

L. DECREUSEFOND

MESURE
INTÉGRATION
ET PROBABILITÉS

Table des matières

I	Préparation à l'écrit	5
1	Mesures & lois	7
§ 1	Dénombrabilité	7
§ 2	Tribu, mesures, etc.	8
§ 3	Fonctions mesurables	11
§ 4	Intégration	12
2	Variables aléatoires	35
§ 1	De la théorie de la mesure aux probabilités	35
§ 2	Calculs de lois	41
§ 3	Exercices	45
§ 4	Vecteurs gaussiens	58
3	Convergences	65
§ 1	Convergence presque-sûre	66
§ 2	Convergence en probabilité	67
§ 3	Convergence en loi	70
§ 4	Exercices	76
4	Modélisation	79
§ 1	Marche aléatoire	79
§ 2	Processus de Poisson	79
II	Préparation aux oraux	81
5	Chaînes de Markov	83
§ 1	Version simple	83
§ 2	Définition et exemples	84
§ 3	Propriété de Markov forte	89
§ 4	Classification des états	91
§ 5	Mesures et probabilité invariantes	99
§ 6	Calcul pratique de la probabilité invariante	113

§ 7 Problèmes	114
§ 8 Notes et commentaires	118
6 Espérance Conditionnelle	121
§ 1 Définition et premiers exemples	121
§ 2 Propriétés de l'espérance conditionnelle	122
§ 3 Conditionnement des vecteurs gaussiens	126
§ 4 Cas intégrable	127
§ 5 Exercices	130

Première partie

Préparation à l'écrit

1

Mesures & lois

§ 1 Dénombrabilité

Ce qui suit sur la dénombrabilité est principalement issu de KUTTNER, *Modern Analysis*.

Le concept de dénombrabilité joue un rôle primordial en probabilités parce que la notion d'ensemble mesurable est stable par réunion dénombrable (voir 1.7).

Définition 1.1 Un ensemble est fini s'il est en bijection avec un certain $\{1, \dots, n\}$.

Définition 1.2 Un ensemble est dit dénombrable s'il est en bijection avec \mathbf{N} , l'ensemble des entiers naturels.

■ **Exemple 1.1** Quelques exemples d'ensembles dénombrables.

- $2\mathbf{N}$, l'ensemble des entiers pairs, est dénombrable.
- $\mathcal{P}(\mathbf{N})$, l'ensemble des parties de \mathbf{N} , ne l'est pas. Il est en bijection avec l'ensemble des réels \mathbf{R} . En effet, pour une partie A de \mathbf{N} , on peut définir « la suite indicatrice » $\mathbf{1}_A(n)$ par

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

Puis é cette suite, on peut associer le réel x_A défini par :

$$x_A = \sum_{n=1}^{\infty} 2^{-n} \mathbf{1}_A(n).$$

Il est évident que $\mathcal{P}(\mathbf{N})$ est alors en bijection avec $\{0, 1\}^{\mathbf{N}}$. Etant donné qu'à tout réel entre 0 et 1, on peut associer son développement diadique, il existe une injection $[0, 1]$ dans $\{0, 1\}^{\mathbf{N}}$ donc $\mathcal{P}(\mathbf{N})$ n'est pas dénombrable.

■

La dénombrabilité est stable par réunion et produit cartésien.

Théorème 1.3 Si X et Y sont deux ensembles dénombrables alors $X \times Y$ et $X \cup Y$ sont dénombrables.

Tout est basé sur le résultat essentiel qui stipule que $\mathbf{N} \times \mathbf{N}$ est en bijection avec \mathbf{N} . La bijection est construite comme indiquée sur la figure 1.1. Par exemple, l'élément $(2, 0)$ est envoyé sur 3 et l'élément $(0, 2)$ est envoyé sur 5.

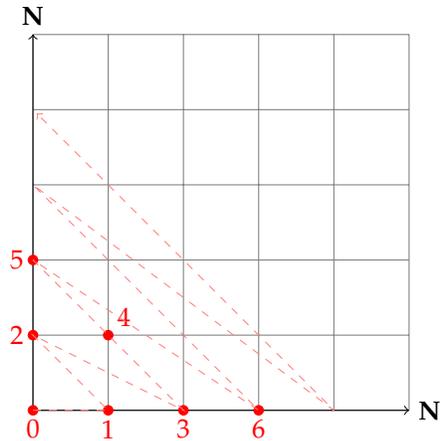


FIGURE 1.1: Bijection de $\mathbf{N} \times \mathbf{N}$ avec \mathbf{N} .

Corollaire 1.4 L'ensemble des entiers relatifs, \mathbf{Z} , est dénombrable.

Le théorème suivant est loin d'être trivial.

Théorème 1.5 S'il existe une injection de l'ensemble X dans l'ensemble Y et une injection de Y dans X alors il existe une bijection de X dans Y .

Corollaire 1.6 L'ensemble des rationnels, \mathbf{Q} est dénombrable.

Démonstration. Il existe évidemment une injection de \mathbf{N} dans \mathbf{Q} et par construction de \mathbf{Q} , il existe une injection de \mathbf{Q} dans $\mathbf{Z} \times \mathbf{Z}$, qui d'après le théorème précédent est en bijection avec \mathbf{N} . Par conséquent, \mathbf{Q} est en bijection avec \mathbf{N} . ■

§ 2 Tribu, mesures, etc.

Il était déjà un ouvrage indispensable de mon temps, il l'est toujours. Pour tout ce qui touche la théorie de la mesure et de l'intégration, RUDIN, *Analyse Réelle et Complexe* est une indémodable ressource.

Sauf à renoncer à l'axiome du choix, on ne peut pas construire de mesure sur toutes les parties de n'importe quel espace. Il faut donc définir l'ensemble des parties « mesurables ». Pour ce faire, on introduit la notion de tribu. Ensuite seulement, viendra la notion de mesure.

Définition 1.7 Soit Ω un ensemble. Un sous-ensemble \mathcal{A} de $\mathcal{P}(\Omega)$ est une tribu lorsque les trois conditions suivantes sont vérifiées :

1. $\emptyset \in \mathcal{A}$,
2. $A \in \mathcal{A} \implies A^c \in \mathcal{A}$,
3. $A_i \in \mathcal{A}$ pour tout $i \in \mathbf{N} \implies \cup_i A_i \in \mathcal{A}$.

Les éléments d'une tribu sont appelés *événements* en probabilités ou (*ensembles*) *mesurables* en théorie de la mesure.

Les exemples suivantes sont d'une utilisation constante.

- La tribu *grossière* est la tribu constituée des seuls éléments \emptyset et Ω .
 - La tribu la plus *fine* est la tribu $\mathcal{P}(\Omega)$.
 - Si \mathcal{A}_1 et \mathcal{A}_2 sont deux tribus alors $\mathcal{A}_1 \cap \mathcal{A}_2$ est encore une tribu.
 - Pour $\mathcal{C} \subset \mathcal{P}(\Omega)$, $\mathcal{P}(\Omega)$ est une tribu qui contient \mathcal{C} donc l'ensemble des tribus qui contiennent \mathcal{C} est non vide. Par conséquent, cet ensemble a un plus petit élément, au sens de l'intersection : c'est la plus petite tribu qui contient \mathcal{C} . On appelle cette tribu la tribu *engendrée par \mathcal{C}* , elle est notée $\sigma(\mathcal{C})$.
- Pour un ensemble A de Ω , $\sigma(\{A\}) = \{A, A^c, \emptyset, \Omega\}$.
- Sur un ensemble topologique, la tribu borélienne est la tribu engendrée par les ouverts. La tribu borélienne sur \mathbf{R} est la plus petite tribu qui contient les intervalles ouverts de la forme $]a, b[$ avec $-\infty \leq a < b$. De même, sur \mathbf{R}^k , la tribu borélienne est la plus petite tribu qui contient les pavés $]a_1, b_1[\times \dots \times]a_k, b_k[$.

Définition 1.8 Une application μ de \mathcal{A} dans \mathbf{R}^+ est une mesure lorsqu'elle satisfait les deux propriétés suivantes :

- $\mu(\emptyset) = 0$,
- μ est une application pour σ -additive : pour toute famille $(A_j, j \in \mathbf{N}^*)$ d'éléments de \mathcal{A} deux à deux disjoints,

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

Une mesure est dite de probabilité lorsque la masse totale, i.e., $\mu(\Omega)$, vaut 1. Dans ce cas, on parle de mesure de probabilités et on la note \mathbf{P} plutôt que μ .

■ **Exemple 1.2** L'exemple le plus simple de mesure est donnée par la

mesure de Dirac en un point $a \in \Omega$:

$$\delta_a(A) = 1 \text{ si } a \in A, \delta_a(A) = 0 \text{ sinon.}$$

■

Dès que Ω n'est pas dénombrable, il est impossible de décrire une mesure en donnant sa valeur pour tous les ensembles mesurables. Arrive à notre secours le théorème de classe monotone qui nous dit, en substance, qu'une mesure est totalement déterminée par sa valeur sur un ensemble d'ensembles suffisamment riche.

Définition 1.9 Un ensemble \mathcal{C} de parties de Ω est un π -système s'il est stable par intersections finie :

$$A, B \in \mathcal{C} \implies A \cap B \in \mathcal{C}.$$

En particulier, on montre que

$$\mathcal{R} = \left\{ \bigcup_{i=1}^n I_i, n \in \mathbf{N}^*, I_i = (a_i, b_i) \right\}$$

où (a, b) est l'une des quatre intervalles $[a, b],]a, b], [a, b[,]a, b[$ est un π -système de $\mathcal{P}(\mathbf{R})$.

Définition 1.10 Un ensemble \mathcal{M} de parties de Ω est une classe monotone si

- Si A et B sont deux éléments de \mathcal{M} tels que $A \subset B$ alors $B \setminus A$ appartient à \mathcal{M} ,
- toute limite croissante (i.e., réunion d'ensembles inclus les uns dans les autres) d'éléments de \mathcal{M} est dans \mathcal{M} .

En particulier, pour deux mesures de probabilité \mathbf{P} et \mathbf{Q} , l'ensemble

$$\mathcal{S} = \{A \in \mathcal{P}(\mathbf{R}), \mathbf{P}(A) = \mathbf{Q}(A)\}$$

est une classe monotone.

On a le résultat suivant :

Théorème 1.11 — classe monotone. Soit \mathcal{C} un π -système et \mathcal{M} une classe monotone contenant \mathcal{C} alors $\sigma(\mathcal{C}) \subset \mathcal{M}$.

En conséquence, on en déduit :

Théorème 1.12 Deux mesures qui coïncident sur \mathcal{R} sont égales.

Corollaire 1.13 Pour identifier une mesure sur \mathbf{R} , il faut et il suffit que l'on connaisse $\mathbf{P}(]-\infty, x])$ pour tout réel x .

■ **Remarque 1** Ce résultat s'étend sans changement aux dimensions supérieures : pour identifier une probabilité sur \mathbf{R}^d , il faut et il suffit que

l'on connait

$$\mathbf{P}([-\infty, x_1] \times \dots \times [-\infty, x_d])$$

pour tout d -uplet (x_1, \dots, x_d) .

Un autre théorème fondamental de la théorie de la mesure est le suivant :

Théorème 1.14 Il existe une unique mesure sur \mathbf{R}^k , notée λ , munie de la tribu des boréliens, qui coïncident avec la mesure de longueur/surface/volume sur les pavés, i.e., telle que

$$\lambda([a_1, b_1] \times \dots \times [a_k, b_k]) = (b_1 - a_1) \dots (b_k - a_k).$$

Cette mesure s'appelle la mesure de Lebesgue.

§ 3 Fonctions mesurables

Les fonctions mesurables sont à la théorie de la mesure ce que les fonctions continues sont à la topologie.

Définition 1.15 Une fonction f de $(\Omega_1, \mathcal{A}_1)$ dans $(\Omega_2, \mathcal{A}_2)$ est mesurable lorsque

$$f^{-1}(C) \in \mathcal{A}_1 \text{ pour tout } C \in \mathcal{A}_2.$$

Par le théorème de classe monotone, on peut se restreindre à prouver cette propriété pour des éléments C d'une algèbre engendrant la tribu \mathcal{A}_2 . Ce qui signifie, que si $\Omega_2 = \mathbf{R}$ et que \mathcal{A}_2 est la tribu borélienne, on peut se contenter de le prouver pour les éléments de \mathcal{R} (voir ci-dessus).

- Une fonction continue est mesurable.
- La somme, le produit de deux fonctions mesurables sont mesurables.
- La composition de deux fonctions mesurables est mesurable.
- Le suprémum et l'infimum d'une famille de fonctions mesurables sont mesurables :

$$\sup_n f_n \text{ et } \inf_n f_n \text{ sont mesurables.}$$

- Par conséquent, les limites inférieures et supérieures d'une suite de fonctions sont mesurables.
- En particulier, si une suite de fonctions converge simplement, ses limites inférieures et supérieures coïncident donc une limite simple de fonctions mesurables est mesurable.

Ce dernier résultat est très intéressant parce que pour les fonctions continues, on est assuré de la continuité d'une limite de fonctions continues que si la convergence est uniforme.

■ **Définition 1.16** Une variable aléatoire est une fonction mesurable.

Partant d'un espace mesuré $(\Omega_1, \mathcal{A}_1, \mu)$ et d'une application mesurable f de $(\Omega_1, \mathcal{A}_1, \mu)$ dans $(\Omega_2, \mathcal{A}_2)$, il est naturel de se demander comment se transforme la mesure μ sous l'effet de f . Par exemple, si Ω_1 est une plaque inhomogène à laquelle, on fait subir divers traitement, on peut se demander comment seront réparties les inhomogénéités de la plaque transformée.

■ **Définition 1.17** Soit $(\Omega_1, \mathcal{A}_1, \mu)$ un espace mesuré et d'une application mesurable f de $(\Omega_1, \mathcal{A}_1, \mu)$ dans $(\Omega_2, \mathcal{A}_2)$, la mesure image de μ par f , notée $f^*\mu$ est définie par :

$$f^*\mu(A) = \mu(f^{-1}(A)) \text{ pour tout } A \in \mathcal{A}_2.$$

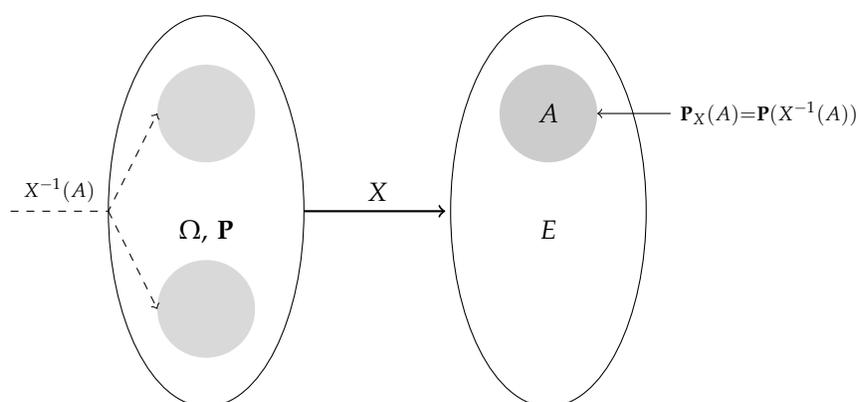


FIGURE 1.2: Mesure image

■ **Remarque 2 — vocabulaire.** La loi d'une v.a. X de Ω dans \mathbf{R}^n , notée \mathbf{P}_X , est la mesure image de \mathbf{P} par X :

$$\mathbf{P}_X(A) = \mathbf{P}(X^{-1}(A)) = \mathbf{P}(X \in A) = \mathbf{P}(\omega : X(\omega) \in A),$$

pour tout A borélien de \mathbf{R}^n . Si on avait respecté la notation de la TdM, on devrait avoir noté \mathbf{P}^*X au lieu de \mathbf{P}_X .

■ **Remarque 3** Si μ et ν sont des lois, c'est-à-dire si μ est la loi d'une v.a. X et ν la loi d'une v.a. Y , ces deux v.a. sont indépendantes si et seulement si

$$\mathbf{P}((X, Y) \in A \times B) = \mathbf{P}(X \in A, Y \in B) = \mathbf{P}(X \in A)\mathbf{P}(Y \in B) = \mu(A)\nu(B),$$

autrement dit, d'après (1.3), si et seulement si la loi du couple (X, Y) est la mesure produit des lois de chacune des composantes, voir (??).

§ 4 Intégration

Définition 1.18 Une fonction mesurable $f : (E, \mathcal{E}) \rightarrow \mathbf{R}^+$ est dite étagée si elle prend un nombre fini de valeurs. On note \mathfrak{E} l'ensemble des fonctions étagées.

Soit $\alpha_1, \dots, \alpha_n$ les valeurs possibles de f et $A_k = \{x : f(x) = \alpha_k\}$ pour $k = 1, \dots, n$. On peut toujours écrire

$$f(x) = \sum_{k=1}^n \alpha_k \mathbf{1}_{A_k}(x).$$

On remarque que les A_k sont disjoints deux à deux et les α_k tous distincts et non nuls.

Dans la suite, on supposera toujours que la décomposition d'une fonction étagée est celle là.

Lemme 1.19 Toute fonction mesurable positive est limite simple croissante de fonctions étagées : $f_1 \leq f_2 \leq \dots \leq f_n \uparrow f$.

Démonstration. Pour une fonction positive mesurable, on pose

$$f_n(x) = \begin{cases} n & \text{si } f(x) \geq n \\ k2^{-n} & \text{si } k2^{-n} \leq f(x) < (k+1)2^{-n}, k \in \{0, \dots, n2^n - 1\}, \end{cases}$$

converge simplement vers f .

Soit x fixé, on considère $f(x)$ et pour n tel que $2^n \geq f(x)$, on pose

$$k_n = [2^n f(x)] \text{ soit } 2^{-n} k_n \leq f(x) \leq (k_n + 1)2^{-n}.$$

Par définition de f_n , on a $f_n(x) = k_n 2^{-n}$, donc

$$f(x) - 2^{-n} \leq f_n(x) \leq f(x)$$

et par conséquent, on a bien que f_n tend vers f pour tout $x \in E$. ■

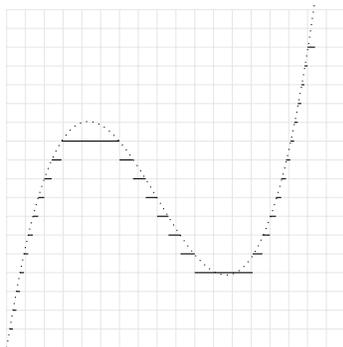


FIGURE 1.3: Une fonction (en pointillés) et son approximation par une fonction étagée. Comme la fonction est très régulière pour les besoins du dessin, les A_i sont des réunions d'intervalle mais pour une fonction mesurable quelconque (non continue), ce ne sera plus le cas.

On peut maintenant définir l'intégrale par rapport à une mesure μ quelconque.

Définition 1.20 L'intégrale de $f \in \mathfrak{E}$ est naturellement définie par

$$\int f \, d\mu = \sum_{i=1}^n \alpha_i \mu(A_i)$$

D'après le lemme 1.19, il est tentant de définir l'intégrale d'une fonction positive comme la limite des intégrales des fonctions étagées qui l'approchent. Il resterait à s'assurer que la limite ne dépend pas de la suite choisie. Pour éviter cet écueil, on définira l'intégrale de la fonction f par un supremum qui permet de s'affranchir d'une suite approximante particulière.

Dans ce qui suit, on manipule des fonctions qui valent éventuellement $\pm\infty$ en certains points. On est donc obligé de travailler avec des fonctions à valeurs dans $\overline{\mathbf{R}} = \mathbf{R} \cup \{\pm\infty\}$. L'arithmétique dans cet espace est l'arithmétique usuelle selon les conventions suivantes.

- $a + (+\infty) = +\infty$, pour tout $a \in \overline{\mathbf{R}}$,
- $\infty \cdot 0 = 0$, i.e. 0 reste absorbant pour la multiplication
- toute série à termes positifs converge dans $\overline{\mathbf{R}}$, c'est-à-dire vers éventuellement $+\infty$.

Définition 1.21 Pour f mesurable positive, on définit

$$\int f \, d\mu = \sup_{g \in \mathfrak{E}} \left\{ \int g \, d\mu, 0 \leq g \leq f \right\} \in \overline{\mathbf{R}}^+.$$

Lemme 1.22 Soient $a, b \geq 0$ et f, g deux fonctions étagées positives.

Alors

$$\int (af + bg) \, d\mu = a \int f \, d\mu + b \int g \, d\mu.$$

Démonstration. Pour $f = \mathbf{1}_A$ et $g = \mathbf{1}_B$,

$$af + bg = a\mathbf{1}_{A \setminus A \cap B} + b\mathbf{1}_{B \setminus A \cap B} + (a+b)\mathbf{1}_{A \cap B}.$$

Par conséquent,

$$\begin{aligned} \int (af + bg) \, d\mu &= a\mu(A \setminus A \cap B) + b\mu(B \setminus A \cap B) + (a+b)\mu(A \cap B) \\ &= a\mu(A) + b\mu(B) = a \int f \, d\mu + b \int g \, d\mu. \end{aligned}$$

Le résultat général s'en déduit par récurrence double sur le nombre de valeurs de f et g . ■

Pour une fonction de signe quelconque, on considère

$$f^+(x) = \max(f(x), 0) \text{ et } f^-(x) = \min(-f(x), 0).$$

Il est évident que

$$f^+ + f^- = |f| \text{ et que } f^+ - f^- = f.$$

L'idée est de définir l'intégrale de f comme la différence des intégrales de f^+ et f^- . Cela n'est possible que si ces deux quantités sont finies or ceci est impliqué par la condition que l'intégrale de $|f|$, qui est bien définie puisque $|f|$ est positive, soit finie. De cette discussion, on déduit la définition suivante.

Définition 1.23 Soit $f : E \rightarrow \mathbf{R}$ mesurable telle que $\int f^+ d\mu$ ou $\int f^- d\mu$ soit finie, on pose

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu.$$

La fonction est dite μ -intégrable si $\int f^+ d\mu$ et $\int f^- d\mu$ sont finies.

Lemme 1.24 Soient f, g deux fonctions mesurable de $E \rightarrow \overline{\mathbf{R}}$, dont les intégrales sont définies. Alors

1. $f \leq g$ implique $\int f d\mu \leq \int g d\mu$.
2. $|\int f d\mu| \leq \int |f| d\mu$.

Démonstration. ÉTAPE 1. Supposons $0 \leq f \leq g$. Toute fonction φ étagée plus petite que f est aussi plus petite que g . On a donc l'inclusion

$$\left\{ \int \varphi d\mu : 0 \leq \varphi \leq f, \varphi \text{ étagée} \right\} \subset \left\{ \int \varphi d\mu : 0 \leq \varphi \leq g, \varphi \text{ étagée} \right\}.$$

Il en découle que le supremum de l'ensemble de gauche est plus petit que le supremum de l'ensemble de droite, ce qui se lit $\int f d\mu \leq \int g d\mu$.

ÉTAPE 2. Traitons le cas où $f \leq g$ sont quelconques. On a $f^+ \leq g^+$ et $f^- \geq g^-$, donc, d'après ci-dessus, $\int f^+ d\mu \leq \int g^+ d\mu$ et $\int f^- d\mu \geq \int g^- d\mu$. En soustrayant, $\int f d\mu \leq \int g d\mu$.

On a $f \leq |f|$ donc $\int f d\mu \leq \int |f| d\mu$. De même $-f \leq |f|$, donc $\int (-f) d\mu \leq \int |f| d\mu$. Or $\int (-f) d\mu = \int (-f)^+ d\mu - \int (-f)^- d\mu = \int f^- d\mu - \int f^+ d\mu = -\int f d\mu$. On a donc $-\int f d\mu \leq \int |f| d\mu$, ce qui conclut la preuve. ■

Théorème 1.25 Soient f, f_n ($n \in \mathbf{N}$) des fonctions mesurables de $E \rightarrow \overline{\mathbf{R}}^+$ telles que $0 \leq f_n \uparrow f$. Alors

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu.$$

Démonstration. Pour tout n , $f_n \leq f_{n+1} \leq f$, donc $\int f_n d\mu$ est une suite croissante majorée par $\int f d\mu$. Sa limite existe, et $\lim \int f_n d\mu \leq \int f d\mu$. Il reste à démontrer l'autre inégalité.

Soit une fonction étagée g telle que $0 \leq g \leq f$. Soit $\varepsilon > 0$ et $B_n = \{f_n \geq (1 - \varepsilon)g\}$. Décrivons g sous la forme $g = \sum_{i=1}^p \alpha_i \mathbf{1}_{A_i}$.

On a :

$$\begin{aligned} \int f_n d\mu &\geq \int f_n \mathbf{1}_{B_n} d\mu \geq \int (1 - \varepsilon) g \mathbf{1}_{B_n} d\mu \\ &= \int \left(\sum_{i=1}^p (1 - \varepsilon) \alpha_i \mathbf{1}_{A_i \cap B_n} \right) d\mu \\ &= \sum_{i=1}^p (1 - \varepsilon) \alpha_i \mu(A_i \cap B_n), \end{aligned}$$

où, dans la dernière égalité, on a utilisé la linéarité de l'intégrale pour les fonctions étagées (Lemme 1.22). Comme $B_n \uparrow E$, $\mu(A_i \cap B_n)$ tend vers $\mu(A_i)$ pour tout i , lorsque $n \rightarrow \infty$. Par conséquent,

$$\lim_{n \rightarrow \infty} \int f_n d\mu \geq (1 - \varepsilon) \sum_{i=1}^p \alpha_i \mu(A_i) = (1 - \varepsilon) \int g d\mu.$$

Cela étant vrai pour tout $\varepsilon > 0$, nous avons $\lim_{n \rightarrow \infty} \int f_n d\mu \geq \int g d\mu$. En prenant le supremum sur l'ensemble des fonctions g étagées positives minorant f , on conclut que $\lim_{n \rightarrow \infty} \int f_n d\mu \geq \int f d\mu$. ■

■ **Remarque 4** On ne peut pas se passer du $(1 - \varepsilon)$ parce que dans le cas limite $f = g = \mathbf{1}_A$, la suite de fonctions $f_n = (1 - n^{-1})\mathbf{1}_A$ satisfait les hypothèses mais pour $\varepsilon = 0$, $A_n = \emptyset$ pour tout n donc dans ce cas, A_n ne tend pas vers A .

Théorème 1.26 — Lemme de Fatou. Soit $(f_n, n \geq 1)$ une suite de fonctions (mesurables) positives,

$$\int \liminf_n f_n d\mu \leq \liminf_n \int f_n d\mu.$$

Démonstration. Par définition,

$$\liminf_n f_n(x) = \sup_k \inf_{n \geq k} f_n(x) = \liminf_k \inf_{n \geq k} f_n(x).$$

La suite $(g_k = \inf_{n \geq k} f_n, k \in \mathbf{N})$ est croissante positive donc le théorème de convergence monotone assure que

$$\int \lim_k g_k d\mu = \lim_k \int g_k d\mu.$$

Par conséquent,

$$\int \liminf_n f_n d\mu = \lim_k \int \inf_{n \geq k} f_n d\mu \leq \lim_k \inf_{n \geq k} \int f_n d\mu = \liminf_n \int f_n d\mu.$$

La preuve est terminée. ■

Si la mesure est finie, on peut évidemment remplacer l'hypothèse de positivité par l'hypothèse que les f_n sont inférieurement bornées. Dans le cas où toutes les fonctions ne sont pas positives, il faut une contrainte de domination.

Théorème 1.27 — Convergence dominée. Soit $(f_n, n \geq 1)$ une suite de fonctions (mesurables) qui converge simplement vers f . Si de plus, il existe g telle que

$$|f_n(x)| \leq g(x), \text{ pour tout } x \text{ et } \int g \, d\mu < \infty$$

alors

$$\lim_{n \rightarrow \infty} \int f_n \, d\mu = \int (\lim_{n \rightarrow \infty} f_n) \, d\mu = \int f \, d\mu.$$

Démonstration. Comme $|f_n| \leq g$ et que l'intégrale de g est finie, celle de $|f_n|$ l'est aussi pour tout n entier. D'autre part, $g - f_n$ et $g + f_n$ sont mesurables positives et $g \pm f_n$ tend vers $g \pm f$ donc

$$\liminf_n g \pm f_n = g \pm f.$$

Appliquons le lemme de Fatou aux deux suite $((g \pm f_n), n \geq 1)$, on obtient

$$\begin{aligned} \int (g - f) \, d\mu &\leq \liminf_n \int (g - f_n) \, d\mu \\ \int (g + f) \, d\mu &\leq \liminf_n \int (g + f_n) \, d\mu. \end{aligned}$$

De plus, le lemme de Fatou implique aussi que

$$\int |f| \, d\mu = \int \liminf_n |f_n| \, d\mu \leq \liminf_n \int |f_n| \, d\mu \leq \int g \, d\mu < +\infty,$$

donc f est intégrable. On peut donc écrire

$$\begin{aligned} \int g \, d\mu - \int f \, d\mu &\leq \int g \, d\mu + \liminf_n \int (-f_n) \, d\mu \\ &= \int g \, d\mu - \limsup_n \int f_n \, d\mu \\ \int g \, d\mu + \int f \, d\mu &\leq \int g \, d\mu + \liminf_n \int f_n \, d\mu. \end{aligned}$$

On tire de ces inégalités que

$$\int f \, d\mu \leq \liminf_n \int f_n \, d\mu \leq \limsup_n \int f_n \, d\mu \leq \int f \, d\mu,$$

d'où l'on conclut que la suite $\int f_n \, d\mu$ converge et que la limite est $\int f \, d\mu$. ■

Si on dispose de deux espaces mesurés $(\Omega_1, \mathcal{A}_1, \mu)$ et $(\Omega_2, \mathcal{A}_2, \nu)$, on veut construire une mesure sur le produit cartésien $\Omega_1 \times \Omega_2$. La première difficulté à surmonter est la définition de la tribu sur $E \times F$. Les éléments de $\mathcal{A}_1 \times \mathcal{A}_2$ sont les produits cartésiens d'un élément de \mathcal{A}_1 et d'un élément de \mathcal{A}_2 . Comme la réunion de deux rectangles n'est pas un rectangle, $\mathcal{A}_1 \times \mathcal{A}_2$ n'est pas une tribu. Qu'à cela ne tienne, on

note $\mathcal{A}_1 \otimes \mathcal{A}_2$ la plus petite tribu qui contient $\mathcal{A}_1 \times \mathcal{A}_2$ et le tour est joué!

Les deux lemmes suivants se démontrent de manière identique en utilisant le théorème de classe monotone. Nous ne donnons donc que la démonstration de l'un d'entre eux.

Lemme 1.28 Soit f une fonction mesurable de $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ dans (E, \mathcal{E}) . Pour tout $\omega_1 \in \Omega_1$, la fonction $\omega_2 \mapsto f(\omega_1, \omega_2)$ est mesurable de $(\Omega_2, \mathcal{A}_2)$ dans (E, \mathcal{E}) .

Lemme 1.29 Soit f une fonction mesurable positive de $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ dans $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$. Soit μ une mesure σ -finie sur $(\Omega_1, \mathcal{A}_1)$. La fonction

$$\omega_2 \mapsto \int f(\omega_1, \omega_2) \, d\mu(\omega_1)$$

est mesurable de $(\Omega_2, \mathcal{A}_2)$ dans $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$.

Démonstration. Si $f = \mathbf{1}_{A \times B}$ avec $A \in \mathcal{A}_1$ et $B \in \mathcal{A}_2$ alors

$$\int f(\omega_1, \omega_2) \, d\mu(\omega_1) = \mathbf{1}_B(\omega_2)\mu(A).$$

Dans ce cas, comme $B \in \mathcal{A}_2$, la fonction $\omega_2 \mapsto \int f(\cdot, \omega_2) \, d\mu$ est mesurable de $(\Omega_2, \mathcal{A}_2)$ dans $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$.

Soit

$$\mathcal{M} = \left\{ C \in \mathcal{A}_1 \otimes \mathcal{A}_2, \omega_2 \mapsto \int \mathbf{1}_C(\omega_1, \omega_2) \, d\mu(\omega_1) \text{ est mesurable} \right\}.$$

On vient de démontrer que $\mathcal{A}_1 \times \mathcal{A}_2 \subset \mathcal{M}$. D'autre part, \mathcal{M} est une classe monotone. En effet, si les $(C_n, n \geq 1)$ sont des ensembles croissants appartenant à \mathcal{M} , d'après le théorème de convergence monotone,

$$\int \mathbf{1}_C(\omega_1, \omega_2) \, d\mu(\omega_1) = \lim_n \int \mathbf{1}_{C_n}(\omega_1, \omega_2) \, d\mu(\omega_1). \quad (1.2)$$

Etant donné que toute limite de fonctions mesurables est mesurable, C appartient à \mathcal{M} . Soit maintenant $(C_n, n \geq 1)$ une suite décroissante d'éléments de \mathcal{M} . Supposons dans un premier temps que μ soit finie. Dans ce cas, le théorème de convergence dominée implique (1.2) et la mesurabilité s'ensuit. Si la mesure μ n'est que σ -finie, il existe une suite $(K_l, l \geq 1)$ de compacts tels que $\cup_l K_l = \Omega_1$ et $\mu(K_l) < +\infty$, pour tout l . On applique alors le raisonnement précédent à $\mu|_{K_l}$ donc

$$\omega_2 \mapsto \int_{K_l} f(\omega_1, \omega_2) \, d\mu(\omega_1)$$

est mesurable. Par convergence monotone, en faisant tendre l vers l'infini, la mesurabilité de $\int f(\cdot, \omega_2) \, d\mu$ s'ensuit.

Par linéarité, le résultat reste vrai pour les fonctions étagées. Par passage à la limite monotone, le résultat est vrai pour les fonctions f mesurables positives. ■

Théorème 1.30 — Fubini. Soit $(\Omega_1, \mathcal{A}_1, \mu_1)$ et $(\Omega_2, \mathcal{A}_2, \mu_2)$ deux espaces mesurés avec μ_1 et μ_2 σ -finies. Il existe une unique mesure, notée $\mu \otimes \nu$, dite mesure produit de μ et ν , sur $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \times \mathcal{A}_2)$ qui soit telle que

$$\mu_1 \otimes \mu_2(A \times B) = \mu_1(A)\mu_2(B), \text{ pour tout } A \in \mathcal{A}_1, B \in \mathcal{A}_2. \quad (1.3)$$

Par ailleurs, soit $f : \Omega_1 \times \Omega_2 \rightarrow \mathbf{R}$, mesurable, satisfaisant l'une des deux conditions suivantes :

- (1) f est positive.
- (2) La fonction f est intégrable par rapport à la mesure $\mu_1 \otimes \mu_2$:

$$\int |f| \, d\mu_1 \otimes \mu_2 < +\infty.$$

L'identité suivante est satisfaite :

$$\begin{aligned} \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) \, d\mu_2(\omega_2) \right) \, d\mu_1(\omega_1) \\ = \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2) \, d\mu_1(\omega_1) \right) \, d\mu_2(\omega_2) \\ = \int f \, d(\mu_1 \otimes \mu_2). \end{aligned}$$

L'unicité découle immédiatement du théorème de classe monotone puisqu'une mesure est toujours parfaitement caractérisée par ses valeurs sur un ensemble d'ensembles qui engendrent la tribu.

Soit maintenant, l'application γ

$$\begin{aligned} \gamma : \mathcal{A}_1 \otimes \mathcal{A}_2 &\longrightarrow \mathbf{R}^+ \\ C &\longrightarrow \int_{\Omega_1} \int_{\Omega_2} \mathbf{1}_C(\omega_1, \omega_2) \, d\mu_2(\omega_2) \, d\mu_1(\omega_1). \end{aligned}$$

Notons que γ est bien définie en vertu des lemmes 1.28 et 1.29. Par convergence monotone, γ est bien une mesure sur $\mathcal{A}_1 \otimes \mathcal{A}_2$: soit $(C_n, n \geq 1)$ une suite de mesurables deux à deux disjoints,

$$\mathbf{1}_{\cup_n C_n} = \lim_n \uparrow \mathbf{1}_{\cup_{k \leq n} C_k} = \lim_n \uparrow \sum_{k \leq n} \mathbf{1}_{C_k},$$

donc

$$\int_{\Omega_2} \mathbf{1}_C(\omega_1, \omega_2) \, d\mu_2(\omega_2) = \lim_n \uparrow \sum_{k \leq n} \int_{\Omega_2} \mathbf{1}_{C_k}(\omega_1, \omega_2) \, d\mu_2(\omega_2).$$

Par conséquent, par convergence monotone, on peut intervertir la limite en n et l'intégrale en μ_1 , d'où

$$\gamma(C) = \lim_n \uparrow \sum_{k \leq n} \gamma(C_k) = \sum_n \gamma(C_n).$$

On note cette mesure $\mu_1 \otimes \mu_2$. D'autre part, pour $C = A \times B$ avec $A \in \mathcal{A}_1$ et $B \in \mathcal{A}_2$,

$$\begin{aligned} \gamma(A \times B) &= \int_{\Omega_1} \int_{\Omega_2} \mathbf{1}_{A \times B}(\omega_1, \omega_2) \, d\mu_2(\omega_2) \, d\mu_1(\omega_1) \\ &= \mu_1(A) \mu_2(B) \\ &= \int_{\Omega_2} \int_{\Omega_1} \mathbf{1}_{A \times B}(\omega_1, \omega_2) \, d\mu_1(\omega_1) \, d\mu_2(\omega_2). \end{aligned}$$

Pour f de la forme $\mathbf{1}_{A \times B}$, on a donc montré

$$\begin{aligned} \int_{\Omega_1} \int_{\Omega_2} \mathbf{1}_{A \times B}(\omega_1, \omega_2) \, d\mu_2(\omega_2) \, d\mu_1(\omega_1) \\ = \int_{\Omega_2} \int_{\Omega_1} \mathbf{1}_{A \times B}(\omega_1, \omega_2) \, d\mu_1(\omega_1) \, d\mu_2(\omega_2). \end{aligned}$$

Selon le même principe que dans la démonstration du lemme 1.29, on étend cette identité d'abord aux fonctions indicatrices d'ensemble de $\mathcal{A}_1 \otimes \mathcal{A}_2$, puis aux fonctions étagées positives, aux fonctions positives et enfin aux fonctions $\mu_1 \otimes \mu_2$ -intégrables.

Ensembles μ -négligeables

Un ensemble μ -négligeable est un ensemble $N \in \mathbf{E}$ qui peut être inclus dans un ensemble mesurable A tel que $\mu(A) = 0$.

