb{P}\{X \in A, Y \in A\} + \mathbb{P}\{X \in A, Y \neq X\} - \mathbb{P}\{Y \in A\} \\
 &\leq \mathbb{P}\{X \in A, Y \in A\} + \mathbb{P}\{X \neq Y\} - \mathbb{P}\{Y \in A\} \\
 &\leq \mathbb{P}\{X \neq Y\}, \tag{1.6.17}
 \end{aligned}$$

ce qui conclut la démonstration. \square

Le corollaire suivant montre l'utilité de la relation (1.6.10).

Corollaire 1.6.6. *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*

$$|\mathbb{E}(f(X)) - \mathbb{E}(f(Y))| \leq Mnq^2. \tag{1.6.18}$$

De plus, pour tout $A \subset \mathbb{N}$, on a

$$|\mathbb{P}\{X \in A\} - \mathbb{P}\{Y \in A\}| \leq nq^2. \tag{1.6.19}$$

DÉMONSTRATION.

$$|\mathbb{E}(f(X)) - \mathbb{E}(f(Y))| \leq \sum_{k=0}^{\infty} |f(k)| |b(k; n, q) - \pi_{nq}(k)| \leq Mnq^2. \tag{1.6.20}$$

La relation (1.6.19) est un cas particulier de (1.6.18), 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 1.6.7. Revenons à l'exemple 1.6.4. Avec $n = 2000$ et $q = 1/1000$, la relation (1.6.19) devient

$$|\mathbb{P}\{X \in A\} - \mathbb{P}\{Y \in A\}| \leq 0.002. \tag{1.6.21}$$

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.

1.7 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 1.7.1 (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

$$G_X(z) = \sum_{k \geq 0} z^k \mathbb{P}\{X = k\}. \quad (1.7.1)$$

On remarque que

$$G_X(1) = \sum_{k \geq 0} \mathbb{P}\{X = k\} = 1, \quad (1.7.2)$$

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

$$\begin{aligned} G'_X(z) &= \sum_{k \geq 0} k z^{k-1} \mathbb{P}\{X = k\} & \Rightarrow & \quad G'_X(1) = \sum_{k \geq 0} k \mathbb{P}\{X = k\} = \mathbb{E}(X), \\ G''_X(z) &= \sum_{k \geq 0} k(k-1) z^{k-2} \mathbb{P}\{X = k\} & \Rightarrow & \quad G''_X(1) = \sum_{k \geq 0} k(k-1) \mathbb{P}\{X = k\} \\ & & & = \mathbb{E}(X(X-1)). \end{aligned} \quad (1.7.3)$$

Nous avons donc le résultat suivant.

Proposition 1.7.2. 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

$$\begin{aligned} \mathbb{E}(X) &= G'_X(1), \\ \text{V}(X) &= G''_X(1) + G'_X(1) - G'_X(1)^2. \end{aligned} \quad (1.7.4)$$

Exemple 1.7.3.

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

$$G_X(z) = \sum_{k \geq 1} z^k q(1-q)^{k-1} = \frac{q}{1-q} \sum_{k \geq 1} [z(1-q)]^k = \frac{qz}{1-(1-q)z}. \quad (1.7.5)$$

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

2. *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:

$$G_X(z) = \sum_{k=0}^n \binom{n}{k} (qz)^k (1-q)^{n-k} = (qz + 1 - q)^n. \quad (1.7.6)$$

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

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

$$G_X(z) = e^{-\lambda} \sum_{k \geq 0} \frac{(\lambda z)^k}{k!} = e^{\lambda(z-1)}. \quad (1.7.7)$$

Le rayon de convergence est infini.

On vérifiera que les relations (1.7.4) donnent bien les espérances et les variances calculées précédemment. Nous les résumons dans le tableau suivant.

Loi	$\mathbb{P}\{X = k\}$	$G_X(z)$	$\mathbb{E}(X)$	$V(X)$
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, en particulier, de retrouver des résultats que nous avons déjà dérivés pour l'espérance de la somme de certaines variables aléatoires.

Proposition 1.7.4. *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

$$\begin{aligned}
 G_{X+Y}(z) &= \sum_{k \geq 0} z^k \mathbb{P}\{X + Y = k\} \\
 &= \sum_{k \geq 0} \sum_{\ell=0}^k z^k \mathbb{P}\{X = \ell, Y = k - \ell\} \\
 &= \sum_{k \geq 0} \sum_{\ell=0}^k z^k \mathbb{P}\{X = \ell\} \mathbb{P}\{Y = k - \ell\} \\
 &= \sum_{\ell \geq 0} \sum_{n \geq 0} z^{\ell+n} \mathbb{P}\{X = \ell\} \mathbb{P}\{Y = n\} \\
 &= G_X(z) G_Y(z).
 \end{aligned} \tag{1.7.8}$$

□

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, comme nous l'avons montré précédemment.

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

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 $[a, b]$:

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

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

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

Le Lemme 1.5.1 montre que $\varphi(x)$ est effectivement une densité de probabilité.

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

$$\varphi(x; \mu, \sigma^2) = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi} \sigma} . \quad (2.1.5)$$

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

$$\int_{-\infty}^{\infty} \varphi(x; \mu, \sigma^2) dx = \int_{-\infty}^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy = 1 . \quad (2.1.6)$$

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

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 , \\ 0 & \text{si } x < 0 . \end{cases} \quad (2.1.7)$$

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

$$f(x) = \frac{c}{\pi} \frac{1}{x^2 + c^2} . \quad (2.1.8)$$

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

$$F(x) = \int_{-\infty}^x f(y) dy . \quad (2.1.9)$$

On a les propriétés suivantes:

1. Si f est une densité, la fonction F définie par (2.1.9) 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,

- 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.

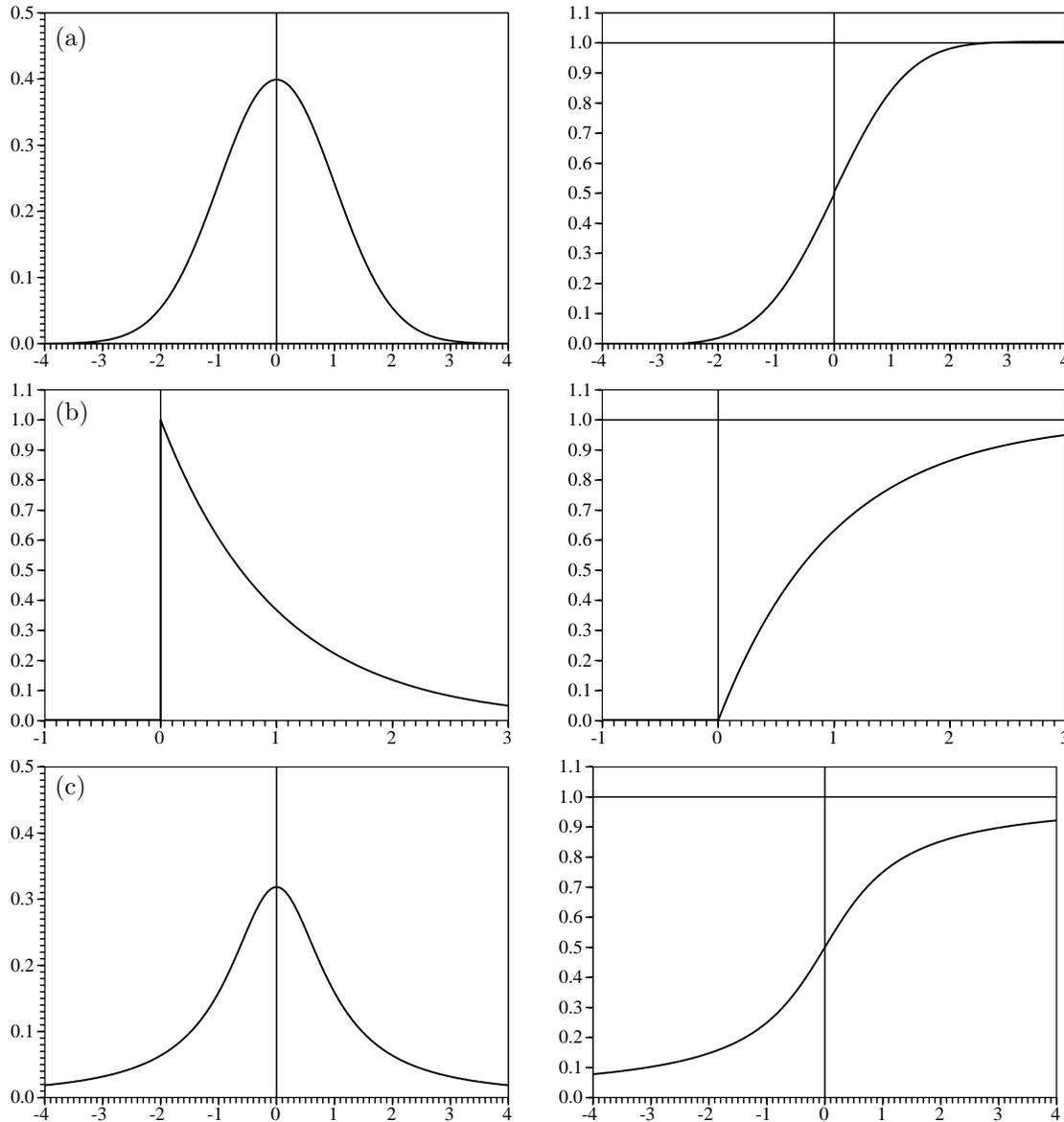


FIGURE 2.1. Densités et fonctions de répartition correspondantes: (a) loi normale standard $\mathcal{N}(0,1)$, (b) loi exponentielle $\text{Exp}(1)$, (c) loi de Cauchy de paramètre 1.

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

$$F_X(t) = \mathbb{P}\{X \leq t\} \quad (2.1.10)$$

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

$$\mathbb{P}\{X \leq t\} = \int_{-\infty}^t f(s) \, ds, \quad (2.1.11)$$

$$\mathbb{P}\{a < X \leq b\} = \mathbb{P}\{X \leq b\} - \mathbb{P}\{X \leq a\} = \int_a^b f(s) \, ds. \quad (2.1.12)$$

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

$$F_X(t) = \sum_{z \in X(\Omega) : z \leq t} \mathbb{P}\{X = z\} = \sum_{\omega \in \Omega : X(\omega) \leq t} p(\omega) = \sum_{\omega \in \Omega} 1_{[X(\omega), \infty[}(t) p(\omega). \quad (2.1.13)$$

$F_X(t)$ 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. Dans la Section 1.5, nous avons vu que pour des X_i i.i.d., d'espérance μ et variance σ , la variable aléatoire

$$\widehat{S}_n = \frac{1}{\sqrt{n}\sigma} (S_n - n\mu), \quad S_n = \sum_{i=1}^n X_i, \quad (2.1.14)$$

satisfait

$$\lim_{n \rightarrow \infty} F_{\widehat{S}_n}(t) = \int_{-\infty}^t \varphi(s) \, ds, \quad (2.1.15)$$

où $\varphi(s) = e^{-s^2/2} / \sqrt{2\pi}$ est la densité de la loi normale standard. Bien que nous n'ayons pas encore défini d'espace probabilisé commun à tous les \widehat{S}_n , on est tenté de dire que $\widehat{S}_\infty = \lim_{n \rightarrow \infty} \widehat{S}_n$ est une variable aléatoire admettant la densité φ . On dira aussi que cette variable aléatoire *suit la loi normale standard*, et on notera $\widehat{S}_\infty \sim \mathcal{N}(0, 1)$.

3. Supposons que X soit une variable aléatoire satisfaisant

$$\mathbb{P}\{X > t\} = \begin{cases} 1 & \text{si } t < 0, \\ e^{-\lambda t} & \text{si } t \geq 0. \end{cases} \quad (2.1.16)$$

Alors la fonction de répartition de X est donnée par

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

On s'aperçoit que

$$F'_X(t) = \begin{cases} 0 & \text{si } t < 0, \\ \lambda e^{-\lambda t} & \text{si } t > 0 \end{cases} \quad (2.1.18)$$

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 (2.1.7). On dira que X suit la loi exponentielle, et on notera $X \sim \text{Exp}(\lambda)$.

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

$$\mathbb{P}\{X > t | X > s\} = \frac{\mathbb{P}\{X > t\}}{\mathbb{P}\{X > s\}} = e^{-\lambda(t-s)} = \mathbb{P}\{X > t - s\}. \quad (2.1.19)$$

La loi exponentielle est donc un analogue continu de la loi géométrique, cf. la Proposition 1.3.7.

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

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x) dx . \quad (2.1.20)$$

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

$$V(X) = \int_{-\infty}^{\infty} (x - \mathbb{E}(X))^2 f(x) dx . \quad (2.1.21)$$

Exemple 2.1.8.

1. *Loi normale standard:* On a

$$\lim_{L \rightarrow \infty} \int_{-L}^L |x| \varphi(x) dx = \lim_{L \rightarrow \infty} 2 \int_0^L x \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \lim_{L \rightarrow \infty} \frac{2}{\sqrt{2\pi}} \left[-e^{-x^2/2} \right]_0^L = \frac{2}{\sqrt{2\pi}} , \quad (2.1.22)$$

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, elle vaut

$$V(X) = \int_{-\infty}^{\infty} x^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \left[-x \frac{e^{-x^2/2}}{\sqrt{2\pi}} \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = 1 . \quad (2.1.23)$$

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 .

2. *Loi normale de moyenne μ et variance σ^2 :* Nous noterons $X \sim \mathcal{N}(\mu, \sigma^2)$ si elle admet la densité $\varphi(x; \mu, \sigma^2)$ introduite en (2.1.5). Le changement de variable $y = (x - \mu)/\sigma$ donne

$$\begin{aligned} \mathbb{E}(X) &= \int_{-\infty}^{\infty} x \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma y + \mu) e^{-y^2/2} dy = \mu , \\ V(X) &= \int_{-\infty}^{\infty} (x - \mu)^2 \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma} dx = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-y^2/2} dy = \sigma^2 . \end{aligned} \quad (2.1.24)$$

3. *Loi exponentielle de paramètre λ :*

$$\begin{aligned} \mathbb{E}(X) &= \int_0^{\infty} x \lambda e^{-\lambda x} dx = \left[-x e^{-\lambda x} \right]_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx = \frac{1}{\lambda} , \\ V(X) &= \int_0^{\infty} \left(x - \frac{1}{\lambda} \right)^2 \lambda e^{-\lambda x} dx = \left[-\left(x - \frac{1}{\lambda} \right)^2 e^{-\lambda x} \right]_0^{\infty} + \int_0^{\infty} 2 \left(x - \frac{1}{\lambda} \right) e^{-\lambda x} dx \\ &= \frac{1}{\lambda^2} + \frac{2}{\lambda} \mathbb{E}(X) - \frac{2}{\lambda} \int_0^{\infty} e^{-\lambda x} dx = \frac{1}{\lambda^2} . \end{aligned} \quad (2.1.25)$$

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

$$\lim_{L \rightarrow \infty} \int_{-L}^L |x| \frac{c}{\pi} \frac{1}{x^2 + c^2} dx = \frac{2c}{\pi} \lim_{L \rightarrow \infty} \int_0^L \frac{x}{x^2 + c^2} dx = \frac{c}{\pi} \lim_{L \rightarrow \infty} \left[\log(x^2 + c^2) \right]_0^L = +\infty , \quad (2.1.26)$$

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).

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

$$G_X(z) = \int_{-\infty}^{\infty} e^{izx} f(x) dx \equiv \mathbb{E}(e^{izX}). \quad (2.1.27)$$

On a $G_X(0) = 1$, et pour z réel,

$$|G_X(z)| \leq \int_{-\infty}^{\infty} |e^{izx}| f(x) dx = 1, \quad (2.1.28)$$

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

$$G'_X(z) = i \int_{-\infty}^{\infty} x e^{izx} f(x) dx, \quad (2.1.29)$$

et par conséquent

$$G'_X(0) = i\mathbb{E}(X), \quad G''_X(0) = -\mathbb{E}(X^2). \quad (2.1.30)$$

La fonction génératrice 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

$$G_X(z) = \lambda \int_0^{\infty} e^{izx - \lambda x} dx = \left[\frac{\lambda}{iz - \lambda} e^{(iz - \lambda)x} \right]_0^{\infty} = \frac{\lambda}{\lambda - iz}. \quad (2.1.31)$$

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

$$G_X(z) = \frac{1}{1 - iz/\lambda} = \sum_{k=0}^{\infty} \left(\frac{iz}{\lambda} \right)^k, \quad (2.1.32)$$

et comme par ailleurs

$$G_X(z) = \mathbb{E}(e^{izX}) = \sum_{k=0}^{\infty} \frac{(iz)^k}{k!} \mathbb{E}(X^k), \quad (2.1.33)$$

on obtient par identification

$$\mathbb{E}(X^k) = \frac{k!}{\lambda^k}. \quad (2.1.34)$$

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

$$G_X(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{izx - x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-iz)^2/2} dx e^{-z^2/2} = e^{-z^2/2}. \quad (2.1.35)$$

En développant l'exponentielle, on obtient

$$G_X(z) = \sum_{\ell=0}^{\infty} \frac{(-1)^\ell}{\ell!} \frac{z^{2\ell}}{2^\ell}, \quad (2.1.36)$$

ce qui donne

$$\mathbb{E}(X^k) = \begin{cases} 0 & \text{si } k \text{ est impair ,} \\ \frac{(2\ell)!}{2^{\ell\ell!}} & \text{si } k = 2\ell \text{ est pair .} \end{cases} \quad (2.1.37)$$

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

$$\begin{aligned} \frac{(2\ell)!}{2^{\ell\ell!}} &= \frac{2\ell(2\ell-2)\dots 4\cdot 2}{2^{\ell\ell!}}(2\ell-1)(2\ell-3)\dots 5\cdot 3\cdot 1 \\ &= (2\ell-1)(2\ell-3)\dots 5\cdot 3\cdot 1, \end{aligned} \quad (2.1.38)$$

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

2.2 Vecteurs aléatoires à densité

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 l'intégrale de f_1 est égale à celle de

$$f_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_1. \quad (2.2.2)$$

On note alors

$$\int_{\mathbb{R}^2} f(x_1, x_2) dx_1 dx_2 := \int_{-\infty}^{\infty} f_1(x_1) dx_1 \equiv \int_{-\infty}^{\infty} f_2(x_2) dx_2. \quad (2.2.3)$$

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

$$\int_B f(x_1, x_2) dx_1 dx_2 = \int_{\mathbb{R}^2} 1_B(x_1, x_2) f(x_1, x_2) dx_1 dx_2, \quad (2.2.4)$$

pour autant que cette dernière intégrale existe, où

$$1_B(x_1, x_2) = \begin{cases} 1 & \text{si } (x_1, x_2) \in B, \\ 0 & \text{sinon} \end{cases} \quad (2.2.5)$$

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

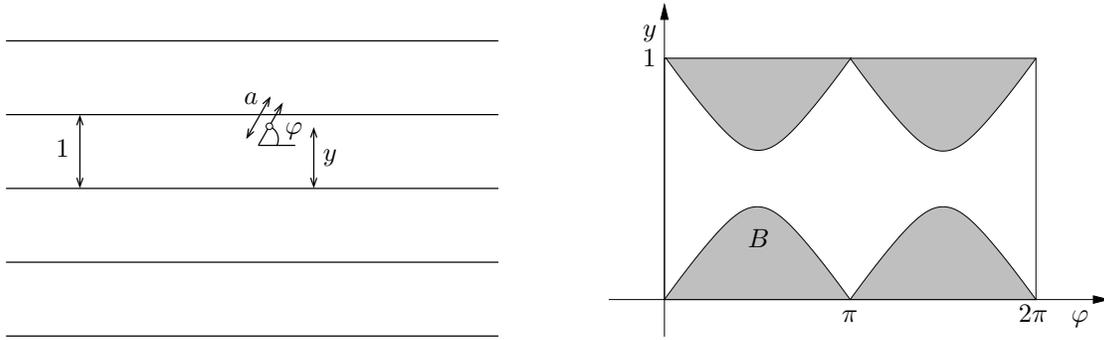


FIGURE 2.2. Expérience de l'aiguille de Buffon. Le domaine B correspond aux valeurs de φ et y pour lesquelles il y a une intersection entre l'aiguille et une ligne horizontale.

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

$$\int_{\mathbb{R}^n} f(x_1, \dots, x_n) dx_1 \dots dx_n = 1. \quad (2.2.6)$$

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

$$\mathbb{P}\{X_1 \leq a_1, \dots, X_n \leq a_n\} = \int_{]-\infty, a_1] \times \dots \times]-\infty, a_n]} f(x_1, \dots, x_n) dx_1 \dots dx_n \quad (2.2.7)$$

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 (2.2.7) 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

$$\mathbb{P}\{X \in B\} = \int_B f(x_1, \dots, x_n) dx_1 \dots dx_n. \quad (2.2.8)$$

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:

$$f(y, \varphi) = \frac{1}{2\pi} 1_{[0, 1[\times [0, 2\pi[}(y, \varphi). \quad (2.2.9)$$

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

$$B = \left\{ (y, \varphi) \in [0, 1[\times [0, 2\pi[: y - \frac{a}{2} |\sin \varphi| < 0 \text{ ou } y + \frac{a}{2} |\sin \varphi| > 1 \right\}. \quad (2.2.10)$$

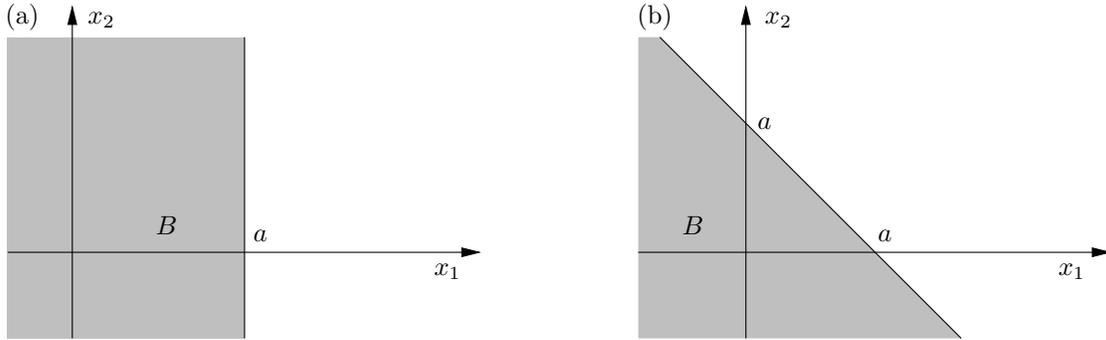


FIGURE 2.3. (a) Domaine d'intégration déterminant la densité marginale d'une densité conjointe. (b) Domaine d'intégration déterminant la densité de la somme $X_1 + X_2$.

La probabilité d'une intersection est donc donnée par

$$\begin{aligned} \mathbb{P}\{(y, \varphi) \in B\} &= \int_B f(y, \varphi) \, dy \, d\varphi = 4 \int_{\{(y, \varphi): 0 \leq \varphi \leq \pi, 0 \leq y \leq \frac{a}{2} \sin \varphi\}} \frac{1}{2\pi} \, dy \, d\varphi \\ &= \frac{4}{2\pi} \int_0^\pi \int_0^{\frac{a}{2} \sin \varphi} \, dy \, d\varphi = \frac{2}{\pi} \int_0^\pi \frac{a}{2} \sin \varphi \, d\varphi = \frac{2a}{\pi}. \end{aligned} \quad (2.2.11)$$

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).

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

$$\mathbb{P}\{X_1 \leq a\} = \mathbb{P}\{X \in B\} = \int_{-\infty}^a \underbrace{\int_{-\infty}^{\infty} f(x_1, x_2) \, dx_2}_{f_1(x_1)} \, dx_1. \quad (2.2.12)$$

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

$$\begin{aligned} \mathbb{P}\{X_1 + X_2 \leq a\} &= \int_{\{(x_1, x_2): x_1 \leq a - x_2\}} f(x_1, x_2) \, dx_1 \, dx_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{a - x_2} f(x_1, x_2) \, dx_1 \, dx_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^a f(z - x_2, x_2) \, dz \, dx_2 \\ &= \int_{-\infty}^a \underbrace{\left[\int_{-\infty}^{\infty} f(z - x_2, x_2) \, dx_2 \right]}_{=: g(z)} \, dz. \end{aligned} \quad (2.2.13)$$

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

$$g(z) = \int_{-\infty}^{\infty} f(z - x_2, x_2) \, dx_2. \quad (2.2.14)$$

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*

$$\mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2) . \quad (2.2.15)$$

DÉMONSTRATION. En utilisant (2.2.15) et le changement de variable $z = x_1 + x_2$, on a

$$\begin{aligned} \mathbb{E}(X_1 + X_2) &= \int_{-\infty}^{\infty} z g(z) dz = \int_{-\infty}^{\infty} z \int_{-\infty}^{\infty} f(z - x_2, x_2) dx_2 dz \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 + x_2) f(x_1, x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} x_1 \underbrace{\int_{-\infty}^{\infty} f(x_1, x_2) dx_2}_{f_1(x_1)} dx_1 + \int_{-\infty}^{\infty} x_2 \underbrace{\int_{-\infty}^{\infty} f(x_1, x_2) dx_1}_{f_2(x_2)} dx_2 \\ &= \mathbb{E}(X_1) + \mathbb{E}(X_2) . \end{aligned} \quad (2.2.16)$$

□

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

$$\mathbb{P}\{X_1 \leq a_1, \dots, X_n \leq a_n\} = \mathbb{P}\{X_1 \leq a_1\} \dots \mathbb{P}\{X_n \leq a_n\} \quad (2.2.17)$$

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*

$$f(x_1, \dots, x_n) := f_1(x_1) \dots f_n(x_n) \quad (2.2.18)$$

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

DÉMONSTRATION.

⇐ Si f est la densité de X , alors

$$\begin{aligned} \mathbb{P}\{X_1 \leq a_1, \dots, X_n \leq a_n\} &= \int_{-\infty}^{a_1} \dots \int_{-\infty}^{a_n} f_1(x_1) \dots f_n(x_n) dx_1 \dots dx_n \\ &= \int_{-\infty}^{a_1} f_1(x_1) dx_1 \dots \int_{-\infty}^{a_n} f_n(x_n) dx_n \\ &= \mathbb{P}\{X_1 \leq a_1\} \dots \mathbb{P}\{X_n \leq a_n\} . \end{aligned} \quad (2.2.19)$$

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

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

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

Remarquons que les $f_i(x_i)$ sont évidemment les densités marginales de $f(x_1, \dots, x_n)$. La relation (2.2.13) 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*

$$g(z) = \int_{-\infty}^{\infty} f_1(z - x_2) f_2(x_2) dx_2. \quad (2.2.21)$$

Ceci motive la définition suivante.

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

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

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

$$u = u(y_2) = \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sigma_1 \sigma_2} y_2 - \frac{\sigma_2}{\sigma_1 \sqrt{\sigma_1^2 + \sigma_2^2}} z. \quad (2.2.23)$$

Nous allons utiliser le fait que

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

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

$$\begin{aligned} g(z) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \sigma_1} e^{-(z-y_2)^2/2\sigma_1^2} \frac{1}{\sqrt{2\pi} \sigma_2} e^{-y_2^2/2\sigma_2^2} dy_2 \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left(\frac{(z-y_2)^2}{\sigma_1^2} + \frac{y_2^2}{\sigma_2^2}\right)\right\} dy_2 \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} e^{-u^2/2} e^{-z^2/2(\sigma_1^2+\sigma_2^2)} \frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} du \\ &= \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} e^{-z^2/2(\sigma_1^2+\sigma_2^2)} = \varphi(z; 0, \sigma_1^2 + \sigma_2^2), \end{aligned} \quad (2.2.25)$$

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

La densité conjointe des n variables gaussiennes indépendantes est donnée par

$$f(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2} \sigma_1 \dots \sigma_n} \exp \left\{ -\frac{1}{2} \left[\frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \dots + \frac{(x_n - \mu_n)^2}{\sigma_n^2} \right] \right\}, \quad (2.2.26)$$

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

$$f(x) = f(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}} e^{-\|x\|^2/2}. \quad (2.2.27)$$

Par l'indépendance, nous aurons $\text{cov}(X_i, X_j) = 0$ pour $i \neq j$, alors que $V(X_i) = 1$, 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

$$Y = TX + \mu, \quad (2.2.28)$$

dont l'espérance vaut μ . Quelle est sa densité? Nous avons, en vertu de la formule de changement de variable,

$$\begin{aligned} \mathbb{P}\{Y \in B\} &= \mathbb{P}\{X \in S(B - \mu)\} = \int_{S(B - \mu)} f(x) dx_1 \dots dx_n \\ &= \int_B f(S(y - \mu)) |\det S| dy_1 \dots dy_n. \end{aligned} \quad (2.2.29)$$

Par conséquent, la densité de Y est donnée par

$$g(y) = \frac{|\det S|}{(2\pi)^{n/2}} e^{-\|S(y - \mu)\|^2/2}. \quad (2.2.30)$$

Remarquons que l'on peut récrire

$$\begin{aligned} \|S(y - \mu)\|^2 &= \langle S(y - \mu) | S(y - \mu) \rangle = \langle y - \mu | A(y - \mu) \rangle \\ &= \sum_{i,j=1}^n a_{ij} (y_i - \mu_i)(y_j - \mu_j), \end{aligned} \quad (2.2.31)$$

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

$$g(y_1, \dots, y_n) = \frac{\sqrt{\det A}}{(2\pi)^{n/2}} \exp \left\{ -\frac{1}{2} \sum_{i,j=1}^n a_{ij} (y_i - \mu_i)(y_j - \mu_j) \right\}, \quad (2.2.32)$$

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

$$\begin{aligned} \mathbb{E}(Y_i Y_j) &= \mathbb{E} \left(\left[\sum_{k=1}^n T_{ik} X_k + \mu_i \right] \left[\sum_{\ell=1}^n T_{j\ell} X_\ell + \mu_j \right] \right) \\ &= \sum_{k,\ell=1}^n T_{ik} T_{j\ell} \mathbb{E}(X_k X_\ell) + \mu_i \mu_j \\ &= \sum_{k=1}^n T_{ik} T_{jk} + \mu_i \mu_j = (TT^T)_{ij} + \mu_i \mu_j, \end{aligned} \quad (2.2.33)$$

d'où, comme $TT^T = A^{-1}$,

$$\text{cov}(Y_i, Y_j) = (A^{-1})_{ij}. \quad (2.2.34)$$

La matrice $C = A^{-1}$ s'appelle la *matrice de covariance* de la densité gaussienne (2.2.32). 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 2.2.11 (Densité gaussienne). *Soit C une matrice $n \times n$ symétrique, définie positive, et $\mu \in \mathbb{R}^n$. La densité*

$$\varphi(x; \mu, C) = \frac{1}{\sqrt{(2\pi)^n \det C}} \exp\left\{-\frac{1}{2}\langle x - \mu | C^{-1}(x - \mu) \rangle\right\} \quad (2.2.35)$$

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

Voici pour terminer une autre propriété remarquable des densités gaussiennes.