Proposition 1.31 Soit $f : \mathbf{E} \rightarrow \overline{\mathbf{R}}$ une fonction mesurable et N un ensemble μ -négligeable. Alors

$$\int f \mathbf{1}_N \, d\mu = 0.$$

Démonstration. Traitons d'abord le cas où $f \geq 0$. Soit g une fonction étagée positive telle que $g \leq f \mathbf{1}_N$, disons de la forme $g = \sum_{i=1}^p \alpha_i \mathbf{1}_{A_i}$. On a $\int g \, d\mu = \sum_i \alpha_i \mu(A_i)$. Il est facile de voir que chaque terme de la somme est nul. En effet, pour i fixé tel que $\alpha_i \neq 0$, on $g(x) = \alpha_i > 0$ et par construction de $g \leq f \mathbf{1}_N$, on a $\mathbf{1}_N(x) \neq 0$. Donc $x \in N$, ce qui montre que $A_i \subset N$, et donc $\mu(A_i) = 0$. Toute fonction étagée g minorant $f \mathbf{1}_N$ est d'intégrale nulle. Donc $f \mathbf{1}_N$ est d'intégrale nulle. On déduit la propriété dans le cas général en utilisant la décomposition $\int f \mathbf{1}_N \, d\mu = \int f^+ \mathbf{1}_N \, d\mu - \int f^- \mathbf{1}_N \, d\mu = 0$. ■

Les ensembles mesurables de mesure nulle jouent un rôle particulier parce qu'ils ne sont pas « visibles » par la mesure même s'ils sont non vides.

Définition 1.32 — Presque-partout, presque-sûre. On dit d'une propriété qu'elle est vraie μ -presque-partout ou μ -presque-sûrement, en abrégé μ -p.p. ou μ -p.s, lorsque son complémentaire est de mesure nulle.

Donnons un exemple. L'écriture " $0 \leq f_n \uparrow f$ μ -p.p." signifie qu'il existe un ensemble $N \in \mathcal{E}$ tel que $\mu(N) = 0$ et tel que pour tout $x \in E \setminus N$, $f_n(x)$ est une suite positive croissante convergeant vers $f(x)$.

Soit (E, \mathcal{E}, μ) un espace probabilisé. Soit f une fonction mesurable positive de μ -intégrale nulle. Pour tout $n > 0$ et $A_n = \{x : f(x) > 1/n\}$,

$$0 = \int f \, d\mu \geq \int_{A_n} f \, d\mu \geq \frac{1}{n} \mu(A_n),$$

donc $\mu(A_n) = 0$ pour tout $n \geq 1$. Comme $\bigcap_{n \geq 1} A_n = A_0$, par monotonie, on obtient $\mu(A_0) = 0$. On a donc démontré le théorème suivant :

Théorème 1.33 Soit f une fonction mesurable positive, de μ -intégrale nulle alors f est nulle μ -presque partout.

Mais une fonction nulle « presque partout » n'est pas une fonction nulle. Par exemple, la fonction indicatrice de \mathbf{Q} , l'ensemble des rationnels, est presque-partout nulle pour la mesure de Lebesgue mais elle est non nulle sur un ensemble dense !

■ **Remarque 5** Ainsi, l'application $f \mapsto \int |f| \, d\mu$ ne définit pas une norme sur l'espace vectoriel $L^1(\mu)$, car la nullité de l'intégrale n'implique pas la nullité de la fonction. Afin de construire une norme, il faut que deux fonctions qui sont égales μ -p.p. soient identifiées comme un seul et même élément. Formellement, cela revient à définir la relation d'équivalence suivante.

f est en relation d'égalité p.p. avec g lorsque $\mu(f \neq g) = 0$. On note $f \mathcal{R} g$.

On définit alors $L^1(\mu)$ comme l'ensemble des classes d'équivalence de la relation \mathcal{R} , et l'application qui à toute classe d'équivalence c associe la valeur $\int |f| \, d\mu$ pour une fonction f quelconque de la classe c définit bien une norme sur $L^1(\mu)$ puisque tous les éléments d'une même classe d'équivalence ont la même intégrale.

Dans la pratique, travailler avec des classes d'équivalence est fastidieux, on peut continuer de penser les fonctions mesurables comme des fonctions ordinaires en prenant garde qu'elles ne sont définies qu'à un ensemble de mesure nulle près.

Cela implique que parler de la valeur d'une fonction mesurable, définie sur \mathbf{R}^n muni de la mesure de Lebesgue, en un point, sans hypothèse de continuité, n'a aucun sens puisqu'un point est de mesure nulle !

Les théorèmes de convergence monotone, Fatou, convergence dominée sont valables sans changement à condition de spécifier des *presque-partout* chaque fois que nécessaire. Ainsi dans le théorème de convergence dominée, la convergence de f_n vers f peut n'avoir lieu que μ -p.p. et la relation de domination peut n'être satisfaite qu'à un ensemble négligeable près.

On peut compléter nos théorèmes fondamentaux par ceux concernant les intégrales (et séries) à paramètres.

■ **Remarque 6** Les plus attentifs remarqueront que le théorème suivant se distingue de celui que vous connaissez pour l'intégrale de Riemann en ce qu'il n'est pas nécessaire de vérifier que la fonction dominante et la fonction limite sont continues !

Théorème 1.34 — Continuité sous le signe somme. Soit I un ouvert de \mathbf{R}^n et $\{f(x, t), t \in I\}$ une famille de fonctions mesurables telle que pour tout $t \in I$, $f(\cdot, t)$ soit μ -intégrable. Si $\mu(x)$ -p.p., $t \mapsto f(x, t)$ est continue sur I , s'il existe G une fonction mesurable telle que

$$— \mu(x) \text{ p.p.}, |f(x, t)| \leq G(x) \text{ pour tout } t \in I,$$

$$— \int G \, d\mu < \infty,$$

alors l'application $t \mapsto \int f(x, t) \, d\mu(x)$ est continue.

Démonstration. Soit $(t_n, n \geq 1)$ une suite d'éléments de I qui tend vers t . Alors, on a évidemment

$$\left| \int f(x, t) \, d\mu(x) - \int f(x, t_n) \, d\mu(x) \right| \leq \int_E |f(x, t) - f(x, t_n)| \, d\mu(x).$$

Les assertions suivantes sont immédiates :

- $g_n : x \mapsto |f(x, t) - f(x, t_n)|$ est mesurable ;
- $g_n(x) = |f(x, t) - f(x, t_n)| \xrightarrow{n \rightarrow \infty} 0, \mu$ -p.p. ;
- $|g_n(x)| = |f(x, t) - f(x, t_n)| \leq 2G(x), \mu$ -p.p. ;
- $\int G \, d\mu < \infty$;

donc on peut appliquer le théorème de convergence dominée à g_n , ce qui induit que

$$\int_E |f(x, t) - f(x, t_n)| \, d\mu(x) \xrightarrow{n \rightarrow \infty} 0.$$

D'où le résultat. ■

Une preuve similaire permet d'obtenir le résultat de dérivation sous le signe « somme ».

Théorème 1.35 Soit I un ouvert de \mathbf{R}^n et $\{f(x, t), t \in I\}$ une famille de fonctions mesurables telles que pour tout $t \in I$, $f(\cdot, t)$ soit μ -intégrable. Si $t \mapsto f(x, t)$ est dérivable sur I , $d\mu$ p.p., s'il existe $G(x)$ une fonction mesurable telle que

- μ p.p., $|\frac{d}{dt}f(x, t)| \leq G(x)$ pour tout $t \in I$,
- $\int G d\mu < \infty$,

alors l'application

$$\begin{aligned} I &\longrightarrow \mathbf{R} \\ t &\longmapsto \int f(x, t) d\mu(x) \end{aligned}$$

est dérivable sur I et

$$\frac{d}{dt} \int f(x, t) d\mu = \int \frac{d}{dt} f(x, t) d\mu(x).$$

Espaces L^p

Définition 1.36 Pour $p \in [1, \infty[$, l'espace $L^p(\mu)$ est l'espace des fonctions mesurables (définies à un ensemble μ négligeable près) telles que

$$\int |f|^p d\mu < \infty.$$

On en fait un espace normé en posant

$$\|f\|_p = \left(\int |f|^p d\mu \right)^{1/p}.$$

Définition 1.37 Pour une fonction mesurable f , on définit son sup-essentiel par

$$\text{ess-sup } f = \inf\{M, |f(x)| \leq M \mu - \text{p.p.}\}.$$

Une fonction dont le sup-essentiel est fini appartient à $L^\infty(\mu)$ et

$$\|f\|_\infty = \text{ess-sup } f.$$

Si l'on instancie μ en la mesure de comptage sur $E = \mathbf{N}$ ou \mathbf{Z} on parle plus volontiers de suite que de fonctions et on note $l^p(\mathbf{N})$, respectivement $l^p(\mathbf{Z})$, l'ensemble des suites p -sommables : les suites de terme général u_n indexées par \mathbf{N} ou \mathbf{Z} telles que

$$\sum_n |u_n|^p < \infty.$$

On a souvent besoin de majorations d'intégrales : pour $p \in [1, \infty[$,

on définit son conjugué, souvent noté q par

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Le conjugué de 2 est lui-même, celui de 1 est $+\infty$ et celui d'un nombre supérieur à 2 est inférieur à 2. De plus le biconjugué de p est égal à p .

Théorème 1.38 — Inégalité de Jensen. Soit μ une mesure de probabilité. Soit ϕ une fonction convexe réelle et f une fonction μ intégrable,

$$\phi\left(\int f \, d\mu\right) \leq \int \phi(f) \, d\mu.$$

Démonstration. La fonction ϕ est convexe donc son graphe est en tout point au dessus de ses droites de contact. On applique ce résultat aux points $t = \int f \, d\mu$ et $s = f(x)$.

$$\phi(f(x)) \geq \phi(t) + \alpha(f(x) - t)$$

pour un certain α . En intégrant cette relation par rapport à μ on obtient

$$\int_E \phi \circ f \, d\mu \geq \phi(t) \int_E d\mu + \alpha \left(\int_E f \, d\mu - t \int_E f \, d\mu \right).$$

Le fait que μ soit une mesure de probabilités joue un rôle fondamental parce que c'est ce qui permet de simplifier le terme de droite en $\phi(t)$, d'où le résultat. ■

La démonstration de l'inégalité de Hölder n'est pas la plus simple mais elle montre comment elle peut se déduire de l'inégalité de Jense et est donc une conséquence de la convexité. C'est l'approche utilisée dans LIEB et LOSS, *Analysis*.

Théorème 1.39 — Inégalités de Hölder. Soit $p \in]1, \infty[$ et $q = p(p - 1)^{-1}$, soient $f \in L^p(\mu)$ et $g \in L^q(\mu)$, on a

$$\left| \int fg \, d\mu \right| \leq \int |f| |g| \, d\mu \leq \left(\int |f|^p \, d\mu \right)^{1/p} \left(\int |g|^q \, d\mu \right)^{1/q},$$

avec égalité ssi il existe une constante λ telle que $f(x) = \lambda g(x)$, μ -p.p..

Démonstration. On peut toujours supposer f et g positives et appartenant respectivement à L^p et L^q (sinon il n'y a rien à prouver). Soit $A = \{x, g(x) > 0\}$. Comme l'intégrale de fg sur E est égale à l'intégrale

sur A , on peut se contenter de supposer que $E = A$. Posons,

$$d\nu(x) = \left(\int g^q d\mu \right)^{-1} g(x)^q d\mu(x)$$

$$F(x) = f(x)g(x)^{-q/p}$$

$$\phi(t) = |t|^p.$$

L'inégalité de Jensen donne

$$\begin{aligned} \frac{1}{\left(\int g^q d\mu \right)^p} \left(\int f(x)g(x)^{-q/p} g(x)^q d\mu(x) \right)^p \\ \leq \frac{1}{\left(\int g^q d\mu \right)} \int f(x)^p g(x)^{-q} g(x)^q d\mu(x). \end{aligned}$$

D'où le résultat après simplification. ■

Théorème 1.40 — Inégalité de Minkowski vectorielle. Soit E et F deux espaces munis de mesures μ et ν supposées σ -finies. Soit $f : E \times F \rightarrow \mathbf{R}^+$ mesurable. Pour $1 \leq p < \infty$,

$$\begin{aligned} \int_F \left(\int_E f(x,y)^p d\mu(x) \right)^{1/p} d\nu(y) \\ \geq \left(\int_E \left(\int_F f(x,y) d\nu(y) \right)^p d\mu(x) \right)^{1/p}. \quad (1.4) \end{aligned}$$

Avant d'en faire la démonstration, donnons tout de suite le corollaire principal usuellement appelé inégalité de Minkowski.

Corollaire 1.41 Soit f et g deux éléments de L^p ,

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

Démonstration. Appliquons le résultat précédent avec $d\nu(x) = \delta_1(x) + \delta_2(x)$ et

$$f(x,1) = f(x) \text{ et } f(x,2) = g(x).$$

On a d'une part,

$$\int_F \left(\int_E f(x,y)^p d\mu(x) \right)^{1/p} d\nu(y) = \|f\|_p + \|g\|_p,$$

et d'autre part,

$$\left(\int_E \left(\int_F f(x,y) d\nu(y) \right)^p d\mu(x) \right)^{1/p} = \|f + g\|_p.$$

■

Preuve du théorème 1.40. Posons

$$H(x) = \int_F f(x, y) \, d\nu(y).$$

On suppose que le terme de gauche est fini sinon on tronque f et l'espace d'intégration puis on conclut par convergence monotone grâce à la σ -finitude. On part de

$$\begin{aligned} \int_E H(x)^p \, d\mu(x) &= \int_E H(x) H(x)^{p-1} \, d\mu(x) \\ &= \int_F \left(\int_E f(x, y) H(x)^{p-1} \, d\mu(x) \right) \, d\nu(y), \end{aligned}$$

d'après Fubini. On applique l'inégalité de Hölder dans l'intégrale intérieure. D'où

$$\begin{aligned} \int_E H(x)^p \, d\mu(x) &\leq \int_F \left(\int_E f(x, y)^p \, d\mu(x) \right)^{1/p} \left(\int_E H(x)^{q(p-1)} \, d\mu(x) \right)^{1/q} \, d\nu(y). \end{aligned}$$

On remarque que $q(p-1) = p$ donc

$$\begin{aligned} \int_E H(x)^p \, d\mu(x) &\leq \left(\int_E H(x)^p \, d\mu(x) \right)^{1/q} \int_F \left(\int_E f(x, y)^p \, d\mu(x) \right)^{1/p} \, d\nu(y). \end{aligned}$$

On conclut en remarquant que $1 - 1/q = 1/p$. ■

Exercice 1.1 On considère la fonction

$$\begin{aligned} f : [0, 1] \times [0, 1] &\longrightarrow \mathbf{R} \\ (x, y) &\longmapsto \frac{x^2 - y^2}{(x^2 + y^2)^2}. \end{aligned}$$

1. Montrer que si le théorème de Fubini est applicable alors nécessairement cette intégrale est nulle.
2. En faisant les calculs explicites des intégrales partielles, montrer que l'on n'obtient pas le même résultat.
3. Montrer enfin, en utilisant Fubini-Tonelli, que la fonction étudiée n'est pas de module intégrable. ■

Exercice 1.2 Cet exercice peut se retrouver dans l'exemple 18.4 de BILLINGSLEY, *Probability and Measure*.

Soit la fonction

$$f : [0, a] \times \mathbf{R}^+ \longrightarrow \mathbf{R} \\ (x, y) \longmapsto e^{-xy} \sin x.$$

1. Montrer que f est intégrable.
2. En utilisant le théorème de Fubini, montrer que

$$\int_0^a \frac{\sin x}{x} dx = \frac{\pi}{2} - \cos a \int_0^\infty \frac{e^{-ay}}{1+y^2} dy - \sin a \int_0^\infty \frac{ye^{-ay}}{1+y^2} dy.$$

3. En déduire que

$$\left| \int_0^a \frac{\sin x}{x} dx - \frac{\pi}{2} \right| \leq \frac{2}{a}.$$

■

Exercice 1.3 — Lemme de Lebesgue.

1. Montrer que

$$\int_a^b \sin(tx) dx \xrightarrow{t \rightarrow \infty} 0.$$

Soit f une fonction de $L^1(\mathbf{R})$. On admet que l'ensemble des fonctions continues à support compact est dense dans L^1 .

2. Déduire de la question précédente que

$$\int_{\mathbf{R}} f(x) \sin(tx) dx \xrightarrow{t \rightarrow \infty} 0.$$

■

Exercice 1.4 On rappelle que

$$\sum_{k=1}^N \frac{1}{k} - \ln N \xrightarrow{N \rightarrow \infty} \gamma$$

où γ est la constante d'Euler qui vaut approximativement

$$\gamma = 0,5772156649015328606 \dots$$

On considère la fonction f définie par

$$f :]0, 1[\longrightarrow \mathbf{R} \\ x \longmapsto \frac{1}{x} - \left[\frac{1}{x} \right].$$

1. Montrer que f mesurable et qu'elle est intégrable sur $]0, 1[$.
2. Montrer que

$$\int_0^1 f(x) \, dx = \sum_{k=1}^{\infty} \ln \left(\frac{k+1}{k} \right) - \frac{1}{k+1}.$$

3. En déduire que

$$\int_0^1 f(x) \, dx = 1 - \gamma.$$

■

Exercice 1.5 — Intégrale fractionnaire. Cet exercice d'un énoncé de l'épreuve d'Analyse et Probabilités de l'agrégation 2014. La référence pour tout le calcul fractionnaire est le livre SAMKO, KILBAS et MARICHEV, *Fractional Integrals and Derivatives*.

Pour $a > 0$ et $f \in L^1(\mathbf{R}^+)$, on pose

$$I^a f(t) = \frac{1}{\Gamma(a)} \int_0^t (t-s)^{a-1} f(s) \, ds$$

1. Montrer que pour tout $T > 0$,

$$\int_0^T \int_0^t (t-s)^{a-1} |f(s)| \, ds \, dt < \infty.$$

2. En déduire que $I^a f$ est une fonction localement intégrable sur \mathbf{R}^+ .
3. Montrer que pour tout $a, b > 0$,

$$I^a \circ I^b = I^{a+b}.$$

Indication : on pourra admettre la relation

$$\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 t^{a-1}(1-t)^{b-1} \, dt.$$

4. Montrer que I^α est continue de \mathcal{C} dans \mathcal{C}^1 .
5. Montrer qu'entre ces deux espaces, I^α est injectif.

La question suivante ne faisait pas partie de l'énoncé, on peut donc supposer que les théorèmes de classe monotone sont considérés comme hors programme.

6. En utilisant le théorème de classe monotone, montrer que I^α est injectif de L^1 dans \mathcal{C} .

■

Exercice 1.6 La partie sur le mélange est dans les pages 32 et suivantes de de COUDÈNE, *Théorie ergodique et systèmes dynamiques*. La partie spécifique à la suite logistique se retrouve page 77 et suivantes.

La suite récurrente $x = (x_n, n \geq 0)$ définie par

$$x_0 \in]0, 1[, x_{n+1} = 4x_n(1 - x_n),$$

appartient à la classe des *systèmes dynamiques* dits chaotiques. En effet, si l'on part d'une condition initiale quelconque, l'orbite (la suite des valeurs prises par la suite pour une condition initiale donnée) présente de façon évidente un caractère hautement erratique.

Qui plus est, pour des conditions initiales très proches, les orbites sont proches au début mais finissent par se séparer franchement au bout de quelques itérations.

Ce comportement est en partie dû au fait qu'à l'intersection de la première bissectrice et de la courbe représentative de $f(x) = 4x(1 - x)$, la dérivée est en module supérieure à 1 donc le potentiel point fixe $x = 3/4$ est de type répulsif : $f'(3/4) = -2$.

On ne peut donc rien dire de déterministe sur le comportement asymptotique de cette suite. En revanche, on peut essayer de voir ce que sont les « zones » de $[0, 1]$ qu'elle visite le plus fréquemment, c'est-à-dire de regarder

$$\hat{\mu}([a, b]) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mathbf{1}_{[a, b]}(x_n)$$

pour tous les intervalles $[a, b]$. Cette quantité correspond à la fréquence empirique des passages dans l'intervalle $[a, b]$. Si l'on fait l'analogie avec les temps de séjour dans un état d'une chaîne de Markov, $\hat{\mu}$ doit se comporter comme une mesure sur $[0, 1]$. L'objectif de cet exercice est de montrer que c'est bien le cas et d'identifier au passage la dite mesure. Comme les techniques mises en place relèvent de ce que l'on appelle la *théorie ergodique*, les premières questions sont génériques et s'appliquent à de nombreux autres systèmes dynamiques.

Soit $(E, \mathcal{E}, \mathbf{P})$ un espace probabilisé et T une application mesurable de E dans lui-même. On suppose que \mathbf{P} est invariante par T , c'est-à-dire que

$$\mathbf{P}(T^{-1}(A)) = \mathbf{P}(A) \text{ pour tout } A \in \mathcal{E}.$$

- a) Montrer que l'ensemble des mesurables invariants par T , c'est-à-dire qui vérifie $T^{-1}(A) = A$, est une tribu (notée \mathcal{I} par la suite).
- b) Soit f une fonction mesurable de E dans \mathbf{R} . Montrer que si f est invariante par T (c'est-à-dire $f \circ T = f$) alors f est mesurable de (E, \mathcal{I}) dans $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$.

Le système dynamique (E, T, \mathbf{P}) est dit ergodique lorsque

$$\mathcal{I} \subset \sigma\{A \subset E, \mathbf{P}(A) = 0 \text{ ou } \mathbf{P}(A) = 1\}.$$

3. Montrer que (E, T, \mathbf{P}) est ergodique si et seulement si les fonctions invariantes par T sont constantes presque partout.

On dit que T est mélangeante si et seulement si pour tout couple f, g d'éléments de $L^2(\mathbf{P})$,

$$\lim_{n \rightarrow +\infty} \int_E f \circ T^n g \, d\mathbf{P} = \int_E f \, d\mathbf{P} \int_E g \, d\mathbf{P}. \quad (1.5)$$

4. Montrer que si T est mélangeante alors (E, T) est ergodique.
5. Montrer pour $f \in L^2(\mathbf{P})$, pour tout entier $n \geq 1$,

$$\int_E f \circ T^n \, d\mathbf{P} = \int_E f \, d\mathbf{P}. \quad (1.6)$$

6. Montrer que si la condition de mélange (1.5) est vérifiée pour f, g appartenant à un sous-ensemble dense de $L^2(\mathbf{P})$ alors T est mélangeante.

On veut maintenant étudier le système dynamique donnée par l'équation d'évolution :

$$x_{n+1}^a = T(x_n^a) \text{ où } T(x) = 4x(1-x), \quad x_0^a = a \in [0, 1].$$

On veut montrer en particulier que pour presque tout $a \in [0, 1]$,

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^n f(x_j^a) = \int_0^1 f(u) (\pi \sqrt{u} \sqrt{1-u})^{-1} \, du.$$

On admet le *théorème de Birkhoff* qui stipule que si (E, T, \mathbf{P}) est un système ergodique alors

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^n f \circ T^j(x) = \int_E f \, d\mathbf{P} \quad (1.7)$$

pour presque tout x . Il nous faut donc trouver une mesure invariante μ par T et montrer que le système dynamique $([0, 1], T, \mu)$

est ergodique. Pour ce faire on considère un autre système dynamique :

$$E_1 = [0, 1[, \quad T_1 x = 2x \text{ si } 0 \leq x \leq 1/2, \\ T_1(x) = 2 - 2x \text{ pour } 1/2 \leq x < 1.$$

(où $[x]$ est la partie entière de x) muni de la mesure de Lebesgue sur $[0, 1[$, notée λ .

7. Montrer que λ est invariante par T_1 .
8. En admettant (ou se souvenant, cf. séries de Fourier) que la famille de fonctions $e_k(x) = \exp(2i\pi kx)$ pour $k \in \mathbf{Z}$ est une famille dense de $L^2(d\lambda)$, montrer que T_1 est mélangeante.

Soit Θ l'application de E_1 dans $[0, 1]$ définie par :

$$\Theta(x) = \sin^2(\pi x/2).$$

9. Identifier μ la mesure image de λ par Θ .
10. Montrer que $([0, 1[, T, \mu)$ est ergodique et conclure.

■

Exercice 1.7 — Inégalité de Hölder généralisée. Soit p, q, r tels que

$$\frac{1}{r} = \frac{1}{p} + \frac{1}{q}.$$

1. Montrer que

$$\|fg\|_r \leq \|f\|_p \|g\|_q.$$

■

Exercice 1.8 Soit $f \in L^\infty \cap L^1$.

1. Etablir que

$$(\|f\|_\infty - \varepsilon)^p \mu(|f| > \|f\|_\infty - \varepsilon) \leq \int |f|^p d\mu \leq \|f\|_\infty^{p-1} \|f\|_1.$$

2. Montrer que

$$\lim_{p \rightarrow \infty} \|f\|_p = \|f\|_\infty.$$

■

Exercice 1.9 — Inégalité de Hardy. Soit $p > 1$ et $f \in L^p(\mathbf{R}^+, \ell)$. On

pose

$$F(x) = \frac{1}{x} \int_0^x f(t) dt$$

On veut établir l'identité

$$\|F\|_p \leq \frac{p}{p-1} \|f\|_p. \quad (1.8)$$

1. Montrer que

$$\frac{1}{x} \int_0^x f(t) dt = \int_0^1 f(sx) ds.$$

2. En utilisant l'inégalité intégrale de Minkowski, établir le résultat.

3. Montrer que la constante $p/(p-1)$ est optimale.

Considérer la fonction $f(x) = x^{-1/p} \mathbf{1}_{[1/A, A]}(x)$.

■

Exercice 1.10 — Inégalité de Clarkson. On peut retrouver cette démonstration dans le théorème IV.10 de BRÉZIS, *Analyse Fonctionnelle*.

On suppose $p \geq 2$.

1. Montrer que pour $x, y \geq 0$,

$$x^p + y^p \leq (x^2 + y^2)^{p/2}.$$

Indication : on pourra considérer la fonction $f(t) = (t^2 + 1)^{p/2} - t^p - 1$.

2. En déduire que

$$\left| \frac{a+b}{2} \right|^p + \left| \frac{a-b}{2} \right|^p \leq \frac{1}{2} |a|^p + \frac{1}{2} |b|^p.$$

3. En déduire que

$$\left\| \frac{1}{2}(f+g) \right\|_p^p + \left\| \frac{1}{2}(f-g) \right\|_p^p \leq \frac{1}{2} \|f\|_p^p + \frac{1}{2} \|g\|_p^p. \quad (1.9)$$

Soit $(f_n, n \geq 1)$ et $(g_n, n \geq 1)$ deux suites d'éléments de L^p telles que

— $\forall n \geq 1, \|f_n\|_p \leq 1$ et $\|g_n\|_p \leq 1$,

— $\|f_n + g_n\|_p \xrightarrow{n \rightarrow \infty} 2$.

4. Montrer que $\|f_n - g_n\|_p \xrightarrow{n \rightarrow \infty} 0$.

■

2

Variables aléatoires

Bourbaki s'est écarté des probabilités, les a rejetées, les a considérées comme non rigoureuses et, par son influence considérable, a dirigé la jeunesse hors du sentier des probabilités.

Laurent Schwartz.

Presque tout probabiliste a commis un jour ou l'autre son livre de cours de probabilités. Ils ont tous un parti pris différent selon que le public visé connaît la théorie de la mesure ou pas, n'est intéressé que par les probabilités dites élémentaires ou vise une plus grande maîtrise, etc.

On trouvera une foule d'exercices dans GRIMMETT et STIRZAKER, *Probability and Random Processes*. Pour une approche plus abstraite, on pourra regarder JACOD et PROTTER, *Probability Essentials*.

Il y en a d'autres mais ils sont manifestement épuisés.

§ 1 De la théorie de la mesure aux probabilités

Stricto sensu, pour faire des probabilités, il suffit d'une mesure de probabilités c'est-à-dire d'une mesure de masse totale égale à 1. Il y a juste un changement de notations à opérer par rapport aux notations de la théorie de la mesure :

- L'espace noté E précédemment devient Ω et est appelé *l'espace des événements ou l'univers des possibles*,
- Les éléments de la tribu \mathcal{A} sont appelés des *événements* plutôt que des mesurables,
- La mesure, nécessairement finie, est notée \mathbf{P} ou \mathbf{Q} au lieu de μ et ν ,
- On dire d'une propriété qu'elle est vraie *presque partout* si elle est vraie en dehors d'un ensemble \mathbf{P} -négligeable,

- La différence la plus sensible apparaît au niveau de l'intégrale. Les fonctions mesurables deviennent des variables aléatoires, notées **exclusivement** par des lettres majuscules X, Y , etc. Les intégrales sont appelées *espérances* :

$$\mathbf{E}[X] := \int_{\Omega} X(\omega) \, d\mathbf{P}(\omega).$$

Il y a cependant deux nouveaux concepts propres aux probabilités : la notion de loi de variable aléatoire et l'indépendance.

Comme on l'a vu, une variable aléatoire X de $(\Omega, \mathcal{A}, \mathbf{P})$ à valeurs dans $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$ n'est rien d'autre qu'une fonction mesurable. La seule information que l'on ait sur une variable aléatoire est la probabilité que ses valeurs soit dans un sous-ensemble donné de \mathbf{R}^d : on ne connaît (ou l'on ne cherche) que

$$\mathbf{P}(X \in A) \text{ pour tout } A \in \mathcal{B}(\mathbf{R}^d).$$

Ce qui nous amène à la définition de la loi de X .

Définition 2.1 La loi d'une variable aléatoire (ou la mesure image de \mathbf{P} par X) est la mesure de probabilités \mathbf{P}_X sur $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$ définie par

$$\begin{aligned} \forall A \in \mathcal{B}(\mathbf{R}^d), \mathbf{P}_X(A) &= \mathbf{P}(X \in A) \\ &= \mathbf{P}(\omega, X(\omega) \in A) \\ &= \mathbf{P}(X^{-1}(A)). \end{aligned}$$

■ **Remarque 7** Le fait que \mathbf{P}_X soit une mesure de probabilités est une conséquence directe des propriétés des images réciproques :

$$X^{-1}\left(\bigcup_{j=1}^{\infty} A_j\right) = \bigcup_{j=1}^{\infty} X^{-1}(A_j).$$

Le résultat le plus utile est alors le théorème de transfert qui permet de calculer une intégrale sur Ω comme une intégrale sur \mathbf{R}^d .

Théorème 2.2 — Théorème de transfert. Soit $f : \mathbf{R}^d \rightarrow \mathbf{R}$. Si f est à valeurs positives alors

$$\mathbf{E}[f(X)] := \int_{\Omega} f(X) \, d\mathbf{P} = \int_{\mathbf{R}^d} f(x) \, d\mathbf{P}_X(x) \in \mathbf{R}^+ \cup \{+\infty\}.$$

Si f prend ses valeurs sur tout \mathbf{R} , $f \in L^1(\mathbf{R}^d, \mathbf{P}_X)$ si

$$\mathbf{E}[|f(X)|] < \infty$$

et alors

$$\mathbf{E}[f(X)] = \int_{\mathbf{R}^d} f(x) \, d\mathbf{P}_X(x). \quad (2.1)$$

Démonstration. Soit $A \in \mathcal{B}(\mathbf{R}^d)$, l'application $\mathbf{1}_A \circ X$ est la composée de deux fonctions mesurables donc mesurable. En vérifiant sur tous les cas possibles, on voit que

$$\mathbf{1}_A \circ X = \mathbf{1}_A(X) = \mathbf{1}_{X^{-1}(A)}.$$

Par construction de l'intégrale

$$\mathbf{E}[\mathbf{1}_A \circ X] := \int_{\Omega} \mathbf{1}_A \circ X \, dP = \mathbf{P}(X^{-1}(A)).$$

Par définition de la mesure image

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

Par définition de l'intégrale par rapport à la mesure \mathbf{P}_X

$$\mathbf{P}_X(A) = \int_{\mathbf{R}^d} \mathbf{1}_A \, d\mathbf{P}_X.$$

Par conséquent,

$$\int_{\Omega} \mathbf{1}_A(X) \, dP := \mathbf{E}[\mathbf{1}_A(X)] = \int_{\mathbf{R}^d} \mathbf{1}_A \, d\mathbf{P}_X.$$

Par linéarité, (2.1) est vraie pour les fonctions étagées puis par limite monotone pour toutes les fonctions mesurables positives. Enfin, en décomposant f en $f^+ - f^-$, on obtient le résultat pour tout f mesurable. ■

■ **Remarque 8** On retiendra de la preuve, l'identité

$$\mathbf{E}[\mathbf{1}_A(X)] = \int_A \, d\mathbf{P}_X = \mathbf{P}_X(A). \quad (2.2)$$

■ **Remarque 9** Les calculs d'espérance ne nécessite donc en fait que la connaissance de \mathbf{P}_X et pas celle de \mathbf{P} . On est donc ramené à travailler avec des mesures sur des espaces bien connus que sont \mathbf{N} , \mathbf{R} ou \mathbf{R}^d . Dans un premier temps, on se moque de la construction du fameux univers des possibles parce qu'il n'apparaît jamais dans les calculs. On a juste besoin de savoir qu'il existe, ce qui peut ne pas être évident mais seulement dans des cas hors du programme de l'agrégation.

■ **Exemple 2.1 — Variables aléatoires discrètes.** Si X ne prend qu'un nombre au plus dénombrable de valeurs, \mathbf{P}_X ne charge qu'un nombre (au plus) dénombrable de points donc caractériser \mathbf{P}_X revient à connaître

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

Par ailleurs, pour $A \subset \mathbf{R}^d$,

$$X \in A \iff X \in A \cap X(\Omega) = \bigcup \{x, x \in X(\Omega) \text{ et } x \in A\}.$$

Par conséquent,

$$\mathbf{P}_X(A) = \mathbf{P}(X \in A) = \sum_{x \in X(\Omega) \cap A} \mathbf{P}(X = x) \quad (2.3)$$

et

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

i.e. on fait la somme sur toutes les valeurs possibles de X de f évaluée en cette valeur multiplié par la probabilités que X vaille cette valeur.

On peut résumer formellement les identités (2.3) et (2.4) par

$$\mathbf{P}_X = \sum_{x \in X(\Omega)} \mathbf{P}(X = x) \delta_x,$$

où δ_x est la masse de Dirac en x . ■

Soit X une variable aléatoire à valeurs dans \mathbf{R} , sa loi \mathbf{P}_X est une mesure sur \mathbf{R} . On sait qu'elle est totalement caractérisée par les valeurs de $\mathbf{P}_X(]-\infty, b])$ pour b parcourant \mathbf{R} .

Définition 2.3 Soit X une variable aléatoire réelle, la fonction

$$\begin{aligned} F_X : \mathbf{R} &\longrightarrow [0, 1] \\ x &\longmapsto \mathbf{P}_X(]-\infty, x]) = \mathbf{P}(X \leq x) \end{aligned}$$

s'appelle la fonction de répartition de X .

Lemme 2.4 En vertu des propriétés de monotonie des mesures, F_X possède les propriétés suivantes :

- i) $\lim_{x \rightarrow -\infty} F_X(x) = 0$,
- ii) $\lim_{x \rightarrow +\infty} F_X(x) = 1$,
- iii) F_X est croissante, continue à droite, i.e., $\lim_{y \downarrow x} F_X(y) = F_X(x)$.

On a, d'après les propriétés de monotonie des mesures,

$$F_X(x_-) = \lim_{n \rightarrow +\infty} \mathbf{P}\left(\bigcup_{n=1}^{\infty}]-\infty, x - \frac{1}{n}]\right) = \mathbf{P}_X(]-\infty, x[).$$

Par conséquent, $F_X(x_-) = \mathbf{P}_X(]-\infty, x[)$ et donc

$$F_X(x) - F_X(x_-) = \mathbf{P}(X = x).$$

En d'autres termes, si F_X est continue en x , $\mathbf{P}(X = x) = 0$.

Lemme 2.5 Le nombre de points de discontinuité d'une fonction de répartition est au plus dénombrable.

Soit $\{x_n, n \in \mathbf{N}^*\}$ les points de discontinuité de F_X . On peut alors parler de F_X^c , la régularisée de F_X :

$$\begin{aligned} F_X^c(x) &= F_X(x) - \sum_{n=1}^{\infty} (F_X(x_n) - F_X(x_{n-})) \mathbf{1}_{[x_n, +\infty[}(x). \\ &= F_X(x) - \sum_{n=1}^{\infty} \Delta F_X(x) \mathbf{1}_{[x_n, +\infty[}(x). \end{aligned}$$

La fonction F_X^c est continue et croissante par définition. Elle est d'après un théorème de Lebesgue, dérivable sauf sur un ensemble de mesure de Lebesgue nulle. Dans la suite, nous ne nous préoccupons pas de savoir ce qui se passe si elle n'est pas dérivable en tout point.

Théorème 2.6 Soit X une variable aléatoire réelle de fonction de répartition F_X . Si F_X^c est dérivable sur \mathbf{R} , alors

$$d\mathbf{P}_X(x) = \sum_{n=1}^{\infty} \Delta F_X(x_n) d\delta_{x_n} + (F_X^c)'(x) dx. \quad (2.5)$$

Si F_X est continue alors

$$d\mathbf{P}_X(x) = (F_X^c)'(x) dx$$

et $(F_X^c)'$ s'appelle la densité de la loi de X .

Démonstration. Remarquons que

$$\mathbf{1}_{[x_n, +\infty[}(x) = \delta_{x_n}(\cdot - \infty, x],$$

puisque le terme de gauche ne vaut 1 que si $x \geq x_n$, soit de manière équivalente $x_n \in \cdot - \infty, x]$. Partant de l'écriture,

$$F_X(x) = F_X^c(x) + \sum_{n=1}^{\infty} \Delta F_X(x) \mathbf{1}_{[x_n, +\infty[}(x),$$

si F_X^c est dérivable en tout point alors on a

$$\mathbf{P}(X \in \cdot - \infty, x] = \int_{-\infty}^x (F_X^c)'(s) ds + \sum_{n=1}^{\infty} \Delta F_X(x_n) \delta_{x_n}(\cdot - \infty, x].$$

Les deux mesures de part et d'autre de l'égalité (2.5) coïncident donc sur les ensembles de la forme $\cdot - \infty, x]$ pour tout x réel. C'est suffisant pour assurer que ces deux mesures sont égales. ■

L'autre élément est la notion d'indépendance, qui est intimement liée à celle de mesure produit.