Proposition 2.2.12. *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

$$a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 = a_{11}\left(x_1 + \frac{a_{12}}{a_{11}}x_2\right)^2 + \frac{a_{11}a_{22} - a_{12}^2}{a_{11}}x_2^2 \quad (2.2.36)$$

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.

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 .

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}):

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

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 probabilisé 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

Introduction aux processus stochastiques

Jusqu'ici, nous avons essentiellement considéré des suites de variables aléatoires indépendantes. Les processus stochastiques sont des objets plus généraux, comprenant des suites de variables aléatoires qui ne sont pas indépendantes. En général, on interprète l'indice numérotant ces variables comme le temps. Un cas important est celui des processus markoviens, pour lesquels chaque variable aléatoire de la suite ne dépend que de la précédente. La marche aléatoire et les chaînes de Markov sont des exemples de processus stochastiques markoviens, que nous allons étudier plus en détail.

3.1 La marche aléatoire unidimensionnelle symétrique

Considérons une expérience de Pile ou Face. Pour N jets successifs, on peut prendre comme univers le produit $\Omega_N = \{P, F\}^N$. Considérons alors la variable aléatoire S_k égale à la différence entre le nombre de Pile et de Face obtenus lors des k premiers jets. On peut écrire

$$S_k = \sum_{j=1}^k X_j \quad \text{où } X_j = \begin{cases} 1 & \text{si le } j\text{-ème jet donne Pile,} \\ -1 & \text{si le } j\text{-ème jet donne Face.} \end{cases} \quad (3.1.1)$$

La *marche aléatoire unidimensionnelle symétrique* est la suite des S_k obtenue dans le cas

$$\mathbb{P}\{X_j = 1\} = \mathbb{P}\{X_j = -1\} = \frac{1}{2}, \quad (3.1.2)$$

les X_j étant supposés indépendants. Les *réalisations* du processus sont des suites $\{S_k(\omega)\}$ d'entiers telles que $S_0(\omega) = 0$ et $S_{k+1}(\omega) - S_k(\omega) = \pm 1$ (Figure 3.1). La probabilité de chaque suite de longueur k est 2^{-k} .

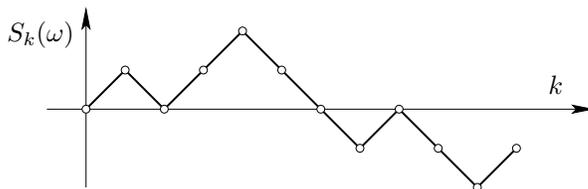


FIGURE 3.1. Une réalisation d'une marche aléatoire $S_k(\omega)$.

Commençons par établir quelques propriétés élémentaires de la variable aléatoire S_k .

Proposition 3.1.1.

1. L'image de S_k est $S_k(\Omega_N) = \{-k, -k+2, \dots, k-2, k\}$.
2. La loi de S_k est binomiale, donnée par

$$\mathbb{P}\{S_k = n\} = b\left(\frac{k+n}{2}; k, \frac{1}{2}\right). \quad (3.1.3)$$

3. On a $\mathbb{E}(S_k) = 0$ et $V(S_k) = k$.

DÉMONSTRATION.

1. Suit du fait que $S_0 = 0$ et $S_{k+1} - S_k \in \{-1, 1\}$.
2. On a

$$\begin{aligned} \mathbb{P}\{S_k = n\} &= \frac{\text{nombre de réalisations avec } \frac{k+n}{2} \text{ Pile et } \frac{k-n}{2} \text{ Face}}{\text{nombre de réalisations de longueur } k} \\ &= \frac{1}{2^k} \binom{k}{\frac{k+n}{2}} = \frac{1}{2^k} \frac{k!}{\left(\frac{k+n}{2}\right)! \left(\frac{k-n}{2}\right)!} = b\left(\frac{k+n}{2}; k, \frac{1}{2}\right). \end{aligned} \quad (3.1.4)$$

3. Comme $\mathbb{E}(X_i) = 0$ et $V(X_i) = 1$, on a $\mathbb{E}(S_k) = 0$ et $V(S_k) = k$. □

La Proposition 3.1.1 implique que S_k est nul en moyenne, mais que ses fluctuations sont d'ordre \sqrt{k} . En particulier, la probabilité que le processus se trouve en 0 au k -ème pas est donnée par

$$\mathbb{P}\{S_k = 0\} = \begin{cases} 0 & \text{si } k \text{ est impair,} \\ b(\ell; 2\ell, \frac{1}{2}) = \frac{(2\ell)!}{2^{2\ell}(\ell!)^2} & \text{si } k = 2\ell \text{ est pair.} \end{cases} \quad (3.1.5)$$

Remarquons que la formule de Stirling implique que pour ℓ grand,

$$\mathbb{P}\{S_{2\ell} = 0\} \sim \frac{1}{2^{2\ell}} \frac{\sqrt{4\pi\ell} e^{-2\ell} (2\ell)^{2\ell}}{2\pi\ell e^{-2\ell} \ell^{2\ell}} = \frac{1}{\sqrt{\pi\ell}}. \quad (3.1.6)$$

Cependant, la loi de chaque S_k ne détermine pas le processus, les S_k n'étant pas indépendants. Voici d'abord quelques propriétés simples du processus considéré dans son ensemble:

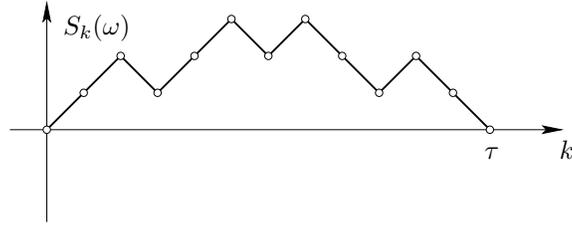
Proposition 3.1.2.

1. Propriété de Markov: Pour tout $k > \ell \geq 0$, on a

$$\mathbb{P}\{S_k = n | S_\ell = m_\ell, S_{\ell-1} = m_{\ell-1}, \dots, S_1 = m_1\} = \mathbb{P}\{S_k = n | S_\ell = m_\ell\}. \quad (3.1.7)$$

2. Incréments indépendants: Pour tout $k > \ell \geq 0$, $S_k - S_\ell$ est indépendant de S_1, \dots, S_ℓ .
3. Incréments stationnaires: Pour tout $k > \ell \geq 0$,

$$\mathbb{P}\{S_k = n | S_\ell = m\} = \mathbb{P}\{S_{k-\ell} = n - m\}. \quad (3.1.8)$$

FIGURE 3.2. Une réalisation d'une marche aléatoire pour laquelle $\tau = 12$.

DÉMONSTRATION. Nous pouvons décomposer

$$S_k = S_\ell + Y, \quad Y = \sum_{j=\ell+1}^k X_j. \quad (3.1.9)$$

L'indépendance des X_i implique que $Y = S_k - S_\ell$ est indépendante de X_1, \dots, X_ℓ , donc aussi de S_1, \dots, S_ℓ . Il suit que

$$\begin{aligned} \mathbb{P}\{S_k = n | S_\ell = m_\ell, \dots, S_1 = m_1\} &= \frac{\mathbb{P}\{Y = n - m_\ell, S_\ell = m_\ell, \dots, S_1 = m_1\}}{\mathbb{P}\{S_\ell = m_\ell, \dots, S_1 = m_1\}} \\ &= \mathbb{P}\{Y = n - m_\ell\}. \end{aligned} \quad (3.1.10)$$

Par le même raisonnement, nous avons également

$$\mathbb{P}\{S_k = n | S_\ell = m_\ell\} = \mathbb{P}\{Y = n - m_\ell\}. \quad (3.1.11)$$

Finalement,

$$\mathbb{P}\{Y = n - m\} = \mathbb{P}\left\{\sum_{j=\ell+1}^k X_j = n - m\right\} = \mathbb{P}\left\{\sum_{j=1}^{k-j} X_j = n - m\right\} = \mathbb{P}\{S_{k-\ell} = n - m\}, \quad (3.1.12)$$

du fait que les X_i sont identiquement distribués. \square

La propriété de Markov signifie que l'évolution du processus ne dépend que de son état présent, indépendamment de son passé. Elle permet d'écrire des relations du genre

$$\begin{aligned} \mathbb{P}\{S_k = n, S_{k-1} = m, (S_1, \dots, S_{k-2}) \in A\} \\ = \mathbb{P}\{S_k = n | S_{k-1} = m\} \mathbb{P}\{S_{k-1} = m, (S_1, \dots, S_{k-2}) \in A\}. \end{aligned} \quad (3.1.13)$$

Voyons maintenant comment déduire de ces propriétés des informations non triviales sur les réalisations du processus. Une première quantité intéressante est le temps τ du premier retour du processus en 0 (Figure 3.2):

$$\tau = \inf\{k \geq 1: S_k = 0\}. \quad (3.1.14)$$

Par exemple, dans l'expérience de jet de pièce de monnaie, τ est le nombre de fois que l'on jette la pièce jusqu'à obtenir pour la première fois autant de Pile que de Face. Quelle est la loi de τ ? Il est clair que τ ne peut prendre que des valeurs paires. De plus, si $\tau = k$ alors $S_k = 0$, donc $\mathbb{P}\{\tau = k\} \leq \mathbb{P}\{S_k = 0\}$. En fait, il nous faut déterminer

$$\mathbb{P}\{\tau = k\} = \mathbb{P}\{S_1 \neq 0, S_2 \neq 0, \dots, S_{k-1} \neq 0, S_k = 0\}. \quad (3.1.15)$$

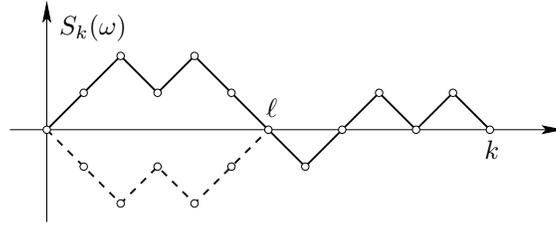


FIGURE 3.3. Pour chaque réalisation d'une marche aléatoire avec $\tau = \ell < k$ telle que $S_1 = 1$, il existe une autre réalisation telle que $\tau = \ell$ et $S_1 = -1$, obtenue par réflexion par rapport à l'axe des abscisses.

Théorème 3.1.3. *La loi de τ est donnée par*

$$\mathbb{P}\{\tau = k\} = \begin{cases} 0 & \text{pour } k \text{ impair,} \\ \frac{1}{k} \mathbb{P}\{S_{k-2} = 0\} & \text{pour } k \text{ pair.} \end{cases} \quad (3.1.16)$$

DÉMONSTRATION. Supposons que $\tau = k$. Comme le processus ne peut pas changer de signe sans passer par 0, on a

$$\begin{aligned} \mathbb{P}\{\tau = k\} &= \mathbb{P}\{S_1 > 0, S_2 > 0, \dots, S_{k-1} > 0, S_k = 0\} \\ &\quad + \mathbb{P}\{S_1 < 0, S_2 < 0, \dots, S_{k-1} < 0, S_k = 0\} \\ &= 2\mathbb{P}\{S_1 > 0, S_2 > 0, \dots, S_{k-1} > 0, S_k = 0\} \\ &= 2\mathbb{P}\{S_1 = 1, S_2 > 0, \dots, S_{k-2} > 0, S_{k-1} = 1, S_k = 0\} \\ &= 2\mathbb{P}\{S_k = 0 | S_{k-1} = 1\} \mathbb{P}\{S_1 = 1, S_2 > 0, \dots, S_{k-2} > 0, S_{k-1} = 1\}, \end{aligned} \quad (3.1.17)$$

où nous avons utilisé la propriété de Markov dans la dernière ligne. La propriété des incréments stationnaires implique

$$\mathbb{P}\{S_k = 0 | S_{k-1} = 1\} = \mathbb{P}\{S_1 = -1\} = \frac{1}{2}. \quad (3.1.18)$$

Il suit que

$$\begin{aligned} \mathbb{P}\{\tau = k\} &= \mathbb{P}\{S_1 = 1, S_2 > 0, \dots, S_{k-2} > 0, S_{k-1} = 1\} \\ &= \mathbb{P}\{S_1 = S_{k-1} = 1\} - \mathbb{P}\{S_1 = S_{k-1} = 1, \exists \ell \in \{2, \dots, k-2\} : S_\ell = 0\}. \end{aligned} \quad (3.1.19)$$

Nous utilisons maintenant un argument important, appelé le *principe de réflexion*: A tout chemin allant de $(1, 1)$ à $(k-1, 1)$ passant par 0, on peut faire correspondre un unique chemin de $(-1, 1)$ à $(k-1, 1)$, obtenu en réfléchissant par rapport à l'axe des abscisses la partie du chemin antérieure au premier passage en 0. On a donc

$$\mathbb{P}\{S_1 = S_{k-1} = 1, \exists \ell \in \{2, \dots, k-2\} : S_\ell = 0\} = \mathbb{P}\{S_1 = -1, S_{k-1} = 1\}. \quad (3.1.20)$$

Finalement, en appliquant de nouveau la propriété des incréments stationnaires, on voit que

$$\begin{aligned} \mathbb{P}\{S_1 = 1, S_{k-1} = 1\} &= \mathbb{P}\{S_{k-1} = 1 | S_1 = 1\} \mathbb{P}\{S_1 = 1\} = \mathbb{P}\{S_{k-2} = 0\} \cdot \frac{1}{2}, \\ \mathbb{P}\{S_1 = -1, S_{k-1} = 1\} &= \mathbb{P}\{S_{k-1} = 1 | S_1 = -1\} \mathbb{P}\{S_1 = -1\} = \mathbb{P}\{S_{k-2} = 2\} \cdot \frac{1}{2}. \end{aligned} \quad (3.1.21)$$

En remplaçant dans (3.1.19), il vient

$$\mathbb{P}\{\tau = k\} = \frac{1}{2} [\mathbb{P}\{S_{k-2} = 0\} - \mathbb{P}\{S_{k-2} = 2\}]. \quad (3.1.22)$$

Le reste de la preuve est un calcul direct. Comme

$$\frac{\mathbb{P}\{S_{k-2} = 2\}}{\mathbb{P}\{S_{k-2} = 0\}} = \frac{\binom{k-2}{k/2}}{\binom{k-2}{k/2-1}} = \frac{(\frac{k}{2}-1)!(\frac{k}{2}-1)!}{(\frac{k}{2})!(\frac{k}{2}-2)!} = \frac{\frac{k}{2}-1}{\frac{k}{2}} = 1 - \frac{2}{k}, \quad (3.1.23)$$

on obtient

$$\mathbb{P}\{\tau = k\} = \frac{1}{2} \mathbb{P}\{S_{k-2} = 0\} \left[1 - 1 + \frac{2}{k} \right] = \frac{1}{k} \mathbb{P}\{S_{k-2} = 0\}, \quad (3.1.24)$$

ce qui conclut la démonstration. \square

Le tableau suivant donne les premières valeurs de la loi de τ :

k	2	4	6	8	10	12	14
$\mathbb{P}\{\tau = k\}$	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{5}{128}$	$\frac{7}{256}$	$\frac{21}{1024}$	$\frac{33}{2048}$
	= 0.5	= 0.125	$\cong 0.063$	$\cong 0.039$	$\cong 0.027$	$\cong 0.021$	$\cong 0.016$

Il est donc assez probable de revenir rapidement en 0, puis la loi prend des valeurs plutôt faibles, tout en décroissant lentement. Il suit de (3.1.6) que pour des grands k , $\mathbb{P}\{\tau = k\}$ décroît comme $1/k^{3/2}$. Ce fait a une conséquence surprenante:

Corollaire 3.1.4. $\mathbb{E}(\tau) = +\infty$.

DÉMONSTRATION. On a

$$\mathbb{E}(\tau) = \sum_{k \geq 1} k \mathbb{P}\{\tau = k\} = \sum_{\ell \geq 1} 2\ell \frac{1}{2\ell} \mathbb{P}\{S_{2\ell-2} = 0\} \sim \sum_{\ell \geq 1} \frac{1}{\sqrt{\pi\ell}} = +\infty. \quad (3.1.25)$$

\square

En d'autres termes, la marche aléatoire finit toujours par revenir en 0, mais la loi de τ décroît trop lentement pour que son espérance soit finie. Cela est lié au fait que si la marche aléatoire s'éloigne beaucoup de 0, il lui faut longtemps pour y revenir.

Considérons une autre application du principe de réflexion. Supposons qu'une machine à sous fonctionne avec une mise d'un euro, et que la machine rende soit deux euros, soit rien, avec la même probabilité. Combien de fois peut-on jouer, si l'on possède initialement 10 euros? La somme que l'on possède après avoir joué k fois est égale à $10 - S_k$, on peut donc jouer jusqu'à ce que $S_k = 10$. Il nous faut donc déterminer la loi de

$$\tau_L = \inf\{k \geq 0: S_k = L\} \quad (3.1.26)$$

pour $L = 10$ dans notre exemple.

Théorème 3.1.5. La loi de τ_L est donnée par

$$\mathbb{P}\{\tau_L = k\} = \begin{cases} \frac{|L|}{k} \mathbb{P}\{S_k = L\} & \text{pour } k \in \{|L|, |L| + 2, \dots\}, \\ 0 & \text{sinon.} \end{cases} \quad (3.1.27)$$

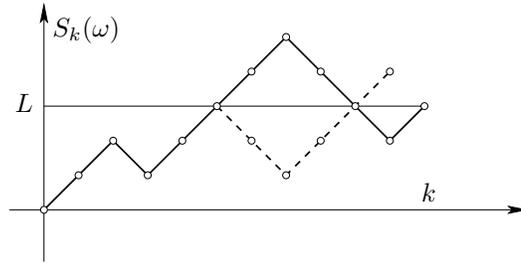


FIGURE 3.4. Pour chaque réalisation d'une marche aléatoire avec $\tau_L < k$, et telle que $S_{k-1} = L - 1$, il existe une autre réalisation telle que $S_{k-1} = L + 1$, obtenue par réflexion par rapport à la droite d'ordonnée L .

DÉMONSTRATION. Considérons le cas $L > 0$. On a

$$\begin{aligned}
 \mathbb{P}\{\tau_L = k\} &= \mathbb{P}\{S_1 < L, S_2 < L, \dots, S_{k-1} = L - 1, S_k = L\} \\
 &= \mathbb{P}\{S_k = L | S_{k-1} = L - 1\} \mathbb{P}\{S_1 < L, S_2 < L, \dots, S_{k-1} = L - 1\} \\
 &= \frac{1}{2} \mathbb{P}\{S_1 < L, S_2 < L, \dots, S_{k-1} = L - 1\} \\
 &= \frac{1}{2} [\mathbb{P}\{S_{k-1} = L - 1\} - \mathbb{P}\{S_{k-1} = L - 1, \exists j \in \{1, \dots, k-2\} : S_j = L\}].
 \end{aligned} \tag{3.1.28}$$

Invoquons alors à nouveau le principe de réflexion (Figure 3.4). A chaque réalisation telle que $S_{k-1} = L - 1$, qui a déjà atteint le niveau L auparavant, on peut associer une réalisation telle que $S_{k-1} = L + 1$. Nous avons donc

$$\begin{aligned}
 \mathbb{P}\{\tau_L = k\} &= \frac{1}{2} [\mathbb{P}\{S_{k-1} = L - 1\} - \mathbb{P}\{S_{k-1} = L + 1\}] \\
 &= \frac{1}{2} \frac{1}{2^{k-1}} \left[\binom{k-1}{\frac{k+L}{2}-1} - \binom{k-1}{\frac{k-L}{2}} \right] \\
 &= \frac{(k-1)!}{2^k} \left[\frac{1}{\left(\frac{k+L}{2}-1\right)! \left(\frac{k-L}{2}\right)!} - \frac{1}{\left(\frac{k+L}{2}\right)! \left(\frac{k-L}{2}-1\right)!} \right] \\
 &= \frac{1}{2^k} \frac{k!}{\left(\frac{k+L}{2}\right)! \left(\frac{k-L}{2}\right)!} \frac{1}{k} \left[\frac{k+L}{2} - \frac{k-L}{2} \right] \\
 &= \mathbb{P}\{S_k = L\} \frac{L}{k}.
 \end{aligned} \tag{3.1.29}$$

Le cas $L < 0$ suit par symétrie. \square

Pour des raisons similaires à celles du cas du retour en 0, la loi de τ_L décroît en $1/k^{3/2}$, et son espérance est donc infinie. On est donc sûr de perdre tôt ou tard, tout en pouvant espérer jouer infiniment longtemps. Le tableau suivant donne les premières valeurs de la loi dans le cas $L = 2$, donc si on ne possède que deux euros au début. On constate que si l'on ne perd pas lors des 6 premiers coups, la probabilité de perdre en k coups change très peu d'un k au suivant.

k	2	4	6	8	10	12	14
$\mathbb{P}\{\tau_2 = k\}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{5}{64}$	$\frac{7}{128}$	$\frac{21}{512}$	$\frac{33}{1024}$	$\frac{429}{16384}$
	= 0.25	= 0.125	\cong 0.078	\cong 0.054	\cong 0.041	\cong 0.032	\cong 0.026

3.2 Le processus ponctuel de Poisson

Le processus ponctuel de Poisson est un processus stochastique d'un type un peu différent, qui associe une distribution de probabilité aux configurations de points sur \mathbb{R}_+ . Ces points modélisent, par exemple, les temps de passage d'un bus à un arrêt, les instants d'arrivée d'appels téléphoniques dans une centrale, et ainsi de suite.

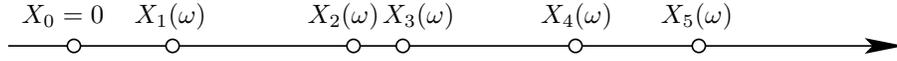


FIGURE 3.5. Une réalisation d'un processus de Poisson.

Le processus peut être caractérisé de plusieurs manières différentes. Une réalisation peut être spécifiée par une suite croissante de nombres réels positifs

$$X_0 = 0 < X_1(\omega) < X_2(\omega) < X_3(\omega) < \dots, \quad (3.2.1)$$

désignant les positions des points dans \mathbb{R}_+ . Alternativement, on peut décrire une réalisation en donnant le nombre de points $N_I(\omega)$ contenus dans chaque intervalle I de la forme $I =]t, t + s]$. Si nous abrégeons $N_{]0, t]}$ par N_t , nous aurons $N_{]t, t+s]} = N_{t+s} - N_t$, et les N_t sont donnés en fonction des X_n par

$$N_t(\omega) = \sup\{n \geq 0 : X_n(\omega) \leq t\}. \quad (3.2.2)$$

Inversement, les X_n se déduisent des N_t par la relation

$$X_n(\omega) = \inf\{t \geq 0 : N_t(\omega) \geq n\}. \quad (3.2.3)$$

Nous allons voir deux constructions équivalentes du processus de Poisson. La première construction part de la distribution des N_t .

Définition 3.2.1 (Processus de Poisson). *Le processus ponctuel de Poisson satisfait les conditions suivantes:*

1. N_I ne dépend que de la longueur de I , i.e. $N_{]t, t+s]}$ a la même loi que N_s .
2. Si I_1, \dots, I_k sont deux à deux disjoints, N_{I_1}, \dots, N_{I_k} sont indépendants.
3. $\mathbb{E}(N_I)$ existe pour tout I (de longueur finie).
4. Il existe un intervalle I tel que $\mathbb{P}\{N_I > 0\} > 0$.
5. Absence de points doubles: $\lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \mathbb{P}\{N_\varepsilon \geq 2\} = 0$.

Supposant qu'un tel processus existe bel et bien, nous pouvons dériver quelques-unes de ses propriétés.

Proposition 3.2.2.

1. Soit $\alpha(t) = \mathbb{E}(N_t)$. Alors il existe $\lambda > 0$ tel que $\alpha(t) = \lambda t$.
2. Pour tout intervalle $I \subset \mathbb{R}_+$, on a $\mathbb{P}\{N_I \geq 1\} \leq \mathbb{E}(N_I)$.

DÉMONSTRATION.

1. Comme $N_0 = 0$, on a $\alpha(0) = 0$. De plus, comme $N_{t+s} = N_t + N_{]t, t+s]}$ on a

$$\alpha(t + s) = \alpha(t) + \mathbb{E}(N_{]t, t+s]}) = \alpha(t) + \alpha(s), \quad (3.2.4)$$

en vertu de la condition 1. Par un résultat d'analyse, ceci implique nécessairement que $\alpha(t) = \lambda t$ pour un $\lambda \geq 0$, et la propriété 4. implique $\lambda > 0$.

2. C'est une conséquence directe du Lemme 1.4.1 et du fait que $N_I \geq 0$. \square

La propriété remarquable du processus de Poisson est alors que les variables aléatoires $N_{]t,t+s]}$ suivent nécessairement une loi de Poisson de paramètre λs .

Théorème 3.2.3. *Si le processus satisfait les 5 conditions de la Définition 3.2.1, alors les variables aléatoires $N_{]t,t+s]}$ suivent des lois de Poisson de paramètre λs :*

$$\mathbb{P}\{N_{]t,t+s]} = k\} = \pi_{\lambda s}(k) = e^{-\lambda s} \frac{(\lambda s)^k}{k!} . \quad (3.2.5)$$

DÉMONSTRATION. Par la propriété 1., il suffit de montrer le résultat pour $t = 0$, c'est-à-dire pour N_s . Partageons $]0, s]$ en k intervalles de longueur égale, de la forme

$$]s_{j-1}, s_j] \quad \text{où } s_j = \frac{j s}{k} \text{ pour } 0 \leq j \leq k . \quad (3.2.6)$$

L'idée de la preuve est que pour k suffisamment grand, il est peu probable d'avoir plus d'un point par intervalle, donc la loi de $Y_j^{(k)} = N_{]s_{j-1}, s_j]}$ est à peu près une loi de Bernoulli. La loi de N_s est donc proche d'une loi binomiale, que l'on peut approximer par la loi de Poisson pour k grand.