Si la variable aléatoire $X = (X_1, X_2)$ est à valeurs dans \mathbf{R}^2 , on a trois mesures qui apparaissent :

- La loi de X , notée \mathbf{P}_X qui est une mesure de probabilités sur \mathbf{R}^2 ,
- Les lois dites *marginales* de X_1 et X_2 , notées respectivement \mathbf{P}_{X_1} et \mathbf{P}_{X_2} , qui sont des mesures sur \mathbf{R} .

On peut calculer les lois marginales de X_1 et X_2 à partir de celle de X puisque pour $A \in \mathcal{B}(\mathbf{R})$,

$$\mathbf{P}(X_1 \in A) = \mathbf{P}(X_1 \in A, X_2 \in \mathbf{R}) = \mathbf{P}_X(A \times \mathbf{R}).$$

En revanche, on ne peut pas forcément calculer la loi de X à partir des lois marginales. Le seul où l'on peut le faire est celui où les variables sont indépendantes :

Définition 2.7 Deux variables aléatoires X et Y à valeurs respectivement dans \mathbf{R}^d et \mathbf{R}^p sont indépendantes si et seulement si l'une (et donc toutes) des propositions équivalentes suivantes est vérifiée

- i) La mesure du couple (X, Y) est la mesure produit de \mathbf{P}_X et \mathbf{P}_Y :

$$\mathbf{P}_{(X,Y)} = \mathbf{P}_X \otimes \mathbf{P}_Y.$$

- ii) Pour tout $A \in \mathcal{B}(\mathbf{R}^d)$ et tout $B \in \mathcal{B}(\mathbf{R}^p)$,

$$\mathbf{P}(X \in A, Y \in B) = \mathbf{P}(X \in A)\mathbf{P}(Y \in B).$$

- iii) Pour toutes fonctions $f \in L^1(\mathbf{P}_X)$, $g \in L^1(\mathbf{P}_Y)$,

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

⚡ Il n'y a qu'une seule façon d'être indépendant, il y en a plein d'être dépendant.

Le cas plus dégénéré est quand $X = Y$ alors $\mathbf{P}_{(X,X)}$ est concentré sur la diagonale de $\mathbf{R}^d \times \mathbf{R}^d$ tandis que si les variables sont indépendantes, le support de $\mathbf{P}_{(X,Y)}$ est le produit cartésien des supports.

On peut définir la notion d'indépendance pour plus que deux variables aléatoires.

Définition 2.8 Les variables aléatoires (X_1, \dots, X_n) sont indépendantes *dans leur ensemble* ssi l'une des propriétés équivalentes suivantes est vérifiée :

- i) La mesure du n-uple est la mesure produit

$$\mathbf{P}_{(X_1, \dots, X_n)} = \mathbf{P}_{X_1} \otimes \dots \otimes \mathbf{P}_{X_n}.$$

- ii) Pour toutes fonctions $f_i \in L^1(\mathbf{P}_{X_i})$,

$$\mathbf{E}\left[\prod_{j=1}^n f_j(X_j)\right] = \prod_{j=1}^n \mathbf{E}[f_j(X_j)].$$

iii) Pour tout $A_i \in \mathcal{B}(\mathbf{R}^{d_i})$,

$$\mathbf{P}(X \in A_1 \times \dots \times A_n) = \prod_{j=1}^n \mathbf{P}(X_j \in A_j).$$

Une famille infinie de variables aléatoires $(X_i, i \geq 1)$ sont indépendantes dans leur ensemble lorsque toute sous-famille finie est composée de variables aléatoires indépendantes.

Exercice 2.1 Soit X de loi uniforme sur $\{0, 1\}$ et Z de loi uniforme sur $\{-1, +1\}$ indépendante de X . Soit $Y = ZX$. Montrer que X et Y sont décorrélées mais ne sont pas indépendantes. ■

§ 2 Calculs de lois

Définition 2.9 Un vecteur aléatoire (ou variable aléatoire vectorielle) est une application mesurable de $(\Omega, \mathcal{A}, \mathbf{P})$ dans \mathbf{R}^n , c'est-à-dire telle que

$$(X \in]-\infty, x_1] \times \dots \times]-\infty, x_n]) \in \mathcal{A},$$

pour tout n -uplet (x_1, \dots, x_n) .

Sa loi est la mesure image de \mathbf{P} par X et sa fonction de répartition est donnée par

$$\begin{aligned} F_X(x_1, \dots, x_n) &= \mathbf{P}(X \in]-\infty, x_1] \times \dots \times]-\infty, x_n]) \\ &= \mathbf{P}(X_1 \leq x_1, \dots, X_n \leq x_n) \end{aligned}$$

où l'on a noté X_1, \dots, X_n les composantes de X , qui sont bien évidemment des variables aléatoires réelles. La loi d'un vecteur à n composantes est une mesure sur \mathbf{R}^n . Cette loi est dite à densité lorsqu'il existe $f_X : \mathbf{R}^n \rightarrow \mathbf{R}^+$ telle que pour toute h continue borné de \mathbf{R}^n dans \mathbf{R} ,

$$\mathbf{E}[h(X_1, \dots, X_n)] = \int_{\mathbf{R}^n} h(x_1, \dots, x_n) f_X(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Définition 2.10 Les lois des X_i pour chaque $i \in \{1, \dots, n\}$ sont appelées les lois marginales.

■ **Remarque 10** Si l'on connaît la loi d'un vecteur aléatoire X à valeurs dans \mathbf{R}^n , on peut calculer toutes les lois marginales, car

$$\mathbf{P}(X_i \in]-\infty, b]) = \mathbf{P}(X_1 \leq +\infty, \dots, X_i \leq b, \dots, X_n \leq +\infty).$$

Réciproquement, on ne peut pas, sans hypothèse supplémentaire, déterminer la loi d'un vecteur à partir de la seule connaissance des

marginales. Le **seul** cas où c'est possible est lorsque les composantes de X sont supposées être indépendantes. Dans ce cas, par définition de l'indépendance

$$\mathbf{P}(X \in]-\infty, x_1] \times \dots \times]-\infty, x_n]) = \mathbf{P}(X_1 \leq x_1) \dots \mathbf{P}(X_n \leq x_n).$$

La loi de X est alors bien entièrement caractérisée par les lois \mathbf{P}_{X_i} .

L'un des types de calcul qui revient régulièrement dans la pratique des probabilités est celui du calcul de la loi de la transformation d'un vecteur aléatoire de loi connue. L'outil principal pour ces calculs est la formule de changement de variables dans les intégrales multiples.

Définition 2.11 Soit $T : O \subset \mathbf{R}^n \rightarrow \mathbf{R}^n$ dont toutes les dérivées partielles existent sur O , la jacobienne de T au point x , est la matrice $J_T(x)$ où

$$J_T(x) = \left(\frac{\partial T_i}{\partial x_j}(x), 1 \leq i \leq n, 1 \leq j \leq n \right)$$

$$= \begin{pmatrix} \frac{\partial T_1}{\partial x_1}(x) & & & \\ & \vdots & & \\ \dots & \frac{\partial T_i}{\partial x_j}(x) & \dots & \\ & \vdots & & \\ & & & \frac{\partial T_n}{\partial x_n}(x) \end{pmatrix}.$$

Le jacobien de T est le déterminant de J_T .

Définition 2.12 Soit O un ouvert de \mathbf{R}^n , $T : O \subset \mathbf{R}^n \rightarrow \mathbf{R}^n$, T est un \mathcal{C}^1 -difféomorphisme de O sur $\Delta \subset \mathbf{R}^n$, lorsque

- les dérivées partielles de T existent et sont continues sur O ,
- T est une bijection de O sur Δ ,
- le jacobien de T ne s'annule pas sur O .

Théorème 2.13 Soit O un ouvert de \mathbf{R}^n , $T : O \subset \mathbf{R}^n \rightarrow \mathbf{R}^n$ un \mathcal{C}^1 -difféomorphisme de T sur Δ . Pour tout fonction continue bornée,

$$\int_O f(T(x)) \, dx = \int_\Delta f(y) \frac{1}{|\det J_T(T^{-1}(y))|} \, dy.$$

■ **Exemple 2.2** Soit (X_1, X_2) deux variables aléatoires réelles indépendantes, de même loi

$$d\mathbf{P}(x) = \mathbf{1}_{]1, \infty[}(x) \frac{1}{x^2} \, dx.$$

On pose $U = X_1.X_2$ et $V = X_1/X_2$.

1. Calculer la loi du vecteur (U, V) .
2. Calculer la loi de U et celle de V .
3. U et V sont-elles indépendantes?

On part de l'hypothèse que la loi du couple (U, V) a une densité par rapport à la mesure de Lebesgue, ce qui en utilisant la caractérisation des mesures induites par le théorème de Riesz ??, revient à trouver $h : \mathbf{R}^2 \rightarrow \mathbf{R}^+$ telle que pour toute fonction f continue bornée de \mathbf{R}^2 dans \mathbf{R} , on ait

$$\mathbf{E}[f(U, V)] = \int_{\mathbf{R}^2} f(x, y)h(x, y) \, dx \, dy.$$

Posons

$$\begin{aligned} T : \mathbf{R}^2 &\rightarrow \mathbf{R}^2 \\ (x, y) &\mapsto (xy, x/y). \end{aligned}$$

On a

$$\mathbf{E}[f(U, V)] = \mathbf{E}[(f \circ T)(X, Y)] = \int (f \circ T)(x, y) \, d\mathbf{P}_{X, Y}(x, y),$$

où la deuxième égalité découle du théorème de transfert. Maintenant, les variables aléatoires X et Y sont indépendantes, ce qui équivaut (cf. (??)) à dire que

$$d\mathbf{P}_{X, Y}(x, y) = d\mathbf{P}_X(x) \otimes d\mathbf{P}_Y(y).$$

Par hypothèse,

$$d\mathbf{P}_X(x) = \mathbf{1}_{[1, \infty[}(x) \frac{1}{x^2} \, dx \text{ et } d\mathbf{P}_Y(y) = \mathbf{1}_{[1, \infty[}(y) \frac{1}{y^2} \, dy,$$

donc

$$d\mathbf{P}_{X, Y}(x, y) = \mathbf{1}_{[1, \infty[}(x) \frac{1}{x^2} \mathbf{1}_{[1, \infty[}(y) \, dx \, dy.$$

On a donc obtenu

$$\mathbf{E}[f(U, V)] = \int_{[1, +\infty[^2} (f \circ T)(x, y) \frac{1}{x^2 y^2} \, dx \, dy.$$

Rappelons-nous que nous voulons aboutir à une identité de la forme

$$\mathbf{E}[f(U, V)] = \int_{\mathbf{R}^2} f(x, y)h(x, y) \, dx \, dy.$$

On est naturellement enclin à utiliser le théorème de changement de variables 2.13, pour cela, il nous faut calculer Δ , l'ensemble image de $[1, +\infty[^2$ par T et le jacobien de T . Posons $u = xy$ et $v = x/y$,

$$\det J_T(x, y) = \det \begin{pmatrix} y & x \\ \frac{1}{y} & -\frac{x}{y^2} \end{pmatrix} = -2 \frac{x}{y} = -2v.$$

Si x et y sont tous deux plus grands que 1 alors u l'est, et v est strictement positif. Par ailleurs,

$$\begin{cases} u = xy \\ v = x/y \end{cases} \iff \begin{cases} x^2 = uv \\ y^2 = u/v \end{cases}.$$

On déduit de ces dernières équations que $u \geq v$ et $uv \geq 1$. On vérifie alors facilement que T est une bijection de $[1, +\infty]^2$ sur

$$\Delta = \{(u, v), u \geq v \geq 0 \text{ et } uv \geq 1\}.$$

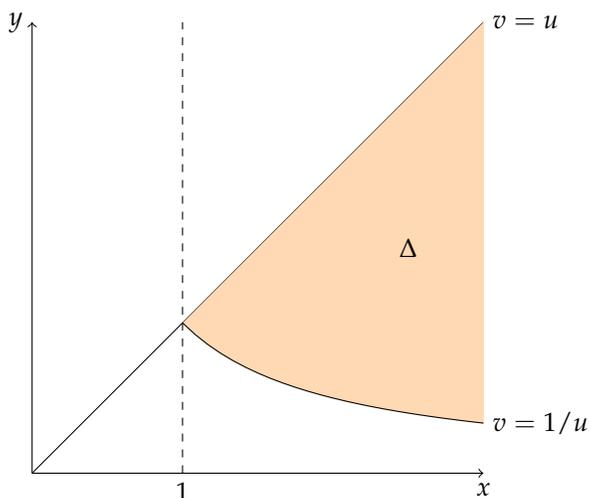


FIGURE 2.1: Le domaine Δ dans le plan (u, v) .

On tire du théorème de changement de variables que

$$\begin{aligned} \int_{[1, +\infty]^2} (f \circ T)(x, y) \frac{1}{x^2 y^2} dx dy &= \int_{\Delta} f(u, v) \frac{1}{uv \cdot u/v} \left| \frac{1}{-2v} \right| du dv \\ &= \int_{\Delta} f(u, v) \frac{1}{2u^2 v} du dv, \end{aligned}$$

d'où par identification,

$$d\mathbf{P}_{(U, V)}(u, v) = \frac{1}{2u^2 v} \mathbf{1}_{\Delta}(u, v) du dv.$$

Pour calculer la loi de U , on veut exprimer $\mathbf{E}[f(U)]$ pour toute fonction continue bornée de \mathbf{R} dans \mathbf{R} . On remarque alors que l'application $\tilde{f}(x, y) = f(x)$ est continue bornée de \mathbf{R}^2 dans \mathbf{R} donc

$$\begin{aligned} \mathbf{E}[f(U)] &= \mathbf{E}[\tilde{f}(U, V)] = \int_{\Delta} f(u) \frac{1}{2u^2 v} du dv \\ &= \int_1^{+\infty} f(u) \frac{1}{2u^2} \left(\int_{1/u}^u \frac{1}{v} dv \right) du, \end{aligned}$$

d'après le théorème de Fubini ?? . Par conséquent,

$$\begin{aligned} d\mathbf{P}_U(u) &= \frac{1}{2u^2} \left(\int_{1/u}^u \frac{1}{v} dv \right) \mathbf{1}_{[1, +\infty[}(u) du \\ &= \frac{\ln u}{u^2} \mathbf{1}_{[1, +\infty[}(u) du. \end{aligned}$$

De même,

$$\begin{aligned} \mathbf{E}[f(V)] &= \int_{\Delta} f(v) \frac{1}{2u^2v} du dv \\ &= \int_0^{+\infty} f(v) \frac{1}{2v} \left(\int_{\Delta_v} \frac{1}{u^2} du \right) dv, \end{aligned}$$

où

$$\begin{aligned} \Delta_v &= \{u : (u, v) \in \Delta\} \\ &= \begin{cases} [1/v, +\infty[& \text{si } 0 \leq v \leq 1 \\ [v, +\infty[& \text{si } v \geq 1. \end{cases} \end{aligned}$$

Par conséquent,

$$\begin{aligned} d\mathbf{P}_V(v) &= \frac{1}{2v} \left(\int_{1/v}^{+\infty} \frac{1}{u^2} du \mathbf{1}_{[0, 1]}(v) + \int_v^{+\infty} \frac{1}{u^2} du \mathbf{1}_{]v, +\infty[}(v) \right) \\ &= \frac{1}{2v} \left(v \mathbf{1}_{[0, 1]}(v) + \frac{1}{v} \mathbf{1}_{]v, +\infty[}(v) \right). \end{aligned}$$

Comme

$$d\mathbf{P}_{(U, V)} \neq d\mathbf{P}_U \otimes d\mathbf{P}_V,$$

les variable aléatoire U et V ne sont pas indépendantes. ■

§ 3 Exercices

Exercice 2.2 En codage correcteur d'erreurs, les erreurs interviennent au hasard sur l'un quelconque des bits. Si on transmet des mots de n bits, on pose $\Omega = \{0, 1\}^n$, que l'on munit de la loi uniforme. On introduit $X_i(\omega) = \omega_i$ pour $i = 1, \dots, n$. La distance de Hamming entre mots de code $x = (x_1, \dots, x_n)$ et $y = (y_1, \dots, y_n)$, est définie par :

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

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

- 1) Montrer que sous la loi uniforme sur Ω , les variables $(X_i, i \in \{1, \dots, n\})$ sont indépendantes et identiquement distribuées de loi de Bernoulli de paramètre $1/2$.
- 2) Quelle est la longueur moyenne d'un mot ?

- 3) Quelle est la variance de la longueur d'un mot ?
- 4) On choisit deux mots au hasard indépendamment l'un de l'autre, soit X et Y les variables aléatoires correspondantes. Calculer

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

■

Exercice 2.3 Les règles du jeu du **not-seven** sont les suivantes : on part d'un score $X_0 = 0$. À chaque coup, on lance deux dés non pipés, si la somme des faces égale 7, le score retourne à 0 et la partie est terminée. Sinon, le score augmente de la somme des faces et on a le droit de rejouer ou pas. Si l'on ne rejoue pas, le score est acquis et la partie est terminée. Si l'on rejoue, on relance les deux dés avec la même règle.

- 1) Calculer la loi de la somme S des deux faces. Calculer son espérance.
On considère une suite $(S_n, n \in \mathbf{N})$ de variables aléatoires indépendantes de même loi que S .
- 2) Soit $\tau = \inf\{n \geq 1, S_n = 7\}$, trouver la loi de τ . Quelle est la moyenne de τ ?
- 3) Quelle est la stratégie d'un Initié (celui qui sait le résultat du prochain lancer de dés) ?
- 4) Calculer son gain moyen.
- 5) On appelle X_n le score au n -ième coup en l'absence de stratégie d'arrêt. Montrer que

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

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

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

■

Exercice 2.4 Un étang contient un nombre de poissons N inconnu. Pour estimer N , on prélève un échantillon de r poissons que l'on

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

- 1) Calculer la loi de X (dite loi hypergéométrique).

On note pour la suite de cet exercice

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

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

- 2) Montrer que $p_k^2 \geq p_{k-1}p_{k+1}$.
 3) En déduire qu'il existe une unique valeur de k telle que $p_k = \max_j p_j$.
 4) Soit k_0 tel cette valeur. Par définition, $p_{k_0+1} < p_{k_0}$ et $p_{k_0-1} < p_{k_0}$. En déduire que

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

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

- 5) En déduire une estimation de N . ■

L'exercice suivant est inspiré du livre ?? qui contient plein de choses intéressantes sur les liens entre probabilités et mathématiques discrètes.

Exercice 2.5 — Erdős et Renyi (1960). On fabrique un graphe sur n sommets en choisissant ses arêtes « au hasard ». Plus précisément, on considère le graphe $G_{n,p}$ obtenu en choisissant chacune des $\binom{n}{2}$ arêtes potentielles indépendamment avec probabilité p . Le but de ce problème est d'étudier la probabilité que $G_{n,p}$ soit connexe. On s'intéressera au cas où p est de la forme

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

où c est une constante fixée.

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

que $r \geq 1$ et $2r + 1 \leq n$ on a :

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

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

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

Suggestion. On pourra montrer que

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

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

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

- b) Que valent $\mathbf{E} [X_i]$ et $\mathbf{E} [X]$?
- c) On suppose dorénavant c fixé. Montrer que la quantité $F^{(k)}$, pour la variable X , converge, lorsque n tend vers l'infini, vers $e^{-ck}/k!$.
- d) Montrer que $\lim_{n \rightarrow \infty} \mathbf{P}(X = 0) = e^{-e^{-c}}$.
- e) Calculer l'espérance du nombre de composantes connexes à 2 sommets, et constater que celle-ci tend vers zéro quand n tend vers l'infini.
- f) Plus généralement, soit C_t le nombre de composantes connexes à t sommets. Montrer que pour $2 \leq t \leq n/2$,

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

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

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

■

Exercice 2.6 On rappelle qu'une suite de variables aléatoires $(X_n, n \in \mathbf{N})$ converge en probabilité vers la variable aléatoire X si et seule-

ment si pour tout $\epsilon > 0$,

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

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

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

■

Exercice 2.7 — Borne de Chernoff. Soit X une variable aléatoire de loi de Poisson de paramètre λ .

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

$$\mathbf{P}(X \geq \eta) \leq e^{-\eta\theta} \mathbf{E}[\exp(\theta X)]. \quad (2.6)$$

- Calculer $\mathbf{E}[\exp(\theta X)]$.
- Trouver θ qui minimise le terme de droite de (3.7).
- Trouver K tel que $\mathbf{P}(X \geq K\lambda) \leq 0,001$.

■

Exercice 2.8 On veut collectionner N images dont une et une seule apparaît dans chaque tablette de chocolat achetée. Les images sont mises au hasard dans les tablettes. On appelle T_i le nombre de tablettes nécessaires avant d'avoir i images distinctes. On pose $T_0 = 0$.

- Montrer que $T_{i+1} - T_i$ suit une loi géométrique de paramètre $1 - i/N$.
- Montrer que les variables aléatoires $T_0, T_1 - T_0, \dots, T_N - T_{N-1}$ sont indépendantes dans leur ensemble.
- Calculer l'espérance et la variance de T_N . Trouver un équivalent de l'espérance et montrer que la variance est un $O(N^2)$ quand N tend vers $+\infty$.
- En utilisant l'exercice 3.1, montrer que $T_N/(N \log N)$ tend vers 1 en probabilité.

■

Fonctions génératrices

Exercice 2.9 Dans une file d'attente à un serveur, on peut montrer sous certaines hypothèses que le nombre de clients dans le système (serveur + salle d'attente) suit la loi d'une géométrique translatée de -1 :

$$\mathbf{P}(N = k) = \mathbf{P}(\mathcal{G}(\rho) - 1 = k)$$

pour $k \geq 0$. Le temps de service d'un client suit une loi exponentielle de paramètre μ .

1. Calculer la fonction génératrice du temps d'attente moyen.
2. En déduire sa loi et sa moyenne.

■

Exercice 2.10 Pour une variable aléatoire X à valeurs dans \mathbf{N} , on définit son p -amincissement comme la variable aléatoire $p \circ X$ définie par

$$p \circ X = \sum_{i=1}^X B_i$$

où $(B_n, n \geq 1)$ est une suite de variables aléatoires indépendantes (et indépendante de X) de loi de Bernoulli de paramètre p .

1. Calculer la fonction génératrice de $p \circ X$ en fonction de celle de X .
2. Quelle est la loi de $p \circ X$ quand X suit une loi de Poisson de paramètre λ ?

■

Exercice 2.11 On note Φ_X la fonction génératrice de la variable aléatoire X .

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

■

Fonctions de répartition et densité

Exercice 2.12 Soit X une variable aléatoire r. à densité et $(a, b) \in \mathbf{R}^2$. Exprimer la densité de la variable aléatoire r. $Y := aX + b$ en fonction de la densité de X . ■

Exercice 2.13 Soit $X \sim \mathcal{N}(0, 1)$ et $Y = X^2$.

1. Calculer la fonction de répartition F_Y de Y en fonction de celle de X .
2. En déduire que Y admet une densité, que l'on exprimera. ■

Exercice 2.14 Soit μ_n la suite de mesure sur $[0, 1]$ donnée par

$$d\mu_n(x) = \frac{1}{n} \sum_{j=0}^{n-1} \delta_{j/n}.$$

Pour f continue sur $[0, 1]$, quelle est la limite de $\int f(t) d\mu_n(t)$ quand n tend vers $+\infty$? ■

Exercice 2.15 1. Soit X une variable aléatoire r. de densité f telle que

$$\lim_{|x| \rightarrow \infty} |x|^p f(x) = 0$$

pour tout $p > 0$. Montrer que X possède tous ses moments. En déduire qu'une variable aléatoire r. gaussienne possède tous ses moments.

2. Soit X une variable aléatoire r. suivant une loi de Cauchy. Montrer que $E(|X|)$ diverge. Plus généralement, montrer que X ne possède aucun moment. ■

Exercice 2.16 — *. Un nombre est choisi au hasard dans l'intervalle $[0, 10]$ suivant une loi \mathbf{P} donnée par

$$d\mathbf{P}(t) = K t \mathbf{1}_{[0, 10]}(t) dt,$$

où K est une constante à calculer. On note par X sa partie entière et par Y sa partie fractionnaire.

1. Calculer la loi du vecteur (X, Y) . Est-ce que les composantes sont indépendantes?
2. Calculer la matrice de covariance de (X, Y) .

■

Exercice 2.17 — **. Pour $a > 0$, on définit

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

Une variable aléatoire r. X est dite de loi gamma de paramètres a et $\lambda > 0$ si sa loi est donnée par

$$dP_X(t) = 1_{[0, \infty[}(t) \frac{\lambda^a}{\Gamma(a)} e^{-\lambda t} t^{a-1} dt,$$

notée par $X \sim G(a, \lambda)$.

1. Calculer l'espérance et la variance de X .
2. Soit Y une autre variable aléatoire r. indépendante de X , de loi $G(b, \lambda)$. Montrer que $X + Y$ et $\frac{X}{X + Y}$ sont indépendantes, calculer leur loi.
3. En déduire que

$$\beta(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

■

Exercice 2.18 — *. Soit X une variable aléatoire réelle de fonction de répartition F_X et F_X^{-1} l'inverse à droite de F_X défini par :

$$F_X^{-1}(y) = \inf\{u; F_X(u) \geq y\}.$$

Soit U une variable aléatoire de loi uniforme sur $[0, 1]$, montrer que $F_X^{-1}(U)$ a la loi de X . Cette relation permet de générer des variable aléatoire de loi arbitraire à partir de variables de loi uniforme sur $[0, 1]$. Ceci est très fréquemment utilisé en simulation et connu sous le nom de méthode d'inversion.

Trouver comment générer des variables de loi exponentielle et de Cauchy avec cette méthode. ■

Exercice 2.19 — *. La difficulté qui apparaît lors de la mise en oeuvre de la méthode précédente est l'inversion de la fonction de répartition. On a fréquemment la densité de façon explicite mais pas la fonction de répartition. Dans ce cas, on applique la méthode de rejet. Soit f_X la densité de X et g une densité qui majore à une constante près f_X et pour laquelle on sait facilement générer des

variable aléatoire dont la loi a pour densité g . On procède de la manière suivante : soit a tel que $f_X(u) \leq ag(u)$ pour tout u . On tire une variable aléatoire de loi de densité g , soit Y le résultat de ce tirage. On tire, indépendamment, une variable aléatoire de loi uniforme sur $[0, 1]$ et on note U le résultat de ce tirage. Si $U \leq f(Y)/ag(Y)$ alors le résultat est Y sinon on recommence au début.

1. Quel est l'espace de probabilité sous-jacent sur lequel sont définies les variable aléatoire Z et Y .
2. Montrer que $\mathbf{P}(Y \leq t) = F_X(t)$.
3. Soit X et Y deux variable aléatoire indépendantes de loi exponentielle de paramètre μ . Calculer la densité de la loi de $Z = X - Y$.
4. En déduire une façon d'engendrer des variable aléatoire de loi de densité :

$$\frac{\mu}{2\gamma(1 + 1/\alpha)} \exp(-\mu|x|^\alpha)$$

où $\alpha \geq 1$ et $\mu > 0$.

■

Exercice 2.20 Soit U et V deux variable aléatoire indépendantes de loi uniforme sur $[0, 1]$. Posons :

$$X = \sqrt{-2 \ln(U)} \cos(2\pi V) \text{ et } Y = \sqrt{-2 \ln(U)} \sin(2\pi V).$$

Montrer que X et Y sont des variable aléatoire gaussiennes centrées, réduites, indépendantes.

■

Exercice 2.21 — Simulation dans un cercle - * 1. Comment utiliser la méthode de rejet pour simuler le tirage d'un point « au hasard » dans un cercle de rayon R ?

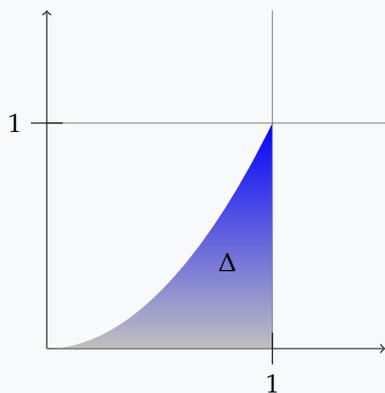
2. Calculer la loi joint de module et de l'argument d'un point « au hasard » dans le même cercle. En déduire une autre façon de simuler le tirage d'un point choisi uniformément dans le cercle.
3. Quelle est la meilleure méthode?

■

Calculs de lois

Exercice 2.22 — **. Soit un couple de variable aléatoire (X, Y) de densité conjointe $f(x, y) = cxy^2$ si $(x, y) \in \Delta$ et 0 sinon, où c est une constante positive et Δ est donné par

$$\Delta = \{(x, y) \in [0, 1]^2, 0 < y < x^2 < 1\}.$$



1. Déterminer la constante c .
2. Déterminer les lois marginales des variables X et Y .
3. Soit U une variable aléatoire de loi exponentielle de paramètre 8, et V une variable aléatoire de loi exponentielle de paramètre 3, indépendante de U . On pose

$$W = \exp(-U), \quad Z = \exp(-V)W^2.$$

Vérifier que f est la densité de la loi du couple (W, Z) .

■

Exercice 2.23 Soit W une variable aléatoire de loi de Poisson de paramètre $\lambda > 0$:

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

1. Montrer que pour toute fonction positive f :

$$\lambda \mathbf{E}[f(W+1)] = \mathbf{E}[Wf(W)]. \quad (2.7)$$

2. Réciproquement, soit W une variable aléatoire discrète, à valeurs dans \mathbf{N} , telle que pour toute fonction positive, l'identité 2.7 soit satisfaite. En appliquant 2.7 à des fonctions f judicieusement choisies, montrer que

$$\mathbf{P}(W = j) = \frac{\lambda}{j} \mathbf{P}(W = j-1),$$

pour tout $j \geq 1$.

3. En déduire la loi de W .

■

Exercice 2.24 On tire un nombre X uniformément sur $[0, 1]$. On tire ensuite des nombres Y_1, Y_2, \dots indépendamment les uns des autres et indépendamment de X , uniformément sur $[0, 1]$. Le jeu s'arrête dès que $Y_i > X$. Vous gagnez alors $(i - 1)\text{€}$. On appelle G le gain. Pour k entier, on définit

$$\varphi_k(x, y_1, \dots, y_{k+1}) = \begin{cases} \mathbf{1}_{\{y_1 > x\}} & \text{si } k = 0 \\ \mathbf{1}_{\{y_1 \leq x, \dots, y_k \leq x, y_{k+1} > x\}} & \text{si } k > 0. \end{cases}$$

1. Pour k entier, montrer que

$$\int_{[0,1]^{k+2}} \varphi_k(x, y_1, \dots, y_{k+1}) dy_1 dy_2 \dots dy_{k+1} dx = \frac{1}{k+1} - \frac{1}{k+2}.$$

On traitera séparément les cas $k = 0$ et $k > 0$.

2. Calculer la loi de G .
3. Calculer l'espérance de G .

■

Exercice 2.25 Pour tout a réel strictement positif, G_a désigne une variable aléatoire de loi gamma de paramètres $(a, 1)$: la densité g_a de sa loi est donnée par

$$g_a(x) = \frac{1}{\Gamma(a)} x^{a-1} e^{-x} \mathbf{1}_{\mathbf{R}^+}(x),$$

où

$$\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx.$$

En particulier, G_1 suit une loi exponentielle de paramètre 1. On admet que

$$\mathbf{E} \left[e^{itG_a} \right] = (1 - it)^{-a}, \text{ pour tout } t \in \mathbf{R}.$$

De plus, pour a, b réels strictement positifs, $B_{a,b}$ désigne une variable aléatoire de loi bêta de paramètres (a, b) : la densité $h_{a,b}$ de sa loi est donnée par

$$h_{a,b}(y) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1} (1-y)^{b-1} \mathbf{1}_{[0,1]}(y).$$

1. Calculer la loi du couple $(G_{a+b}B_{a,b}, G_{a+b})$ lorsque les variable aléatoire G_{a+b} et $B_{a,b}$ sont indépendantes.
2. En déduire que pour deux variables $G_{a+b}, B_{a,b}$ indépendantes, la loi de $B_{a,b}G_{a+b}$ est identique à celle de G_a .
3. Soit $n \geq 0$. Montrer par récurrence, que lorsque les variables aléatoires $B_{a,1}, \dots, B_{a+n,1}, G_{a+n+1}$ sont indépendantes, la loi de

$$P_n = G_{a+n+1} \prod_{j=0}^n B_{a+j,1}$$

est la même que celle de G_a .

On utilisera la question précédente et les hypothèses d'indépendance.

On évitera les longs calculs.

4. Soit X une variable aléatoire de loi exponentielle de paramètre 1 indépendante de G_a , montrer que $G_a + X$ a la même loi que G_{a+1} .
5. En déduire que pour tout entier n , G_{a+n} a même loi que

$$H_n = G_a + X_1 + X_2 + \dots + X_n,$$

où les X_i sont des variable aléatoire dont on précisera les propriétés.

On pose $W_n = G_a + X_1 + X_2 + \dots + X_n$ où les X_i sont indépendantes, identiquement distribuées de loi exponentielle de paramètre 1. On suppose de plus que les variable aléatoire G_a et $\{X_k, k \geq 1\}$ sont définies sur le même espace de probabilité.

6. Quelle est la limite presque-sûre de $(n^{-1}W_n, n \geq 1)$?
7. Montrer que la suite $(n^{-1}G_{a+n}, n \geq 1)$ converge en loi, vers une loi que l'on précisera.

■

Exercice 2.26 On rappelle que

$$\int_0^1 u^{-1/2}(1-u)^{-1/2} du = \pi.$$

Soit $X = (X_1, X_2)$ un vecteur gaussien de \mathbf{R}^2 , centré, de matrice de covariance (ou dispersion) $\Gamma = \text{Id}$. On pose

$$U = \frac{X_1^2}{X_1^2 + X_2^2} \text{ et } V = X_1^2 + X_2^2.$$

1. Calculer la densité de la loi de (U, V) .

2. Donner les densités marginales de U et V . On précisera les constantes de normalisation.
3. Soit $Z = \frac{X_2^2}{X_1^2}$. Exprimer Z en fonction de U puis calculer la densité de la loi de Z .

On note R_θ la rotation d'angle θ dans \mathbf{R}^2 . Si $x \in \mathbf{R}^2$,

$$R_\theta x = \begin{pmatrix} x_1 \cos \theta - x_2 \sin \theta \\ x_1 \sin \theta + x_2 \cos \theta \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

où x_1 et x_2 sont les composantes de x dans la base canonique de \mathbf{R}^2 .

Soit $X = (X_1, X_2)$ une variable aléatoire à valeurs dans \mathbf{R}^2 telle que pour tout $\theta \in [-\pi, \pi]$, $R_\theta X$ a même loi que X . C'est-à-dire que

$$\mathbf{E}[g(R_\theta X)] = \mathbf{E}[g(X)], \quad (2.8)$$

pour toute fonction g mesurable bornée de \mathbf{R}^2 dans \mathbf{R} . On suppose que la loi de X a une densité par rapport à la mesure de Lebesgue, notée v .

4. Montrer que pour toute fonction g mesurable bornée de \mathbf{R}^2 dans \mathbf{R} , pour tout $\theta \in [-\pi, \pi]$,

$$\int_{\mathbf{R}^2} g(x)v(x) \, dx = \int_{\mathbf{R}^2} g(y)v(R_\theta y) \, dy.$$

On admet qu'alors il existe $w : \mathbf{R}^+ \rightarrow \mathbf{R}^+$, mesurable, telle que

$$v(x) = w(\|x\|) \text{ pour tout } x \in \mathbf{R}^2.$$

5. Montrer que dans ce cas,

$$\int_0^{+\infty} w(r) r \, dr = \frac{1}{2\pi}.$$

On suppose maintenant que $X = (X_1, X_2)$ est un vecteur gaussien centré de matrice de covariance (ou dispersion) Γ .

6. Soit $\theta \in [-\pi, \pi]$, quelle est la loi de $R_\theta X$?
7. Montrer que $R_\theta X$ a même loi que X pour tout θ si et seulement si $\Gamma R_\theta = R_\theta \Gamma$.
8. Supposons que $\Gamma R_\theta = R_\theta \Gamma$ pour tout $\theta \in [-\pi, \pi]$. En écrivant les équations satisfaites par les coefficients de Γ , montrer que Γ est la matrice d'une homothétie positive (c'est-à-dire qu'il existe σ^2 tel que $\Gamma = \sigma^2 \text{Id}$).

■

§ 4 Vecteurs gaussiens

Les vecteurs gaussiens ont une importance toute particulière pour deux raisons : d'une part, le théorème de la limite centrée montre que la loi de Gauss est le domaine d'attraction de nombreuses limites et d'autre part, les calculs sur les lois normales se ramènent à de l'algèbre linéaire. Nous aurons en particulier besoin de considérer la transposée d'une matrice A , que nous noterons A^T . Le produit scalaire de deux vecteurs x et y de \mathbf{R}^k est noté $x \cdot y$. On rappelle que

$$x \cdot y = \sum_j x_j y_j = x^T y.$$

En particulier, $Ax \cdot y = x \cdot A^T y$, pour une matrice réelle A quelconque.

Rappelons d'abord la définition d'une variable aléatoire gaussienne réelle.

Définition 2.14 $X : \Omega \rightarrow \mathbf{R}$ est une variable aléatoire gaussienne réelle de paramètres m et σ^2 lorsque

$$d\mathbf{P}_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx.$$

La fonction caractéristique est donnée par :

$$\mathbf{E} \left[e^{itX} \right] = \exp\left(itm - \frac{\sigma^2 t^2}{2}\right).$$

En dimension supérieure, la définition d'un vecteur gaussien ne repose pas sur la densité de sa loi mais sur une caractérisation différente.

Définition 2.15 $X : \Omega \rightarrow \mathbf{R}^n$ est un vecteur gaussien lorsque $t \cdot X$ est une variable aléatoire gaussienne réelle pour tout $t \in \mathbf{R}^n$.

■ **Remarque 11** En particulier, chacune des composantes est une variable aléatoire gaussienne réelle. Réciproquement, si (X_1, \dots, X_n) sont des variables aléatoires gaussiennes réelles indépendantes alors $X = (X_1, \dots, X_n)$ est un vecteur gaussien.

Théorème 2.16 Soit $X = (X_1, \dots, X_n)$ un vecteur gaussien, on note

$$m = \mathbf{E}[X] = (\mathbf{E}[X_1], \dots, \mathbf{E}[X_n])$$

$$\Gamma_X = (\text{cov}(X_i, X_j), 1 \leq i, j \leq n)$$

où $\text{cov}(X, Y)$ est la covariance des variables aléatoires X et Y :

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

La fonction indicatrice de X est donnée par :

$$\begin{aligned}\mathbf{E} \left[e^{is \cdot X} \right] &= \exp \left(it \cdot m - \frac{1}{2} s^T \Gamma_X s \right) \\ &= \exp \left(i \sum_{j=1}^n t_j X_j - \frac{1}{2} \sum_{k=1}^n \sum_{l=1}^n \Gamma_X(k, l) s_k s_l \right),\end{aligned}$$

pour tout $s = (s_1, \dots, s_n) \in \mathbf{R}^n$.

Démonstration. Par définition, d'un vecteur gaussien, $s \cdot X$ est une variable aléatoire gaussienne réelle dont on sait que la loi est indicatrice par la moyenne et la variance. Par linéarité de l'espérance,

$$\mathbf{E} [s \cdot X] = \mathbf{E} \left[\sum_{k=1}^n s_k X_k \right] = \sum_{k=1}^n s_k \mathbf{E} [X_k] = s \cdot m.$$

D'autre part,

$$\begin{aligned}\text{var}(s \cdot X) &= \mathbf{E} \left[\left(\sum_{k=1}^n s_k (X_k - m_k) \right)^2 \right] \\ &= \mathbf{E} \left[\sum_{k=1}^n s_k^2 (X_k - m_k)^2 + 2 \sum_{1 \leq k < l \leq n} s_k s_l (X_k - m_k)(X_l - m_l) \right] \\ &= \sum_{k=1}^n s_k^2 \text{var}(X_k) + 2 \sum_{1 \leq k < l \leq n} s_k s_l \text{cov}(X_k, X_l) \\ &= \Gamma_X s \cdot s,\end{aligned}$$

d'où le résultat. ■

Théorème 2.17 Soit X un vecteur gaussien de \mathbf{R}^n , de vecteur moyen m et matrice de covariance Γ_X , soit A une matrice à r lignes et n colonnes et B un vecteur colonne de r lignes. Le vecteur aléatoire

$$Y = AX + B$$

est un vecteur gaussien de vecteur moyen $Am + B$ et de matrice de covariance $A \Gamma_X A^T$.

Démonstration. Vérifions d'abord que Y est un vecteur gaussien de \mathbf{R}^r . Soit $s \in \mathbf{R}^r$,

$$s \cdot Y = s \cdot AX + s \cdot B = A^T s \cdot X + s \cdot B$$

est une variable aléatoire gaussienne réelle puisque X est un vecteur gaussien. Il reste à calculer moyenne et variance de $s \cdot Y$ pour tout $s \in \mathbf{R}^r$. Le calcul de la moyenne est immédiat. Pour la variance, remarquons d'abord que

$$\text{var}(s \cdot Y) = \text{var}(A^T s \cdot X),$$

puisque la partie $s.B$ est déterministe donc a une variance nulle. Les calculs du théorème précédent appliqués à $A^T s$ montrent que

$$\text{var}(A^T s.X) = \Gamma_X A^T s.A^T s = A \Gamma_X A^T s.s$$

■

Théorème 2.18 — Représentation canonique. Soit X un vecteur gaussien de \mathbf{R}^n , de vecteur moyen m et de matrice de covariance Γ_X . Il existe une matrice A symétrique, positive telle que $A A^T = \Gamma_X$. Si Y est un vecteur gaussien de \mathbf{R}^n de vecteur moyen nul et de matrice de covariance l'identité alors en loi, on a l'égalité suivante :

$$X = AY + m.$$

Démonstration. Comme

$$\Gamma_X(k, l) = \text{cov}(X_k, X_l) = \text{cov}(X_l, X_k) = \Gamma_X(l, k),$$

Γ_X est une matrice symétrique. Comme

$$\begin{aligned} \Gamma_X s.s &= \sum_{k,l} \Gamma_X(k, l) s_k s_l \\ &= \text{var}(s.X) \geq 0, \end{aligned}$$

la forme bilinéaire associée à Γ_X est positive donc les valeurs propres de Γ_X sont positives ou nulles. Il existe une matrice orthogonale O telle que

$$O \Gamma_X O^{-1} = \begin{pmatrix} 0 & & & & & \\ & \ddots & & & & \\ & & 0 & & & \\ & & & \lambda_1 & & \\ & & & & \ddots & \\ & & & & & \lambda_r \end{pmatrix}$$

où $(\lambda_i, i = 1, \dots, r)$ sont les valeurs propres non nulles de Γ_X . La matrice

$$A = O^{-1} \begin{pmatrix} 0 & & & & & \\ & \ddots & & & & \\ & & 0 & & & \\ & & & \sqrt{\lambda_1} & & \\ & & & & \ddots & \\ & & & & & \sqrt{\lambda_r} \end{pmatrix} O$$

satisfait $AA^T = \Gamma_X$. En vertu du théorème 2.17, $AY + m$ est bien un vecteur gaussien de vecteur moyen m et de matrice de covariance Γ_X donc a la loi de X . ■

■ **Remarque 12** On peut illustrer le rôle de la matrice Γ dans la répartition spatiale des valeurs du vecteur gaussien. Considérons un vecteur gaussien de dimension 2, de matrice $\Gamma = \text{Id}$ alors la distribution se répartit de façon isotrope et pour une matrice de covariance Γ avec

$$\Gamma^{1/2} = \begin{pmatrix} 1 & 1 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 1 & 1 \\ \frac{1}{\sqrt{2}} & \frac{1}{10\sqrt{2}} \end{pmatrix},$$

les points se répartissent autour une ellipse d'axes les premières bissectrices, voir Figure 2.2.

```
import numpy as np
import scipy
from pylab import *
N=5000
x=np.random.randn(N,2)
gamma=np.array([[1/sqrt(2.0),1/sqrt(2.0)],[1/sqrt(2.0),0.1/
sqrt(2.0)]])
g=np.dot(gamma,x)
p=point([(g[0,i],g[1,i]) for i in range(0,N-1)],pointsize=1)
p.show()
```

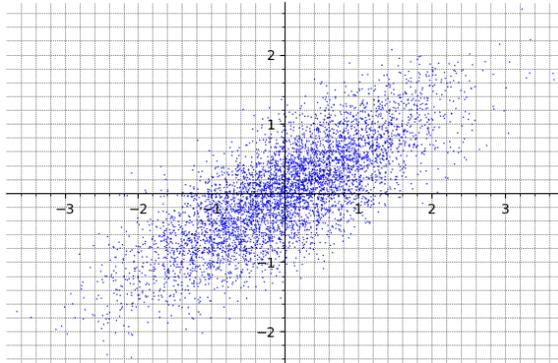


FIGURE 2.2: Si la matrice de covariance n'est pas l'identité, la distribution gaussienne associée n'est plus isotrope.

■ **Remarque 13** Le théorème précédent implique que si Γ_X est non inversible, où $r < n$, X prend ses valeurs dans un sous-espace affine strict (de dimension r strictement inférieure à n) de \mathbf{R}^n donc sa loi ne peut avoir de densité par rapport à la mesure de Lebesgue. En revanche, si Γ_X est inversible, le théorème précédent permet le calcul de la densité de la loi de X .

Théorème 2.19 Soit X un vecteur gaussien de \mathbf{R}^n , si sa matrice de

covariance Γ_X est inversible alors

$$d\mathbf{P}_X(x) = \frac{1}{\sqrt{(2\pi)^n \det \Gamma_X}} \exp\left(-\frac{1}{2} \Gamma_X^{-1}(x-m).(x-m)\right) dx.$$

Démonstration. Soit X est gaussienne de moyenne m et matrice de covariance Γ_X . Soit S symétrique telle que $SS^t = \Gamma_X$ et $Y = (Y_1, \dots, Y_n)$ un vecteur composé de variable aléatoire gaussiennes centrées réduites indépendantes. D'après le théorème précédent, $m + SY$ et X ont même loi. Remarquons que $\det(S)^2 = \det \Gamma_X$ dont S est inversible. Comme Γ_X est une matrice symétrique positive, son déterminant est positif et $\det S = \sqrt{\det \Gamma_X}$.

Considérons le changement de variable

$$\begin{aligned} \Theta : \mathbf{R}^n &\longrightarrow \mathbf{R}^n \\ y = (y_1, \dots, y_n) &\longmapsto x = Sy + m. \end{aligned}$$

Comme S est inversible, Θ est une bijection de \mathbf{R}^n dans \mathbf{R}^n et sa jacobienne est S . Enfin, remarquons que

$$\sum_{j=1}^n y_j^2 = y.y = S^{-1}(x-m).S^{-1}(x-m) = (S^{-1})^t S^{-1}(x-m).(x-m).$$

Or,

$$(S^{-1})^t S^{-1} = (S^t)^{-1} S^{-1} = (SS^t)^{-1} = \Gamma_X^{-1}.$$

Par conséquent, $y.y = \Gamma_X^{-1}(x-m).(x-m)$. Pour toute fonction ψ bornée, on a donc

$$\begin{aligned} \mathbf{E}[\psi(X)] &= \mathbf{E}[\psi(m + SY)] \\ &= \frac{1}{\sqrt{2\pi}^n} \int_{\mathbf{R}^n} \psi(m + Sy) \exp\left(-\frac{1}{2} \sum_{j=1}^n y_j^2\right) dy_1 \dots dy_n \\ &= \frac{1}{\sqrt{2\pi}^n} \int_{\mathbf{R}^n} \psi(x) \exp\left(-\frac{1}{2} \Gamma_X^{-1}(x-m).(x-m)\right) \frac{1}{\sqrt{\det \Gamma_X}} dx_1 \dots dx_n. \end{aligned}$$

D'où le résultat par identification. ■

Théorème 2.20 Soit $X = (Y, Z)$ un vecteur gaussien de \mathbf{R}^n avec $Y \in \mathbf{R}^{n_Y}$ et $Z \in \mathbf{R}^{n_Z}$ ($n = n_Y + n_Z$). Les vecteurs gaussiens Y et Z sont indépendants si et seulement si ils sont non-corrélés :

$$\text{cov}(Y_k, Z_l) = 0, \text{ pour tout } k \in \{1, \dots, n_Y\}, l \in \{1, \dots, n_Z\}.$$

Démonstration. Notons Γ_{YZ} la matrice de covariance de Y et Z , de taille $n_Y \times n_Z$ définie par :

$$\Gamma_{YZ}(k, l) = \text{cov}(Y_k, Z_l).$$

Par définition, la matrice de covariance de X se décompose en matrices-blocs sous la forme

$$\Gamma_X = \begin{pmatrix} \Gamma_Y & \vdots & \Gamma_{YZ} \\ \dots & \dots & \dots \\ \Gamma_{YZ} & \vdots & \Gamma_Z \end{pmatrix}.$$

En décomposant chaque vecteur $s \in \mathbf{R}^n$ sous la forme $s = (s_Y, s_Z)$ avec $s_Y \in \mathbf{R}^{n_Y}$ et $s_Z \in \mathbf{R}^{n_Z}$, on a d'une part

$$s.m = \sum_{k=1}^n s_k m_k = \sum_{k=1}^{n_Y} s_k m_k + \sum_{k=n_Y+1}^n s_k m_k = s_Y.m_Y + s_Z.m_Z,$$

et d'autre part

$$\begin{aligned} \Gamma_X s.s &= (\Gamma_Y s_Y + \Gamma_{YZ} s_Z).s_Y + (\Gamma_{YZ} s_Z + \Gamma_Z s_Z).s_Z \\ &= \Gamma_Y s_Y.s_Y + \Gamma_Z s_Z.s_Z + 2\Gamma_{YZ} s_Z.s_Y. \end{aligned}$$

Par conséquent,

$$\begin{aligned} \mathbf{E} \left[e^{is.X} \right] &= \mathbf{E} \left[e^{i(s_Y.Y + s_Z.Z)} \right] = \exp \left(is_Y.m_Y + is_Z.m_Z \right) \\ &\quad \cdot \exp \left(-\frac{1}{2} (\Gamma_Y s_Y.s_Y + \Gamma_Z s_Z.s_Z + 2\Gamma_{YZ} s_Z.s_Y) \right) \\ &= \mathbf{E} \left[e^{is_Y.Y} \right] \mathbf{E} \left[e^{is_Z.Z} \right] \exp \left(-\Gamma_{YZ} s_Z.s_Y \right). \end{aligned}$$

D'après la caractérisation (??) de l'indépendance, on en déduit que Y et Z sont indépendantes si et seulement si

$$\Gamma_{YZ} s_Z.s_Y = 0 \text{ pour tout } s_Y, s_Z. \quad (2.9)$$

Soit $(e_k, k = 1, \dots, n_Y)$, respectivement $(f_l, l = 1, \dots, n_Z)$, la base canonique de \mathbf{R}^{n_Y} , respectivement de \mathbf{R}^{n_Z} . Comme $\Gamma_{YZ} e_k.f_l = \Gamma_{YZ}(k, l)$, il s'ensuit que (2.9) est équivalent à $\Gamma_{YZ} = 0$. ■

■ **Remarque 14** En conséquence, les composantes d'un vecteur gaussien sont indépendantes si et seulement si la matrice de covariance est diagonale.

Exercice 2.27 Soit deux variable aléatoire indépendantes $X \sim \mathcal{N}(0, 1)$ et Y de loi $d\mathbf{P}_Y = \frac{1}{2}(\delta_{-1} + \delta_1)$.

1. Montrer que $Z = YX$ est une variable aléatoire gaussienne.
2. Montrer que X et Z sont non corrélées.
3. Si (X, Z) était un vecteur gaussien, quelle serait sa loi?
4. Calculer la loi de (X, Z) .
5. Est-ce que (X, Z) est un vecteur gaussien?
6. Est-ce que X et Z sont indépendantes?

■

Exercice 2.28 Soit X et Y deux gaussiennes centrées réduites indépendantes. Montrer que les variables aléatoires $X + Y$ et $\sin(X - Y)$ sont indépendantes. ■

Exercice 2.29 — Sphere hardening. Soit X_N un vecteur gaussien de \mathbf{R}^N , centré, réduit. Soit $\|X_N\|$, la norme euclidienne de X_N et $X'_N = \|X_N\|/\sqrt{N}$.

1. Calculer $\mathbf{E}[(X'_N)^2]$.
2. Calculer $\text{var}[(X'_N)^2]$.
3. Montrer que, pour tout $\eta > 0$,

$$\mathbf{P}(|X'_N - 1| \geq \eta) \xrightarrow{N \rightarrow +\infty} 0.$$

On pourra utiliser l'inégalité de Bienaymé-Tchebycev.

■

3

Convergences

Définition 3.1 Pour une suite d'événements $(A_n, n \geq 1)$, on note

$$\limsup_n A_n = \bigcap_k \bigcup_{n \geq k} A_n$$
$$\liminf_n A_n = \bigcup_k \bigcap_{n \geq k} A_n$$

■ **Remarque 15** L'élément ω appartient à $\limsup_n A_n$ si et seulement si ω appartient à une infinité des A_n .

Par ailleurs, $\omega \in \liminf_n A_n$ si et seulement si ω appartient à tous les A_n sauf un nombre fini. Autrement dit, il existe un rang dépendant de ω tel que ω appartienne à tous les A_n au delà de cet indice.

Enfin, on a de façon évidente

$$\left(\limsup_n A_n \right)^c = \liminf_n A_n^c.$$

Lemme 3.2 — Borel-Cantelli. Soit $(A_n, n \in \mathbf{N}^*)$ une suite d'événements. Si $\sum \mathbf{P}(A_n) < \infty$ alors $\mathbf{P}(\limsup_n A_n) = 0$.

Si de plus, ces événements sont indépendants et si $\sum \mathbf{P}(A_n) = +\infty$ alors $\mathbf{P}(\limsup_n A_n) = 1$.

Démonstration. Les ensembles $B_k = \bigcup_{n \geq k} A_n$ sont décroissants donc

$$\mathbf{P}(\limsup_n A_n) = \lim_{k \rightarrow +\infty} \mathbf{P}(B_k) \leq \lim_{k \rightarrow +\infty} \sum_{n=k}^{+\infty} \mathbf{P}(A_k) = 0,$$

en tant que reste d'une série convergente.

Si maintenant les ensembles sont indépendants, on peut écrire

$$\begin{aligned}
 \mathbf{P}(\liminf_n A_n^c) &= \lim_k \mathbf{P}(\cap_{n \geq k} A_n^c) \\
 &= \lim_k \lim_N \mathbf{P}(\cap_{k \leq n \leq N} A_n^c) \\
 &= \lim_k \lim_N \prod_{n=k}^N \mathbf{P}(A_n^c) \\
 &= \lim_k \lim_N \prod_{n=k}^N (1 - \mathbf{P}(A_n)).
 \end{aligned}$$

On utilise maintenant l'inégalité $1 - x \leq \exp(-x)$ pour conclure que

$$\begin{aligned}
 \mathbf{P}(\liminf_n A_n^c) &\leq \lim_k \lim_N \prod_{n=k}^N \exp(-\mathbf{P}(A_k)) \\
 &\leq \lim_k \lim_N \exp(-\sum_{n=k}^N \mathbf{P}(A_k)) = 0,
 \end{aligned}$$

d'où le résultat. ■

§ 1 Convergence presque-sûre

Définition 3.3 La suite $(X_n, n \in \mathbf{N}^*)$ converge presque sûrement vers X lorsqu'il existe un mesurable $A \in \mathcal{A}$ tel que $\mathbf{P}(A^c) = 0$ et

$$\forall \omega \in A, \forall \epsilon > 0, \exists n(\omega, \epsilon) \text{ tel que } n \geq n(\omega, \epsilon) \implies d(X_n(\omega), X(\omega)) \leq \epsilon.$$

Théorème 3.4 La suite $(X_n, n \in \mathbf{N}^*)$ converge presque sûrement vers X si et seulement si

$$\mathbf{P}\left(\limsup_n (d(X_n, X) > \epsilon)\right) = 0. \quad (3.1)$$

Démonstration. Soit A l'ensemble des ω pour lesquels $X_n(\omega)$ converge vers $X(\omega)$. Si ω appartient à A alors pour tout $\epsilon > 0$, il existe un indice $n(\omega, \epsilon)$ tel que

$$n \geq n(\omega, \epsilon) \implies d(X_n(\omega), X(\omega)) \leq \epsilon,$$

ce qui revient à dire que

$$\omega \in \bigcap_{n \geq n(\omega, \epsilon)} (d(X_n(\omega), X(\omega)) \leq \epsilon).$$

Donc, pour tout $\epsilon > 0$,

$$A \subset \liminf_n (\omega : d(X_n(\omega), X(\omega)) \leq \epsilon).$$

Réciproquement, supposons que pour tout $\epsilon > 0$,

$$\mathbf{P}(\liminf_n(\omega : d(X_n(\omega), X(\omega)) \leq \epsilon)) = 1 \quad (3.2)$$

et prenons $A = \liminf_n(\omega : d(X_n(\omega), X(\omega)) \leq \epsilon)$. Si ω appartient à A alors à partir d'un certain rang

$$d(X_n(\omega), X(\omega)) \leq \epsilon$$

donc $X_n(\omega)$ tend vers $X(\omega)$.

Il y a donc convergence presque sûre si et seulement si l'identité (3.2) est vérifiée. En passant au complémentaire, on trouve la caractérisation (3.1). ■

Le théorème principal concernant la convergence presque-sûre est la loi forte des grands nombres.

Théorème 3.5 — Loi forte des grands nombres. Soit $(X_n, n \in \mathbf{N}^*)$ une suite de variables aléatoires (à valeurs dans \mathbf{R}^p) indépendantes et de même loi telle que $\mathbf{E}[\|X_1\|_{\mathbf{R}^p}] < \infty$ alors

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \mathbf{E}[X_1].$$

§ 2 Convergence en probabilité

Définition 3.6 Une suite $(X_n, n \in \mathbf{N}^*)$ converge en probabilité vers X lorsque pour tout $\epsilon > 0$,

$$\lim_{n \rightarrow +\infty} \mathbf{P}(d(X_n, X) > \epsilon) = 0.$$

Théorème 3.7 Une suite $(X_n, n \in \mathbf{N}^*)$ converge en probabilité vers X si et seulement si

$$\mathbf{E}[d(X_n, X) \wedge 1] \xrightarrow[n \rightarrow +\infty]{} 0.$$

Démonstration. D'après l'inégalité de Markov,

$$\mathbf{P}(d(X_n, X) \wedge 1 > \epsilon) \leq \epsilon^{-1} \mathbf{E}[d(X_n, X) \wedge 1].$$

Par ailleurs, d'après le théorème de Fubini,

$$\mathbf{E}[d(X_n, X) \wedge 1] = \int_0^1 \mathbf{P}(d(X_n, X) \wedge 1 > \epsilon) \, d\epsilon.$$

Le résultat découle des relations précédentes. ■

Théorème 3.8 La convergence presque-sûre implique la convergence en probabilité mais la réciproque est fautive.

Démonstration. On a clairement

$$(d(X_n, X) > \varepsilon) \subset \bigcup_{k \geq n} (d(X_k, X) > \varepsilon).$$

Or la convergence p.s. de X_n vers X équivaut à

$$0 = \mathbf{P}(\limsup_k (d(X_k, X) > \varepsilon)) = \lim_{k \rightarrow +\infty} \mathbf{P}(\bigcup_{k \geq n} (d(X_k, X) > \varepsilon)),$$

donc X_n converge en probabilité.

Réciproquement, soit $(X_n, n \in \mathbf{N}^*)$ la suite de v.a. indépendantes de loi définie par

$$\mathbf{P}(X_n = n) = \frac{1}{n} \text{ et } \mathbf{P}(X_n = 0) = 1 - \frac{1}{n}.$$

Pour $0 < \varepsilon < 1$, $\mathbf{P}(X_n > \varepsilon) = n^{-1}$ donc X_n converge en probabilité vers 0. En revanche,

$$\sum_n \mathbf{P}(X_n > \varepsilon) = +\infty,$$

d'après le lemme de Borel-Cantelli, $\mathbf{P}(\liminf_n |X_n| \leq \varepsilon) = 0$, il n'y a donc pas convergence presque-sûre. ■

Théorème 3.9 Une suite $(X_n, n \in \mathbf{N}^*)$ converge en probabilité vers X si et seulement de toute sous-suite, on peut extraire une sous-suite qui converge presque-sûrement vers X .

Démonstration. Si $(X_n, n \in \mathbf{N}^*)$ converge en probabilité vers X , il existe n_1 tel que

$$n \geq n_1 \implies \mathbf{P}(d(X_n, X) > 2^{-1}) \leq 2^{-1}.$$

On construit ensuite $n_2 < \dots < n_k$ tels que

$$n \geq n_k \implies \mathbf{P}(d(X_n, X) > 2^{-k}) \leq 2^{-k}.$$

D'après le lemme de Borel-Cantelli,

$$\mathbf{P}\left(\limsup_k (d(X_{n_k}, X) > 2^{-k})\right) = 0.$$

Pour \mathbf{P} -presque tout ω , ω appartient à $\liminf (d(X_{n_k}, X) \leq 2^{-k})$ donc pour presque tout ω , quel que soit $\varepsilon > 0$, il existe k_0 tel que

$$k \geq k_0 \implies d(X_{n_k}, X) \leq \varepsilon.$$

Ceci signifie que la suite $(X_{n_k}, k \in \mathbf{N}^*)$ converge p.s. vers X .

Réciproquement, si X_n ne converge pas en probabilité vers X , il existe une suite extraite $(X_{n_k}, k \in \mathbf{N}^*)$ telle que

$$\mathbf{P}(d(X_{n_k}, X) > \varepsilon) \geq \varepsilon, \text{ pour tout } k. \quad (3.3)$$

D'après l'hypothèse, on peut extraire de cette sous-suite, une suite qui converge p.s. donc en probabilité. Ce dernier point est en contradiction avec (3.3) donc X converge en probabilité vers X . ■

Corollaire 3.10 Soit S et T deux espaces métriques et $(X_n, n \in \mathbf{N}^*)$ une suite de v.a. à valeurs dans S qui converge en probabilité vers X . Soit $f : S \rightarrow T$ une fonction mesurable, p.s. continue sur $X(\Omega)$. La suite $(f(X_n), n \in \mathbf{N}^*)$ converge en probabilité vers $f(X)$ dans T .

Démonstration. Soit $(f(X_n), k \in \mathbf{N}')$ une sous-suite, on sait qu'il en existe une suite extraite $(X_n, n \in \mathbf{N}'')$ qui converge p.s. vers X . Comme f est continue, $f(X_n)$ converge vers $f(X)$ pour $n \in \mathbf{N}''$. D'après le théorème 3.9, cela induit la convergence en probabilité de la suite $(f(X_n), n \in \mathbf{N}^*)$. ■

Considérons maintenant une suite dénombrable d'espaces polonais (S_k, d_k) et soit S le produit cartésien de tous ces ensembles, muni de la topologie produit, c'est-à-dire de la topologie associée à la distance

$$d(x, y) = \sum_k 2^{-k} (d_k(x_k, y_k) \wedge 1).$$

Comme S_k est séparable, $\mathcal{B}(S) = \otimes_k \mathcal{B}(S_k)$ et une v.a. à valeurs dans S est juste une suite de v.a. à valeurs dans S_k .

Théorème 3.11 Soit une suite dénombrable d'espaces polonais (S_k, d_k) et soit S le produit cartésien de tous ces ensembles. Soit $X = (X_k, k \in \mathbf{N}^*)$ et $X^n = (X_k^n, k \in \mathbf{N}^*)$ des v.a. de S . X^n converge en probabilité vers X si et seulement si pour tout k , X_k^n converge en probabilité vers X_k dans S_k .

Démonstration. Compte-tenu de la définition de d , on a, pour tout n ,

$$\mathbf{E}[d(X, X^n) \wedge 1] = \mathbf{E}[d(X, X^n)] = \sum_k 2^{-k} \mathbf{E}[d_k(X_k, X_k^n) \wedge 1].$$

Par convergence dominée, il s'ensuit que

$$\mathbf{E}[d(X, X^n) \wedge 1] \xrightarrow{n \rightarrow +\infty} 0 \iff \mathbf{E}[d_k(X_k, X_k^n) \wedge 1] \xrightarrow{n \rightarrow +\infty} 0,$$

pour tout k . ■

Corollaire 3.12 — Lemme de Slutsky. Soit $(X_n, n \in \mathbf{N}^*)$ et $(Y_n, n \in \mathbf{N}^*)$ des suites de v.a.r. telles que $X_n \xrightarrow{\mathbf{P}} X$ et $Y_n \xrightarrow{\mathbf{P}} Y$. Il découle des théorèmes précédents que

$$\begin{aligned} aX_n + bY_n &\xrightarrow{\mathbf{P}} aX + bY \\ X_n Y_n &\xrightarrow{\mathbf{P}} XY \\ X_n / Y_n &\xrightarrow{\mathbf{P}} X/Y, \end{aligned}$$

lorsque presque-sûrement, $Y \neq 0$ et $Y_n \neq 0$ pour tout n .

Démonstration. En vertu du théorème 3.11, (X^n, Y^n) converge en probabilité vers (X, Y) . Compte-tenu du corollaire 3.10, les deux premiers points s'ensuivent aisément. Quant au troisième, il suffit d'appliquer le même corollaire à la fonction $(x, y) \mapsto (x/y)\mathbf{1}_{\{y \neq 0\}}$ qui est continue p.s. au point (X, Y) d'après les hypothèses. ■

§ 3 Convergence en loi

Nous abordons maintenant le mode de convergence le plus « faible » mais par conséquent, le plus fréquent. Pour de plus amples détails sur cette convergence, on pourra consulter BILLINGSLEY, *Convergence of Probability Measures*; KALLENBERG, *Foundations of modern probability*.

Dans tout ce qui suit les variables aléatoires sont à valeurs dans \mathbf{R}^d donc les mesures de probabilités que l'on considère sont définies sur \mathbf{R}^d muni de sa tribu borélienne.

Définition 3.13 La suite $(\mathbf{P}_n, n \in \mathbf{N}^*)$ converge faiblement vers la mesure de probabilités \mathbf{P} lorsque $\int f d\mathbf{P}_n$ converge vers $\int f d\mathbf{P}$ pour toute fonction f continue bornée.

On dit que la suite de variables aléatoires $(X_n, n \in \mathbf{N}^*)$ converge en loi vers la variable aléatoire X lorsque $(\mathbf{P}_{X_n}, n \geq 1)$ converge faiblement vers \mathbf{P}_X : pour toute fonction f continue bornée de \mathbf{R}^d dans \mathbf{R} , on a

$$\mathbf{E}[f(X_n)] \xrightarrow{n \rightarrow \infty} \mathbf{E}[f(X)].$$

La démonstration du théorème suivant est longue et fastidieuse et se trouve dans les ouvrages sus-mentionnés. Nous l'omettrons donc. On rappelle qu'une fonction de répartition n'est discontinue qu'en au plus un nombre dénombrable de points. Pour un ensemble A , on note ∂A sa frontière définie par

$$\partial A = \bar{A} - \overset{\circ}{A}.$$

Théorème 3.14 Les propriétés suivantes sont équivalentes :

- (i) $(\mathbf{P}_{X_n}, n \geq 1)$ converge faiblement vers \mathbf{P}_X ,
- (ii) Pour tout borélien $A \in \mathbf{R}^d$ tel que $\mathbf{P}(X \in \partial A) = 0$,

$$\mathbf{P}(X_n \in A) \xrightarrow{n \rightarrow \infty} \mathbf{P}(X \in A),$$

- (iii) Les fonctions de répartition $(F_{X_n}, n \geq 1)$ converge vers F_X en tout point de continuité de F_X ,
- (iv) Pour tout $t \in \mathbf{R}^d$,

$$\mathbf{E} \left[e^{it \cdot X_n} \right] \xrightarrow{n \rightarrow \infty} \mathbf{E} \left[e^{it \cdot X} \right].$$

Le dernier point est connu sous le nom de théorème de Lévy.

Théorème 3.15 Si la suite $(X_n, n \in \mathbf{N}^*)$ converge en probabilité vers X alors elle converge en loi. Si X est constante alors la réciproque est vraie.

Démonstration. Supposons que $(X_n, n \in \mathbf{N}^*)$ converge en probabilité vers X mais ne converge pas en loi. Cela revient à dire qu'il existe f continue bornée, qu'il existe $\varepsilon > 0$ tel que pour tout $n \geq 1$, il existe $k \geq n$ et

$$|\mathbf{E}[f(X_n)] - \mathbf{E}[f(X)]| \geq \varepsilon. \quad (3.4)$$

Cela revient à dire qu'il existe une sous-suite $(n_k, k \geq 1)$ tels que (3.4) soit vrai le long de cette sous-suite. En vertu du théorème 3.9, on peut alors extraire une sous-sous-suite de cette sous-suite qui converge presque-sûrement. Mézalor, le théorème de convergence dominée implique que pour cette sous-sous-suite

$$\mathbf{E} \left[f(X_{n'_k}) \right] \xrightarrow{k \rightarrow \infty} \mathbf{E} [f(X)],$$

ce qui est en contradiction avec l'hypothèse (3.4). Par conséquent, la convergence en loi a bien lieu.

Si maintenant, $(X_n, n \in \mathbf{N}^*)$ converge en loi vers m déterministe. La fonction $(x \mapsto \|x - m\| \wedge 1)$ est continue et bornée. Par conséquent,

$$\mathbf{E} [\|X_n - m\| \wedge 1] \xrightarrow{n \rightarrow \infty} \mathbf{E} [\|m - m\|] = 0.$$

Compte-tenu de la définition 3.6, cela signifie que $(X_n, n \in \mathbf{N}^*)$ converge en probabilité vers m . ■

Nous avons déjà vu que la convergence en probabilité impliquait l'existence d'une sous-suite convergeant presque-sûrement. Rien de tel pour la convergence en loi puisqu'elle n'impose même pas que les variables aléatoires soient définies sur le même espace de probabilités. Par contre,

nous disposons du théorème de couplage de Skorohod qui peut s'avérer fort utile. Nous l'énonçons dans \mathbf{R}^d mais ne ferons la preuve que dans le cas réel (voir le théorème 3.30 du Kallenberg KALLENBERG, *Foundations of modern probability*).

Définition 3.16 Soit F_X la fonction de répartition d'une variable aléatoire. Pour tout $t \in [0, 1]$, on pose

$$F_X^{-1}(t) = \inf\{x, F_X(x) \geq t\}.$$

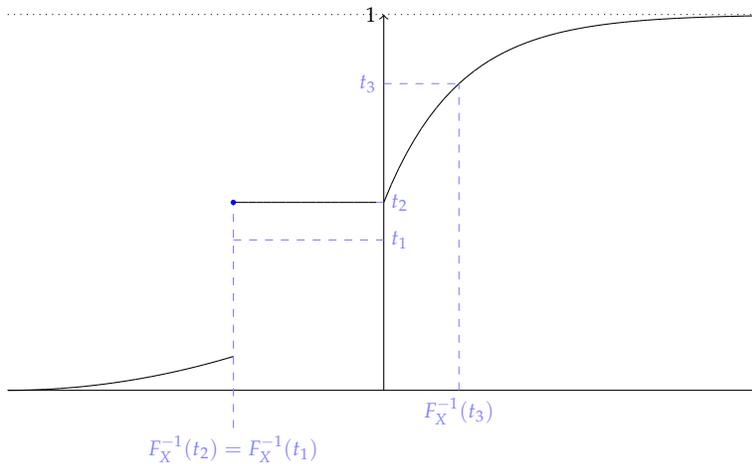


FIGURE 3.1: Inverse d'une fonction de répartition. On remarque que les sauts de F_X engendrent des intervalles de constance de F_X^{-1} et que les plateaux de F_X engendrent des discontinuités de F_X^{-1} .

En regardant attentivement la figure 3.1, on se convainc aisément du lemme suivant.

Lemme 3.17 Pour tout réel x , on a l'égalité suivante :

$$(u : F_X^{-1}(u) \leq x) = (u : u \leq F_X(x)).$$

Le théorème suivant permet de construire n'importe quelle variable aléatoire réelle sur l'espace de probabilité $([0, 1], \mathcal{B}([0, 1]), \text{Leb})$. Sa preuve est immédiate en vue du lemme précédent.

Théorème 3.18 Soit U une variable aléatoire de loi uniforme sur $[0, 1]$ alors $F_X^{-1}(U)$ a la loi de X .

Lemme 3.19 L'ensemble des intervalles ouverts sur lesquels F_X est constante est au plus dénombrable.

Démonstration. Pour tout réel x , considérons A_x , l'intérieur de $F_X^{-1}(\{x\})$. Notons que par construction, $A_x \cap A_y = \emptyset$ dès que $x \neq y$. Dans chacun des A_x non vide, on choisit un rationnel r_x , on a donc une injection de $(x, A_x \neq \emptyset)$ dans \mathbf{Q} donc le nombre de plateaux de F_X est au plus dénombrable. Comme les A_x sont des ouverts, chacun d'entre eux est

une réunion (au plus dénombrable) d'intervalles ouverts mais en raison de la croissance de F_X , il ne peut à chaque fois, n'y avoir qu'un seul intervalle dans A_X . ■

Lemme 3.20 Soit $(X_n, n \in \mathbf{N}^*)$ une suite de variables aléatoires convergeant en loi vers X . La suite de fonctions $(F_{X_n}^{-1}, n \geq 1)$ converge Lebesgue presque-partout vers F_X^{-1} .

■ **Remarque 16** Une preuve rapide et intuitive marche dans le cas où F_X est continue et strictement croissante. D'après le lemme de Dini, la convergence des F_n est alors uniforme, ce qui veut dire que pour tout $\varepsilon > 0$, à partir d'un certain n les graphes Γ_n des fonctions $(x, F_n(x))$ sont dans un tube de hauteur ε autour de Γ , celui de $(x, F(x))$. Comme le graphe de $(y, F_{X_n}^{-1}(y))$ (respectivement Γ) est le symétrique par rapport à la première bissectrice de Γ_n (respectivement Γ) donc pour tout $u \in \mathbf{R}$, $F_{X_n}^{-1}(u)$ tend vers $F_X^{-1}(u)$.

Démonstration. Posons

$$\begin{aligned} X^+(u) &= \inf\{x, F_X(x) > u\} \\ X^-(u) &= F_X^{-1}(u) = \inf\{x, F_X(x) > u\}. \end{aligned}$$

■

Théorème 3.21 — Théorème de couplage de Skorohod. Soit $(X_n, n \in \mathbf{N}^*)$ une suite de variables aléatoires convergeant en loi vers X . Il existe $(X'_n, n \in \mathbf{N}^*)$ et X' construites sur le même espace de probabilité telles que

$$X'_n \stackrel{\text{loi}}{=} X_n, \quad X' \stackrel{\text{loi}}{=} X \quad \text{et} \quad X'_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} X'. \quad (3.5)$$

Démonstration. On pose

$$X'_n = F_{X_n}^{-1}(U), \quad X' = F_X^{-1}(U)$$

où U est une variable aléatoire uniforme sur $[0, 1]$. Compte-tenu du lemme 3.20 et du théorème 3.18, les propriétés de (3.5) sont vérifiées. ■

Delta méthode

Soit $(X_n, n \in \mathbf{N}^*)$ une suite de vecteurs aléatoires (dans \mathbf{R}^d) indépendants et identiquement distribués. On note $\mathbf{E}[X]$ le vecteur moyen et Γ_X la matrice de covariance :

$$\Gamma_X(i, j) = \text{cov}(X(i), X(j)).$$

D'après la loi forte des grands nombres,

$$S_n = \frac{1}{n} \sum_{j=1}^n X_j \xrightarrow[\text{p.s.}]{n \rightarrow \infty} \mathbf{E}[X] \in \mathbf{R}^d$$

et le TCL indique que

$$\sqrt{n}(S_n - \mathbf{E}[X]) \xrightarrow[\text{loi}]{n \rightarrow \infty} \mathcal{N}(0, \Gamma_X).$$

On peut être amené à considérer des fonctions des composantes de X . Soit $\psi : \mathbf{R}^d \rightarrow \mathbf{R}$ différentiable au voisinage de $\mathbf{E}[X]$. D'après le principe de continuité,

$$\psi(S_n) \xrightarrow[\text{p.s.}]{n \rightarrow \infty} \psi(\mathbf{E}[X]) \in \mathbf{R}.$$

La delta-méthode est destinée à trouver des intervalles de confiance pour cette convergence.