Il suit des conditions 1. et 2. que les $Y_j^{(k)}$ sont i.i.d., de même loi que $N_{s_1} = N_{s/k}$, et on a

$$N_s = \sum_{j=1}^k Y_j^{(k)} . \quad (3.2.7)$$

Introduisons alors des variables aléatoires

$$\bar{Y}_j^{(k)} = \begin{cases} 0 & \text{si } Y_j^{(k)} = 0 , \\ 1 & \text{si } Y_j^{(k)} \geq 1 . \end{cases} \quad (3.2.8)$$

Les $\bar{Y}_j^{(k)}$ sont également i.i.d.. La variable aléatoire

$$\bar{N}_s^{(k)} = \sum_{j=1}^k \bar{Y}_j^{(k)} , \quad (3.2.9)$$

satisfaisant $\bar{N}_s^{(k)} \leq N_s$ pour tout k , on a

$$\mathbb{P}\{\bar{N}_s^{(k)} \geq m\} \leq \mathbb{P}\{N_s \geq m\} \quad (3.2.10)$$

pour tout k et tout m . De plus, $\bar{N}_s^{(k)}$ suit une loi binomiale de paramètre

$$p_k = \mathbb{P}\{\bar{Y}_j^{(k)} = 1\} = \mathbb{P}\{Y_j^{(k)} \geq 1\} = \mathbb{P}\{N_{s/k} \geq 1\} . \quad (3.2.11)$$

Estimons maintenant la différence entre les lois de $\bar{N}_s^{(k)}$ et N_s . Nous avons

$$\begin{aligned} \mathbb{P}\{\bar{N}_s^{(k)} \neq N_s\} &= \mathbb{P}\{\exists j \in \{1, \dots, k\} : Y_j^{(k)} \geq 2\} \\ &\leq \sum_{j=1}^k \mathbb{P}\{Y_j^{(k)} \geq 2\} \\ &= k \mathbb{P}\{Y_1^{(k)} \geq 2\} = k \mathbb{P}\{N_{s/k} \geq 2\} . \end{aligned} \quad (3.2.12)$$

La condition 5. avec $\varepsilon = s/k$ implique alors

$$\lim_{k \rightarrow \infty} \mathbb{P}\{\bar{N}_s^{(k)} \neq N_s\} = 0. \quad (3.2.13)$$

Comme on a d'une part la minoration

$$\mathbb{P}\{N_s = m\} \geq \mathbb{P}\{N_s = \bar{N}_s^{(k)} = m\} \geq \mathbb{P}\{\bar{N}_s^{(k)} = m\} - \mathbb{P}\{\bar{N}_s^{(k)} \neq N_s\}, \quad (3.2.14)$$

et d'autre part la majoration

$$\begin{aligned} \mathbb{P}\{N_s = m\} &= \mathbb{P}\{\bar{N}_s^{(k)} = N_s = m\} + \mathbb{P}\{\bar{N}_s^{(k)} \neq N_s = m\} \\ &\leq \mathbb{P}\{\bar{N}_s^{(k)} = m\} + \mathbb{P}\{N_s \neq \bar{N}_s^{(k)}\}, \end{aligned} \quad (3.2.15)$$

il suit que

$$\lim_{k \rightarrow \infty} \mathbb{P}\{\bar{N}_s^{(k)} = m\} = \mathbb{P}\{N_s = m\}. \quad (3.2.16)$$

Il reste à montrer que kp_k tend vers λs pour $k \rightarrow \infty$. Si c'est le cas, alors la Proposition 1.6.3 montre que N_s suit une loi de Poisson de paramètre λs . Or nous avons

$$\begin{aligned} kp_k &= \mathbb{E}(\bar{N}_s^{(k)}) = \sum_{j=1}^{\infty} j \mathbb{P}\{\bar{N}_s^{(k)} = j\} = \sum_{j=1}^{\infty} \sum_{\ell=1}^j \mathbb{P}\{\bar{N}_s^{(k)} = j\} \\ &= \sum_{\ell=1}^{\infty} \sum_{j=\ell}^{\infty} \mathbb{P}\{\bar{N}_s^{(k)} = j\} = \sum_{\ell=1}^{\infty} \mathbb{P}\{\bar{N}_s^{(k)} \geq \ell\}. \end{aligned} \quad (3.2.17)$$

Un calcul analogue montre que

$$\lambda s = \mathbb{E}(N_s) = \sum_{\ell=1}^{\infty} \mathbb{P}\{N_s \geq \ell\}. \quad (3.2.18)$$

Il suit de (3.2.10) que $kp_k \leq \lambda s$ pour tout k . Par un théorème d'analyse, on peut alors intervertir limite et somme, et écrire

$$\lim_{k \rightarrow \infty} \sum_{\ell=1}^{\infty} \mathbb{P}\{\bar{N}_s^{(k)} \geq \ell\} = \sum_{\ell=1}^{\infty} \lim_{k \rightarrow \infty} \mathbb{P}\{\bar{N}_s^{(k)} \geq \ell\} = \sum_{\ell=1}^{\infty} \mathbb{P}\{N_s \geq \ell\} = \lambda s, \quad (3.2.19)$$

en vertu de (3.2.16). Ceci montre que kp_k converge bien vers λs , et par conséquent que la loi de N_s , étant la limite d'une loi binomiale de paramètre kp_k , est une loi de Poisson de paramètre λs . \square

La seconde construction du processus ponctuel de Poisson se base sur la distribution des différences de position $Z_n = X_n - X_{n-1}$. Celles-ci caractérisent également de manière univoque le processus, via la relation

$$X_n(\omega) = \sum_{j=1}^n Z_j(\omega). \quad (3.2.20)$$

Théorème 3.2.4. *Pour tout n , les variables aléatoires Z_1, \dots, Z_n sont indépendantes, et suivent la même loi exponentielle $\mathcal{E}xp(\lambda)$.*

DÉMONSTRATION. Fixons des instants

$$t_0 = 0 < s_1 < t_1 < s_2 < t_2 < \dots < s_n < t_n . \quad (3.2.21)$$

Nous pouvons alors calculer

$$\begin{aligned} & \mathbb{P}\{X_1 \in]s_1, t_1], X_2 \in]s_2, t_2], \dots, X_n \in]s_n, t_n]\} \\ &= \mathbb{P}\{N_{]0, s_1]} = 0, N_{]s_1, t_1]} = 1, N_{]t_1, s_2]} = 0, \dots, N_{]t_{n-1}, s_n]} = 0, N_{]s_n, t_n]} \geq 1\} \\ &= \prod_{k=1}^n \mathbb{P}\{N_{]t_{k-1}, s_k]} = 0\} \prod_{k=1}^{n-1} \mathbb{P}\{N_{]s_k, t_k]} = 1\} \mathbb{P}\{N_{]s_n, t_n]} \geq 1\} \\ &= \prod_{k=1}^n e^{-\lambda(s_k - t_{k-1})} \prod_{k=1}^{n-1} \lambda(t_k - s_k) e^{-\lambda(t_k - s_k)} [1 - e^{-\lambda(t_n - s_n)}] \\ &= \lambda^{n-1} \prod_{k=1}^{n-1} (t_k - s_k) [e^{-\lambda s_n} - e^{-\lambda t_n}] \\ &= \int_{s_1}^{t_1} \int_{s_2}^{t_2} \dots \int_{s_n}^{t_n} \lambda^n e^{-\lambda x_n} dx_n dx_{n-1} \dots dx_2 dx_1 . \end{aligned} \quad (3.2.22)$$

La loi conjointe de (X_1, \dots, X_n) admet donc la densité

$$f(x_1, \dots, x_n) = \begin{cases} \lambda^n e^{-\lambda x_n} & \text{si } 0 < x_1 < \dots < x_n , \\ 0 & \text{sinon .} \end{cases} \quad (3.2.23)$$

Nous pouvons alors calculer la fonction de répartition des Z_k :

$$\begin{aligned} \mathbb{P}\{Z_1 \leq z_1, \dots, Z_n \leq z_n\} &= \mathbb{P}\{X_1 \leq z_1, X_2 - X_1 \leq z_2, \dots, X_n - X_{n-1} \leq z_n\} \\ &= \mathbb{P}\{X_1 \leq z_1, X_2 \leq z_2 + X_1, \dots, X_n \leq z_n + X_{n-1}\} \\ &= \int_0^{z_1} \int_{x_1}^{z_2 + x_1} \dots \int_{x_{n-1}}^{z_n + x_{n-1}} f(x_1, \dots, x_n) dx_n \dots dx_1 . \end{aligned} \quad (3.2.24)$$

La densité conjointe de (Z_1, \dots, Z_n) s'obtient alors en calculant la dérivée

$$\begin{aligned} \frac{\partial^n}{\partial z_1 \dots \partial z_n} \mathbb{P}\{Z_1 \leq z_1, \dots, Z_n \leq z_n\} &= f(z_1, z_1 + z_2, \dots, z_1 + \dots + z_n) \\ &= \lambda^n e^{-\lambda(z_1 + \dots + z_n)} , \end{aligned} \quad (3.2.25)$$

Or cette densité est bien la densité conjointe de n variables exponentielles indépendantes de paramètre λ . \square

Ce résultat fournit une méthode permettant de construire le processus de Poisson: chaque X_n est obtenu à partir de X_{n-1} en lui ajoutant une variable aléatoire de loi exponentielle, indépendante des variables précédentes. Ceci présuppose l'existence d'un espace probabilisé pouvant contenir une infinité de variables aléatoires exponentielles indépendantes: C'est l'espace produit que nous avons évoqué (sans prouver son existence) dans l'Exemple 2.3.4.

3.3 Les chaînes de Markov

Dans cette section et les suivantes, nous allons étudier des chaînes de Markov sur un ensemble fini. En fait, tout processus stochastique sur un ensemble fini, à temps discret, et jouissant de la propriété de Markov, est une chaîne de Markov. Nous commençons par étudier quelques exemples, avant de donner une définition et des propriétés générales.

Exemple 3.3.1. Dans des temps très reculés, il n'y avait en France que deux chaînes de télévision. Admettons qu'un téléspectateur choisisse entre ces chaînes de la manière suivante:

- il commence par regarder la première chaîne;
- au bout de 10 minutes, il change de chaîne avec probabilité p ;
- 10 minutes plus tard, il change de chaîne avec probabilité p s'il regarde la première chaîne, et avec probabilité q s'il regarde la seconde chaîne;
- le processus se répète indéfiniment.

On peut associer le graphe suivant au processus:

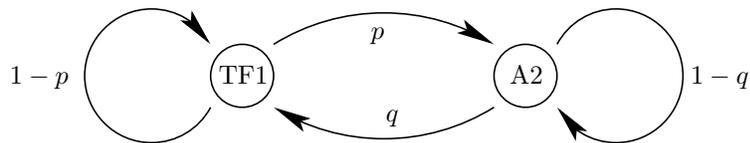


FIGURE 3.6. Une chaîne de Markov à deux états.

Au bout d'un temps très long, avec quelle probabilité le téléspectateur regarde-t-il la première chaîne? Quelle proportion du temps a-t-il passée à regarder chaque chaîne?

Soit X_n la chaîne regardée pendant le $(n+1)$ -ème intervalle de temps. On peut calculer la loi des X_n par récurrence. Si $(x_n, y_n) = (\mathbb{P}\{X_n = 1\}, \mathbb{P}\{X_n = 2\})$, on peut écrire sous forme matricielle

$$\begin{aligned} (x_0 \ y_0) &= (1 \ 0) , \\ (x_1 \ y_1) &= (1 \ 0) \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} = (1-p \ p) , \\ (x_2 \ y_2) &= (1-p \ p) \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} = ((1-p)^2 + pq \ p(2-p-q)) , \end{aligned} \quad (3.3.1)$$

et ainsi de suite. Après n itérations, on aura donc

$$(x_n \ y_n) = (x_0 \ y_0) P^n , \quad (3.3.2)$$

où P est la matrice

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} . \quad (3.3.3)$$

P est appelée la *matrice de transition* de la chaîne. Une méthode de calculer P^n est basée sur la diagonalisation de P . On constate que P admet les vecteurs propres

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{et} \quad \begin{pmatrix} p \\ -q \end{pmatrix} , \quad (3.3.4)$$

avec valeurs propres respectives 1 et $1 - p - q$. On a donc $P = SDS^{-1}$, avec

$$S = \begin{pmatrix} 1 & p \\ 1 & -q \end{pmatrix} \quad \text{et} \quad D = \begin{pmatrix} 1 & 0 \\ 0 & 1 - p - q \end{pmatrix}. \quad (3.3.5)$$

Par conséquent,

$$P^n = (SDS^{-1})^n = SD^nS^{-1} = \frac{1}{p+q} \begin{pmatrix} q & p \\ q & p \end{pmatrix} + \frac{(1-p-q)^n}{p+q} \begin{pmatrix} p & -p \\ -q & q \end{pmatrix}. \quad (3.3.6)$$

Si $0 < p + q < 2$, alors $\lim_{n \rightarrow \infty} (1 - p - q)^n = 0$, et donc

$$\lim_{n \rightarrow \infty} P^n = \frac{1}{p+q} \begin{pmatrix} q & p \\ q & p \end{pmatrix}. \quad (3.3.7)$$

Il suit que

$$\lim_{n \rightarrow \infty} (x_n \ y_n) = \left(\frac{q}{p+q} \quad \frac{p}{p+q} \right). \quad (3.3.8)$$

Au bout d'un temps très long, le téléspectateur regarde donc la première chaîne avec probabilité $q/(p+q)$. Nous verrons que $q/(p+q)$ est également la proportion du temps passée à regarder la première chaîne.

Exemple 3.3.2. Un ivrogne titube le long d'une rue. A chacun des angles de la rue, numérotés 2, 3, 4, il change de direction avec probabilité $1/2$. Aux extrémités de la rue se trouvent le bar (en 1) et sa maison (en 5). S'il arrive soit au bar, soit à la maison, il y passe le reste de la nuit. Partant d'un angle de rue donné, avec quelle probabilité finit-il dans le bar ou dans son lit?

Le graphe du processus est le suivant:

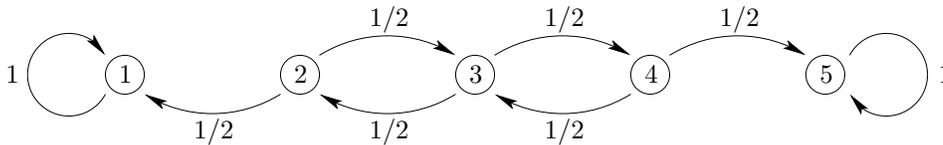


FIGURE 3.7. La chaîne de Markov associée à la marche de l'ivrogne.

Sa matrice de transition est donnée par

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (3.3.9)$$

Calculer P^n par la méthode donnée dans l'exemple précédent est laborieux. Nous introduirons une méthode plus efficace permettant de déterminer les probabilités cherchées, ainsi que d'autres quantités intéressantes.

Le point commun des deux exemples précédents est que la position au temps $k + 1$ est choisie avec une distribution de probabilité ne dépendant que de l'endroit où l'on se trouve au temps k , indépendamment de ce qui s'est passé aux temps précédents: C'est la *propriété de Markov*.

Définition 3.3.3 (Chaîne de Markov). Soit $\mathcal{X} = \{1, \dots, m\}$ un ensemble fini. Une chaîne de Markov sur \mathcal{X} est une suite de variables aléatoires (X_0, X_1, X_2, \dots) , prenant leurs valeurs dans \mathcal{X} , telles que

$$\mathbb{P}\{X_k = j | X_1 = i_1, \dots, X_{k-1} = i_{k-1}\} = \mathbb{P}\{X_k = j | X_{k-1} = i_{k-1}\}, \quad (3.3.10)$$

cette dernière probabilité ne dépendant que de j et i_{k-1} , mais pas du temps k . La matrice de transition de la chaîne est la matrice P de taille $m \times m$ dont les éléments sont donnés par

$$p_{ij} = \mathbb{P}\{X_k = j | X_{k-1} = i\}. \quad (3.3.11)$$

Dans la suite, il sera utile de noter, pour tout événement A ,

$$\begin{aligned} \mathbb{P}^{k,i}\{A\} &:= \mathbb{P}\{A | X_k = i\}, \\ \mathbb{P}^i\{A\} &:= \mathbb{P}^{0,i}\{A\} = \mathbb{P}\{A | X_0 = i\}. \end{aligned} \quad (3.3.12)$$

En particulier, on aura $\mathbb{P}^{k,i}\{X_{k+1} = j\} = \mathbb{P}^i\{X_1 = j\} = p_{ij}$ pour tout temps k .

Proposition 3.3.4. Soit P la matrice de transition d'une chaîne de Markov. Alors

1. Pour chaque ligne de P , on a $\sum_{j=1}^m p_{ij} = 1$.
2. $\mathbb{P}^{k,i}\{X_{k+2} = j\}$ est donnée par l'élément de matrice ij de P^2 . Plus généralement, $\mathbb{P}^{k,i}\{X_{k+\ell} = j\}$ est donnée par l'élément de matrice ij de P^ℓ .

DÉMONSTRATION.

1. On a

$$1 = \mathbb{P}^{k,i}\{X_{k+1} \in \mathcal{X}\} = \sum_{j=1}^m \mathbb{P}^{k,i}\{X_{k+1} = j\} = \sum_{j=1}^m p_{ij}. \quad (3.3.13)$$

2. Comme $X_{k+1} \in \mathcal{X}$, on peut décomposer

$$\begin{aligned} \mathbb{P}^{k,i}\{X_{k+2} = j\} &= \mathbb{P}^{k,i}\left\{\bigcup_{\ell=1}^m \{X_{k+2} = j, X_{k+1} = \ell\}\right\} \\ &= \sum_{\ell=1}^m \mathbb{P}^{k,i}\{X_{k+2} = j, X_{k+1} = \ell\} \\ &= \sum_{\ell=1}^m \frac{\mathbb{P}\{X_{k+2} = j, X_{k+1} = \ell, X_k = i\}}{\mathbb{P}\{X_k = i\}} \\ &= \sum_{\ell=1}^m \mathbb{P}\{X_{k+2} = j | X_{k+1} = \ell, X_k = i\} \frac{\mathbb{P}\{X_{k+1} = \ell, X_k = i\}}{\mathbb{P}\{X_k = i\}} \\ &= \sum_{\ell=1}^m \mathbb{P}^{k+1,\ell}\{X_{k+2} = j\} \mathbb{P}^{k,i}\{X_{k+1} = \ell\} \\ &= \sum_{\ell=1}^m p_{\ell j} p_{i \ell}. \end{aligned} \quad (3.3.14)$$

Par définition du produit matriciel, c'est exactement l'élément ij de la matrice P^2 . Le cas $\ell > 2$ suit par récurrence. \square

Une matrice P dont tous les éléments sont non-négatifs, et telle que la somme des éléments de chaque ligne vaut 1 est appelée une *matrice stochastique*. La relation (3.3.14) s'appelle *équation de Chapman–Kolmogorov*. Elle signifie que les éléments de matrice de P^2 donnent les probabilités de passer d'un état à l'autre en deux pas, celles de P^ℓ donnent les probabilités de passer d'un état à l'autre en ℓ pas.

Définition 3.3.5 (État atteignable). *L'état j est dit atteignable depuis i s'il existe un $k \geq 1$ tel que l'élément de matrice ij de P^k soit strictement positif.*

Exemple 3.3.6. Revenons à l'exemple 3.3.2 de la marche de l'ivrogne. On trouve

$$P^2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1/2 & 1/4 & 0 & 1/4 & 0 \\ 1/4 & 0 & 1/2 & 0 & 1/4 \\ 0 & 1/4 & 0 & 1/4 & 1/2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad P^3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 5/8 & 0 & 1/4 & 0 & 1/8 \\ 1/4 & 1/4 & 0 & 1/4 & 1/4 \\ 1/8 & 0 & 1/4 & 0 & 5/8 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (3.3.15)$$

Par exemple, partant du point 2, l'ivrogne se retrouve après deux pas au bar avec probabilité $1/2$, et en 2 ou 4 avec probabilité $1/4$. Après trois pas, il se retrouve au bar avec probabilité $5/8$, en 3 avec probabilité $1/4$, et à la maison avec probabilité $1/8$. Le bar est atteignable depuis 2 (en un pas), la maison également (en 3 pas). Par contre, la maison n'est pas atteignable depuis le bar, puisque l'ivrogne reste au bar dès qu'il s'y trouve.

Dans la suite, nous allons étudier plus en détail deux types de chaînes de Markov:

- les chaînes de Markov *absorbantes*, dans lesquelles il est impossible de sortir de certains états, comme dans l'exemple de l'ivrogne;
- les chaînes de Markov *ergodiques*, dans lesquelles on tend vers une distribution stationnaire, comme c'est le cas dans l'exemple du téléspectateur.

3.4 Chaînes de Markov absorbantes

Définition 3.4.1 (Chaîne de Markov absorbante).

- Un état $i \in \mathcal{X}$ d'une chaîne de Markov est dit absorbant si $p_{ii} = 1$ (et donc nécessairement $p_{ij} = 0$ pour $j \neq i$).
- Une chaîne de Markov est absorbante s'il existe, pour tout état, un état absorbant atteignable depuis cet état.
- Une chaîne non absorbante est dite transiente.

Il est commode de renuméroter les états d'une chaîne absorbante, en plaçant d'abord les états non absorbants, ensuite les états absorbants. Dans ce cas, on dira que la matrice de transition est écrite sous forme *canonique*

$$P = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}. \quad (3.4.1)$$

S'il y a r états absorbants, Q est une matrice de taille $(m-r) \times (m-r)$, R est une matrice de taille $(m-r) \times r$ et I est la matrice identité de taille r .

Exemple 3.4.2. Dans l'Exemple 3.3.2 de l'ivrogne, les états 1 et 5 sont absorbants, les états 2, 3 et 4 ne le sont pas. La matrice s'écrira donc, sous forme canonique

$$P = \begin{array}{ccccc|cc} & 2 & 3 & 4 & 1 & 5 & \\ \hline & 0 & 1/2 & 0 & 1/2 & 0 & 2 \\ & 1/2 & 0 & 1/2 & 0 & 0 & 3 \\ & 0 & 1/2 & 0 & 0 & 1/2 & 4 \\ \hline & 0 & 0 & 0 & 1 & 0 & 1 \\ & 0 & 0 & 0 & 0 & 1 & 5 \end{array} \quad (3.4.2)$$

Les matrices Q et R sont données par

$$Q = \begin{pmatrix} 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{pmatrix}, \quad R = \begin{pmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 1/2 \end{pmatrix}. \quad (3.4.3)$$

Pour la matrice sous forme canonique (3.4.1), on vérifie facilement par récurrence que

$$P^n = \begin{pmatrix} Q^n & [I + Q + \dots + Q^{n-1}]R \\ 0 & I \end{pmatrix}. \quad (3.4.4)$$

Pour comprendre le comportement à grand temps de la chaîne, il faut étudier l'évolution de cette matrice pour $n \rightarrow \infty$.

Proposition 3.4.3. *Pour une chaîne de Markov absorbante,*

$$\lim_{n \rightarrow \infty} Q^n = 0. \quad (3.4.5)$$

DÉMONSTRATION. Soit i un état non absorbant ($i \leq m - r$). L'élément de matrice $(Q^n)_{ij}$ de Q^n est la probabilité de se trouver dans l'état non absorbant j , après n pas, partant de i . Par conséquent, $(Q^n)_{ij}$ est inférieur ou égal à la probabilité de ne pas avoir atteint d'état absorbant en n pas. Soit

$$m_i = \min\{k \geq 1 : \exists \ell > m - r, (P^k)_{i\ell} > 0\} \quad (3.4.6)$$

le nombre minimal de pas nécessaire à atteindre un état absorbant ℓ depuis i . Soit

$$p_i = \mathbb{P}^i\{X_{m_i} \leq m - r\} < 1 \quad (3.4.7)$$

la probabilité de ne pas atteindre d'état absorbant en m_i pas, partant de i . Soit enfin

$$M = \max_{i=1, \dots, m-r} m_i, \quad p = \max_{i=1, \dots, m-r} p_i. \quad (3.4.8)$$

Alors la probabilité de ne pas atteindre d'état absorbant en M pas, partant de n'importe quel état non absorbant, est bornée par p . Il suit que la probabilité de ne pas atteindre d'état absorbant en Mk pas est bornée par p^k . Cette probabilité tend vers 0 lorsque k tend vers l'infini. La probabilité de ne pas être absorbé après un nombre arbitraire ℓ de pas étant une fonction décroissante de ℓ , elle tend nécessairement vers 0. Par conséquent, $(Q^n)_{ij}$ tend vers zéro lorsque n tend vers l'infini, pour tout $j \in \{1, \dots, r\}$. \square

Montrons maintenant que la matrice Q permet de déterminer le nombre moyen de passages dans chaque état non absorbant, avant de finir par atteindre un état absorbant.

Théorème 3.4.4. *Soit P la matrice de transition d'une chaîne de Markov absorbante, écrite sous forme canonique. Alors*

1. *La matrice $I - Q$ est inversible, et*

$$N := (I - Q)^{-1} = I + Q + Q^2 + \dots \quad (3.4.9)$$

2. *L'élément de matrice n_{ij} de N est l'espérance du nombre de passages dans l'état j , en partant de l'état i :*

$$n_{ij} = \mathbb{E}^i \left(\sum_{k \geq 0} 1_{\{X_k=j\}} \right) := \mathbb{E} \left(\sum_{k \geq 0} 1_{\{X_k=j\}} \mid X_0 = i \right). \quad (3.4.10)$$

DÉMONSTRATION.

1. Supposons que x soit un vecteur tel que $Qx = x$, c'est-à-dire $(I - Q)x = 0$. Alors

$$x = Qx = Q^2x = \dots = Q^n x = \dots, \quad (3.4.11)$$

donc

$$x = \lim_{n \rightarrow \infty} Q^n x = 0. \quad (3.4.12)$$

Par conséquent, $(I - Q)$ n'admet pas 0 comme valeur propre, donc est inversible. Comme

$$(I - Q)(I + Q + Q^2 + \dots + Q^n) = I - Q^{n+1}, \quad (3.4.13)$$

on a

$$I + Q + Q^2 + \dots + Q^n = N(I - Q^{n+1}), \quad (3.4.14)$$

et donc

$$\lim_{n \rightarrow \infty} \sum_{j=1}^n Q^j = N. \quad (3.4.15)$$

2. Fixons deux états non absorbants $i, j \leq m - r$. Soit $Y_k = 1_{\{X_k=j\}}$. On a

$$\mathbb{P}^i \{Y_k = 1\} = (Q^k)_{ij} \quad \text{et} \quad \mathbb{P}^i \{Y_k = 0\} = 1 - (Q^k)_{ij}, \quad (3.4.16)$$

donc

$$\mathbb{E}^i(Y_k) = (Q^k)_{ij}. \quad (3.4.17)$$

Par conséquent,

$$\mathbb{E}^i \left(\sum_{k \geq 0} Y_k \right) = \sum_{k \geq 0} (Q^k)_{ij} = (N)_{ij} = n_{ij}. \quad (3.4.18)$$

□

Définition 3.4.5 (Matrice fondamentale). N est appelée la matrice fondamentale de la chaîne de matrice de transition P .

En utilisant les relations (3.4.5) et (3.4.9) dans l'expression (3.4.4) de P^n , on obtient

$$\lim_{n \rightarrow \infty} P^n = \begin{pmatrix} 0 & NR \\ 0 & I \end{pmatrix}. \quad (3.4.19)$$

La matrice NR devrait donc décrire les probabilités de transition d'un état non absorbant vers un état absorbant, en un nombre arbitraire de pas. Il nous reste également à caractériser le temps s'écoulant jusqu'à ce que le processus ait atteint un état absorbant. Son espérance peut aussi être déduite de la matrice N .

Théorème 3.4.6. *Soit N la matrice fondamentale d'une chaîne de Markov absorbante.*

1. Soit

$$\tau = \inf\{k \geq 1 : X_k > m - r\} \quad (3.4.20)$$

le temps s'écoulant jusqu'à ce qu'un état absorbant soit atteint, et soit $t_i = \mathbb{E}^i(\tau)$ l'espérance de ce temps, si l'on part de l'état i . Alors

$$t_i = \sum_{j=1}^{m-r} n_{ij}, \quad (3.4.21)$$

ou encore, sous forme matricielle,

$$\begin{pmatrix} t_1 \\ \vdots \\ t_{m-r} \end{pmatrix} = Nc \quad \text{où} \quad c = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}. \quad (3.4.22)$$

2. Soit $b_{iq} = \mathbb{P}^i\{X_\tau = q\}$ la probabilité d'atteindre l'état absorbant q si l'on part de i . Alors b_{iq} est l'élément iq de la matrice

$$B = NR. \quad (3.4.23)$$

DÉMONSTRATION.

1. On a

$$\begin{aligned} \sum_{j=1}^{m-r} n_{ij} &= \sum_{j=1}^{m-r} \mathbb{E}^i \left(\sum_{k \geq 1} 1_{\{X_k = j\}} \right) = \mathbb{E}^i \left(\sum_{k \geq 1} 1_{\{X_k \leq m-r\}} \right) \\ &= \sum_{k \geq 1} \mathbb{P}^i\{\tau > k\} = \sum_{k \geq 1} \sum_{\ell \geq k+1} \mathbb{P}^i\{\tau = \ell\} \\ &= \sum_{\ell \geq 1} \sum_{k=1}^{\ell} \mathbb{P}^i\{\tau = \ell\} = \sum_{\ell \geq 1} \ell \mathbb{P}^i\{\tau = \ell\} = \mathbb{E}^i(\tau). \end{aligned} \quad (3.4.24)$$

2. Si le processus atteint l'état absorbant q , il existe un temps k tel que $X_k \leq m - r$ et $X_{k+1} = q$. On a donc

$$\begin{aligned} b_{iq} &= \mathbb{P}^i\{X_\tau = q\} = \sum_{k \geq 0} \mathbb{P}^i\{X_k \leq m - r, X_{k+1} = q\} \\ &= \sum_{k \geq 0} \sum_{\ell=1}^{m-r} \mathbb{P}^i\{X_k = \ell\} \mathbb{P}^{k,\ell}\{X_{k+1} = q\} \\ &= \sum_{k \geq 0} \sum_{\ell=1}^{m-r} (Q^k)_{i\ell} (R)_{\ell q} \\ &= \sum_{k \geq 0} (Q^k R)_{iq} = (NR)_{iq}, \end{aligned} \quad (3.4.25)$$

ce qui montre qu'on a bien $B = NR$. □

Exemple 3.4.7. Rappelons que pour l'exemple de la marche de l'ivrogne, on a

$$Q = \begin{pmatrix} 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{pmatrix}, \quad R = \begin{pmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 1/2 \end{pmatrix}. \quad (3.4.26)$$

Un calcul direct donne alors

$$N = \begin{pmatrix} 1 & -1/2 & 0 \\ -1/2 & 1 & -1/2 \\ 0 & -1/2 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{pmatrix}. \quad (3.4.27)$$

Par conséquent, si nous gardons la numérotation originale des états,

$$\begin{pmatrix} t_2 \\ t_3 \\ t_4 \end{pmatrix} = Nc = (3 \quad 4 \quad 3) \quad \text{et} \quad B = NR = \begin{pmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix}. \quad (3.4.28)$$

On peut donc dire qu'en partant du point 2, par exemple,

- le temps moyen jusqu'à absorption est de 3 pas;
- l'ivrogne a une probabilité de 3/4 de finir la soirée au bar, et de 1/4 de la terminer dans son lit;
- il passe en moyenne 3/2 fois en 2, 1 fois en 3 et 1/2 fois en 4.