Théorème 3.22 Sous les hypothèses qui précèdent,

$$\sqrt{n}(\psi(S_n) - \psi(\mathbf{E}[X])) \xrightarrow[\text{loi}]{n \rightarrow \infty} \mathcal{N}\left(0, \nabla \psi(\mathbf{E}[X]) \Gamma_X (\nabla \psi(\mathbf{E}[X]))^t\right).$$

Démonstration. D'après la formule de Taylor, on peut écrire

$$\psi(x) = \psi(\mathbf{E}[X]) + \nabla \psi(\mathbf{E}[X]) \cdot (x - \mathbf{E}[X]) + R_n(x) \quad (3.6)$$

avec

$$\|R_n(x)\| = o(\|x - \mathbf{E}[X]\|).$$

D'après le théorème 3.21, on peut construire S'_n de même loi que S_n et une variable aléatoire G de loi gaussienne centrée, de covariance Γ_X telle que $n^{1/2}(S'_n - \mathbf{E}[X])$ tende p.s. vers G (et donc S'_n tend p.s. vers $\mathbf{E}[X]$). D'après (3.6),

$$\|\sqrt{n}R(S'_n)\| = o(\sqrt{n}\|S'_n - \mathbf{E}[X]\|)$$

et $\sqrt{n}(S'_n - \mathbf{E}[X])$ tend vers G donc

$$\sqrt{n}R(S'_n) = o(1) \xrightarrow[\text{p.s.}]{n \rightarrow \infty} 0.$$

Par conséquent,

$$\sqrt{n}(\psi(S'_n) - \psi(\mathbf{E}[X])) \xrightarrow[\text{p.s.}]{n \rightarrow \infty} \nabla \psi(\mathbf{E}[X]) \cdot G.$$

D'après le cours sur les gaussiennes, on sait que

$$\nabla \psi(\mathbf{E}[X]) \cdot G \stackrel{\text{loi}}{=} \mathcal{N}(0, \nabla \psi(\mathbf{E}[X]) \Gamma_X (\nabla \psi(\mathbf{E}[X]))^t).$$

Le résultat s'en déduit. ■

Application aux chaînes de Markov

On sait que la probabilité stationnaire d'une chaîne de Markov irréductible, récurrente peut s'écrire

$$\pi(y) = \frac{1}{\mathbf{E}[\tau_x^1]} \mathbf{E}_x \left[\sum_{j=0}^{\tau_x^1-1} \mathbf{1}_{X_j=y} \right]$$

où τ_x^1 est l'instant de premier retour en x partant de x . De manière plus générale

$$\bar{z} = \sum_{y \in E} f(y)\pi(y) = \frac{1}{\mathbf{E}[\tau_x^1]} \mathbf{E}_x \left[\sum_{j=0}^{\tau_x^1-1} f(X_j) \right].$$

Simulons R cycles de x à lui-même, soit (d_1, \dots, d_R) leur longueur, i.e. $d_j = \tau_x^j - \tau_x^{j-1}$. Soit

$$Y_l = \sum_{j=\tau_x^{l-1}}^{\tau_x^l} f(X_j).$$

Les $((d_k, Y_k), k \in \{1, \dots, R\})$ sont indépendants et identiquement distribués donc

$$\begin{aligned} \hat{\tau}_R &= \frac{1}{R} \sum_{k=1}^R d_k \xrightarrow{\text{p.s.}} \mathbf{E}_x [\tau_x] = \bar{\tau} \\ \hat{Y}_R &= \frac{1}{R} \sum_{k=1}^R Y_k \xrightarrow{\text{p.s.}} \mathbf{E}_x \left[\sum_{j=0}^{\tau_x-1} f(X_j) \right] = \bar{y} \\ \sqrt{R}((\hat{\tau}_R, \hat{Y}_R) - (\bar{\tau}, \bar{y})) &\xrightarrow{\text{loi}} \mathcal{N}(0, \Gamma) \end{aligned}$$

avec

$$\Gamma = \begin{pmatrix} \text{var}(\tau_x^1) & \text{cov}(\tau_x^1, Y_1) \\ \text{cov}(\tau_x^1, Y_1) & \text{var}(Y_1^2) \end{pmatrix}.$$

Théorème 3.23 Un estimateur consistant mais biaisé de \bar{z} est donné par

$$\hat{z}_R = \frac{\hat{Y}_R}{\hat{\tau}_R}.$$

La précision de cet estimateur est donné par

$$\sqrt{R}(\hat{z}_R - \bar{z}) \xrightarrow{\text{Loi}} \mathcal{N}(0, \eta^2)$$

où

$$\eta^2 = \gamma_{11} \frac{\bar{y}^2}{\bar{\tau}^4} - 2\gamma_{12} \frac{\bar{y}}{\bar{\tau}^3} + \gamma_{22} \frac{1}{\bar{\tau}^2}.$$

L'intervalle de confiance à 95% est donc donné par

$$\bar{z} \in]\hat{z}_R - \frac{1,95 \eta_R}{\sqrt{R}}, \hat{z}_R + \frac{1,95 \eta_R}{\sqrt{R}}[.$$

Démonstration. Soit

$$\begin{aligned} \varphi : \mathbf{R}^* \times \mathbf{R} &\longrightarrow \mathbf{R} \\ (x, y) &\longmapsto y/x. \end{aligned}$$

Son jacobien est

$$\nabla \varphi(x, y) = \left(-\frac{y}{x^2}, \frac{1}{x} \right).$$

D'après le théorème 3.22, $\sqrt{R}(\hat{z}_R - \bar{z})$ converge en loi vers une gaussienne de matrice de covariance

$$\begin{aligned} \eta_R^2 &= \left(-\frac{\bar{y}}{\bar{\tau}^2}, \frac{1}{\bar{\tau}} \right) \Gamma \left(-\frac{\bar{y}}{\bar{\tau}^2}, \frac{1}{\bar{\tau}} \right)^t \\ &= \left(-\frac{\bar{y}}{\bar{\tau}^2}, \frac{1}{\bar{\tau}} \right) \left(-\gamma_{11} \frac{\bar{y}}{\bar{\tau}^2} + \gamma_{12} \frac{1}{\bar{\tau}}, -\gamma_{12} \frac{\bar{y}}{\bar{\tau}^2} + \gamma_{22} \frac{1}{\bar{\tau}} \right)^t \\ &= \gamma_{11} \frac{\bar{y}^2}{\bar{\tau}^4} - 2\gamma_{12} \frac{\bar{y}}{\bar{\tau}^3} + \gamma_{22} \frac{1}{\bar{\tau}^2}. \end{aligned}$$

D'où le résultat. ■

Comme on ne connaît pas a priori ni les γ_{ij} , ni \bar{y} , ni $\bar{\tau}$ on les remplace par leur version « empirique » : $\bar{\tau}$ est remplacé par $\hat{\tau}_R$, \bar{y} est remplacé par \hat{Y}_R et γ_{ij} par

$$\begin{aligned} \hat{\gamma}_{12}^R &= \frac{1}{R-1} \sum_{k=1}^R d_k Y_k - \hat{\tau}_R \hat{Y}_R \\ \hat{\gamma}_{11} &= \frac{1}{R-1} \sum_{k=1}^R d_k^2 - \hat{\tau}_R^2 \\ \hat{\gamma}_{22} &= \frac{1}{R-1} \sum_{k=1}^R Y_k^2 - \hat{Y}_R^2 \end{aligned}$$

§ 4 Exercices

Exercice 3.1 Soit $(X_n, n \in \mathbf{N})$ une suite de v.a. de moyenne μ_n et de variance σ_n^2 . Soit $(b_n, n \in \mathbf{N})$ une suite de réels positifs tels

que σ_n^2/b_n^2 tende vers 0. Montrer que

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

■

Exercice 3.2 — Borne de Chernoff. Soit X une v.a. de loi de Poisson de paramètre λ .

1. Montrer que $\mathbf{P}(X \geq \eta) = \mathbf{P}(\exp(\theta X) \geq \exp(\theta \eta))$ pour tout $\theta > 0$.
2. Montrer que, pour tout $\theta \geq 0$,

$$\mathbf{P}(X \geq \eta) \leq e^{-\eta\theta} \mathbf{E}[\exp(\theta X)]. \quad (3.7)$$

3. Calculer $\mathbf{E}[\exp(\theta X)]$.
4. Trouver θ qui minimise le terme de droite de (3.7).
5. Trouver K tel que $\mathbf{P}(X \geq K\lambda) \leq 0,001$.

■

Exercice 3.3 On veut collectionner N images dont une et une seule apparaît dans chaque tablette de chocolat achetée. Les images sont mises au hasard dans les tablettes. On appelle T_i le nombre de tablettes nécessaires avant d'avoir i images distinctes. On pose $T_0 = 0$.

1. Montrer que $T_{i+1} - T_i$ suit une loi géométrique de paramètre $1 - i/N$.
2. Montrer que les variables aléatoires $T_0, T_1 - T_0, \dots, T_N - T_{N-1}$ sont indépendantes dans leur ensemble.
3. Calculer l'espérance et la variance de T_N . Trouver un équivalent de l'espérance et montrer que la variance est un $O(N^2)$ quand N tend vers $+\infty$.
4. En utilisant l'exercice 3.1, montrer que $T_N/(N \log N)$ tend vers 1 en probabilité.

■

Exercice 3.4 — Inversion de la transformée de Laplace. Soit X_1, \dots, X_n des variables aléatoires réelles, indépendantes, de même loi exponentielle de paramètre λ .

1. Calculer la loi de leur somme.

2. À l'aide de ce résultat et du théorème central limite, montrer que l'on a

$$\lim_{n \rightarrow \infty} \int_{[0, n/\alpha]} x e^{-yx} \frac{(xy)^{n-1}}{(n-1)!} dy = \frac{1}{2} \mathbf{1}_{(x=\alpha)} + \mathbf{1}_{(x>\alpha)}.$$

Pour X variable aléatoire positive, on pose

$$G(s) = \mathbf{E} [\exp -sX], \text{ pour } s \geq 0.$$

3. En déduire que

$$\lim_{n \rightarrow \infty} \int_{[0, n/\alpha]} \frac{(-1)^n}{(n-1)!} s^n \frac{d^n G(s)}{ds^n} ds = \frac{1}{2} \mathbf{P}(X = \alpha) + \mathbf{P}(X > \alpha).$$

yy

■

4

Modélisation

§ 1 *Marche aléatoire*

On suit la présentation des pages 80 et suivantes de GRIMMETT et STIRZAKER, *Probability and Random Processes*. On peut obtenir bien d'autres résultats, notamment des théorèmes limites, par la théorie des martingales.

§ 2 *Processus de Poisson*

Les résultats de cette partie sont extraits de DECREUSEFOND et MOYAL, *Stochastic Modeling and Analysis of Telecom Networks*.

Deuxième partie

Préparation aux oraux

5

Chaînes de Markov

§ 1 Version simple

Définition 5.1 Soit $(X_n, n \in \mathbf{N})$ une suite de variables aléatoires à valeurs dans E fini ou dénombrable. La suite $(X_n, n \in \mathbf{N})$ est une chaîne de Markov si pour tout entier n , pour tout $x_0, x_1, \dots, x_{n+1} \in E$,

$$\begin{aligned} \mathbf{P}(X_{n+1} = x_{n+1} \mid X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \\ = \mathbf{P}(X_{n+1} = x_{n+1} \mid X_n = x_n). \end{aligned} \quad (5.1)$$

On dit d'une chaîne de Markov qu'elle « oublie son passé » : conditionnellement à la variable aléatoire X_n (l'état présent), la variable aléatoire X_{n+1} (l'état futur) ne dépend pas des variables aléatoires X_0, \dots, X_{n-1} (les états passés). Plus généralement, toute l'évolution future d'une chaîne de Markov ne dépend que de son état présent :

Théorème 5.2 Pour tout $n \in \mathbf{N}$, pour tout $k \geq 1$, pour tout $x_0, x_1, \dots, x_{n+k} \in E$, on a :

$$\begin{aligned} \mathbf{P}(X_{n+1} = x_{n+1}, \dots, X_{n+k} = x_{n+k} \mid X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \\ = \mathbf{P}(X_{n+1} = x_{n+1}, \dots, X_{n+k} = x_{n+k} \mid X_n = x_n). \end{aligned}$$

Démonstration. Par récurrence sur k . La propriété est vraie pour $k = 1$ par définition. On suppose qu'elle est vraie pour un certain $k \geq 1$. Alors,

pour tout $n \in \mathbf{N}$, pour tout $x_0, x_1, \dots, x_{n+k} \in E$, on a

$$\begin{aligned}
& \mathbf{P}(X_{n+1} = x_{n+1}, \dots, X_{n+k+1} = x_{n+k+1} \mid X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \\
&= \mathbf{P}(X_{n+1} = x_{n+1}, \dots, X_{n+k} = x_{n+k} \mid X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \\
&\quad \times \mathbf{P}(X_{n+k+1} = x_{n+k+1} \mid X_0 = x_0, X_1 = x_1, \dots, X_{n+k} = x_{n+k}), \\
&= \mathbf{P}(X_{n+1} = x_{n+1}, \dots, X_{n+k} = x_{n+k} \mid X_n = x_n) \\
&\quad \times \mathbf{P}(X_{n+k+1} = x_{n+k+1} \mid X_{n+k} = x_{n+k}), \\
&= \mathbf{P}(X_{n+1} = x_{n+1}, \dots, X_{n+k} = x_{n+k} \mid X_n = x_n) \\
&\quad \times \mathbf{P}(X_{n+k+1} = x_{n+k+1} \mid X_n = x_n, X_{n+1} = x_{n+1}, \dots, X_{n+k} = x_{n+k}), \\
&\quad = \mathbf{P}(X_{n+1} = x_{n+1}, \dots, X_{n+k+1} = x_{n+k+1} \mid X_n = x_n),
\end{aligned}$$

où l'on a utilisé (5.1), l'hypothèse de récurrence et la définition des chaînes de Markov pour obtenir la seconde égalité, et à nouveau la définition des chaînes de Markov pour obtenir la troisième égalité. ■

Pour prédire l'évolution d'une chaîne de Markov, on peut simplement considérer les maillons de cette chaîne, correspondant aux transitions successives d'un état à l'autre :

Théorème 5.3 On a : $\forall n \in \mathbf{N}, \forall k \geq 1, \forall x_n, x_{n+1}, \dots, x_{n+k} \in E,$

$$\begin{aligned}
& \mathbf{P}(X_{n+1} = x_{n+1}, \dots, X_{n+k} = x_{n+k} \mid X_n = x_n) = \mathbf{P}(X_{n+1} = x_{n+1} \mid X_n = x_n) \times \\
& \mathbf{P}(X_{n+2} = x_{n+2} \mid X_{n+1} = x_{n+1}) \times \dots \times \mathbf{P}(X_{n+k} = x_{n+k} \mid X_{n+k-1} = x_{n+k-1}).
\end{aligned}$$

Démonstration. Par la formule des probabilités composées,

$$\begin{aligned}
& \mathbf{P}(X_{n+1} = x_{n+1}, \dots, X_{n+k} = x_{n+k} \mid X_n = x_n) \\
&= \mathbf{P}(X_{n+1} = x_{n+1} \mid X_n = x_n) \times \mathbf{P}(X_{n+2} = x_{n+2} \mid X_{n+1} = x_{n+1}, X_n = x_n) \\
&\quad \times \dots \times \mathbf{P}(X_{n+k} = x_{n+k} \mid X_n = x_n, \dots, X_{n+k-1} = x_{n+k-1}).
\end{aligned}$$

La preuve découle alors de la définition des chaînes de Markov. ■

Ce qui suit est la copie quasi conforme du chapitre correspondant dans DECREUSEFOND et MOYAL, *Stochastic Modeling and Analysis of Telecom Networks*.

§ 2 Définition et exemples

Considérons donc une suite de variables aléatoires $X = (X_n, n \geq 0)$ à valeurs dans E fini ou dénombrable et indexée par les entiers, et la filtration $\mathcal{F}_n = \sigma\{X_j, 0 \leq j \leq n\}$ engendrée par cette suite.

Les trajectoires sont des éléments de $E^{\mathbf{N}}$, c'est-à-dire des suites d'éléments de E . Le décalage est alors défini par :

$$\begin{aligned}
& \theta : E^{\mathbf{N}} \longrightarrow E^{\mathbf{N}} \\
& (x_0, x_1, \dots) \longmapsto (x_1, x_2, \dots).
\end{aligned}$$

On aura souvent besoin du n^e itéré de θ , noté θ^n et défini par :

$$\begin{aligned} \theta^n : E^{\mathbf{N}} &\longrightarrow E^{\mathbf{N}} \\ (x_0, x_1, \dots) &\longmapsto (x_n, x_{n+1}, x_{n+2}, \dots). \end{aligned}$$

Par la suite, on identifiera θ et θ^1 . Par la même méthode que dans le théorème 5.3, on peut montrer que

$$\begin{aligned} &\mathbf{P}(X_0 = x_0, \dots, X_{n+k+1} = x_{n+k+1} \mid X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \\ &= \mathbf{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n \mid X_n = x_n) \\ &\quad \times \mathbf{P}(X_{n+1} = x_{n+1}, \dots, X_{n+k+1} = x_{n+k+1} \mid X_n = x_n). \end{aligned}$$

Cela se traduit par la définition suivante.

Définition 5.4 La suite X est une chaîne de Markov lorsque pour tout $n \leq m$, la tribu \mathcal{F}_n est indépendante de la tribu $\sigma(X_m)$ conditionnellement à $\sigma(X_n)$. En d'autres termes, pour toute fonction F et G bornées, on a :

$$\mathbf{E}[F(X_0, \dots, X_n) G(X_m) \mid \mathcal{F}_n] = \mathbf{E}[F(X_0, \dots, X_n) \mid \mathcal{F}_n] \mathbf{E}[G(X_m) \mid X_n]. \quad (5.2)$$

En vertu du corollaire 6.21, on sait que cette propriété est équivalente à l'indépendance du passé et du futur conditionnellement au présent, ce qui s'exprime par :

$$\mathbf{E}[F(X_0, \dots, X_n) G \circ \theta^n \mid \mathcal{F}_n] = F(X_0, \dots, X_n) \mathbf{E}[G \circ \theta^n \mid X_n]. \quad (5.3)$$

En particulier, pour $G = \mathbf{1}_{\{y\}}(X_1)$, pour tout entier n , on obtient :

$$\mathbf{P}(X_{n+1} = y \mid \mathcal{F}_n) = \mathbf{P}(X_{n+1} = y \mid X_n).$$

Définition 5.5 La chaîne de Markov X est dite homogène lorsque :

$$\mathbf{P}(X_{n+1} = y \mid X_n = x)$$

ne dépend pas de n mais seulement de x et y . On notera cette quantité $p(x, y)$ et on appelle $P = (\mathbf{P}(X_1 = y \mid X_0 = x), x, y \in E)$ l'opérateur de transition de X . Si E est de cardinal fini, P s'identifie à une matrice qui a autant de lignes et de colonnes que d'éléments dans E .

■ **Exemple 5.1** Un rat se déplace dans le labyrinthe à sept cases représenté dans la figure 5.1. Il passe d'une case à l'autre uniformément suivant les possibilités qui lui sont offertes, c'est-à-dire que lorsqu'il y a deux (respectivement trois) sorties dans la case où il se trouve, il va dans chacune des cases possibles avec une probabilité d'un demi (respectivement d'un tiers). Son évolution est sans mémoire : chaque changement

ne dépend que de la situation courante, pas du passé. On appelle X_n la position du rat après son n^e mouvement, X_0 est sa position initiale.

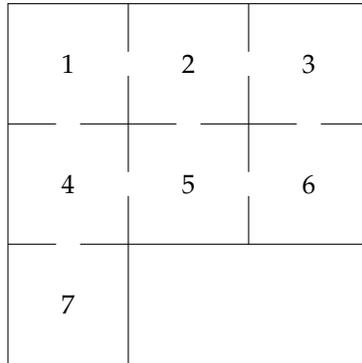


FIGURE 5.1: Le labyrinthe

Ici, $E = \{1, 2, 3, 4, 5, 6, 7\}$ et la matrice de transition se déduit aisément de la figure 5.1 :

$$P = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

■ **Exemple 5.2** Partant d'un score vierge, on lance deux dés non pipés. Si la somme est différente de 7, on ajoute cette somme au score courant et on rejoue. Sinon le score s'annule et la partie s'arrête. On appelle X_n le score après le n^e lancer. Il faut distinguer ici deux états 0 si l'on veut que X soit une chaîne homogène. En effet, on peut quitter le 0 du départ mais on ne peut pas quitter le 0 consécutif à un 7. On prend donc espace d'états, $E = \mathbf{N} \cup \{\delta\}$, δ est ce qu'il est convenu d'appeler un point cimetièrre. Les transitions sont alors données pour tout $i \neq \delta$ par :

$$\begin{aligned} \mathbf{P}(X_{n+1} = i + 2 \mid X_n = i) &= \mathbf{P}(X_{n+1} = i + 12 \mid X_n = i) = 1/36 \\ \mathbf{P}(X_{n+1} = i + 3 \mid X_n = i) &= \mathbf{P}(X_{n+1} = i + 11 \mid X_n = i) = 2/36 \\ \mathbf{P}(X_{n+1} = i + 4 \mid X_n = i) &= \mathbf{P}(X_{n+1} = i + 10 \mid X_n = i) = 3/36 \\ \mathbf{P}(X_{n+1} = i + 5 \mid X_n = i) &= \mathbf{P}(X_{n+1} = i + 9 \mid X_n = i) = 4/36 \\ \mathbf{P}(X_{n+1} = i + 6 \mid X_n = i) &= \mathbf{P}(X_{n+1} = i + 8 \mid X_n = i) = 5/36 \\ \mathbf{P}(X_{n+1} = \delta \mid X_n = i) &= 1/6 \\ \mathbf{P}(X_{n+1} = \delta \mid X_n = \delta) &= 1. \end{aligned}$$

■

La définition même d'une chaîne de Markov implique que toute son évolution est déterminée par la loi de la position initiale que l'on notera ν et l'opérateur de transition P .

Théorème 5.6 Pour tout n , la loi jointe de (X_0, \dots, X_n) est déterminée par la loi de X_0 et P à partir de la formule suivante :

$$\mathbf{P}(X_0 = x_0, \dots, X_m = x_m) = \nu(\{x_0\}) \prod_{l=0}^{m-1} p(x_l, x_{l+1}),$$

pour tout n et tout x_0, \dots, x_n dans E .

■ **Remarque 17** Par la suite, nous noterons \mathbf{P}_ν la loi d'une chaîne de Markov avec loi initiale ν . Par abus de notation, \mathbf{P}_x représentera la loi de la chaîne si le point de départ est fixe, égal à $x \in E$. Comme E est au plus dénombrable, on peut numéroter les états : utiliser une injection qui existe entre E et \mathbf{N} . On peut alors supposer que $E \subset \mathbf{N}$. On retrouve alors le formalisme des vecteurs et matrices même si l'on est amené à manipuler de tels objets avec un nombre infini de composantes... On considère souvent le « vecteur » π_n défini par $\pi_n(x) = \mathbf{P}(X_n = i)$ pour $i \in E \subset \mathbf{N}$. Il est usuel de le considérer comme un vecteur ligne. Pour tout n , la relation :

$$\begin{aligned} \mathbf{P}(X_{n+1} = j) &= \sum_{i \in E} \mathbf{P}(X_{n+1} = j | X_n = i) \mathbf{P}(X_n = i) \\ &= \sum_{i \in E} \mathbf{P}(X_n = i) p(i, j), \quad j \in E, \end{aligned}$$

se lit en notation matricielle :

$$\pi_{n+1} = \pi_n \cdot P \text{ soit } \pi_n = \pi_0 \cdot P^n \quad (5.4)$$

où P^n est la n^{e} puissance de P . En particulier, si π_0 n'est composé que de 0 sauf un 1 en i^{e} position (ce qui revient à travailler sous \mathbf{P}_i) alors pour tout $j \in E$, on a :

$$\mathbf{P}_i(X_n = j) = p^{(n)}(i, j)$$

où $p^{(n)}(i, j)$ est le terme en i^{e} ligne et j^{e} colonne de P^n .

Comme $P^{n+m} = P^n P^m$, on déduit de (5.4) l'équation de Chapman-Kolmogorov :

$$p^{(n+m)}(x, y) = \sum_{z \in E} p^{(n)}(x, z) p^{(m)}(z, y), \quad (5.5)$$

valable pour tout n, m , toute condition initiale et tout état final. On l'a écrite ici sous forme « intrinsèque », c'est-à-dire sans tenir compte de l'injection mentionnée plus haut. Les deux exemples qui suivent illustrent l'intérêt de l'opérateur *shift*.

■ **Exemple 5.3 — Suite de 5.1.** Supposons qu'il y ait un bout de fromage en case 3 et une batterie en case 7. On veut calculer la probabilité que le rat puisse manger avant d'être électrocuté. Introduisons :

$$\tau_3 = \inf\{n \geq 0, X_n = 3\} \text{ et } \tau_7 = \inf\{n \geq 0, X_n = 7\}.$$

On pose $u_i = \mathbf{P}_i(\tau_3 < \tau_7)$. Il est clair que $u_3 = 1$ et que $u_7 = 0$. Pour $i \notin \{3;7\}$,

$$u_i = \sum_{j=1}^7 \mathbf{P}_i(\tau_3 < \tau_7 | X_1 = j) \mathbf{P}_i(X_1 = j).$$

Puisque i est différent de 3 et 7, l'événement $(\tau_3 < \tau_7)$ est \mathbf{P}_i presque sûrement égal à A_1 où :

$$\begin{aligned} A_1 &= \{\omega, \exists i \geq l \text{ tel que } \omega_i = 3 \text{ et } \omega_j \in \{1, 2, 4, 5, 6\} \text{ pour tout } l \leq j < i\} \\ &= \{\text{postérieurement à l'instant } l, \text{ on atteint 3 avant 7}\}. \end{aligned}$$

Comme $\mathbf{1}_{A_1} = \mathbf{1}_{A_0} \circ \theta$, on a :

$$\mathbf{P}_i(\tau_3 < \tau_7 | X_1 = j) = \mathbf{P}_j(\tau_3 < \tau_7).$$

Compte tenu du fait que $\mathbf{P}_i(X_1 = j) = p(i, j)$ on voit que les u_i sont solutions du système linéaire :

$$u_3 = 1, u_7 = 0, u_i = \sum_{j=1}^6 p(i, j) u_j \text{ pour } i \notin \{3;7\}.$$

La résolution de ce système donne $u_1 = 7/12, u_2 = 3/4, u_4 = 5/12, u_5 = 2/3, u_6 = 5/6$.

Sans fromage et batterie, calculons maintenant le temps moyen d'atteinte de la case 3. Posons $v_i = \mathbf{E}_i[\tau_3]$. Il est clair $v_3 = 0$. Par ailleurs, pour $i \neq 3$, on a :

$$\mathbf{E}_i[\tau_3] = \sum_{j=1}^7 \mathbf{E}_i[\tau_3 | X_1 = j] p(i, j).$$

Si l'on a comme trajectoire $\omega = (1, 2, 5, 2, 5, 6, 3, \dots)$ alors $\tau_3(\omega) = 6$ mais $\tau_3(\theta\omega) = 5$. Plus généralement, on a, conditionnellement à $X_0 \neq 3$, $\tau_3 = \tau_3 \circ \theta + 1$. Par conséquent, on a :

$$v_i = \sum_{j=1}^7 (\mathbf{E}_i[\tau_3 \circ \theta | X_1 = j] + 1) p(i, j) = \sum_{j=1}^7 p(i, j) v_j + 1,$$

d'après la relation (5.3). Les v_i sont donc les solutions d'un système linéaire à six équations et six inconnues qu'il ne vous reste plus, cher lecteur, qu'à résoudre. ■

Simulation

Rappelons d'abord comment simuler une loi ν sur un ensemble au plus dénombrable E . Les états sont numérotés grâce à une bijection ϕ entre E et un sous-ensemble de \mathbf{N} . On pose ensuite :

$$r_0 = \nu(\{\phi^{-1}(0)\}) \text{ et } r_n = \sum_{j=0}^n \nu(\{\phi^{-1}(j)\}) = \nu(\phi^{-1}(\{0, \dots, n\})).$$

Algorithme 5.1 : Réalisation d'une variable aléatoire de loi ν

Données : r_0, r_1, \dots

Résultat : un élément de E choisi selon la loi ν

$x \leftarrow$ réalisation d'une loi uniforme sur $[0, 1]$;

$n \leftarrow 0$;

tant que $x > r_n$ **faire**

 | $n \leftarrow n + 1$

fin

retourner $\phi^{-1}(n)$

Dans une chaîne de Markov, lorsque l'on est à l'état x , on passe à l'état y avec probabilité $p(x, y)$. Pour passer d'une étape à l'autre, il suffit d'appliquer l'algorithme précédent à la loi $\mu = (p(x, y), y \in E)$.

Algorithme 5.2 : Simulation d'une trajectoire d'une chaîne de Markov (ν, P)

Données : ν, P, N

Résultat : une trajectoire à N pas de la chaîne de Markov (ν, P)

Choisir x_0 condition initiale selon ν ;

pour compteur $\leftarrow 1$ à N **faire**

 | Choisir x_{compteur} selon la loi $(p(x_{\text{compteur}-1}, y), y \in E)$;

fin

retourner x_0, x_1, \dots, x_N

§ 3 Propriété de Markov forte

Pour T temps d'arrêt, sur $(T < \infty)$, on définit θ^T par :

$$\theta^T(\omega) = (\omega_{T(\omega)}, \omega_{T(\omega)+1}, \dots).$$

Pour $x \in E$, on définit :

$$\tau_x^1 = \begin{cases} \infty & \text{si } X_n \neq x \text{ pour tout } n > 0; \\ \inf\{n > 0, X_n = x\} & \text{sinon.} \end{cases}$$

$$\tau_x^k = \begin{cases} \infty & \text{si } \tau_x^{k-1} = \infty \text{ ou } X_n \neq x \text{ pour tout } i > \tau_x^{k-1} \\ \inf\{n > \tau_x^{k-1}, X_n = x\} & \text{sinon.} \end{cases}$$

Pour tout k , τ_x^k est l'instant du k^e passage de la chaîne X à l'état x .

Lemme 5.7 Pour x fixé dans E , sur l'événement $\{\tau_x^1 < \infty\}$, on a :

$$\tau_x^k = \tau_x^{k-1} + \tau_x^1 \circ \theta^{\tau_x^{k-1}}. \quad (5.6)$$

Démonstration. Si $\tau_x^{k-1} = \infty$ alors on a ∞ des deux côtés de l'égalité. Sinon, le résultat est immédiat à partir du moment où l'on se persuade que $\theta^{\tau_x^{k-1}}(\omega)$ représente la partie de la trajectoire postérieure à la $(k-1)^e$ visite à l'état x . Par conséquent, la première visite après la $(k-1)^e$ (si elle existe) est la k^e visite depuis le début. ■

Théorème 5.8 Soit T un temps d'arrêt presque sûrement fini et $F : \Omega \rightarrow \mathbf{R}^+$ une variable aléatoire intégrable. On a l'identité suivante :

$$\mathbf{E} [F \circ \theta^T | \mathcal{F}_T] = \mathbf{E} [F | X_0 = X_T]. \quad (5.7)$$

Pour calculer le terme de droite, on calcule $\mathbf{E} [F | X_0 = x] = \phi(x)$ et l'on a :

$$\mathbf{E} [F | X_0 = X_T] = \phi(X_T).$$

Démonstration. Pour $A \in \mathcal{F}_T$, puisque $A \cap \{T = n\} \in \mathcal{F}_n$ et en utilisant (5.3) et les propriétés de l'espérance conditionnelle, il vient :

$$\begin{aligned} \mathbf{E} [F \circ \theta^T \cdot \mathbf{1}_A] &= \sum_{n=0}^{\infty} \mathbf{E} [F \circ \theta^n \cdot \mathbf{1}_{A \cap \{T=n\}}] \\ &= \sum_{n=0}^{\infty} \mathbf{E} [\mathbf{E} [F \circ \theta^n | \mathcal{F}_n] \mathbf{1}_{A \cap \{T=n\}}] \\ &= \sum_{n=0}^{\infty} \mathbf{E} [\mathbf{E} [F | X_0 = X_n] \mathbf{1}_{A \cap \{T=n\}}] \\ &= \mathbf{E} [\mathbf{E} [F | X_0 = X_T] \mathbf{1}_A] \end{aligned}$$

L'égalité est vraie par linéarité pour toutes les fonctions étagées et donc pour toutes les fonctions positives. ■

■ **Remarque 18** Soit la chaîne de Markov à deux états 0 et 1, et de matrice de transition $p_{0,0} = 0,9, p_{1,1} = 1$. Soit $T = \sup\{n \geq 1, X_n = 0\}$. Sous \mathbf{P}_0 , $T = Y - 1$ où Y suit une loi géométrique de paramètre 0,1. Il s'ensuit T est presque sûrement fini. Or, $\mathbf{P}_0(X_{T+1} = 1 | X_T = 0) = 1$, qui est différent de $\mathbf{P}_0(X_{n+1} = 1 | X_n = 0) = 0,1$.

Cet exemple illustre que l'on ne peut pas supprimer l'hypothèse « T temps d'arrêt » dans la propriété de Markov forte. Il est clair qu'ici T n'est pas un temps d'arrêt car savoir si T est inférieur à n nécessite de connaître la trajectoire après l'instant n pour être sur que l'on ne repassera pas par 0 après n .

§ 4 Classification des états

On note N_x le nombre de visites à l'état x après le départ :

$$N_x = \sum_{n=1}^{\infty} \mathbf{1}_{\{X_n=x\}}.$$

Lemme 5.9 Pour tout k , les deux événements $\{N_x \geq k\}$ et $\{\tau_x^k < \infty\}$ coïncident.

Démonstration. $N_x \geq k$ signifie qu'il y a eu plus de k visites à l'état x , ce qui est très exactement équivalent à dire que $\tau_x^k < \infty$. ■

Définition 5.10 Un état x est dit récurrent lorsque $\mathbf{P}_x(\tau_x^1 < \infty) = 1$. Sinon, x est dit transient. La chaîne X est dite récurrente (respectivement transiente) si tous ses états sont récurrents (respectivement transients).

Lemme 5.11 Pour tout couple (x, y) de E , on a :

$$\mathbf{P}_y(\tau_x^k < \infty) = \mathbf{P}_x(\tau_x^1 < \infty)^{k-1} \mathbf{P}_y(\tau_x^1 < \infty). \quad (5.8)$$

En particulier, si $x = y$, $\mathbf{P}_x(\tau_x^k < \infty) = \mathbf{P}_x(\tau_x^1 < \infty)^k$. Par ailleurs, on a :

$$\mathbf{E}_y[N_x] = \frac{\mathbf{P}_y(\tau_x^1 < \infty)}{1 - \mathbf{P}_x(\tau_x^1 < \infty)} = \sum_{n \geq 1} p^{(n)}(y, x). \quad (5.9)$$

☞ Prenant comme instant présent celui de la k^e visite à l'état x , d'après la propriété de Markov forte, le passé et le futur conditionnellement à cette visite sont indépendants. Par conséquent, sachant que l'on a déjà visité k fois l'état x , la probabilité que l'on revienne en x une $(k+1)^e$ fois est la même que lors de la première visite en x on revienne au moins une fois. De plus, ces deux événements sont indépendants.

Démonstration. Pour $k > 2$, d'après (5.6) et (5.7), on a :

$$\begin{aligned} \mathbf{P}_y(\tau_x^k < \infty) &= \mathbf{P}_y(\tau_x^{k-1} < \infty, \tau_x^1 \circ \theta^{\tau_x^{k-1}} < \infty) \\ &= \mathbf{E}_y \left[\mathbf{1}_{\{\tau_x^{k-1} < \infty\}} \mathbf{P}_y(\tau_x^1 \circ \theta^{\tau_x^{k-1}} < \infty \mid \mathcal{F}_{\tau_x^{k-1}}) \right] \\ &= \mathbf{E}_y \left[\mathbf{1}_{\{\tau_x^{k-1} < \infty\}} \mathbf{P}_y(\tau_x^1 < \infty \mid X_0 = X_{\tau_x^{k-1}}) \right] \\ &= \mathbf{E}_y \left[\mathbf{1}_{\{\tau_x^{k-1} < \infty\}} \mathbf{P}_y(\tau_x^1 < \infty \mid X_0 = x) \right] \\ &= \mathbf{E}_y \left[\mathbf{1}_{\{\tau_x^{k-1} < \infty\}} \mathbf{P}_y(\tau_x^1 < \infty) \right] \\ &= \mathbf{P}_y(\tau_x^{k-1} < \infty) \mathbf{P}_y(\tau_x^1 < \infty), \end{aligned}$$

et on retrouve (5.8) par récurrence.

Maintenant, d'après le théorème de Fubini, on peut écrire :

$$\mathbf{E}_y [N_x] = \sum_{k \geq 1} \mathbf{P}_y(N_x \geq k) = \sum_{k \geq 1} \mathbf{P}_y(\tau_x^k < \infty).$$

La première égalité de (5.9) en découle. Toujours d'après le théorème de Fubini et (5.9), on a :

$$\mathbf{E}_y [N_x] = \mathbf{E}_y \left[\sum_{n=1}^{\infty} \mathbf{1}_{\{X_n=x\}} \right] = \sum_{n=1}^{\infty} \mathbf{E}_y \left[\mathbf{1}_{\{X_n=x\}} \right] = \sum_{n=1}^{\infty} p^{(n)}(y, x),$$

d'où le résultat. ■

Le théorème suivant permet de donner les différentes caractérisations de la récurrence et de la transience.

Théorème 5.12 Soit x un état fixé.

1. Les assertions suivantes sont équivalentes :

- (a) x est récurrent,
- (b) $\mathbf{P}_x(N_x = \infty) = 1$,
- (c) $\mathbf{E}_x [N_x] = \infty$.

2. Les assertions suivantes sont équivalentes :

- (a) x est transient,
- (b) $\mathbf{P}_x(N_x < \infty) = 1$,
- (c) $\mathbf{E}_x [N_x] < \infty$.

Démonstration. Montrons d'abord que $a \Rightarrow b$. D'après (5.9) et le lemme [5.11] :

$$\mathbf{P}_x(N_x > k) = \mathbf{P}_x(\tau_x^k < \infty) = \mathbf{P}_x(\tau_x < \infty)^k, \quad (5.10)$$

et d'après le théorème de convergence monotone :

$$\mathbf{P}_x(N_x = \infty) = \lim_{k \rightarrow \infty} \mathbf{P}_x(N_x > k). \quad (5.11)$$

La récurrence de x signifie $\mathbf{P}_x(\tau_x < \infty)$ et implique donc $N_x = \infty$, \mathbf{P}_x presque sûrement. Par conséquent, x récurrent implique que $\mathbf{P}_x(\tau_x^1 < \infty) = 1$. Par le même raisonnement, x transient implique $\mathbf{P}_x(\tau_x^1 < \infty) < 1$.

$b \Rightarrow c$. Immédiat dans le cas de x récurrent. Pour l'autre cas, utilisons la relation :

$$\mathbf{E}_x [N_x] = \sum_{k \geq 0} \mathbf{P}_x(\tau_x < \infty)^k. \quad (5.12)$$

Comme N_x est fini presque sûrement, d'après (5.11), $\mathbf{P}_x(N_x > k)$ tend vers 0 quand k tend vers l'infini. D'après (5.10) ceci implique que $\mathbf{P}_x(\tau_x < \infty) < 1$ donc que la série converge.

$c \Rightarrow a$. Dans les deux cas, la relation (5.12) permet de conclure. ■

Définition 5.13 On dit qu'un état x conduit à un état y et on le note $x \longrightarrow y$, s'il existe un entier strictement positif m tel que $p^{(m)}(x, y) > 0$. Ce qui revient à dire que $\mathbf{P}_x(\tau_y^1 < \infty) = 1$.

Théorème 5.14 Si x est un état récurrent et $x \longrightarrow y$, alors $y \longrightarrow x$ et y est récurrent.

▮ Partant de x on sait que l'on va presque sûrement en y , si de y il y a un risque de ne pas revenir en x on va finir par effectivement ne pas y revenir; on ne fera donc qu'un nombre fini de visites à x ce qui est incompatible avec l'hypothèse de récurrence. De plus, si de y on est presque sûr de revenir en x et que l'on passe une infinité de fois par x on passera vraisemblablement une infinité de fois en y aussi.

Démonstration. Montrons par l'absurde que y conduit à x en écrivant que la probabilité de ne jamais revenir en x en étant parti de x est supérieure à la probabilité de la même chose mais en passant une fois par y :

$$\begin{aligned} \mathbf{P}_x(\tau_x = \infty) &\geq \mathbf{P}_x(\tau_x \circ \theta^{\tau_y} = \infty, \tau_y < \infty) \\ &= \mathbf{P}_x(\tau_y < \infty) \mathbf{P}_y(\tau_x = \infty), \end{aligned}$$

d'après la propriété de Markov forte. Si y ne conduit pas à x , cette quantité est strictement positive ce qui est en contradiction avec la récurrence de x . De même, on a :

$$\begin{aligned} \mathbf{P}_y(\tau_y < \infty) &\geq \mathbf{P}_y(\tau_y \circ \theta^{\tau_x} < \infty, \tau_x < \infty) \\ &= \mathbf{P}_y(\tau_x < \infty) \mathbf{P}_x(\tau_y < \infty) = 1, \end{aligned}$$

donc y est récurrent. ■

Théorème 5.15 La relation \longrightarrow restreinte aux états récurrents est une relation d'équivalence.

Démonstration. La réflexivité, c'est-à-dire $x \longrightarrow x$, est induite par la définition même d'un état récurrent. La symétrie, c'est-à-dire $x \longrightarrow y \implies y \longrightarrow x$, découle du théorème [5.14]. Soit x, y et z trois états de E tels que $x \longrightarrow y$ et $y \longrightarrow z$. Par définition, il existe deux entiers positifs que nous appelons r et s tels que $p^{(r)}(x, y) > 0$ et $p^{(s)}(y, z) > 0$. L'équation de Chapman-Kolmogorov implique que :

$$p^{(r+s)}(x, z) = \sum_{\ell \in E} p^{(r)}(x, \ell) p^{(s)}(\ell, z).$$

Tous les termes de cette somme sont positifs et il existe au moins un terme strictement positif : $p^{(r)}(x, y) p^{(s)}(y, z)$. Nous avons donc trouvé un entier positif, $r + s$, tel que $p^{(r+s)}(x, z) > 0$, d'où le résultat. ■

L'ensemble des points récurrents peut donc être partitionné en classes d'équivalence. Par définition, un état appartenant à une classe conduit à tous les autres états de cette classe et ne conduit à aucun état récurrent appartenant à une autre classe ni à un état transient. En revanche, un état transient peut conduire aussi bien à un état transient qu'à un état récurrent.

Définition 5.16 Un sous-ensemble F de E est dit fermé, si pour tout x et y :

$$(x \in F \text{ et } x \longrightarrow y) \implies y \in F.$$

Autrement dit, $\sum_{y \in F} p(x, y) = 1$ pour tout $x \in F$.

Théorème 5.17 Tout ensemble fermé de cardinal fini contient au moins un point récurrent.

Démonstration. Soit F un ensemble fermé, si tous ses états sont transients, on a :

$$\mathbf{E}_y [N_x] = \mathbf{P}_y(\tau_x^1 < \infty) \mathbf{E}_x [N_x] < \infty$$

pour tout couple (x, y) de F . Comme F est de cardinal fini, $\sum_{x \in F} \mathbf{E}_y [N_x] < \infty$. Or, on a la suite d'identités :

$$\begin{aligned} \sum_{x \in F} \mathbf{E}_y [N_x] &= \sum_{x \in F} \mathbf{E}_y \left[\sum_{n \geq 0} \mathbf{1}_{\{X_n = x\}} \right] = \sum_{n \geq 0} \mathbf{E}_y \left[\sum_{x \in F} \mathbf{1}_{\{X_n = x\}} \right] \\ &= \sum_{n \geq 0} 1 = \infty, \end{aligned}$$

puisque F est fermé. On a abouti à une absurdité, il existe donc au moins un point récurrent. ■

■ **Exemple 5.4** Il est souvent simple d'avoir une représentation graphique de la matrice de transition d'une chaîne de Markov. Pour ce faire, on construit un graphe orienté dont les sommets correspondent aux états. L'arête (orientée) x, y a comme poids la probabilité de transition de x vers y . Si cette probabilité est nulle, l'arête n'existe pas. Considérons la chaîne de Markov de matrice de transition :

$$P = \begin{pmatrix} 0.8 & 0.1 & 0 & 0.1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & \frac{1}{1000} & \frac{999}{1000} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \end{pmatrix}.$$

La représentation graphique est alors celle de la figure 5.2.

Il est clair que les ensembles $\{2, 3\}$ et $\{4, 5, 6\}$ sont des ensembles fermés. Comme ils sont tous deux de cardinal fini, ils contiennent tous

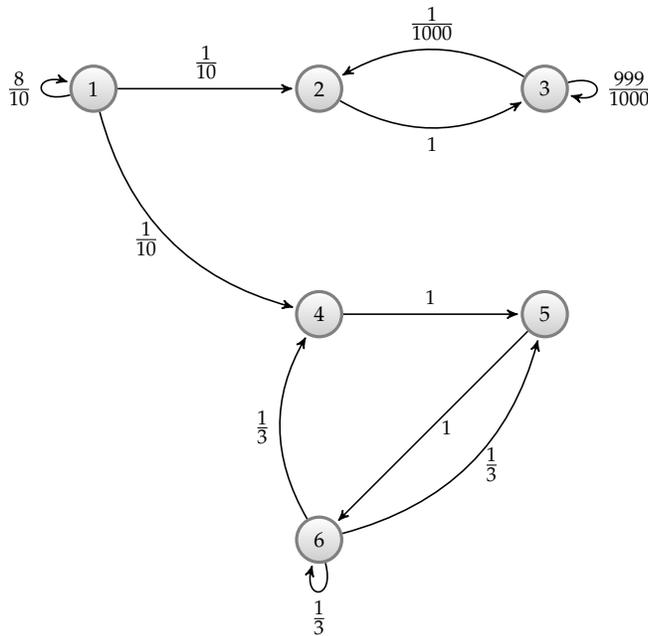


FIGURE 5.2: Représentation graphique d'une chaîne de Markov

deux au moins un point récurrent. A l'intérieur de chacun d'entre eux, les points conduisent tous les uns aux autres, donc ils sont tous récurrents. Si l'on quitte le point 1 pour le point 2 ou 4, on n'est sûr de ne pas y revenir donc la probabilité de ne pas revenir en 1 est $2 * 1/10 > 0$ donc 1 est transient.

Remarquons que même si l'ensemble E est fermé, le théorème [5.17] n'induit pas de contradiction quant au statut de l'état 1. En effet, on sait qu'il existe au moins un point récurrent dans E mais on ne sait pas lequel et on ne peut, *a priori*, rien dire de plus puisque tous les états ne communiquent pas entre eux.

■ **Exemple 5.5 — Suite de l'exemple 5.1.** Tous les états conduisent les uns aux autres donc le seul sous-ensemble fermé est E lui-même. Comme il est de cardinal fini, il existe au moins un état récurrent donc ils sont tous récurrents.

■ **Exemple 5.6 — Suite de l'exemple 5.2.** Tant qu'il n'atteint pas δ , le score ne peut que croître donc tous les états de \mathbb{N} sont transients. δ est lui récurrent.

■ **Exemple 5.7** Un jeu de N cartes est mélangé en le coupant en deux parts qui sont ensuite interverties. Chaque mélange du jeu est représenté par une permutation de $\{1, \dots, N\}$. Si $N = 3$ et que le mélange est représenté par $(3, 2, 1)$, cela signifie que la carte 3 est en position 1, la carte 2 en position 2 et la carte 1 en position 3. On note X_n l'état

du paquet de cartes après la n^e opération de mélange. L'espace d'états est donc le groupe des permutations de $\{1, \dots, N\}$ dans lui-même, noté \mathfrak{S}_N . Si $X_0 = (3, 2, 1)$ et que la coupe se fait entre la première et la deuxième carte, on a $X_1 = (2, 1, 3)$. En d'autres termes, on a juste fait une permutation circulaire sur les cartes mais on n'a pas changé leur ordre relatif. Pour définir les probabilités de transition, considérons l'ensemble à N éléments :

$$E_1 = \left\{ \sigma \in \mathfrak{S}_N, \exists k \in \{1, \dots, N\}, \sigma = (k+1, k+2, \dots, N, 1, \dots, k) \right\}.$$

Lorsque l'on coupe le paquet au niveau de la k^e carte, on applique le cycle $(k+1, k+2, \dots, N, 1, \dots, k)$ à la situation courante. Comme le choix de l'endroit de la coupe est supposé être uniforme sur $\{1, \dots, N\}$, on a :

$$\mathbf{P}(X_1 = \tau \mid X_0 = \sigma) = \frac{1}{N} \text{ si } \tau\sigma^{-1} \in E_1.$$

Les classes d'équivalence de la relation \longrightarrow sont celles de la relation $\sigma \mathfrak{R} \tau \equiv \tau\sigma^{-1} \in E_1$. En d'autres termes, σ conduit à τ si et seulement s'il existe $\rho \in E_1$ tel que $\tau = \rho\sigma$. Il y a donc $(n-1)!$ classes d'équivalence de cardinal n chacune. Toutes ces classes forment des ensembles fermés de cardinal fini qui contiennent donc toutes au moins un point récurrent. Comme à l'intérieur de ces classes les états communiquent tous entre eux, ils sont tous récurrents. La chaîne est donc récurrente. ■

Lorsque l'espace d'état est infini, on ne peut pas appliquer le théorème [5.17]. On introduit alors la notion suivante.

Définition 5.18 Une chaîne de Markov est dite irréductible lorsque tous les états conduisent les uns aux autres. En particulier, le plus petit sous-espace fermé est E lui-même et tous les états ont même nature.

■ **Remarque 19** Si le nombre d'états transients est fini, comme l'on ne passe qu'un nombre fini de fois en chacun d'eux, la chaîne de Markov sera inexorablement contrainte à aller dans une classe de récurrence et à y rester. Remarquons qu'en vertu du lemme [5.14], une classe de récurrence est forcément un sous-ensemble irréductible. Si le nombre d'états transients est infini, le raisonnement précédent ne s'applique plus automatiquement mais les cas dans lesquels on n'atterrit pas automatiquement dans un sous-ensemble fermé irréductible sont hors de notre propos. Pour ce qui nous intéresse (le comportement asymptotique des chaînes de Markov), il n'y a donc pas de perte de généralité à supposer que les chaînes de Markov étudiées sont irréductibles.

Lorsque x est récurrent, on sait que partant de x on reviendra nécessairement en x en un temps fini mais *quid* du temps moyen de retour en x ?

Définition 5.19 Un état récurrent x est dit récurrent positif, si $\mathbf{E}_x [\tau_x^1] < \infty$; récurrent nul, si $\mathbf{E}_x [\tau_x^1] = \infty$.

La chaîne X est alors dite récurrente positive (respectivement récurrente nulle) si tous ses états sont récurrents positifs (respectivement récurrents nuls).

La construction suivante est utilisée plusieurs fois par la suite.

Définition 5.20 Soit X , une chaîne de Markov irréductible et récurrente sur E et F , un sous-ensemble de E . On note :

$$\tau_F^1 = \inf \{ n \geq 1, X_n \in F \} \text{ et } \tau_F^{k+1} = \tau_F^k + \tau_F^1 \circ \theta^{\tau_F^k},$$

les instants de visite successifs de la chaîne à l'ensemble F . On considère la suite aléatoire X^F , définie par $X_n^F = X_{\tau_F^n}$, $n \in \mathbf{N}$. On vérifie facilement que X^F est une chaîne de Markov sur F , appelée chaîne de Markov restreinte à F .

Théorème 5.21 Soit X une chaîne de Markov irréductible et F un sous-ensemble fini de E . Si pour tout $x \in F$, $\mathbf{E}_x [\tau_F^1] < \infty$ alors X est récurrente positive.

Démonstration. Soit pour tout $x \in F$, $\sigma_x = \inf \{ n > 0, X_n^F = x \}$ et pour tout $k \in \mathbf{N}^*$, $Y_k = \tau_F^k - \tau_F^{k-1}$. Comme F est fini, X^F est récurrente positive donc $\mathbf{E}_x [\sigma_x] < \infty$ pour tout $x \in F$. Il nous faut prouver que $\mathbf{E}_x [\tau_x] < \infty$. Par construction des variables Y_k , on a :

$$\begin{aligned} \mathbf{E}_x [\tau_x] &= \mathbf{E}_x \left[\sum_{k=1}^{\sigma_x} Y_k \right] \\ &= \sum_{n \geq 1} \mathbf{E}_x \left[\sum_{k=1}^{\sigma_x} Y_k \mathbf{1}_{\{\sigma_x=n\}} \right] \\ &= \sum_{k=1}^{\infty} \mathbf{E}_x \left[Y_k \sum_{n \geq k} \mathbf{1}_{\{\sigma_x=n\}} \right] \\ &= \sum_{k=1}^{\infty} \mathbf{E}_x \left[Y_k \mathbf{1}_{\{\sigma_x \geq k\}} \right]. \end{aligned}$$

En utilisant la propriété de Markov forte, on obtient :

$$\begin{aligned} \mathbf{E}_x \left[Y_k \mathbf{1}_{\{\sigma_x \geq k\}} \right] &= \mathbf{E}_x \left[\mathbf{E}_x \left[Y_k \mid \mathcal{F}_{\tau_{k-1}^F} \right] \mathbf{1}_{\{\sigma_x \geq k\}} \right] \\ &= \mathbf{E}_x \left[\mathbf{E}_{X_{\tau_{k-1}^F}} [Y_1] \mathbf{1}_{\{\sigma_x \geq k\}} \right] \\ &\leq \sup_{y \in F} \mathbf{E}_y [Y_1] \mathbf{P}_x(\sigma_x \geq k). \end{aligned}$$

Le *supremum* est fini par hypothèse puisque F est fini. On obtient donc :

$$\mathbf{E}_x [\tau_x] \leq c \sum_{k=1}^{\infty} \mathbf{P}_x(\sigma_x \geq k) = c \mathbf{E}_x [\sigma_x].$$

En vertu de la remarque initiale, cela prouve la positive récurrence de X . ■

Lemme 5.22 Soit X une chaîne de Markov et $h : E \times E \rightarrow \mathbf{R}$, bornée. Pour tout entier n , on a :

$$\mathbf{E}[h(X_n, X_{n+1}) | \mathcal{F}_n] = P(h(X_n, \cdot))(X_n) = \sum_{y \in E} p(X_n, y) h(X_n, y). \quad (5.13)$$

Démonstration. Comme h est bornée, seul reste le calcul de l'espérance conditionnelle. D'après la propriété de Markov :

$$\mathbf{E}[h(X_n, X_{n+1}) | \mathcal{F}_n] = \mathbf{E}[h(X_n, X_{n+1}) | X_n].$$

Soit maintenant $\phi : E \rightarrow \mathbf{R}$ bornée. On a :

$$\begin{aligned} \mathbf{E}[h(X_n, X_{n+1})\phi(X_n)] &= \int \phi(x) \int h(x, y) d\mathbf{P}_{X_{n+1} | X_n=x}(y) d\mathbf{P}_{X_n}(x) \\ &= \int \phi(x) \sum_{y \in E} h(x, y)p(x, y) d\mathbf{P}_{X_n}(x) \\ &= \mathbf{E}[\phi(X_n)P(h(X_n, \cdot))(X_n)]. \end{aligned}$$

Comme la précédente équation est vraie pour toute fonction ϕ , on en déduit (5.13). ■

Théorème 5.23 — Critère de Foster. Soit E_0 une partie finie de E . S'il existe une fonction $h : E \rightarrow \mathbf{R}$ telle que l'ensemble $\{x \in E, h(x) < K\}$ soit fini pour tout entier K et que :

$$h(y) \geq \mathbf{E}_y[h(X_1)] \text{ pour tout } y \in E_0^c,$$

alors X est récurrente.

Démonstration. En particulier, h est minorée donc quitte à rajouter une constante, on peut supposer $h \geq 0$. Soit le temps d'arrêt $\tau = \inf\{n, X_n \in E_0\}$ et Y définie par $Y_n = h(X_n)\mathbf{1}_{\{\tau > n\}}$. Montrons que Y est une sur-martingale positive dès que $X_0 \in E_0^c$. Soit $x \in E_0^c$, on a :

$$\begin{aligned} \mathbf{E}_x [h(X_{n+1})\mathbf{1}_{\{\tau > n+1\}} | \mathcal{F}_n] &= \mathbf{1}_{\{\tau > n+1\}} \mathbf{E}_{X_n} [h(X_{n+1})] \\ &\leq \mathbf{1}_{\{\tau > n\}} h(X_n) = Y_n, \end{aligned}$$

car sur $(\tau > n + 1)$, X_n n'appartient pas à E_0 . Y converge donc presque sûrement vers une variable aléatoire Y_∞ .

Supposons que X soit transiente. Soit $x \notin E_0$, pour tout entier K , l'ensemble $\{x, h(x) < K\}$ est fini donc n'est visité qu'un nombre fini de fois par X donc X n'est pas bornée. Comme Y_∞ est finie, nécessairement τ est fini presque sûrement. Ce qui revient à dire que pour $x \notin E_0$, $\mathbf{P}_x(\tau < \infty) = 1$. Partant de E_0^c , on arrive forcément dans E_0 . Soit on reste dans E_0 pour toujours et comme E_0 est fini, E_0 est récurrent et par irréductibilité la chaîne l'est. Soit la chaîne quitte E_0 mais en vertu de ce que l'on vient de démontrer, elle y reviendra. Le nombre de visites à E_0 est donc infini, ce qui implique encore une fois que E_0 est récurrent donc la chaîne l'est aussi. ■

§ 5 Mesures et probabilité invariantes

Définition 5.24 Soit E un ensemble dénombrable et P un opérateur de transition sur $E \times E$. Une mesure positive finie ν sur E est dite invariante par rapport à P , si et seulement si :

$$\nu = \nu P \text{ c'est-à-dire } \nu(y) = \sum_{x \in E} \nu(x)p(x, y) \text{ pour tout } y \in E. \quad (5.14)$$

Si de plus $\sum \nu(x) = 1$, ν est une probabilité invariante.

■ **Remarque 20** Si $\pi_0 = \nu$ alors $\pi_n = \pi_0 P^n = \pi_0$.

Théorème 5.25 Soit x un état récurrent, alors la mesure ν définie par :

$$\nu(y) = \mathbf{E}_x \left[\sum_{n=0}^{\tau_x^1 - 1} \mathbf{1}_{\{X_n=y\}} \right] = \sum_{n=0}^{\infty} \mathbf{P}_x(X_n = y, \tau_x^1 > n)$$

est une mesure invariante.

Démonstration. Montrons d'abord l'égalité des deux expressions de ν . Puisque x est récurrent, τ_x^1 est presque sûrement fini donc $\cup_{n \geq 1} \{\tau_x^1 = n\}$ est une partition de Ω . En déconditionnant sur toutes les valeurs possibles de τ_x^1 et en utilisant le théorème de Fubini, on obtient :

$$\begin{aligned} \mathbf{E}_x \left[\sum_{n=0}^{\tau_x^1 - 1} \mathbf{1}_{\{X_n=y\}} \right] &= \sum_{\ell=1}^{\infty} \mathbf{E}_x \left[\sum_{n=0}^{\ell-1} \mathbf{1}_{\{X_n=y\}} \mathbf{1}_{\{\tau_x^1=\ell\}} \right] \\ &= \sum_{n=0}^{\infty} \mathbf{E}_x \left[\sum_{\ell > n} \mathbf{1}_{\{\tau_x^1=\ell\}} \mathbf{1}_{\{X_n=y\}} \right] \\ &= \sum_{n=0}^{\infty} \mathbf{E}_x \left[\mathbf{1}_{\{\tau_x^1 > n\}} \mathbf{1}_{\{X_n=y\}} \right]. \end{aligned}$$

Comme sous \mathbf{P}_x , $X_0 = X_{\tau_x^1} = x$, on peut aussi écrire $\nu(y) = \mathbf{E}_x \left[\sum_{n=1}^{\tau_x^1} \mathbf{1}_{\{X_n=y\}} \right]$,

ce qui donne par les mêmes calculs avec des bornes différentes :

$$v(y) = \sum_{n=1}^{\infty} \mathbf{E}_x \left[\mathbf{1}_{\{\tau_x^1 \geq n\}} \mathbf{1}_{\{X_n=y\}} \right]. \quad (5.15)$$

Remarquons d'ores et déjà que l'événement $\{\tau_x^1 \geq n\}$ appartient à \mathcal{F}_{n-1} puisque c'est le complémentaire de l'événement $\{\tau_x^1 \leq n-1\}$. Pour $y \neq x$, en appliquant les propriétés de l'espérance conditionnelle et la propriété de Markov forte :

$$\begin{aligned} & \sum_{n=0}^{\infty} \mathbf{E}_x \left[\mathbf{1}_{\{\tau_x^1 \geq n\}} \mathbf{1}_{\{X_n=y\}} \right] \\ &= \sum_{n=1}^{\infty} \mathbf{E}_x \left[\mathbf{1}_{\{\tau_x^1 \geq n\}} \mathbf{E}_x \left[\mathbf{1}_{\{X_n=y\}} \mid \mathcal{F}_{n-1} \right] \right] \\ &= \sum_{n=1}^{\infty} \mathbf{E}_x \left[\mathbf{1}_{\{\tau_x^1 \geq n\}} \mathbf{E}_x \left[\mathbf{1}_{\{X_n=y\}} \mid X_{n-1} \right] \right] \\ &= \sum_{n=1}^{\infty} \sum_{z \in E} \mathbf{E}_x \left[\mathbf{1}_{\{\tau_x^1 \geq n\}} \mathbf{1}_{\{X_{n-1}=z\}} \mathbf{E}_x \left[\mathbf{1}_{\{X_n=y\}} \mid X_{n-1}=z \right] \right] \\ &= \sum_{z \in E} p(z, y) \sum_{n=1}^{\infty} \mathbf{E}_x \left[\mathbf{1}_{\{\tau_x^1 \geq n\}} \mathbf{1}_{\{X_{n-1}=z\}} \right] \\ &= \sum_{z \in E} v(z) p(z, y). \end{aligned}$$

Pour $y = x$, il est clair que $v(x) = 1$ et d'autre part :

$$\begin{aligned} \sum_{z \in E} v(z) p(z, x) &= \sum_{n=0}^{\infty} \sum_{z \in E} p(z, x) \mathbf{P}_x(X_n = z, \tau_x^1 > n) \\ &= \sum_{n=0}^{\infty} \sum_{z \in E} \mathbf{P}_x(X_n = z, X_{n+1} = x, \tau_x^1 > n) \\ &= \sum_{n=0}^{\infty} \sum_{z \in E} \mathbf{P}_x(X_n = z, \tau_x^1 = n+1) \\ &= \sum_{n=0}^{\infty} \mathbf{P}_x(\tau_x^1 = n+1) \\ &= \mathbf{P}_x(\tau_x^1 < \infty) = 1. \end{aligned}$$

On a donc bien $v = vP$, et il ne reste qu'à vérifier que $v(y) < \infty$ pour tout y . C'est vrai pour $x = y$. Pour $y \neq x$, de deux choses l'une : soit x ne conduit pas à y et alors $v(y) = 0$, soit x conduit à y et comme x est récurrent, d'après le théorème [5.14], y conduit à x , c'est-à-dire qu'il existe m tel que $p^{(m)}(y, x) > 0$. Comme v est invariante, $v.P^m = v$, ce qui implique que :

$$1 = v(x) = \sum_{z \in E} v(z) p^{(m)}(z, x) \geq v(y) p^{(m)}(y, x),$$

et donc $v(y) < \infty$. ■

Corollaire 5.26 Soit X une chaîne de Markov irréductible et récurrente de mesure invariante ν . Soit F un sous-ensemble de E et X^F la chaîne restreinte à F . Alors, X^F est irréductible et récurrente et admet, comme X , pour mesure invariante celle donnée par le théorème [5.25].

Démonstration. Les deux premiers points sont évidents. Pour $y \in F$, le nombre de visites à y de X^F et de X sont les mêmes donc X et X^F la même mesure invariante donnée par le théorème [5.25]. ■

Théorème 5.27 Si la chaîne de Markov X est irréductible et récurrente alors il existe une unique (à un coefficient près) mesure invariante ν telle que pour tout y , $0 < \nu(y) < \infty$. L'unicité à un coefficient multiplicatif près signifie que si ν et ν' sont deux telles mesures alors il existe $c > 0$ tel que $\nu(x) = c\nu'(x)$ pour tout $x \in E$.

Démonstration. Soit μ une mesure invariante et soit $a \in E$. Soit ν la mesure invariante construite dans le théorème [5.25] avec a comme point de départ. Par construction, $\nu(a) = 1$ donc pour toute mesure invariante μ , $\mu(a) = \nu(a)\mu(a)$. Par définition, pour $z \in E \setminus \{a\}$:

$$\mu(z) = \sum_{y \in E} \mu(y)p(y, z) = \mu(a)p(a, z) + \sum_{y \neq a} \mu(y)p(y, z).$$

En itérant cette relation, on obtient :

$$\mu(z) = \mu(a)p(a, z) + \mu(a) \sum_{y \neq a} p(a, y)p(y, z) + \sum_{i \neq a} \sum_{y \neq a} \mu(x)p(x, y)p(y, z),$$

ce qui peut se récrire de la façon suivante :

$$\begin{aligned} \mu(z) &= \mu(a)\mathbf{P}_a(X_1 = z) \\ &\quad + \mu(a)\mathbf{P}_a(X_1 \neq a, X_2 = z) + \mathbf{P}_\mu(X_0 \neq a, X_1 \neq a, X_2 = z). \end{aligned}$$

Par récurrence sur n , on montre alors que pour tout n :

$$\mu(z) = \mu(a) \sum_{m=1}^n \mathbf{P}_a(\tau_a^1 > m, X_m = z) + \mathbf{P}_\mu \left(\bigcap_{y=0}^n (X_y \neq a) \cap X_n = z \right).$$

La dernière probabilité est un terme positif et l'on reconnaît, quand n tend vers l'infini, dans la première somme la définition de ν donc :

$$\mu(z) \geq \mu(a)\nu(z) \text{ pour tout } z \in E.$$

D'autre part, puisque pour tout n , $\mu.P^n = \mu$, on a aussi :

$$\mu(a) = \sum_x \mu(x)p^{(n)}(x, a) \geq \mu(a) \sum_x \nu(x)p^{(n)}(x, a) = \mu(a)\nu(a) = \mu(a).$$

Par conséquent, l'inégalité intermédiaire est une égalité et comme $\mu(x) \geq \mu(a)v(x)$, on doit avoir $\mu(x) = \mu(a)v(x)$ chaque fois que n est tel que $p^{(n)}(x, a) > 0$. Étant donné que X est irréductible, un tel entier n existe toujours donc $\mu(x) = \mu(a)v(x)$ pour tout $x \in E$. ■

Théorème 5.28 S'il existe une probabilité invariante ν , alors tous les états vérifiant $\nu(y) > 0$ sont récurrents.

Démonstration. Comme $\nu = \nu P^n$, le théorème de Fubini implique que :

$$\sum_{x \in E} \nu(x) \sum_{n \geq 1} p^{(n)}(x, y) = \sum_{n \geq 1} \nu(y) = \infty \text{ si } \nu(y) > 0.$$

D'autre part, d'après le lemme [5.11] :

$$\sum_{n \geq 1} p^{(n)}(x, y) = \frac{\mathbf{P}_x(\tau_y^1 < \infty)}{1 - \mathbf{P}_y(\tau_y^1 < \infty)}.$$

Comme $\mathbf{P}_x(\tau_y^1 < \infty) \leq 1$, on a :

$$\infty \leq \sum_{x \in E} \nu(x) \cdot \frac{1}{1 - \mathbf{P}_y(\tau_y^1 < \infty)},$$

donc $\mathbf{P}_y(\tau_y^1 < \infty) = 1$ puisque ν est finie, ce qui signifie que y est récurrent. ■

Théorème 5.29 Si X est irréductible et admet ν comme probabilité invariante alors la relation suivante est satisfaite :

$$\nu(x) = \frac{1}{\mathbf{E}_x[\tau_x^1]}.$$

Démonstration. S'il existe x tel que $\nu(x) = 0$ alors comme pour tout n , on a :

$$\nu(x) = \sum_{y \in E} p^{(n)}(y, x)\nu(y),$$

cela signifie que pour tout n et tout y , le produit de $\nu(y)$ et de $p^{(n)}(y, x)$ est nul. Or la chaîne est irréductible donc pour tout y , il existe n_y tel que $p^{(n_y)}(y, x) > 0$, donc $\nu(y) = 0$. Mais alors ν n'est pas une probabilité donc pour tout $x \in E$, $\nu(x) > 0$. D'après le théorème précédent, tous les états sont donc récurrents. On sait donc que ν est à un coefficient c près donné par le théorème [5.25]. Ce coefficient vérifie $c \sum_{y \in E} \nu(y) = 1$, or l'on sait que :

$$\sum_{y \in E} \nu(y) = \sum_{y \in E} \mathbf{E}_x \left[\sum_{n=0}^{\tau_x^1 - 1} \mathbf{1}_{\{X_n=y\}} \right] = \mathbf{E}_x[\tau_x^1],$$

d'après le théorème de Fubini et pour x choisi de façon quelconque dans E . ■

Le théorème suivant résume les principaux résultats précédents.

Théorème 5.30 Si X est irréductible, les trois assertions suivantes sont équivalentes :

1. l'un des états est récurrent positif ;
2. il existe une probabilité invariante ;
3. tous les états sont récurrents positifs.

De plus, la probabilité invariante est donnée par :

$$\nu(y) = \frac{1}{\mathbf{E}_x [\tau_x^1]} \mathbf{E}_x \left[\sum_{n=0}^{\tau_x^1 - 1} \mathbf{1}_{\{X_n=y\}} \right].$$

Démonstration. 1) \Rightarrow 2). En combinant les théorèmes 5.25 et 5.29, on voit que :

$$\nu(y) = \frac{1}{\mathbf{E}_x [\tau_x^1]} \mathbf{E}_x \left[\sum_{n=0}^{\tau_x^1 - 1} \mathbf{1}_{\{X_n=y\}} \right]$$

définit une probabilité invariante. Comme le terme de gauche ne dépend pas de x , on peut choisir $x = y$ et on retrouve bien $\nu(y) = \mathbf{E}_y [\tau_y^1]^{-1}$.

2) \Rightarrow 3). Puisque X est irréductible, on sait que la probabilité invariante est un multiple de celle construite dans le théorème [5.25] et donc que $\pi(y) > 0$ pour tout $y \in E$. D'après le théorème [5.29], cela signifie que tous les états sont récurrents positifs.

3) \Rightarrow 1) est trivial. ■

Corollaire 5.31 Toute chaîne de Markov irréductible sur E de cardinal fini est récurrente positive.

Démonstration. Il existe une mesure invariante μ . Comme l'espace d'états est fini, on peut toujours la normaliser en posant :

$$\nu(x) = \frac{1}{\sum_{y \in E} \mu(y)} \mu(y),$$

et l'on obtient une probabilité invariante. D'après le point 2) du théorème précédent, on en déduit qu'elle est récurrente positive. ■

Quand l'espace d'états est de cardinal infini, on peut utiliser le théorème suivant qui se démontre avec les mêmes outils que le théorème [5.23].

Théorème 5.32 — Critère de Foster. Supposons qu'il existe $h : E \rightarrow \mathbf{R}$ et $\epsilon > 0$ tels que :

— $\liminf_y h(y) > -\infty$;

- $h(X_1)$ est intégrable ;
- pour tout $y \in E_0^c$, on ait :

$$h(y) - \epsilon \geq \mathbf{E} [h(X_1) \mid X_0 = y].$$

Dans ces conditions, X est récurrente positive.

Soit X une chaîne de Markov irréductible récurrente sur E polonais. Pour tout $x \in E$, on pose $(Y_k^n, k \geq 0) = (X_{k \wedge \tau_x^1})$ et pour tout $n \geq 1$, on définit la n^{e} excursion Y^n de X depuis x par $(Y_k^n, k \geq 0)$, où :

$$Y_k^n = X_{k \wedge \tau_x^1} \circ \theta^{\tau_x^n}.$$

⚡ La 0^{e} excursion coïncide avec X jusqu'à la première visite à x , après Y^0 reste à x . La n^{e} excursion est une chaîne de Markov qui part de x et suit le comportement de la chaîne initiale jusqu'au temps d'atteinte de x suivant. Ensuite, elle reste constamment égale à x . L'évolution de la chaîne est reflétée par Y^{n+1} .

D'après la propriété de Markov forte, les processus $(Y^n, n \geq 0)$ sont indépendants les uns des autres et à partir de $n = 1$, ils ont tous la même loi : pour toute fonction $\psi : E^{\mathbf{N}} \rightarrow \mathbf{R}$:

$$\mathbf{E} [\psi(Y^n)] = \mathbf{E} [\psi(Y^1 \circ \theta^{\tau_x^n})] = \mathbf{E} [\psi(Y^1)].$$

Théorème 5.33 Soit X récurrente, irréductible de loi invariante ν . Quelle que soit la condition initiale $x \in E$, pour toute fonction f dans $L^1(\nu)$, pour toute fonction $g \geq 0$ telle que $\sum_y g(y)\nu(y) > 0$, on a :

$$\frac{\sum_{j=0}^n f(X_j)}{\sum_{j=0}^n g(X_j)} \xrightarrow{n \rightarrow \infty} \frac{\sum_{y \in E} f(y)\nu(y)}{\sum_{y \in E} g(y)\nu(y)}, \mathbf{P}_x \text{ presque sûrement.}$$

En particulier, pour $f \in L^1(\nu)$, on a :

$$\frac{1}{n} \sum_{j=0}^n f(X_j) \xrightarrow{n \rightarrow \infty} \sum_{y \in E} f(y)\nu(y), \mathbf{P}_x \text{ presque sûrement.}$$

⚡ On peut découper toute fonctionnelle additive en morceaux dépendant de chaque excursion. D'après l'indépendance et l'équidistribution de celles-ci, on peut appliquer la loi forte des grands nombres. Il reste à prouver que les termes de « bord », c'est-à-dire qui dépendent de Y^0 et ceux qui dépendent de l'excursion incomplète disparaissent dans la division par n .

Démonstration. La probabilité invariante est proportionnelle à la mesure invariante donnée dans le théorème [5.25]. En particulier, il existe $c > 0$

tel que pour toute fonction $g \geq 0$, on ait :

$$c \sum_{y \in E} g(y) \nu(y) = \mathbf{E}_x \left[\sum_{n=0}^{\tau_x^1 - 1} g(X_n) \right].$$

Par homogénéité, on peut supposer que $c = 1$.