Supposons maintenant que l'ivrogne parte avec probabilité 1/3 aux points 2, 3 et 4 respectivement. Alors

$$\begin{pmatrix} 1/3 & 1/3 & 1/3 \end{pmatrix} N = \begin{pmatrix} 1 & 4/3 & 1 \end{pmatrix} \quad (3.4.29)$$

donne les temps moyens passés, respectivement, en 2, 3 et 4.

$$\begin{pmatrix} 1/3 & 1/3 & 1/3 \end{pmatrix} \begin{pmatrix} t_2 \\ t_3 \\ t_4 \end{pmatrix} = \frac{10}{3} \quad (3.4.30)$$

donne le temps moyen jusqu'à absorption. Et enfin,

$$\begin{pmatrix} 1/3 & 1/3 & 1/3 \end{pmatrix} B = \begin{pmatrix} 1/2 & 1/2 \end{pmatrix} \quad (3.4.31)$$

donne les probabilités respectives de finir au bar ou à la maison.

3.5 Chaînes de Markov ergodiques

Définition 3.5.1 (Chaîne de Markov ergodique). Une chaîne de Markov est dite ergodique ou irréductible si tout état est atteignable depuis tout autre état. Elle est dite régulière s'il existe une puissance P^k de sa matrice de transition P dont tous les éléments sont strictement positifs. Une chaîne régulière est donc ergodique.

La particularité des chaînes régulières est que l'on peut aller de n'importe quel état vers n'importe quel autre état en un nombre fixé k de pas, où k est indépendant de l'état de départ. Pour les chaînes ergodiques, on demande simplement que tout état soit atteignable depuis tout autre, mais le nombre de pas n'est pas nécessairement fixé.

Exemple 3.5.2 (Modèle d'Ehrenfest). Quatre boules sont réparties sur deux urnes. À chaque étape du processus, une boule parmi les quatre est choisie au hasard, de manière équiprobable, et changée d'urne. Soit X_k le nombre de boules dans la première urne après k tirages. Alors la suite des X_k forme une chaîne de Markov, dont le graphe est le suivant:

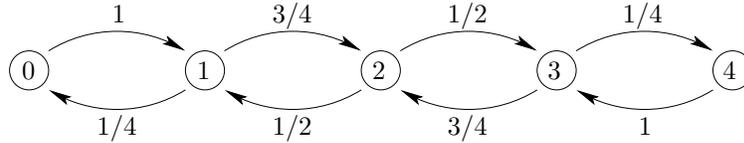


FIGURE 3.8. La chaîne de Markov associée au modèle d'Ehrenfest.

Sa matrice de transition est donnée par

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/4 & 0 & 3/4 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}. \quad (3.5.1)$$

Depuis tout état, on peut trouver un chemin dans le graphe vers tout autre état. La chaîne est donc ergodique. Cependant, partant, par exemple, de l'état 0, la chaîne se trouvera dans l'un des états 0, 2 ou 4 après un nombre pair de tirages, et en 1 ou 3 après un nombre impair de tirages. La chaîne n'est donc pas régulière, puisque l'élément $(P^n)_{ij}$ est nul chaque fois que $i + j + n$ est impair.

Nous allons d'abord étudier de plus près les chaînes régulières, dont le comportement asymptotique est plus simple. En particulier, la suite des puissances P^n de leur matrice de transition converge toujours vers une matrice particulière, dont toutes les lignes sont égales.

Théorème 3.5.3. *Soit P la matrice de transition d'une chaîne de Markov régulière. Alors*

$$\lim_{n \rightarrow \infty} P^n = W = \begin{pmatrix} w_1 & w_2 & \dots & w_m \\ w_1 & w_2 & \dots & w_m \\ \vdots & \vdots & & \vdots \\ w_1 & w_2 & \dots & w_m \end{pmatrix}, \quad (3.5.2)$$

où tous les w_i sont strictement positifs.

DÉMONSTRATION. Si la chaîne n'a qu'un état, le résultat est immédiat, donc nous pouvons admettre que $m \geq 2$. Nous pouvons supposer que tous les éléments de P^k sont positifs pour $k = 1$, sinon il suffit de considérer la chaîne dont la matrice de transition est P^k au lieu de P . Soit $d > 0$ le plus petit élément de P . Alors $d \leq 1/2$, puisque $md \leq 1$. Soit y un vecteur colonne tel que

$$0 \leq m_0 \leq y_i \leq M_0 \quad \forall i \in \{1, \dots, m\}. \quad (3.5.3)$$

Soit $z = Py$. La plus grande valeur possible d'une composante z_j de z est obtenue si $y^T = (m_0, M_0, \dots, M_0)$ et $p_{k1} = d$. Dans ce cas, la somme des $m - 1$ derniers éléments

de la ligne j de P vaut $1 - d$, et par conséquent $z_j = dm_0 + (1 - d)M_0$. On a donc nécessairement

$$z_j \leq dm_0 + (1 - d)M_0 =: M_1 \quad \forall j \in \{1, \dots, m\}. \quad (3.5.4)$$

Un raisonnement similaire montre que

$$z_j \geq dM_0 + (1 - d)m_0 =: m_1 \quad \forall j \in \{1, \dots, m\}. \quad (3.5.5)$$

Par conséquent, nous avons $m_1 \leq z_j \leq M_1$, avec

$$M_1 - m_1 = (1 - 2d)(M_0 - m_0). \quad (3.5.6)$$

De plus, on voit facilement que $m_1 \geq m_0$ et $M_1 \leq M_0$. Après n itérations, les composantes de $P^n y$ seront comprises entre des nombres m_n et M_n , satisfaisant

$$M_n - m_n = (1 - 2d)^n (M_0 - m_0) \quad (3.5.7)$$

et

$$m_0 \leq m_n \leq M_n \leq M_0. \quad (3.5.8)$$

Les suites $\{m_n\}_{n \geq 1}$ et $\{M_n\}_{n \geq 1}$ sont donc adjacentes, et convergent vers une même limite u . On a donc

$$\lim_{n \rightarrow \infty} P^n y = \begin{pmatrix} u \\ \vdots \\ u \end{pmatrix}, \quad (3.5.9)$$

où u dépend de y . Appliquons cette relation sur les vecteurs de base e_1, \dots, e_m . Il existe des nombres w_i tels que

$$\lim_{n \rightarrow \infty} P^n e_i = \begin{pmatrix} w_i \\ \vdots \\ w_i \end{pmatrix} \quad (3.5.10)$$

pour chaque i . Or $P^n e_i$ est la i -ème colonne de P^n , nous avons donc prouvé la relation (3.5.2). Par ailleurs, comme dans le cas $y = e_i$ on a $m_0 = 0$ et $M_0 = 1$, la relation (3.5.5) donne $m_1 \geq d$, donc $w_j \geq d$. Par conséquent, tous les w_j sont strictement positifs. \square

L'interprétation des grandeurs w_1, \dots, w_m est la suivante: On a

$$\lim_{n \rightarrow \infty} \mathbb{P}^i \{X_n = j\} = \lim_{n \rightarrow \infty} (P^n)_{ij} = w_j. \quad (3.5.11)$$

La probabilité que la chaîne de Markov se trouve dans l'état j tend donc vers w_j lorsque le temps tend vers l'infini, indépendamment de l'état de départ.

Pour déterminer les w_j , il n'est pas nécessaire de calculer P^n explicitement. En effet, le résultat suivant montre que w est un vecteur propre (à gauche) de P . On peut donc le calculer en résolvant un système linéaire.

Théorème 3.5.4. *Soit P la matrice de transition d'une chaîne de Markov régulière, et $w = (w_1, \dots, w_m)$ le vecteur ligne tel que $(W)_{ij} = w_j$ pour tout i . Alors*

1. $wP = w$, et tout vecteur propre à gauche de P de valeur propre 1 est colinéaire à w .

2. Soit

$$c = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}. \quad (3.5.12)$$

Alors $Pc = c$, et tout vecteur propre à droite de P de valeur propre 1 est colinéaire à c .

3. Si $v = (v_1, \dots, v_m)$ est un vecteur ligne tel que

$$v_i \geq 0 \quad \forall i \in \{1, \dots, m\} \quad \text{et} \quad \sum_{i=1}^m v_i = 1, \quad (3.5.13)$$

alors

$$\lim_{n \rightarrow \infty} vP^n = w. \quad (3.5.14)$$

DÉMONSTRATION.

1. Comme $P^n \rightarrow W$ et $P^{n+1} \rightarrow WP$ pour $n \rightarrow \infty$, on a nécessairement $WP = W$, et donc $wP = w$. De plus, si v est un vecteur ligne tel que $vP = v$, alors $vP^n = v$ pour tout n et donc $vW = v$. Soit alors r la somme des composantes v_i de v . On a

$$(vW)_i = \sum_{j=1}^m v_j (W)_{ji} = \sum_{j=1}^m v_j w_i = r w_i, \quad (3.5.15)$$

ce qui implique $vW = r w$, donc $v = r w$.

2. Si x est un vecteur colonne tel que $x = Px$, alors $x = Wx$. Par conséquent, $x_i = w x$ est indépendant de i , donc x est un multiple de c . Par ailleurs, nous avons déjà montré que

$$(Pc)_i = \sum_{j=1}^m p_{ij} = 1, \quad (3.5.16)$$

donc $Pc = c$.

3. Si v satisfait (3.5.13), alors on a

$$\lim_{n \rightarrow \infty} vP^n = vW = (v_1 \quad \dots \quad v_m) \begin{pmatrix} w \\ \vdots \\ w \end{pmatrix} = \sum_{i=1}^m v_i w = w. \quad (3.5.17)$$

□

L'équation $wP = w$ a une interprétation importante. Supposons qu'à un instant k , on ait $\mathbb{P}\{X_k = i\} = w_i$ pour tout i . À l'instant $k+1$ on aura

$$\mathbb{P}\{X_{k+1} = j\} = \sum_{i=1}^m w_i P_{ij} = (wP)_j = w_j, \quad (3.5.18)$$

ce qui signifie que X_{k+1} et X_k ont la même loi.

Définition 3.5.5 (Distribution stationnaire). *Le vecteur $w = (w_1, \dots, w_m)$ tel que $(W)_{ij} = w_j$ pour tout i est appelé distribution stationnaire de la chaîne de Markov. C'est donc l'unique solution de l'équation aux valeurs propres $wP = w$ telle que $\sum_{i=1}^m w_i = 1$.*

Exemple 3.5.6. Dans l'Exemple 3.3.1 du téléspectateur, la matrice de transition est donnée par

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}. \quad (3.5.19)$$

Si p et q sont différents de 0 et 1, la chaîne est régulière. Sa distribution stationnaire est donnée par le vecteur

$$w = \left(\frac{q}{p+q}, \frac{p}{p+q} \right), \quad (3.5.20)$$

dont les composantes donnent la probabilité asymptotique que le téléspectateur regarde la première ou la seconde chaîne. On vérifie que l'on a bien $wP = w$ et $w_1 + w_2 = 1$.

Si la chaîne de Markov est ergodique mais non régulière, la suite des P^n ne tend en général pas vers une matrice constante. Par exemple, les P^n peuvent osciller entre plusieurs valeurs. Cependant, certaines propriétés du Théorème 3.5.4 restent vraies pour les chaînes ergodiques générales.

Proposition 3.5.7. *Soit P la matrice de transition d'une chaîne de Markov ergodique.*

1. *Il existe un vecteur ligne w , à composantes positives et de somme 1, tel que $wP = w$.*
2. *Tout vecteur propre à gauche de P de valeur propre 1 est colinéaire à w .*
3. *Tout vecteur propre à droite de P de valeur propre 1 est colinéaire à c .*

DÉMONSTRATION. Considérons la matrice

$$Q = \frac{1}{2}[I + P]. \quad (3.5.21)$$

On vérifie facilement que c'est une matrice stochastique. Soit k le nombre maximal de pas nécessaires pour joindre deux états quelconques de la chaîne. Considérons la matrice

$$Q^k = I + \binom{k}{1}P + \binom{k}{2}P^2 + \dots + \binom{k}{k-1}P^{k-1} + P^k. \quad (3.5.22)$$

Pour tout (i, j) , il existe au moins une puissance ℓ de P telle que $(P^\ell)_{ij}$ soit strictement positif. Comme tous les autres éléments sont positifs ou nuls, la matrice Q^k n'a que des éléments strictement positifs, et par conséquent Q est la matrice de transition d'une chaîne régulière. On peut donc appliquer le Théorème 3.5.4, pour obtenir l'existence d'un vecteur w tel que $wQ = w$. Mais ceci implique

$$\frac{1}{2}[w + wP] = w, \quad (3.5.23)$$

donc $wP = w$. Les autres points se démontrent de la même manière. \square

Nous appellerons encore dans le cas général *distribution stationnaire* l'unique vecteur w tel que $wP = w$ et $\sum_{i=1}^m w_i = 1$. Ce vecteur satisfait toujours la relation (3.5.18).

Exemple 3.5.8. Revenons au modèle d'Ehrenfest du l'Exemple 3.5.2. Pour trouver la distribution stationnaire w , on peut commencer par chercher un vecteur propre à gauche v , de valeur propre 1. Comme tous les multiples de w sont de tels vecteurs propres, il suffit d'en trouver un tel que $v_0 = 1$. Par symétrie, on peut chercher le vecteur v de la forme $v = (1, v_1, v_2, v_1, 1)$. L'équation $vP = v$ devient alors

$$\frac{1}{4}v_1 = 1, \quad 1 + \frac{1}{2}v_2 = v_1, \quad \frac{3}{2}v_1 = v_2, \quad (3.5.24)$$

dont la solution est

$$v = (1, 4, 6, 4, 1) . \quad (3.5.25)$$

Comme la somme des v_i vaut 16, la distribution stationnaire w est donné par

$$w = \frac{1}{16}(1, 4, 6, 4, 1) . \quad (3.5.26)$$

On remarque que c'est une loi binomiale. Ceci n'est pas un hasard, et on peut le généraliser à des modèles d'Ehrenfest avec un nombre arbitraire de boules.