Soit $Z = (Z_k, k \geq 1)$ définie par :

$$Z_k = \sum_{n=\tau_x^k}^{\tau_x^{k+1}-1} f(X_n) = \sum_{n=0}^{\tau_x^1-1} f(Y_n^k) = Z_1 \circ \theta^{\tau_x^k}.$$

D'après la propriété de Markov forte, les variables aléatoires $(Z_k, k \geq 1)$ sont indépendantes et identiquement distribuées. De plus, on a :

$$\mathbf{E}_x [|Z_1|] \leq \mathbf{E}_x \left[\sum_{n=0}^{\tau_x^1-1} |f(X_n)| \right] = \sum_{y \in E} |f(y)| \nu(y) < \infty,$$

puisque $f \in L^1(\nu)$. On peut donc appliquer la loi forte des grands nombres, qui stipule que :

$$\frac{1}{n} \sum_{k=0}^{\tau_x^n-1} f(X_k) = \frac{1}{n} \sum_{k=1}^n Z_k \xrightarrow{n \rightarrow \infty} \mathbf{E}_x [Z_1] = \sum_{y \in E} f(y) \nu(y), \quad \mathbf{P}_x \text{ p.s.} \quad (5.16)$$

Si l'on applique ce résultat à $f \equiv 1$, on obtient :

$$\frac{\tau_x^n}{n} \xrightarrow{n \rightarrow \infty} 1, \quad \mathbf{P}_y \text{ presque sûrement.} \quad (5.17)$$

Soit $e(n)$ le nombre de visites à x entre les instants 0 et n . Notons que $e(n)$ est aussi le nombre d'excursions partant de x complètes avant l'instant n . Par définition, $\tau_x^{e(n)} \leq n < \tau_x^{e(n)+1}$, donc :

$$\frac{\tau_x^{e(n)}}{e(n)} \leq \frac{n}{e(n)} < \frac{\tau_x^{e(n)+1}}{e(n)+1} = \frac{e(n)+1}{e(n)}.$$

En vertu de (5.17), les termes extrêmes de la ligne précédente tendent p.s. vers 1 donc $n^{-1}e(n)$ aussi. On écrit alors :

$$\begin{aligned} \frac{1}{n} \sum_{k=0}^n f(X_k) &= \\ \frac{1}{n} \sum_{k=0}^{\tau_x^1-1} f(X_k) + \frac{e(n)}{n} \frac{1}{e(n)} (Z_1 + \dots + Z_{e(n)}) + \frac{1}{n} \sum_{k=\tau_x^{e(n)}+1}^n f(X_k). \end{aligned} \quad (5.18)$$

Supposons que $f \geq 0$. Le premier terme tend vers 0 \mathbf{P}_y -presque sûrement pour tout $y \in E$. D'après la définition de la convergence presque sûre, cela équivaut à ce que l'on ait pour tout $\epsilon > 0$:

$$\mathbf{P}_y \left(\limsup_n \left(\frac{1}{n} \sum_{k=0}^{\tau_x^1-1} f(X_k) > \epsilon \right) \right) = 0. \quad (5.19)$$

Par conséquent, compte tenu de la propriété de Markov forte, pour tout $\epsilon > 0$, on a :

$$\begin{aligned} \mathbf{P}_y \left(\limsup_n \left(\frac{1}{n} \sum_{k=\tau_x^{\epsilon(n)}}^{\tau_x^{\epsilon(n+1)}-1} f(X_k) > \epsilon \right) \right) \\ = \mathbf{P}_x \left(\limsup_n \left(\frac{1}{n} \sum_{k=0}^{\tau_x^1-1} f(Y_k^n) > \epsilon \right) \right) \\ = \mathbf{P}_x \left(\limsup_n \left(\frac{1}{n} \sum_{k=0}^{\tau_x^1-1} f(X_k) > \epsilon \right) \right) = 0, \end{aligned}$$

en vertu de (5.19). Comme :

$$\sum_{k=\tau_x^{\epsilon(n)}+1}^n f(X_k) \leq \sum_{k=\tau_x^{\epsilon(n)}}^{\tau_x^{\epsilon(n+1)}-1} f(X_k),$$

on a bien :

$$\frac{1}{n} \sum_{k=\tau_x^{\epsilon(n)}}^n f(X_k) \xrightarrow{n \rightarrow \infty} 0, \mathbf{P}_y \text{ presque sûrement.}$$

Pour f de signe quelconque, en appliquant le raisonnement précédent à $|f|$, on montre que les premier et troisième termes de (5.18) tendent presque sûrement vers 0.

D'après la première partie de la preuve (voir (5.16)), le terme médian de (5.18) tend presque sûrement vers $\sum_{y \in E} f(y)\nu(y)$.

Le cas particulier s'obtient en prenant $g = 1$. ■

Définition 5.34 Un état x est dit périodique s'il existe un entier $\delta \geq 2$ tel que :

$$\sum_{k=1}^{\infty} \mathbf{P}_x(\tau_x^1 = \delta k) = 1. \quad (5.20)$$

Le plus petit δ vérifiant (5.20) est appelé la période de x et nous la noterons $d(x)$. Les états qui ne sont pas périodiques sont appelés apériodiques.

■ **Exemple 5.8 — Suite de l'exemple 5.1.** Dans ce cas, un peu de réflexion montre que l'on n'atteindra une case de numéro impair que tous les deux pas, de même pour les cases de numéro pair : si le rat part de 1 il ne peut être en case 3, 5 ou 7 au coup d'après. La période est donc 2. On voit que l'on peut classer les états en deux paquets, les cases $\{1, 3, 5, 7\}$ d'une part, les cases $\{2, 4, 6\}$ d'autre part. Le rat sera tous les pas pairs dans le sous-ensemble de départ et tous les pas impairs dans l'autre sous-ensemble. ■

Plus généralement, on a le théorème suivant.

Théorème 5.35 Soit X une chaîne de Markov récurrente irréductible de période d . Soit x fixé dans E , il existe une partition de E en d ensembles C_0, C_1, \dots, C_{d-1} tels que :

- x appartient à C_0 ;
- soit $y \in C_r$ et $z \in C_s$, si $p^{(n)}(y, z) > 0$ alors $n = (s - r) \pmod{d}$;
- C_0, \dots, C_{d-1} sont des classes de récurrence irréductibles aperiodiques pour la chaîne de Markov de matrice de transition P^d .

La décomposition est unique à une renumérotation près. La chaîne de Markov de matrice de transition P^d est irréductible, récurrente, aperiodique. Si sa condition initiale est dans C_r pour $r \in \llbracket 0, d - 1 \rrbracket$, alors tous ces éléments sont dans C_r .

La démonstration de ce théorème nécessite deux lemmes techniques.

Lemme 5.36 Soit a_1, \dots, a_n des nombres entiers premiers dans leur ensemble, tout entier $m \geq \prod_x (1 + a_x)$ peut s'écrire sous la forme :

$$m = \sum_k x_k a_k \text{ avec } x_k \geq 0 \text{ pour tout } k. \quad (5.21)$$

Démonstration. Montrons par récurrence sur n que si a_1, \dots, a_n sont n entiers (non nécessairement premiers entre eux) et que $m \in \mathbf{N}$ s'écrit $m = \sum_k x_k a_k$ alors on peut toujours trouver une autre écriture satisfaisant les conditions de (5.21). Plus précisément, il existe une permutation σ de $\{1, \dots, N\}$ dans lui-même telle que :

$$x_{\sigma(i)} \leq \min_{l \neq \sigma(1), \dots, \sigma(i-1)} (a_l) \text{ pour tout } l \leq n - 1.$$

Supposons d'abord que $n = 2$. Comme $m \geq 0$, l'un des deux coefficients x_1 ou x_2 est positif. A une renumérotation près, on peut toujours supposer qu'il s'agit de x_1 . Montrons que l'on peut toujours supposer que $x_1 < a_2$. Si ce n'est pas le cas, on écrit alors $x_1 = ka_2 + r$ avec $0 \leq r < a_2$:

$$\begin{aligned} m &= x_1 a_1 + x_2 a_2 + ka_1 a_2 - ka_1 a_2 \\ &= (x_1 - ka_2) a_1 + (x_2 + ka_1) a_2 = r a_1 + (x_2 + ka_1) a_2. \end{aligned}$$

En conclusion, tout entier m peut s'écrire sous la forme $m = x_1 a_1 + x_2 a_2$ avec $0 \leq x < b$. En particulier, si $m \geq a_1 a_2$, x_2 doit être positif.

Supposons le résultat démontré pour $(n - 1)$. A une renumérotation près, on peut toujours supposer que x_1 est positif et appliquer l'hypothèse de récurrence à $m - x_1 a_1$ et aux $(n - 1)$ nombres restants. La

renumérotation que l'on a appliquée au cours de cette opération définit la permutation σ .

Maintenant, si a_1, \dots, a_n sont premiers dans leur ensemble, le lemme de Bezout garantit l'existence de la représentation $m = \sum_k x_k a_k$ pour tout entier. D'après la première partie de la démonstration, on peut toujours supposer que $\sum_{k \leq n-1} x_k a_k$ est positive et inférieure à :

$$\sup_x \left(a_1 \dots a_n + \prod_{y \neq x} a_y + \dots \right) \leq a_1 \dots a_n + \sum_x \prod_{y \neq x} a_y + \dots = \prod (1 + a_x) - 1.$$

Par conséquent, pour m supérieur ou égal à $\prod (1 + a_x)$, il existe toujours une écriture de la forme (5.21). ■

Lemme 5.37 Si x est apériodique alors il existe n_0 tel que si $n \geq n_0$ alors $p^{(n)}(x, x) > 0$.

Démonstration. Définissons l'ensemble :

$$I_x = \left\{ n \in \mathbf{N}, p^{(n)}(x, x) > 0 \right\}.$$

D'après la propriété de Markov, I_x est un semi-groupe : si m et n appartiennent à I_x alors $m + n$ aussi. En effet, on a :

$$p^{(m+n)}(x, x) \geq p^{(m)}(x, x)p^{(n)}(x, x).$$

On ordonne I_x par l'ordre naturel. Soit u_n le nombre de diviseurs communs des n premiers éléments de I_x . $(u_n, n \in \mathbf{N})$ est une suite décroissante positive donc convergente, et puisque x est apériodique, cette limite est 1. Comme u_n est à valeurs entières, il y a nécessairement un rang à partir duquel elle est constante, soit n_0 ce rang et soit a_1, \dots, a_{n_0} les n_0 premiers éléments de I_x . D'après le lemme précédent, pour n assez grand, $n \in I_x$. ■

Démonstration du théorème 5.35. Soit $K_y = \left\{ n, p^{(n)}(x, y) > 0 \right\}$. Pour k et l deux entiers, d'après la propriété de Markov, on a l'identité suivante :

$$\mathbf{P}_x(X_{k+l} = x) \geq \mathbf{P}_x(X_k = y)\mathbf{P}_y(X_l = x).$$

Par conséquent, n ne peut appartenir à K_y que si d divise $n + l$, c'est-à-dire si n s'écrit $\alpha d + r$ où $r \in \{0, \dots, d-1\}$ est le reste de la division de l par d . On définit C_r comme l'ensemble des points de E qui ont le même r . Ces ensembles forment clairement une partition et en prenant $l = 0$, on prouve que $x \in C_0$.

Soit m et n tels que $p^{(m)}(y, z) > 0$ et $p^{(n)}(x, y) > 0$. Comme $p^{(n+m)}(x, z) > 0$, il en découle d'après i) que $n + m \equiv s \pmod{d}$ et comme $n \equiv r \pmod{d}$, le résultat s'ensuit.

L'irréductibilité découle immédiatement du point précédent, l'apériodicité de la définition de la période. ■

Nous pouvons maintenant énoncer le résultat.

Théorème 5.38 Soit X une chaîne de Markov irréductible, récurrente positive, apériodique, de matrice de transition P et de probabilité invariante ν . Alors, on a :

$$\lim_{n \rightarrow \infty} p^{(n)}(x, y) = \nu(y), \text{ pour tout } x \text{ et tout } y.$$

La démonstration se fait par couplage : on va montrer que deux chaînes de Markov indépendantes de même matrice de transition mais de conditions initiales différentes finissent toujours par se rencontrer. Notons qu'à partir de cet instant de rencontre, elles coïncident en loi.

Démonstration. Sur $E \times E$, on définit la chaîne de Markov $Z_n = (W_n, y_n)$ de matrice de transition :

$$\hat{p}((x_1, x_2), (y_1, y_2)) = p(x_1, y_1)p(x_2, y_2).$$

En d'autres termes, les deux coordonnées W et Y évoluent indépendamment l'une de l'autre selon la loi de la chaîne de Markov originelle.

On va montrer premièrement que Z est une chaîne de Markov irréductible. Comme tous les états sont apériodiques, d'après le lemme [5.37], à partir d'un certain rang M :

$$p^{(l)}(y_1, y_1) > 0 \text{ et } p^{(l)}(x_2, x_2) > 0.$$

Comme X est irréductible et récurrent, il existe $K \geq M$ et $L \geq M$ tels que :

$$p^{(K)}(x_1, x_2) > 0 \text{ et } p^{(L)}(y_1, y_2) > 0.$$

Par conséquent, le chemin :

$$(x_1, y_1) \rightarrow (x_2, y_1) \rightarrow (x_2, y_2)$$

est de probabilité strictement positive pour l'indice $K + L + M$. En effet, d'après la propriété de Markov :

$$\begin{aligned} p^{(K+L)}((x_1, y_1), (x_2, y_2)) \\ \geq p^{(K)}(x_1, x_2)p^{(K)}(y_1, y_1) \cdot p^{(L)}(x_2, x_2)p^{(L)}(y_1, y_2) > 0. \end{aligned}$$

Il est clair que $\hat{\nu}(x, y) = \nu(x)\nu(y)$ définit une probabilité invariante pour la chaîne de Markov Z . Par conséquent d'après le théorème [5.30], tous les états sont récurrents positifs. Soit T le temps d'atteinte de la diagonale de $E \times E$ par Z :

$$\begin{aligned} \Delta &= \{(x, y) \in E \times E, x = y\}; \\ T &= \inf \{n > 0, Z_n \in \Delta\}. \end{aligned}$$

Comme Z est irréductible, récurrente, le temps d'atteinte d'un état (x, x) de la diagonale est presque sûrement fini. Comme T est le minimum de tous ces temps, il est aussi presque sûrement fini. Montrons que sur $\{T \leq n\}$, W_n et Y_n ont même loi :

$$\begin{aligned} \mathbf{P}(W_n = y, T \leq n) &= \sum_x \mathbf{E} \left[\mathbf{1}_{\{W_n=y\}} \mathbf{1}_{\{W_T=x\}} \mathbf{1}_{\{T \leq n\}} \right] \\ &= \sum_x \mathbf{E} \left[\mathbf{1}_{\{W_T=x\}} \mathbf{1}_{\{T \leq n\}} \mathbf{E} \left[\mathbf{1}_{\{W_n=y\}} \mid \mathcal{F}_T \right] \right] \\ &= \sum_x \mathbf{E} \left[\mathbf{1}_{\{W_T=x\}} \mathbf{1}_{\{T \leq n\}} \mathbf{E}_x \left[\mathbf{1}_{\{W_{n-T}=y\}} \right] \right] \\ &= \sum_x \mathbf{E} \left[\mathbf{1}_{\{Y_T=x\}} \mathbf{1}_{\{T \leq n\}} \mathbf{E}_x \left[\mathbf{1}_{\{Y_{n-T}=y\}} \right] \right] \\ &= \mathbf{P}(Y_n = y, T \leq n). \end{aligned}$$

D'après ce qui précède :

$$\begin{aligned} \mathbf{P}(W_n = y) &= \mathbf{P}(W_n = y, T \leq n) + \mathbf{P}(W_n = y, T > n) \\ &= \mathbf{P}(Y_n = y, T \leq n) + \mathbf{P}(W_n = y, T > n) \\ &\leq \mathbf{P}(Y_n = y) + \mathbf{P}(W_n = y, T > n). \end{aligned}$$

Symétriquement, on a :

$$\mathbf{P}(Y_n = y) \leq \mathbf{P}(W_n = y) + \mathbf{P}(Y_n = y, T > n),$$

dont on déduit que :

$$|\mathbf{P}(W_n = y) - \mathbf{P}(Y_n = y)| \leq \mathbf{P}(Y_n = y, T > n) + \mathbf{P}(W_n = y, T > n).$$

D'où en sommant sur toutes les valeurs possibles de y , on obtient :

$$\sum_y |\mathbf{P}(Y_n = y) - \mathbf{P}(W_n = y)| \leq 2\mathbf{P}(T > n).$$

Comme T est presque sûrement fini, le terme de droite tend vers 0 quand n croît indéfiniment. Si l'on prend $W_0 = x$ et Y ayant la loi ν , on en déduit :

$$\sum_y |p^{(n)}(x, y) - \nu(y)| \xrightarrow{n \rightarrow \infty} 0,$$

ce qui implique le résultat annoncé. ■

■ **Remarque 21** On remarque que l'hypothèse d'apériodicité ne sert que dans la démonstration de l'irréductibilité de la chaîne de Markov Z . Pour se convaincre que cela est essentiel, regardons encore l'exemple du rat dans son labyrinthe. Formons la chaîne de Markov $Z_n = (X_n, Y_n)$ qui représente les positions de deux rats lâchés dans le même labyrinthe, qui évoluent indépendamment l'un de l'autre selon les mêmes règles que précédemment. Soit C_1 la classe cyclique de 1 et C_2 celle de 2 pour la chaîne de Markov X . Si Z part d'un état de $C_1 \times C_2$ alors Z évolue entre des états de cet ensemble et des états de $C_2 \times C_1$ mais n'atteint jamais d'états de $C_1 \times C_1$, par conséquent Z n'est pas irréductible.

Dans le cas périodique, on a toutefois le résultat suivant :

Théorème 5.39 Soit X une chaîne de Markov irréductible, récurrente positive, périodique de période d et de probabilité invariante ν . Soit $x \in E$ et C_0, \dots, C_{d-1} les classes cycliques associées à x . Si $y \in C_r$, on a :

$$\lim_{n \rightarrow \infty} p^{(nd+r)}(x, y) = d\nu(y).$$

✎ L'idée est d'appliquer le théorème précédent à la chaîne de matrice de transition P^d . Il faut pour cela déterminer la probabilité invariante de cette chaîne de Markov. On remarque d'après le théorème [5.25] qu'à un coefficient près la probabilité invariante d'un état y est égal à la proportion du nombre de visites à cet état entre deux visites à un état fixe x . Etant donné que dans la chaîne de Markov de matrice P^d on divise par d le nombre de pas, cette proportion est multiplié par d .

Démonstration. Par définition de la période et de C_k , C_k est un sous-ensemble fermé pour la chaîne X^k définie par $X_n^k = X_{nd+k}$ pour $k = 0, \dots, d-1$. Ces chaînes sont irréductibles et récurrentes positives. En vertu du corollaire [5.26], la probabilité invariante ν^k de X^k est proportionnelle à ν , c'est-à-dire qu'il existe α_k tel que $\nu^k(y) = \alpha_k \nu(y)$ pour tout $y \in C_k$. Par ailleurs, puisque ν est la probabilité invariante de X , pour tout k et tout l appartenant à $0, \dots, d-1$, on a :

$$\alpha_k = \mathbf{P}_\nu(X_{nd+k} \in C_k) = \mathbf{P}_\nu(X_{nd+k} \in C_k \cup C_l) = \mathbf{P}_\nu(X_{nd+l} \in C_l) = \alpha_l.$$

Il s'ensuit que $\alpha_k = d^{-1}$. Le dernier point découle du théorème [5.38]. ■

Enfin, le dernier résultat utile pour les simulations est le théorème central limite suivant.

Théorème 5.40 Soit X une chaîne de Markov récurrente positive de probabilité invariante ν . Pour $f : E \times E \rightarrow \mathbf{R}$, on note :

$$Pf : \begin{cases} E & \longrightarrow \mathbf{R} \\ x & \longmapsto Pf(x) = \sum_y f(x, y)p(x, y) = \mathbf{E}_x[f(X_0, x_1)]. \end{cases}$$

Pour toute fonction f telle que $\mathbf{E}_\nu[P(f^2)] < \infty$ on a la convergence en loi suivante :

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n (f(X_{k-1}, X_k) - Pf(X_{k-1})) \xrightarrow[n \rightarrow \infty]{\text{Loi}} \mathcal{N}(0, \sigma^2(f)),$$

où $\mathcal{N}(0, \sigma^2(f))$ représente une loi gaussienne centrée de variance

donnée par :

$$\sigma^2(f) = \mathbf{E}_\nu[P(f^2)] - \mathbf{E}_\nu[(Pf)^2].$$

Démonstration. Le lemme [5.13] implique que pour f bornée, la suite :

$$M_n^f = \sum_{j=0}^n f(X_j, X_{j+1}) - \sum_{j=0}^n Pf(X_j, \cdot)(X_j)$$

est une martingale. D'autre part, son processus croissant est défini par :

$$\begin{aligned} \Delta \langle M^f \rangle_n &= \mathbf{E} \left[(\Delta M_n^f)^2 \mid \mathcal{F}_n \right] \\ &= \mathbf{E} \left[\left(f(X_n, X_{n+1}) - Pf(X_n, \cdot) \right)^2 \mid \mathcal{F}_n \right] \\ &= Pf^2(X_n) + Pf(X_n)^2 - 2Pf(X_n)^2 \\ &= \Gamma f(X_n), \end{aligned}$$

où $\Gamma f = Pf^2 - (Pf)^2$ est l'opérateur carré du champ associé à P . Par conséquent, on a :

$$\langle M^f \rangle_n = \sum_{j=0}^n \Gamma f(X_j).$$

Par hypothèse, Γf est intégrable par rapport à ν , le théorème [5.33] implique que :

$$\frac{\langle M^f \rangle_n}{n} \xrightarrow{n \rightarrow \infty} \sigma^2(f), \mathbf{P}_x \text{ presque sûrement.}$$

Le résultat découle du théorème de la limite centrée pour les incréments de martingales. ■

Si l'on prend comme cas particulier, $f(X_{k-1}, X_k) = \mathbf{1}_{\{X_k=x\}}$, on obtient :

$$\mathbf{P}(\sqrt{n}(N_x^n - \pi(x)) \in [a, b]) \xrightarrow{n \rightarrow \infty} \int_a^b \exp(-x^2/2\sigma^2) \frac{dx}{\sigma\sqrt{2\pi}},$$

avec $\sigma^2 = \nu(y) - \sum_x p(x, y)^2 \nu(x)$.

■ **Exemple 5.9 — Suite de l'exemple 5.1.** C'est le cas le plus simple dans lequel on n'a qu'à résoudre le système $\nu = \nu P$ et $\sum \pi(x) = 1$. Tous calculs faits, on trouve :

$$\nu = \left(\frac{1}{8}, \frac{3}{16}, \frac{1}{8}, \frac{3}{16}, \frac{3}{16}, \frac{1}{8}, \frac{1}{16} \right).$$

■ **Exemple 5.10 — Suite de l'exemple 5.7.** Il faut restreindre la chaîne de Markov à une quelconque classe d'équivalence de la relation « communique ». Dans ce cas, il est clair que la probabilité invariante est la mesure uniforme sur ces états. ■

§ 6 Calcul pratique de la probabilité invariante

Le principe est simple : la probabilité invariante est l'unique vecteur à composantes positives ou nulles, de poids total 1 qui satisfait l'équation $\nu(P - \text{Id}) = 0$. Si l'on veut résoudre un tel système par ordinateur, il faut prendre garde au fait que ce système est de co-rang 1 : il faut donc supprimer une colonne de P (par exemple, la dernière) et la remplacer par une colonne composée uniquement de 1. Soit \hat{P} la matrice ainsi obtenue. Il nous faut alors résoudre le système :

$$\pi(\hat{P} - \hat{I}) = b, \text{ avec } b = (0, \dots, 0, 1) \text{ et } \hat{I} = \begin{pmatrix} 1 & & & 0 \\ & 1 & (0) & 0 \\ & (0) & \ddots & \\ & & & 0 \end{pmatrix}.$$

Pratiquement, les chaînes que l'on utilise ont un espace d'états fini mais de cardinal très grand (plusieurs milliers d'états). Se pose alors le problème de la détermination de la probabilité invariante. Il s'agit *a priori* « simplement » de résoudre un système linéaire mais la taille de ce système oblige à utiliser des méthodes d'analyse numérique.

Méthode itérative

On a donc à résoudre l'équation $\pi = \pi P$ où P est la matrice de transition. D'après le théorème [5.38], si la chaîne est apériodique alors $\pi_{n+1} = \pi_n P$ tend vers la probabilité invariante. Pratiquement, on prend un π_0 quelconque et on itère. Ce procédé peut s'avérer coûteux si le calcul des coefficients de P est long. Néanmoins, la convergence est extrêmement rapide puisque géométrique de raison le module de la deuxième plus grande valeur propre de P .

Dans le cas où la chaîne est périodique (voir l'exemple du rat) de période d , il faut être plus précautionneux. Le théorème [5.38] nous indique que la suite π_n a d valeurs d'adhérence. Précisément, par définition même des classes cycliques, si π_0 est une masse de Dirac en x , les termes π_{kn} ont des composantes positives seulement pour les états de la classe cyclique de x , les termes π_{kn+j} ont des composantes positives seulement pour les états de la classe cyclique C_j , pour tout $j \in \{1, \dots, d-1\}$.

■ **Exemple 5.11** — Suite de l'exemple 5.1. Si la condition initiale est $\pi_0 =$

$(1, 0, \dots)$ alors on a :

$$v(2) = \left(0, \frac{1}{2}, 0, \frac{1}{2}, 0, 0, 0\right)$$

$$v(3) = \left(\frac{1}{3}, 0, \frac{1}{6}, 0, \frac{1}{3}, 0, \frac{1}{6}\right)$$

$$v(4) = \left(0, \frac{13}{36}, 0, \frac{4}{9}, 0, \frac{7}{36}, 0\right)$$

$$v(5) = \left(\frac{29}{108}, 0, \frac{47}{216}, 0, \frac{79}{216}, 0, \frac{4}{27}\right)$$

$$v(6) = \left(0, \frac{473}{1296}, 0, \frac{131}{324}, 0, \frac{299}{1296}, 0\right)$$

$$v(7) = \left(\frac{997}{3888}, 0, \frac{1843}{7776}, 0, \frac{2891}{7776}, 0, \frac{131}{972}\right)$$

■

Un moyen pour éviter ce désagrément est de considérer les sommes de Césaro, $\hat{\pi}_n = d^{-1} \sum_{i=n-d}^n \pi_i$. Cela exige de connaître la période, si c'est impossible on peut alors utiliser la méthode ergodique.

Méthode régénérative

On se reportera à la Section § 3 qui trouve sa source dans ASMUSSEN, *Applied Probability and Queues*.

§ 7 Problèmes

Exercice 5.1 Sur un échiquier de 8 par 8, on place un cavalier dans le coin $A1$, on suppose que le cavalier se déplace au hasard (il choisit une direction au hasard parmi celles possibles à chaque coup) et sans mémoire. On rappelle qu'un cavalier se déplace de deux cases dans une direction (horizontale ou verticale) et d'une case dans l'autre direction. En utilisant la réversibilité et des considérations de symétrie, calculer le temps moyen de retour à la case $A1$.

Même question, si l'on identifie les bords opposés de l'échiquier; le cavalier se déplace alors sur un tore!

■

Exercice 5.2 Construire (dans les cas où c'est possible) une chaîne de Markov à deux états telle que :

- les deux états soient récurrents;
- les deux états soient transients;

- l'un des états soit transient, l'autre récurrent ;
- les deux soient transients ;
- les deux soient récurrents nuls.

■

Exercice 5.3 On considère la chaîne de Markov à valeurs dans $\{1, 2, 3\}$ dont la matrice de transition est donnée par :

$$\begin{pmatrix} 0 & 1/2 & 1/2 \\ f(p) & 0 & 1-f(p) \\ 1-f(p) & 0 & f(p) \end{pmatrix}$$

où $p \in [0, 1]$ et $f(p)$ est définie par :

$$f(p) = \begin{cases} 0 & \text{si } p \leq 1/4 \\ 2p - 1/2 & \text{si } 1/4 < p \leq 3/4 \\ 1 & \text{si } p \geq 3/4. \end{cases}$$

1. Donner la classification des états en fonction des valeurs de p .
2. Pour quelles valeurs de p existe-il une probabilité invariante ? La calculer lorsqu'elle existe.
3. Partant de 2, quelle est, en fonction de p , la valeur du temps moyen de retour en 2 ?
4. Soit h la fonction définie par :

$$h(1) = -1, h(2) = 1, h(3) = 1.$$

Que vaut la limite :

$$\frac{1}{n} \sum_{j=1}^n h(X_j)$$

quand n tend vers $+\infty$ pour $p < 3/4$?

5. Si l'on dispose d'un nombre arbitrairement grand d'exemples de trajectoires, comment sait-on si p est supérieur à $3/4$? Comment sait-on si $p < 1/4$? Comment peut-on estimer p dans le cas où il est compris entre $1/4$ et $3/4$?

■

Exercice 5.4 Soit X , une chaîne de Markov irréductible et récurrente sur E , et soit F un sous-ensemble de E . Montrer que la chaîne restreinte à F ($X_n^F, n \in \mathbb{N}$) (voir définition 5.20) est une chaîne de Markov sur E .

■

Exercice 5.5 On considère la chaîne de Markov homogène X à deux états A et B de matrice de transition :

$$P = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

On cherche le temps de première apparition de la séquence ABA . Pour ce faire, on construit le processus $Y_n = (X_n, X_{n+1}, X_{n+2})$.

1. Montrer que Y est une chaîne de Markov homogène dont on donnera la matrice de transition (sous forme de matrice ou de graphe).
2. Cette chaîne est-elle irréductible ? apériodique ? récurrente positive ?
3. Calculer la probabilité invariante de Y . On pourra numérotter les états dans l'ordre lexicographique : $AAA = 1, AAB = 2, \dots$
4. En déduire le temps moyen (à l'état invariante) qui s'écoule entre deux occurrences de ABA .
5. On suppose que $X_0 = A, X_1 = B$. Donner les équations qui permettent de calculer $\mathbf{E}[\tau_{ABA}^1]$. Il n'est pas demandé de les résoudre.

■

Exercice 5.6 Soit un équipement qui émet, sur une ligne de transmission, des paquets de taille constante. On note T le temps de transmission d'un paquet. Par la suite on considère un modèle à temps discret du système, c'est-à-dire, un modèle pour lequel le temps est divisé en intervalles de longueur constante, que nous supposons égale à T . On appelle chaque intervalle un slot. La ligne de transmission peut introduire des erreurs et on définit une suite $(Y_n, n \in \mathbf{N})$ tel que $Y_n = 1$ si, à l'instant $n + 1$, la ligne est dans un état pour lequel elle introduit des erreurs et $Y_n = 0$ si, à l'instant $n + 1$, elle est dans un état où elle n'introduit pas des erreurs. On suppose que $(Y_n, n \in \mathbf{N})$ est une chaîne de Markov invariante et que $P(Y_1 = 1 | Y_0 = 1) = 0,9$ et $P(Y_1 = 0 | Y_0 = 0) = 0,1$.

L'émission se fait avec le protocole « arrêt et attente » (*stop and wait*). Selon ce protocole, chaque paquet doit être acquitté. S'il n'y a pas d'erreur, le paquet est acquitté positivement et le paquet suivant peut être transmis. En cas contraire, le paquet doit être retransmis. Pour simplifier le problème nous considérons que l'acquiescement arrive instantanément.

1. Calculer la distribution de probabilité invariante de Y_n .

2. On suppose que les paquets arrivent selon un processus géométrique. C'est-à-dire qu'au n^{e} slot il y a une arrivée avec probabilité q et aucune arrivée avec probabilité $1 - q$. Un paquet peut être transmis au même slot que celui où il arrive. Soit X_n le nombre de paquets dans le système, au slot n , après la transmission (s'il y a un paquet à transmettre). Le couple (X_n, Y_n) est une chaîne de Markov. On ordonne les états dans l'ordre lexicographique, c'est-à-dire :

0	1	2	3	4	5	6	7	8	9	...
00	01	10	11	20	21	30	31	40	41	...

3. Trouver Q la matrice de transition de (X_n, Y_n) .
4. Montrer que :

$$\begin{aligned} v_0 &= 1, \\ v_{2n} &= 9 \left(3\sqrt{q/1-q} \right)^{2n}, \\ v_{2n+1} &= 9v_{2n} \end{aligned}$$

est une mesure invariante pour la chaîne de Markov (X_n, y_n) .

5. Trouver l'ensemble de valeurs de q pour lesquelles tous les états sont récurrents positifs. Comparer le résultat obtenu avec le résultat de 1. Conclure.

■

Exercice 5.7 On considère un paquet de N cartes. Pour le mélanger, on procède de la manière suivante : on choisit une carte au hasard parmi les N et on met cette carte sur le dessus du paquet sans bouger les ordres relatifs des autres.

1. Comment représenter l'état du paquet, noté X_n , après la n^{e} opération ?
2. En introduisant les permutations particulières :

$$\tau_k = \begin{pmatrix} 1 & 2 & \dots & k-1 & k & k+1 & \dots & N \\ 2 & 3 & \dots & k & 1 & k+1 & \dots & N \end{pmatrix},$$

pour $k \in \{1, \dots, N\}$; écrire les probabilités de transition de X .

L'écriture de τ_k signifie que $\tau_k(1) = 2, \tau_k(2) = 3, \dots$, ce que l'on peut aussi noter :

$$\tau_k = (2, 3, \dots, k, 1, k+1, \dots, n).$$

3. Montrer que cette chaîne est irréductible (on pourra avantageusement raisonner pour de petites valeurs de N comme $N = 4$ par exemple).
4. Montrer qu'au bout d'un nombre suffisamment grand d'opérations on obtient un « bon » mélange, caractérisé par l'équiprobabilité de tous les états possibles du paquet de cartes. (Cette question comporte deux parties : formaliser mathématiquement le problème puis le résoudre.)

■

Exercice 5.8 On pose $E = \{1, \dots, 10\}$. On définit sur E , l'addition comme l'addition modulo 10, c'est-à-dire que $10 + 1 = 1$. On considère X , la chaîne de Markov de matrice de transition $P = (p_{i,j})$ donnée par :

$$p_{i,i+1} = p, \quad p_{i,i-1} = 1 - p.$$

On suppose que p n'est égal ni à 0, ni à 1.

1. Cette chaîne est-elle irréductible ? récurrente ? apériodique ?
2. Quelle est sa probabilité invariante ?

On considère maintenant X_1 et X_2 deux copies indépendantes de cette chaîne. On pose $Y = (X_1, X_2)$.

3. Cette chaîne est-elle irréductible ?
4. Quelle est sa probabilité invariante ?
5. On pose maintenant $Z_n = Y_{2n}$. Cette chaîne est-elle irréductible ? Quels sont ces sous-ensembles fermés ? Est-elle récurrente ? apériodique ?

■

Exercice 5.9 Soit $A = \{A_n : n \geq 1\}$ une suite de variables aléatoires indépendantes et identiquement distribuées à valeurs dans \mathbf{R}^k , soit h une application de $E \times \mathbf{R}^k$ dans E et soit X_0 une variable aléatoire indépendante de la suite A . On définit la suite $X = \{X_n : n \in \mathbf{N}\}$ par X_0 pour $n = 0$ et par $X_n = h(X_{n-1}, A_n)$, pour $n \geq 1$. Montrer que X est une chaîne de Markov. ■

§ 8 Notes et commentaires

Le nombre d'ouvrages traitant des chaînes de Markov est incalculable, on ne saurait les citer tous. Parmi les plus récents et qui

s'approchent le plus ou complètent notre approche, on peut noter **Baldi:2001uq**; **Graham:2008fk**. Les chaînes de Markov constituent toujours un champ d'investigation très actif en raison de leur universalité. Les problèmes actuels se focalisent autour du calcul de la vitesse de convergence vers la probabilité stationnaire et son lien avec le « trou spectral », la réduction de l'espace d'états pour calculer plus facilement une approximation de la probabilité invariante, les applications en simulation et statistiques à travers les méthodes MCMC.

Résumé

- Une chaîne de Markov est définie par sa loi initiale ν et son opérateur de transition P .
- Un état récurrent est un état visité une infinité de fois. Un état transient est un état visité un nombre fini de fois.
- Deux états reliés entre eux ont même nature, c'est-à-dire transient ou récurrent.
- Une chaîne est irréductible si tous les états sont liés entre eux.
- Une mesure stationnaire s'identifie à un vecteur ligne solution de l'équation $\pi P = \pi$.
- Si l'on peut trouver π tel que $\sum_{x \in E} \pi(x) = 1$ alors π est une probabilité invariante, la chaîne est récurrente.
- Dans ce cas, quelle que soit la condition initiale, $\mathbf{P}(X_n = x) \xrightarrow{n \rightarrow \infty} \pi(x)$.
- Pour calculer π , on peut au choix résoudre le système $\pi P = \pi, \sum_{x \in E} \pi(x) = 1$ ou considérer la limite de la suite $\pi_{n+1} = \pi_n P, \pi_0$ quelconque.

6

Espérance Conditionnelle

§ 1 Définition et premiers exemples

Pour (E, \mathcal{E}) polonais, on note $L^2(E, \mathcal{E}, \mathbf{P})$ l'espace des variables aléatoires de carré intégrable : les variables aléatoires telles que $\mathbf{E}[X^2] < \infty$.

Lemme 6.1 Pour \mathcal{B} sous-tribu de \mathcal{E} , $L^2(E, \mathcal{B}, \mathbf{P})$ l'espace des variables aléatoires de carré intégrable qui sont \mathcal{B} -mesurables est un sous-espace vectoriel fermé de $L^2(E, \mathcal{E}, \mathbf{P})$.

Démonstration. Soit $(X_n, n \geq 1)$ une suite de variables aléatoires, \mathcal{B} -mesurables qui converge vers X dans L^2 . On sait qu'il existe une sous-suite qui converge presque partout. Comme une limite presque partout de fonctions \mathcal{B} -mesurables est \mathcal{B} -mesurable, $X \in L^2(\Omega, \mathcal{B}, \mathbf{P})$. ■

Définition 6.2 Soit $X \in L^2(E, \mathcal{E}, \mathbf{P})$, on note $\mathbf{E}[X | \mathcal{B}]$, l'espérance conditionnelle de X sachant \mathcal{B} , définie comme la projection orthogonale de X sur l'espace de Hilbert $L^2(E, \mathcal{B}, \mathbf{P})$. Cela signifie que $\mathbf{E}[X | \mathcal{B}]$ est définie par

- $\mathbf{E}[X | \mathcal{B}]$ est \mathcal{B} -mesurable,
- pour toute variable aléatoire Z bornée, \mathcal{B} -mesurable,

$$\mathbf{E}[ZX] = \mathbf{E}[Z \mathbf{E}[X | \mathcal{B}]]. \quad (6.1)$$

Par analogie avec le cas non conditionnel, on introduit la probabilité conditionnelle sachant (une tribu) \mathcal{B} , par :

$$\mathbf{P}(A | \mathcal{B}) = \mathbf{E}[\mathbf{1}_A | \mathcal{B}].$$

Ainsi définie, l'espérance conditionnelle et la probabilité conditionnelle sont des variables aléatoires.

Lemme 6.3 — Lemme de Doob. Soit $X : E \rightarrow (F, \mathcal{B})$ et $Y : E \rightarrow (\mathbf{R}, \mathcal{B}(\mathbf{R}))$ où E et F sont deux espaces polonais. La variable aléatoire Y est $\sigma(X)$ mesurable si et seulement s'il existe $\psi : F \rightarrow \mathbf{R}$ mesurable telle que $Y = \psi(X)$.