On sait déterminer d'autres propriétés d'une chaîne de Markov que son comportement asymptotique. Une quantité importante est la fréquence moyenne de passage dans les différents états de la chaîne.

Définition 3.5.9 (Temps de premier passage et de récurrence). *Soit une chaîne de Markov ergodique partant de $X_0 = i$.*

- *La variable aléatoire*

$$\tau_{ij} = \inf\{k > 0 : X_k = j\} \quad (3.5.27)$$

est appelée temps de premier passage de i à j . Son espérance

$$m_{ij} = \mathbb{E}^i(\tau_{ij}) \quad (3.5.28)$$

s'appelle le temps de premier passage moyen de i à j . Par convention, on pose $m_{ii} = 0$. La matrice M de composantes m_{ij} est la matrice de premier passage moyen.

- *La variable aléatoire*

$$\tau_{ii} = \inf\{k > 0 : X_k = i\} \quad (3.5.29)$$

est appelée temps de premier retour en i . Son espérance

$$r_i = \mathbb{E}^i(\tau_{ii}) \quad (3.5.30)$$

s'appelle le temps de récurrence moyen en i . La matrice diagonale D d'éléments $d_{ii} = r_i$ est la matrice de récurrence moyenne.

Une première méthode de calculer m_{ij} est la suivante. On modifie la matrice de transition P en rendant l'état j absorbant, c'est-à-dire que l'on remplace P_{jj} par 1, et tous les autres éléments de la ligne j par 0. La chaîne ainsi modifiée est alors absorbante. Le temps de premier passage moyen m_{ij} est alors égal au temps d'absorption t_i du Théorème 3.4.6. Cette méthode est néanmoins assez fastidieuse à mettre en oeuvre, et nous allons introduire une méthode plus efficace pour calculer M .

Théorème 3.5.10. *Soit P la matrice de transition d'une chaîne ergodique.*

1. *Soit C la matrice de taille $m \times m$ dont tous les éléments valent 1. Alors*

$$(I - P)M = C - D . \quad (3.5.31)$$

2. *Les composantes w_i de la distribution stationnaire sont toutes strictement positives.*
3. *Le temps de récurrence moyen est donné par*

$$r_i = \frac{1}{w_i} , \quad (3.5.32)$$

où w_i est la i -ème composante de la distribution stationnaire.

DÉMONSTRATION.

1. Pour $i \neq j$, on a $\mathbb{P}^i\{\tau_{ij} = 1\} = p_{ij}$, donc

$$\begin{aligned}
 m_{ij} &= \mathbb{E}^i(\tau_{ij}) = \mathbb{P}^i\{\tau_{ij} = 1\} + \sum_{k \geq 2} k \mathbb{P}^i\{\tau_{ij} = k\} \\
 &= p_{ij} + \sum_{k \geq 2} k \sum_{\ell \neq j} \underbrace{\mathbb{P}^i\{X_1 = \ell\}}_{p_{i\ell}} \mathbb{P}^{1,\ell}\{\tau_{\ell j} = k - 1\} \\
 &= p_{ij} + \sum_{\ell \neq j} p_{i\ell} \sum_{k' \geq 1} (k' + 1) \mathbb{P}^{\ell}\{\tau_{\ell j} = k'\} \\
 &= p_{ij} + \sum_{\ell \neq j} p_{i\ell} (m_{\ell j} + 1) = 1 + \sum_{\ell \neq j} p_{i\ell} m_{\ell j}. \tag{3.5.33}
 \end{aligned}$$

De manière similaire on obtient, en utilisant $m_{ii} = 0$,

$$r_i = p_{ii} + \sum_{\ell \neq j} p_{i\ell} (m_{\ell i} + 1) = 1 + \sum_{\ell \neq j} p_{i\ell} m_{\ell i}. \tag{3.5.34}$$

Sous forme matricielle, ces deux identités peuvent se résumer par

$$M + D = C + PM, \tag{3.5.35}$$

qui est équivalente à (3.5.31).

2. Comme $wP = P$, on a $w(I - P) = 0$, donc

$$0 = w(I - P)M = wC - wD. \tag{3.5.36}$$

Or comme on a $wC = (1, \dots, 1)$ et $wD = (w_1 r_1, \dots, w_m r_m)$, il suit $w_i r_i = 1$ pour tout $i = 1, \dots, m$. Ceci montre à la fois la positivité des w_i et la relation (3.5.32). \square

Exemple 3.5.11. Les temps de récurrence moyens du modèle d'Ehrenfest sont donnés par

$$r_0 = r_4 = 16, \quad r_1 = r_3 = 4, \quad r_2 = \frac{8}{3}. \tag{3.5.37}$$

Il se passe donc en moyenne 16 itérations entre deux visites au site 0, contre seulement $8/3$ entre deux visites au site 2.

L'équation (3.5.31) ne permet pas encore de calculer M , car $I - P$ n'est pas inversible puisque P admet la valeur propre 1. Il s'avère cependant que $I - P + W$ est inversible, et que son inverse permet de calculer les éléments de matrice de M .

Théorème 3.5.12. Soit P la matrice de transition d'une chaîne de Markov ergodique, w sa distribution stationnaire, et W la matrice dont toutes les lignes sont égales à w . Alors

1. La matrice $I - P + W$ est inversible.
2. Soit $Z = (I - P + W)^{-1}$, et c le vecteur colonne dont tout les éléments sont égaux à 1. Alors $Zc = c$, $wZ = w$ et

$$Z(I - P) = I - W. \tag{3.5.38}$$

3. Les temps de premier passage moyens sont donnés par

$$m_{ij} = \frac{z_{jj} - z_{ij}}{w_j}. \tag{3.5.39}$$

DÉMONSTRATION.

1. Soit x un vecteur colonne tel que $(I - P + W)x = 0$. Alors

$$0 = w(I - P + W)x = \underbrace{w(I - P)x}_{=0} + wWx = wWx . \quad (3.5.40)$$

Comme la somme des w_i vaut 1, on voit facilement que $wW = w$. On a donc obtenu $wx = 0$. Il suit que $Wx = 0$, puisque chaque composante de Wx est égale à wx . Nous voyons donc que

$$(I - P)x = 0 , \quad (3.5.41)$$

ou encore $Px = x$. Nous savons que ceci implique que x est colinéaire au vecteur constant c , mais comme $wx = 0$, ceci n'est possible que si $x = 0$. La matrice $I - P + W$ est donc inversible.

2. Nous savons que $Pc = c$, et on voit immédiatement que $Wc = c$, à cause du fait que la somme des w_i vaut 1. Il suit que

$$(I - P + W)c = c , \quad (3.5.42)$$

et donc $c = Zc$. De manière similaire, on voit que

$$w(I - P + W) = w , \quad (3.5.43)$$

et donc $w = wZ$. Enfin, on a

$$(I - P + W)(I - W) = I - P + WP - W^2 = I - P , \quad (3.5.44)$$

puisque $WP = W$ et $W^2 = W$. Ceci prouve la relation (3.5.38).

3. D'une part, on a l'égalité

$$Z(I - P)M = Z(C - D) = C - ZD . \quad (3.5.45)$$

Comme d'autre part, on a

$$Z(I - P)M = M - WM , \quad (3.5.46)$$

nous voyons que

$$M = C - ZD + WM . \quad (3.5.47)$$

En composantes, cette équation donne

$$m_{ij} = 1 - z_{ij}r_j + (WM)_{ij} . \quad (3.5.48)$$

En particulier, pour $i = j$ on obtient $0 = 1 - z_{jj}r_j + (WM)_{jj}$. Or $(WM)_{ij} = (wM)_j$ ne dépend pas de i , donc en remplaçant $(WM)_{ij}$ par $(WM)_{jj}$ dans (3.5.48), on trouve

$$m_{ij} = 1 - z_{ij}r_j + z_{jj}r_j - 1 = (z_{jj} - z_{ij})r_j , \quad (3.5.49)$$

ce qui prouve (3.5.39) puisque $r_j = 1/w_j$. \square

Remarque 3.5.13. La relation (3.5.38) a l'interprétation suivante en termes d'algèbre linéaire. Soit v un vecteur ligne tel que $\sum_{i=1}^m v_i = 1$, décrivant par exemple la loi de X_k en un certain temps k . On peut décomposer

$$v = w + y \quad \text{avec} \quad \sum_{i=1}^m y_i = 0 . \quad (3.5.50)$$

Nous savons que $w(I - W) = 0$ et un calcul élémentaire montre que $yW = 0$. Il suit que

$$vZ(I - P) = y , \quad (3.5.51)$$

donc $Z(I - P)$ décrit la projection sur le sous-espace supplémentaire à w , décrivant la déviation entre la loi de X_k et la distribution stationnaire. La restriction de la matrice $I - P$ à ce sous-espace est inversible, et comme

$$yZ(I - P) = y , \quad (3.5.52)$$

l'inverse de cette restriction est précisément la restriction de Z au même sous-espace. Comme par ailleurs $wZ = w$, Z agit trivialement dans le sous-espace engendré par w .

Exemple 3.5.14. Dans le cas de la chaîne à deux états de l'Exemple 3.3.1, les temps de récurrence moyens sont donnés par

$$r_1 = \frac{1}{w_1} = 1 + \frac{p}{q} , \quad r_2 = \frac{1}{w_2} = 1 + \frac{q}{p} . \quad (3.5.53)$$

Le téléspectateur revient donc d'autant plus rapidement à la première chaîne que p/q est petit. Par ailleurs, le calcul donne

$$Z = \frac{1}{(p+q)^2} \begin{pmatrix} p+q(p+q) & p(1-p-q) \\ -q(1-p-q) & q+p(p+q) \end{pmatrix} . \quad (3.5.54)$$

On en déduit les temps moyens de premier passage

$$\begin{aligned} m_{12} &= \frac{z_{22} - z_{21}}{w_2} = \frac{1}{p} \\ m_{21} &= \frac{z_{11} - z_{12}}{w_1} = \frac{1}{q} . \end{aligned} \quad (3.5.55)$$

Ce résultat est naturel: $1/p$ est l'espérance de la loi géométrique, qui décrit le premier "succès" du passage de l'état 1 à l'état 2.

Finalement, on peut également caractériser la fréquence des passages en un état donné. Le nombre de passages en j parmi n itérations est donné par la variable aléatoire

$$S_n^j = \sum_{k=0}^{n-1} 1_{\{X_k=j\}} . \quad (3.5.56)$$

La variable $\frac{1}{n} S_n^j$ donne donc la proportion du temps passée dans l'état j .

Théorème 3.5.15. *On a*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}^i(S_n^j) = w_j, \quad (3.5.57)$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} V^i(S_n^j) = 2w_j z_{jj} - w_j - w_j^2. \quad (3.5.58)$$

De plus, S_n^j obéit à un théorème de la limite centrale,

$$\lim_{n \rightarrow \infty} \mathbb{P}^i \left\{ a < \frac{S_n^j - nw_j}{\sqrt{n V^i(S_n^j)}} < b \right\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx. \quad (3.5.59)$$

DÉMONSTRATION. Nous allons nous borner à prouver la relation (3.5.57). Un calcul direct montre que $PW = W$, et on a donc

$$\begin{aligned} (I + P + P^2 + \dots + P^{n-1})(I - P + W) &= I - P^n + W + PW + \dots + P^{n-1}W \\ &= I - P^n + nW. \end{aligned} \quad (3.5.60)$$

En multipliant à droite par Z , il vient

$$\begin{aligned} I + P + P^2 + \dots + P^{n-1} &= (I - P^n + nW)Z \\ &= (I - P^n)Z + nW. \end{aligned} \quad (3.5.61)$$

Or l'espérance de S_n^j est donnée par

$$\begin{aligned} \mathbb{E}^i(S_n^j) &= \sum_{k=0}^{n-1} \mathbb{E}^i(1_{\{X_k=j\}}) = \sum_{k=0}^{n-1} \mathbb{P}^i\{X_k = j\} = \sum_{k=0}^{n-1} (P^k)_{ij} \\ &= \left(\sum_{k=0}^{n-1} P^k \right)_{ij} = ((I - P^n)Z + nW)_{ij} = z_{ij} - (P^n Z)_{ij} + nw_j. \end{aligned}$$

On a donc

$$\frac{1}{n} \mathbb{E}^i(S_n^j) = w_j + \frac{1}{n} [z_{ij} - (P^n Z)_{ij}]. \quad (3.5.62)$$

Comme les éléments de P^n sont compris entre 0 et 1, cette expression converge vers w_j lorsque n tend vers l'infini. \square

Exemple 3.5.16. Considérons un modèle d'Ehrenfest à deux boules, dont le graphe est le suivant:

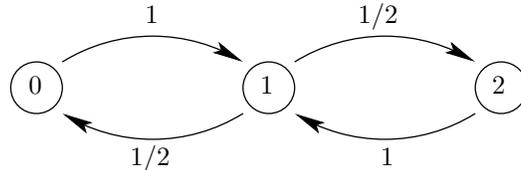


FIGURE 3.9. La chaîne de Markov associée au modèle d'Ehrenfest à deux boules.

Sa matrice de transition est donnée par

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix}. \quad (3.5.63)$$

En résolvant l'équation $wP = w$, on obtient la distribution stationnaire

$$w = \left(\frac{1}{4}, \frac{2}{4}, \frac{1}{4} \right). \quad (3.5.64)$$

Les temps de récurrence moyens sont donc donnés par $r_0 = r_2 = 4$ et $r_1 = 2$. Un calcul fournit la matrice Z ,

$$Z = \frac{1}{8} \begin{pmatrix} 7 & 2 & -1 \\ 1 & 6 & 1 \\ -1 & 2 & 7 \end{pmatrix}, \quad (3.5.65)$$

de laquelle nous déduisons les temps moyens de premier passage

$$\begin{aligned} m_{01} &= \frac{z_{11} - z_{01}}{w_1} = 1 = m_{21}, \\ m_{02} &= \frac{z_{22} - z_{02}}{w_2} = 4 = m_{20}, \\ m_{12} &= \frac{z_{22} - z_{12}}{w_3} = 3 = m_{10}. \end{aligned} \quad (3.5.66)$$

Finalement, les variances des temps relatifs de séjour sont asymptotiquement

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} V(S_n^0) &= 2w_0z_{00} - w_0 - w_0^2 = \frac{1}{4}, \\ \lim_{n \rightarrow \infty} \frac{1}{n} V(S_n^1) &= 2w_1z_{11} - w_1 - w_1^2 = 0. \end{aligned} \quad (3.5.67)$$

Nous voyons donc que le temps relatif de séjour en 1 est égal à $1/2$ presque sûrement. Ceci s'explique par le fait que $P^n = P$ si n est impair, alors que

$$P^n = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix} \quad (3.5.68)$$

si P est pair. La chaîne partant de 1 se retrouve donc toujours en 1 aux temps pairs, ce qui est évident lorsqu'on regarde le graphe du processus.