Démonstration. La tribu engendrée par X contient nécessairement les ensembles de la forme $X^{-1}(A)$ pour $A \in \mathcal{B}$. Comme cet ensemble d'ensembles constitue une tribu, $\sigma(X) = \{X^{-1}(A), A \in \mathcal{B}\}$. Si Y est de la forme $\mathbf{1}_B$ alors $Y^{-1}(\{1\}) = B$ appartient à $\sigma(X)$ donc il existe $C \in \mathcal{B}$ tel que $B = X^{-1}(C)$. Par conséquent, on a $Y = \mathbf{1}_C(X)$.

Soit maintenant Y étagée :

$$Y = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$$

avec $A_i \cap A_j = \emptyset$ et $\alpha_i - \alpha_j \neq 0$ pour $i \neq j$. Puisque $A_i = Y^{-1}(\{\alpha_i\})$, par le même raisonnement, on construit C_1, \dots, C_n ensembles \mathcal{B} mesurables tels que $Y = (\sum_{i=1}^n \alpha_i \mathbf{1}_{C_i})(X)$. Soit Y variable aléatoire positive \mathcal{B} mesurable, on sait qu'il existe Y_n étagée qui converge presque sûrement vers Y . On a précédemment construit ψ_n tel que $Y_n = \psi_n(X)$. Comme Y_n converge vers Y , ψ_n converge vers quelque chose sur l'image de E par X . Malheureusement, rien ne garantit que cet ensemble soit mesurable. Pour éviter cet écueil, on pose $\psi = \limsup_n \psi_n$. Comme toute limite supérieure de fonctions mesurables est mesurable, cette fonction est mesurable. De plus, aux endroits où ψ_n converge vers ψ , la limite supérieure aussi. En conclusion, on a bien construit ψ mesurable telle que $Y = \psi(X)$. ■

§ 2 Propriétés de l'espérance conditionnelle

Dans cette section nous donnons les propriétés essentielles de l'espérance conditionnelle. Bien qu'elles soient semblables à celles de l'espérance (quitte à ajouter le terme « presque sûrement »), il ne faut pas oublier que l'espérance conditionnelle est un opérateur de l'ensemble des variables aléatoires dans lui-même. Le résultat suivant est une conséquence directe de la notion de projection orthogonale.

Théorème 6.4 Si Y est une variable aléatoire bornée et \mathcal{B} -mesurable (i.e., $Y \in L^\infty$), alors nous avons

$$\mathbf{E}[XY | \mathcal{B}] = Y\mathbf{E}[X | \mathcal{B}], \mathbf{P} \text{ p.s.,}$$

pour tout $X \in L^2$.

Démonstration. Il est clair que le terme de droite est \mathcal{B} -mesurable. Par ailleurs pour Z \mathcal{B} -mesurable, puisque l'espérance conditionnelle est

une projection

$$\mathbf{E}[ZX \mathbf{E}[Y | \mathcal{B}]] = \mathbf{E}[\mathbf{E}[ZX | \mathcal{B}] Y] = \mathbf{E}[ZXY] = \mathbf{E}[\mathbf{E}[Z | \mathcal{B}] XY] = \mathbf{E}[Z \mathbf{E}[XY | \mathcal{B}]].$$

D'où le résultat. ■

Théorème 6.5 Si X est une variable aléatoire presque sûrement positive, alors $\mathbf{E}[X | \mathcal{B}]$ est aussi une variable aléatoire presque sûrement positive.

Démonstration. Nous avons, d'après l'hypothèse

$$0 \leq \mathbf{E} \left[X \mathbf{1}_{\{\mathbf{E}[X | \mathcal{B}] \leq 0\}} \right] = \mathbf{E} \left[\mathbf{E}[X | \mathcal{B}] \mathbf{1}_{\{\mathbf{E}[X | \mathcal{B}] \leq 0\}} \right] \leq 0,$$

par conséquent

$$\mathbf{E} \left[X \mathbf{1}_{\{\mathbf{E}[X | \mathcal{B}] \leq 0\}} \right] = 0.$$

Comme $X \mathbf{1}_{\{\mathbf{E}[X | \mathcal{B}] \leq 0\}}$ est une variable aléatoire presque sûrement positive, la nullité de son espérance implique qu'elle même est égale à zéro presque sûrement. Donc

$$X = X \mathbf{1}_{\{\mathbf{E}[X | \mathcal{B}] > 0\}}$$

presque sûrement. En prenant l'espérance conditionnelle de deux cotés, d'après la proposition précédente, nous avons

$$\mathbf{E}[X | \mathcal{B}] = \mathbf{E}[X | \mathcal{B}] \mathbf{1}_{\{\mathbf{E}[X | \mathcal{B}] > 0\}}$$

presque sûrement. ■

Comme la projection orthogonale est un opérateur linéaire, la propriété suivante est évidente :

Théorème 6.6 Si $U, V \in L^2(\Omega, \mathcal{B}, \mathbf{P})$ et $a, b \in \mathbf{R}$, nous avons

$$\mathbf{E}[aU + bV | \mathcal{B}] = a\mathbf{E}[U | \mathcal{B}] + b\mathbf{E}[V | \mathcal{B}], \text{ p.s.}$$

La proposition suivante est une conséquence de la première proposition :

Théorème 6.7 Si $U \geq V$ p.s. alors $\mathbf{E}[U | \mathcal{B}] \geq \mathbf{E}[V | \mathcal{B}]$ p.s.

Théorème 6.8 Si pour tout n , $X_n \geq 0$ p.s. et si $(X_n(\omega), n \in \mathbf{N})$ converge presque sûrement en croissant vers $X(\omega)$, $X \in L^2$, alors

$$\mathbf{E}[X_n | \mathcal{B}] \rightarrow \mathbf{E}[X | \mathcal{B}],$$

presque sûrement en croissant.

Démonstration. D'après le théorème 6.5, $(E[X_n|\mathcal{B}], n \geq 0)$ est une suite monotone, croissante, bornée par $E[X|\mathcal{B}]$, par conséquent la limite existe presque sûrement. Soit $Y = \lim_n E[X_n|\mathcal{B}] \leq E[X|\mathcal{B}]$. Pour achever la preuve il suffit de montrer que cette limite est égale à $E[X|\mathcal{B}]$. Soit $Z \in L^2(\Omega, \mathcal{B}, \mathbf{P})$, $Z \geq 0$ p.s., alors nous avons

$$\begin{aligned} E[Z(E[X|\mathcal{B}] - Y)] &= E[ZE[X|\mathcal{B}]] - E[Z \lim_n E[X_n|\mathcal{B}]] \\ &= E[ZX] - \lim_n E[ZE[X_n|\mathcal{B}]] \\ &= E[ZX] - \lim_n E[ZX_n] \\ &= \lim_n E[Z(X - X_n)] \\ &= E[Z \lim_n (X - X_n)] = 0. \end{aligned}$$

Comme $Z(E[X|\mathcal{B}] - Y) \geq 0$ p.s., nous avons $Z(E[X|\mathcal{B}] - Y) = 0$ p.s et il suffit de prendre $Z = 1$ pour conclure. ■

Théorème 6.9 Si $X_n \geq 0$ et si $\liminf_n X_n \in L^2$ alors

$$E[\liminf_n X_n|\mathcal{B}] \leq \liminf_n E[X_n|\mathcal{B}].$$

Démonstration. D'après la proposition précédente, nous avons :

$$\begin{aligned} E[\liminf_n X_n|\mathcal{B}] &= \lim_{n \rightarrow \infty} E[\inf_{k \geq n} X_k|\mathcal{B}] \\ &\leq \lim_{n \rightarrow \infty} \inf_{k \geq n} E[X_k|\mathcal{B}]. \end{aligned}$$

Théorème 6.10 Si $X_n \rightarrow X$ p.s. et si $|X_n| \leq Y$, pour tout n , où Y est dans $L^2(\mathcal{B})$, alors

$$E[X_n|\mathcal{B}] \rightarrow E[X|\mathcal{B}] \text{ p.s. et dans } L^2.$$

Démonstration. D'après le théorème de convergence dominée, si $X_n \rightarrow X$ p.s. et si $|X_n| \leq Y$ p.s., alors $X_n \rightarrow X$ dans $L^2(\mathcal{B})$; comme $E[.|\mathcal{B}]$ est un opérateur continu sur $L^2(\Omega, \mathcal{B}, P)$, elle commute avec la convergence dans L^2 . Pour la convergence p.s. nous avons :

$$\begin{aligned} \limsup E[X_n|\mathcal{B}] &\leq E[\limsup X_n|\mathcal{B}] = E[X|\mathcal{B}] \\ &= E[\liminf X_n|\mathcal{B}] \leq \liminf E[X_n|\mathcal{B}]. \end{aligned}$$

Théorème 6.11 Si $\mathcal{B}_1, \mathcal{B}_2$ sont deux sous-tribus de \mathcal{F} avec $\mathcal{B}_1 \subset \mathcal{B}_2$ alors

$$E[E[X|\mathcal{B}_2]|\mathcal{B}_1] = E[X|\mathcal{B}_1],$$

pour tout $X \in L^2(\mathcal{F})$.

Démonstration. Si $Y \in L^2(\mathcal{B}_1)$, nous avons

$$\begin{aligned} E[YE[E[X|\mathcal{B}_2]|\mathcal{B}_1]] &= E[YE[X|\mathcal{B}_2]] \\ &= E[YZ] \text{ car } Y \in L^2(\mathcal{B}_1) \subset L^2(\mathcal{B}_2) \\ &= E[YE[X|\mathcal{B}_1]]. \end{aligned}$$

■

Théorème 6.12 Si $\mathcal{B}_1 = \{\emptyset, \Omega\}$, alors

$$E[X|\mathcal{B}_1] = E[X],$$

pour toute variable aléatoire X dans $L^2(\mathcal{F})$.

Démonstration. $E[X]$ est constante donc \mathcal{B} mesurable. Par ailleurs, toute variable aléatoire \mathcal{B} mesurable est constante donc égale à son espérance, par conséquent :

$$E[XY] = E[XE[Y]] = E[X]E[Y] = E[YE[X]],$$

ce qui signifie bien que $E[X|\mathcal{B}_1] = E[X]$.

■

Théorème 6.13 Si X est indépendante de la tribu \mathcal{B} alors

$$E[X|\mathcal{B}] = E[X], \text{ p.s.}$$

Démonstration. $E[X]$ est mesurable par rapport à la tribu grossière donc mesurable aussi par rapport à n'importe quelle tribu. Par ailleurs, si Y est \mathcal{B} mesurable, X est en particulier indépendant de Y donc

$$E[XY] = E[X]E[Y] = E[YE[X]].$$

■

En combinant les deux théorèmes précédents, on obtient

Théorème 6.14 Soient \mathcal{A} une sous-tribu de \mathcal{F} et que $F \in L^2(\mathcal{F})$. On suppose que $\sigma(F) \vee \mathcal{A}$ est indépendante de \mathcal{B} . Alors

$$E[F|\mathcal{A} \vee \mathcal{B}] = E[F|\mathcal{A}]$$

presque sûrement.

■ **Exemple 6.1** Soit $\mathcal{B} = \{\emptyset, A, A^c, \Omega\}$ où $A \in \mathcal{F}$. Donc \mathcal{B} est une sous-tribu de \mathcal{F} car $A \in \mathcal{F}$. Calculons $E[X|\mathcal{B}]$ pour un $X \in L^2(\Omega, \mathcal{F}, P)$ en

utilisant la propriété essentielle écrite ci-dessus. Remarquons que si Z est une variable aléatoire \mathcal{B} -mesurable alors elle est forcément du type

$$Z = a\mathbf{1}_A + b\mathbf{1}_{A^c}, a, b \in \mathbf{R},$$

par conséquent $E[X|\mathcal{B}]$ sera de la forme aussi $c\mathbf{1}_A + d\mathbf{1}_{A^c}$, et il suffit de déterminer les constantes c et d . En remplaçant $E[X|\mathcal{B}]$ par $c\mathbf{1}_A + d\mathbf{1}_{A^c}$, nous obtenons

$$E[(a\mathbf{1}_A + b\mathbf{1}_{A^c})(c\mathbf{1}_A + d\mathbf{1}_{A^c})] = acP(A) + bdP(A^c),$$

donc

$$E[X(a\mathbf{1}_A + b\mathbf{1}_{A^c})] = acP(A) + bdP(A^c)$$

alors nous avons

$$\begin{aligned} aE[X\mathbf{1}_A] &= acP(A) \\ bE[X\mathbf{1}_{A^c}] &= bdP(A^c) \end{aligned}$$

donc

$$c = \frac{E[X\mathbf{1}_A]}{P(A)} \text{ et } d = \frac{E[X\mathbf{1}_{A^c}]}{P(A^c)},$$

finalement, nous obtenons

$$E[X|\mathcal{B}] = \frac{E[X\mathbf{1}_A]}{P(A)}\mathbf{1}_A + \frac{E[X\mathbf{1}_{A^c}]}{P(A^c)}\mathbf{1}_{A^c}.$$

Comme on le voit, le résultat est une combinaison linéaire de fonctions indicatrices d'ensembles mesurables, donc c'est une variable aléatoire. Nous constatons de plus, dans ce cas précis, que les coefficients de cette combinaison linéaire sont les espérances conditionnelles élémentaires. En particulier si nous prenons $X = \mathbf{1}_C, C \in \mathcal{F}$ alors nous notons $E[\mathbf{1}_C|\mathcal{B}]$ par $P(C|\mathcal{B})$ et nous avons

$$P(C|\mathcal{B}) = P(C|A)\mathbf{1}_A + P(C|A^c)\mathbf{1}_{A^c}.$$

■

§ 3 Conditionnement des vecteurs gaussiens

Théorème 6.15 Soit $(X, Y_1, \dots, Y_d) \in \mathbf{R}^{d+1}$ un vecteur gaussien.

1. Il existe des constantes a, b_1, \dots, b_d telles que

$$E[X|Y_1, \dots, Y_d] = a + \sum_{i=1}^d b_i Y_i.$$

2. Il existe une variable aléatoire gaussienne, réelle, centrée Z

indépendante de (Y_1, \dots, Y_d) telle que

$$X = a + \sum_{i=1}^d b_i Y_i + Z.$$

Démonstration. Soit H le sous-espace vectoriel de $L^2(\Omega, \mathcal{F}, \mathbf{P})$ défini par

$$H = \left\{ u + \sum_{i=1}^d v_i Y_i; u, v_i \in \mathbf{R} \right\}.$$

H est fermé car tout sous-espace vectoriel de dimension finie est fermé. Soit π l'opérateur de projection orthogonale correspondant et notons $\pi(X) = X_0$. Alors $X_0 \in H$ et $X - X_0$ est orthogonal à H et

$$X_0 = a + \sum_{i=1}^d b_i Y_i.$$

Notons $Z = X - X_0$, alors Z est orthogonal à H . Le vecteur $(Z, (Y_1, \dots, Y_d))$ est gaussien, de plus Z est orthogonal au vecteur (Y_1, \dots, Y_d) donc Z est indépendante de (Y_1, \dots, Y_d) . Comme $\mathbf{1} \in H$ et $Z \perp H$ nous avons $E[Z] = 0$ et donc

$$\mathbf{E}[Z | Y_1, \dots, Y_d] = \mathbf{E}[Z] = 0.$$

Par conséquent,

$$\mathbf{E}[X | Y_1, \dots, Y_d] = \mathbf{E}[X_0 + Z | Y_1, \dots, Y_d] = X_0.$$

D'où le résultat. ■

On peut calculer explicitement les constantes a, b_i avec du calcul matriciel.

Corollaire 6.16 Soit $Z = (X, Y)$ un vecteur gaussien. On suppose que Γ_Y est inversible. On note Γ_0 la matrice des covariances croisées :

$$\Gamma_0(i, j) = \text{cov}(X_i, Y_j), \text{ i.e. } \Gamma_0 = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])^t].$$

On a

$$\mathbf{E}[X | Y] = \mathbf{E}[X] + \Gamma_0 \Gamma_Y^{-1} (Y - \mathbf{E}[Y]).$$

§ 4 Cas intégrable

Nous avons supposé jusqu'à présent que les v.a. dont on voulait calculer l'espérance conditionnelle était de carré intégrable. En pratique, c'est une hypothèse trop forte, nous allons donc développer la notion d'espérance conditionnelle dans le cas de v.a. uniquement intégrables.

Définition 6.17 Soit $X \in L^1$ et \mathcal{B} une sous-tribu de \mathcal{F} , on appelle espérance conditionnelle de X sachant \mathcal{B} , toute v.a. Y satisfaisant les deux conditions suivantes : Y est \mathcal{B} -mesurable et

$$E[XZ] = E[YZ], \text{ pour tout } Z \in L^\infty(\Omega, \mathcal{B}, \mathbf{P}). \quad (6.2)$$

Théorème 6.18 Pour $X \in L^1$ et \mathcal{B} une sous-tribu de \mathcal{F} , il existe une unique v.a. (définie à un négligeable près) qui satisfait ces deux conditions. On la note $E[X | \mathcal{B}]$.

Démonstration. Supposons dans un premier temps que X appartienne à L^2 . La v.a. $E[X | \mathcal{B}]$ est bien définie et l'événement $A = \{\omega : E[X | \mathcal{B}] \geq 0\}$ est \mathcal{B} -mesurable donc

$$\begin{aligned} E[|E[X | \mathcal{B}]|] &= E[E[X | \mathcal{B}]\mathbf{1}_A] - E[E[X | \mathcal{B}]\mathbf{1}_{A^c}] \\ &= E[X\mathbf{1}_A] - E[X\mathbf{1}_{A^c}] \\ &\leq E[|X|], \end{aligned} \quad (6.3)$$

car $\mathbf{1}_A - \mathbf{1}_{A^c}$ prend deux valeurs -1 ou 1 donc $X(\mathbf{1}_A - \mathbf{1}_{A^c}) \leq |X|$.

L'application $(X \mapsto E[X | \mathcal{B}])$ définie sur $L^2 \subset L^1$ est à valeurs dans L^1 donc on peut l'étendre à L^1 . Précisément, comme L^2 est dense dans L^1 , pour tout $X \in L^1$, il existe une suite $(X_n, n \geq 1)$ d'éléments de L^2 qui converge dans L^1 vers X , $E[X | \mathcal{B}]$ est alors définie comme la limite dans L^1 de la suite $(E[X_n | \mathcal{B}], n \geq 1)$. L'inégalité (6.3) implique que la limite ne dépend pas de la suite approximante.

S'il existe deux v.a., Y_1 et Y_2 , \mathcal{B} -mesurables satisfaisant (6.2) alors prenons $B = \{Y_1 - Y_2 > 0\}$, c'est un événement \mathcal{B} -mesurable et

$$E[(Y_1 - Y_2)\mathbf{1}_B] = 0,$$

donc B est de mesure nulle. De même pour l'ensemble $\{Y_2 - Y_1 > 0\}$ donc $Y_1 = Y_2$ \mathbf{P} -p.s. ■

Par densité, toutes les propriétés énoncées ci-dessus dans le cas L^2 s'étendent au cas où X est dans L^1 .

Chaînes de Markov

On trouvera les deux théorèmes suivants dans KALLENBERG, *Foundations of modern probability*, ils sont abstraits mais utiles (même si hors des exigences de l'agrégation).

Définition 6.19 Soit $\mathcal{F}_1, \dots, \mathcal{F}_n, \mathcal{G} \subset \mathcal{E}$ des tribus. Les tribus $\mathcal{F}_1, \dots, \mathcal{F}_n$ sont indépendantes conditionnellement à \mathcal{G} lorsque :

$$\mathbf{P}\left(\bigcap_{j=1}^n B_j \mid \mathcal{G}\right) = \prod_{j=1}^n \mathbf{P}(B_j \mid \mathcal{G}) \text{ presque sûrement,} \quad (6.4)$$

pour tout $B_j \in \mathcal{F}_j, j = 1, \dots, n$.

Une famille infinie de tribus $(\mathcal{F}_r, r \in T)$ est indépendante conditionnellement à \mathcal{G} si (6.4) est vraie pour toute sous-famille finie.

Théorème 6.20 Soit \mathcal{F}, \mathcal{G} et \mathcal{H} trois tribus de (E, \mathcal{E}) . Il y a équivalence entre les trois propriétés suivantes :

1. les tribus \mathcal{F} et \mathcal{H} sont indépendantes conditionnellement à \mathcal{G} ;
2. pour tout $H \in \mathcal{H}, \mathbf{P}(H | \mathcal{F} \vee \mathcal{G}) = \mathbf{P}(H | \mathcal{G})$, presque sûrement ;

Démonstration. Supposons que \mathcal{F} et \mathcal{H} sont indépendantes conditionnellement à \mathcal{G} . Pour $F \in \mathcal{F}, G \in \mathcal{G}$ et $H \in \mathcal{H}$, on a par définition de l'indépendance conditionnelle :

$$\begin{aligned} \mathbf{E}[\mathbf{P}(H | \mathcal{G}) \mathbf{1}_F \mathbf{1}_G] &= \mathbf{E}[\mathbf{P}(H | \mathcal{G}) \mathbf{P}(F | \mathcal{G}) \mathbf{1}_G] \\ &= \mathbf{E}[\mathbf{P}(H \cap F | \mathcal{G}) \mathbf{1}_G] \\ &= \mathbf{E}[\mathbf{1}_H \mathbf{1}_{F \cap G}], \end{aligned}$$

en utilisant l'indépendance conditionnelle dans le passage de la première à la deuxième ligne. On note :

$$\mathcal{D} = \{M \in \mathcal{F} \vee \mathcal{G}, \mathbf{E}[\mathbf{P}(H | \mathcal{G}) \mathbf{1}_M] = \mathbf{E}[\mathbf{1}_H \mathbf{1}_M]\}.$$

De ce qui précède, $\mathcal{C} = \{M = F \cap G, F \in \mathcal{F}, G \in \mathcal{G}\} \subset \mathcal{D}$. Il est évident que \mathcal{C} est un π -système. Par linéarité et convergence monotone, il apparaît que \mathcal{D} est un λ -système. En vertu du théorème 1.11, \mathcal{D} contient la tribu engendrée par \mathcal{C} . Or $\mathcal{F} \subset \mathcal{C}$ et $\mathcal{G} \subset \mathcal{C}$ donc \mathcal{C} contient $\mathcal{F} \vee \mathcal{G}$. Ceci signifie que pour tout $M \in \mathcal{F} \vee \mathcal{G}$, on a :

$$\mathbf{E}[\mathbf{P}(H | \mathcal{G}) \mathbf{1}_M] = \mathbf{E}[\mathbf{1}_H \mathbf{1}_M].$$

Comme $\mathbf{P}(H | \mathcal{G})$ est $\mathcal{F} \vee \mathcal{G}$ mesurable, on conclut de cette identité que le point 2 est vérifié.

Réciproquement si le point 2 est vérifié, pour tout $F \in \mathcal{F}$ et tout $H \in \mathcal{H}$, on obtient :

$$\begin{aligned} \mathbf{P}(F \cap H | \mathcal{G}) &= \mathbf{E}[\mathbf{P}(F \cap H | \mathcal{F} \vee \mathcal{G}) | \mathcal{G}] \\ &= \mathbf{E}[\mathbf{1}_F \mathbf{P}(H | \mathcal{F} \vee \mathcal{G}) | \mathcal{G}] \\ &= \mathbf{E}[\mathbf{1}_F \mathbf{P}(H | \mathcal{G}) | \mathcal{G}] \\ &= \mathbf{P}(H | \mathcal{G}) \mathbf{P}(F | \mathcal{G}). \end{aligned}$$

Ce qui prouve l'indépendance de \mathcal{F} et \mathcal{H} conditionnellement à \mathcal{G} . ■

Corollaire 6.21 Soit $\mathcal{G}, \mathcal{H}, \mathcal{L}_1, \dots, \mathcal{L}_n, \dots$ des tribus. Les propositions suivantes sont équivalentes :

1. les tribus \mathcal{H} et $\bigvee_n \mathcal{L}_n$ sont indépendantes conditionnellement à \mathcal{G} ;
2. pour tout entier n , les tribus \mathcal{H} et \mathcal{L}_{n+1} sont indépendantes conditionnellement à $\mathcal{G} \vee \mathcal{L}_1 \vee \dots \vee \mathcal{L}_n$.

Démonstration. Si \mathcal{H} et $\bigvee_n \mathcal{L}_n$ sont indépendantes conditionnellement à \mathcal{G} alors \mathcal{H} et toute tribu engendrée par une sous-famille finie des \mathcal{L}_j sont indépendantes conditionnellement à \mathcal{G} . Appliquons le théorème 6.20 avec $\mathcal{L} = \bigvee_{j=1}^n \mathcal{L}_j$ puis $\mathcal{L} = \bigvee_{j=1}^{n+1} \mathcal{L}_j$, on obtient :

$$\mathbf{P}(H | \mathcal{G}) = \mathbf{P}(H | \mathcal{G} \vee \bigvee_{j=1}^n \mathcal{L}_j) \text{ et } \mathbf{P}(H | \mathcal{G}) = \mathbf{P}(H | \mathcal{G} \vee \bigvee_{j=1}^{n+1} \mathcal{L}_j).$$

En appliquant une nouvelle fois le théorème 6.20 avec $\mathcal{L} = \mathcal{L}_{n+1}$, on en déduit le point 2.

Réciproquement, supposons que pour tout $n \geq 0$, pour tout $H \in \mathcal{H}$, on ait :

$$\mathbf{P}(H | \mathcal{G} \vee \bigvee_{j=1}^n \mathcal{L}_j) = \mathbf{P}(H | \mathcal{G} \vee \bigvee_{j=1}^{n+1} \mathcal{L}_j).$$

Par transitivité de la relation d'égalité, on a alors :

$$\mathbf{P}(H | \mathcal{G}) = \mathbf{P}(H | \mathcal{G} \vee \bigvee_{j=1}^m \mathcal{L}_j) \text{ pour tout } m.$$

Toujours d'après le théorème 6.20, on en déduit que \mathcal{H} et $\bigvee_{j=1}^m \mathcal{L}_j$ sont indépendantes conditionnellement à \mathcal{G} . Par définition de l'indépendance conditionnelle pour un nombre infini de tribus, cela suffit à montrer le point 1. ■

§ 5 Exercices

Exercice 6.1 Les règles du jeu du **not-seven** sont les suivantes : on part d'un score $X_0 = 0$. à chaque coup, on lance deux dés non pipés, si la somme des faces égale 7, le score retourne à 0 et la partie est terminée. Sinon, le score augmente de la somme des faces et on a le droit de rejouer ou pas. Si l'on ne rejoue pas, le score est acquis et la partie est terminée. Si l'on rejoue, on relance les deux dés avec la même règle.

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

On considère une suite $(S_n, n \in \mathbf{N})$ de variables aléatoires

indépendantes de même loi que S .

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

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

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

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

■

7

Martingales à temps discret

Soit $(\eta_i; i \in \mathbf{N})$ une suite de Bernouilli, i.e. des variables aléatoires indépendantes, de même loi telle que

$$\mathbf{P}(\eta_i = 1) = p, \mathbf{P}(\eta_i = -1) = q = 1 - p.$$

Supposons qu'un joueur mise sur le résultat d'un tirage à l'instant n une quantité V_n . Alors le gain total du joueur à l'instant n est

$$\begin{aligned} X_n &= \sum_{i=1}^n V_i \eta_i = X_{n-1} + V_n \eta_n \\ X_0 &= 0. \end{aligned}$$

Il est naturel de supposer que la mise V_n à l'instant n ne dépend que des événements qui ont eu lieu jusqu'à l'instant $n - 1$, c'est-à-dire que l'on doit supposer que V_n est mesurable par rapport à la tribu $\mathcal{F}_{n-1} = \sigma\{\eta_1, \dots, \eta_{n-1}\}$. Alors nous avons

$$\begin{aligned} \mathbf{E}[X_{n+1} - X_n | \mathcal{F}_n] &= \mathbf{E}[V_{n+1} \eta_{n+1} | \mathcal{F}_n] \\ &= V_{n+1} \mathbf{E}[\eta_{n+1} | \mathcal{F}_n] \\ &= V_{n+1} \mathbf{E}[\eta_{n+1}], \end{aligned}$$

où la dernière égalité provient de l'indépendance. Nous dirons que le jeu est équitable ou bien que $(X_n, n \geq 0)$ est une martingale si $\mathbf{E}[\eta_1] = 0$, que le jeu est favorable ou $(X_n, n \geq 0)$ est une sous-martingale si $\mathbf{E}[\eta_1] \geq 0$ et finalement que le jeu est défavorable ou $(X_n, n \geq 0)$ est une surmartingale si $\mathbf{E}[\eta_1] \leq 0$.

§ 1 Définitions et propriétés

Définition 7.1 Soit (Ω, \mathcal{F}, P) un espace de probabilité et $(\mathcal{F}_n; n \in \mathbf{N})$ une suite croissante de sous tribus de \mathcal{F} . Une suite (X_n) de variables aléatoires réelles, intégrables s'appelle une martingale (respectivement une sous-martingale ou une surmartingale) si

1. Pour tout n , X_n est \mathcal{F}_n -mesurable,

2. Pour tout n , X_n est intégrable,
 3. $\mathbf{E}[X_{n+1} | \mathcal{F}_n] = X_n$ presque sûrement (respectivement $\mathbf{E}[X_{n+1} | \mathcal{F}_n] \geq X_n$ ou $\mathbf{E}[X_{n+1} | \mathcal{F}_n] \leq X_n$).

■ **Exemple 7.1** — Si $(\zeta_i, i \geq 1)$ sont des variables aléatoires IID¹, avec $\mathbf{E}[\zeta_i] = 0$, alors la suite définie par

$$X_0 = 0, \quad X_n = \sum_{i=1}^n \zeta_i$$

est une $\mathcal{F}_n = \sigma(\zeta_0, \dots, \zeta_n)$ martingale, car

$$\begin{aligned} \mathbf{E} \left[\sum_{i=1}^{n+1} \zeta_i \middle| \mathcal{F}_n \right] &= \mathbf{E}[X_n + \zeta_{n+1} | \mathcal{F}_n] \\ &= X_n + \mathbf{E}[\zeta_{n+1}] = X_n, \end{aligned}$$

car ζ_{n+1} est indépendante de \mathcal{F}_n .

— Si $(\zeta_i; i \geq 1)$ sont des variables aléatoires i.i.d., avec $\mathbf{E}[\zeta_i] = 1$, alors $X_0 = 1$, $X_n = \prod_{i=1}^n \zeta_i$ est une $\mathcal{F}_n = \sigma(\zeta_0, \dots, \zeta_n)$ martingale.

$$\begin{aligned} \mathbf{E} \left[\prod_{i=1}^{n+1} \zeta_i \middle| \mathcal{F}_n \right] &= \mathbf{E}[X_n \zeta_{n+1} | \mathcal{F}_n] \\ &= X_n \mathbf{E}[\zeta_{n+1} | \mathcal{F}_n] \\ &= X_n \mathbf{E}[\zeta_{n+1}] = X_n \end{aligned}$$

car d'une part X_n est \mathcal{F}_n mesurable et que d'autre part ζ_{n+1} est indépendante de \mathcal{F}_n .

— Soit $\zeta \in L^1(\mathbf{P})$, alors $(X_n = \mathbf{E}[\zeta | \mathcal{F}_n])$ est une martingale.

$$\begin{aligned} \mathbf{E}[X_{n+1} | \mathcal{F}_n] &= \mathbf{E}[\mathbf{E}[\zeta | \mathcal{F}_{n+1}] | \mathcal{F}_n] \\ &= \mathbf{E}[\zeta | \mathcal{F}_n], \end{aligned}$$

car $\mathcal{F}_n \subset \mathcal{F}_{n+1}$.

— Si $(X_n, n \geq 0)$ est une martingale et g est une fonction convexe, $(g(X_n), n \geq 0)$ est une sous-martingale.

$$\begin{aligned} \mathbf{E}[g(X_{n+1}) | \mathcal{F}_n] &\geq g(\mathbf{E}[X_{n+1} | \mathcal{F}_n]) \\ &\geq g(X_n) \end{aligned}$$

d'après l'inégalité de Jensen (valable aussi pour les espérances conditionnelles) et la propriété de martingale de X .

1. Indépendantes et identiquement distribuées

■ **Définition 7.2** Soit T une variable aléatoire à valeurs dans $\bar{\mathbf{N}} := \mathbf{N} \cup \{+\infty\}$. La variable aléatoire T est un (\mathcal{F}_n) -temps d'arrêt si

pour tout $n \in \mathbf{N}$,

$$\{\omega : T(\omega) = n\} \in \mathcal{F}_n.$$

■ **Exemple 7.2** Avec l'exemple du joueur, soit $a \geq 0$, et

$$\begin{aligned} T(\omega) &= \inf(k : X_k(\omega) \geq a) \\ &= +\infty \text{ si } \emptyset. \end{aligned}$$

Alors T est un temps d'arrêt :

$$\{\omega : T(\omega) = n\} = \{\omega : X_k(\omega) < a, k < n, X_n \geq a\} \in \mathcal{F}_n.$$

■

Si $(X_n : n \in \mathbf{N})$ sont des variables aléatoires, nous définissons X_T comme

$$X_T(\omega) = \sum_{n=0}^{\infty} X_n(\omega) \mathbf{1}_{\{T=n\}}(\omega).$$

Pour tout $B \in \mathcal{B}(R)$, nous avons

$$\{X_T \in B\} = \bigcup_{n=0}^{\infty} \{X_n \in B, T = n\} \in \mathcal{F},$$

donc X_T est bien une variable aléatoire.

■ **Remarque 22**— Notons que $\{T > n\}, \{T \leq n\} \in \mathcal{F}_n$, $\{T < n\} \in \mathcal{F}_{n-1}$.

— Si S et T sont deux temps d'arrêt, $\sup(S, T) = S \vee T$ et $\inf(S, T) = S \wedge T$ sont aussi des temps d'arrêt.

Théorème 7.3 Soit (X_n) une martingale et T un temps d'arrêt. Alors (X_n^T) , où X_n^T est définie par $X_{T \wedge n}$, est une martingale.

Démonstration. Nous avons

$$X_{T \wedge n} = \sum_{m=0}^{n-1} X_m \mathbf{1}_{\{T=m\}} + X_n \mathbf{1}_{\{T \geq n\}},$$

par conséquent $X_{T \wedge n}$ est \mathcal{F}_n -mesurable. En ce qui concerne l'intégrabilité de $X_{T \wedge n}$, comme $|X_n|$ est une sous-martingale, pour $m \leq n$, on a

$$\mathbf{E}[|X_m| | Y_m] \leq \mathbf{E}[|X_n| | Y_m]$$

pour toute variable aléatoire Y_m \mathcal{F}_m mesurable positive. Par conséquent,

$$\begin{aligned} \mathbf{E}[|X_{T \wedge n}|] &\leq \sum_{m=0}^{n-1} \mathbf{E}[|X_m| \mathbf{1}_{\{T=m\}}] + \mathbf{E}[|X_n| \mathbf{1}_{\{T \geq n\}}] \\ &\leq \sum_{m=0}^{n-1} \mathbf{E}[|X_n| \mathbf{1}_{\{T=m\}}] + \mathbf{E}[|X_n| \mathbf{1}_{\{T \geq n\}}] \leq \mathbf{E}[|X_n|] < +\infty. \end{aligned}$$

De plus

$$\begin{aligned} X_{T \wedge (n+1)} - X_{T \wedge n} &= X_{n+1} \mathbf{1}_{\{T \geq n+1\}} - X_n \mathbf{1}_{\{T > n\}} \\ &= \mathbf{1}_{\{T > n\}} (X_{n+1} - X_n). \end{aligned}$$

Donc

$$\mathbf{E} [X_{T \wedge (n+1)} - X_{T \wedge n} | \mathcal{F}_n] = \mathbf{1}_{\{T > n\}} \mathbf{E} [X_{n+1} - X_n | \mathcal{F}_n] = 0.$$

■

Définition 7.4 Si T est un temps d'arrêt, on note par \mathcal{F}_T la tribu définie par

$$A \in \mathcal{F}_T \iff A \cap \{\omega : T(\omega) \leq n\} \in \mathcal{F}_n,$$

pour tout $n \geq 0$.

Donnons un lemme dont l'utilisation est importante :

Lemme 7.5 Soit T un temps d'arrêt et X une variable aléatoire réelle \mathcal{F}_T -mesurable. Alors

$$X \mathbf{1}_{\{T \leq n\}}$$

est \mathcal{F}_n -mesurable pour tout $n \in \mathbf{N}$.

Démonstration. Soit $A \in \mathcal{F}_T$ et $X = \mathbf{1}_A$. Alors la conclusion est vraie trivialement. Par linéarité, elle est aussi vraie pour des variables aléatoires étagées. Si X est quelconque, alors il existe une suite de variables aléatoires étagées, \mathcal{F}_T -mesurables, qui converge vers X p.s. Par conséquent

$$X \mathbf{1}_{\{T \leq n\}} = \lim_k X_k \mathbf{1}_{\{T \leq n\}} \in \mathcal{F}_n.$$

■

Pour une martingale, nous avons

$$\mathbf{E} [X_n] = \mathbf{E} [X_0].$$

Si T est un temps d'arrêt, est-il vrai que

$$\mathbf{E} [X_T] = \mathbf{E} [X_0]?$$

Théorème 7.6 Soit T un temps d'arrêt borné : il existe $N \in \mathbf{N}$ tel que $T \leq N$ presque-sûrement. Soit $(X_n, n \geq 1)$ une martingale. On a

$$\mathbf{E} [X_T] = \mathbf{E} [X_0].$$

Démonstration. On a vu que le processus X^T était une martingale donc pour tout entier $n \geq 0$,

$$\mathbf{E} \left[X_n^T \right] = \exp X_0^T = \mathbf{E} [X_0].$$

Pour $n = N$,

$$X_N^T = X_{T \wedge N} = X_T$$

puisque $T \leq N$. Le résultat s'ensuit. ■

L'énoncé ci-dessous est connu sous l'appellation **identité de Wald** et possède de nombreuses applications :

Théorème 7.7 — Identité de Wald. Soient (ξ_i) des variables aléatoires réelles, i.i.d., avec $\mathbf{E} [|\xi_1|] < \infty$ et T un temps d'arrêt par rapport à la filtration des tribus $\mathcal{F}_n = \sigma(\xi_i; i \leq n)$. Si $\mathbf{E} [T] < \infty$, nous avons

$$\mathbf{E} \left[\sum_{i=1}^{T(\omega)} \xi_i \right] = \mathbf{E} [\xi_1] \mathbf{E} [T].$$

Si, de plus $\mathbf{E} [\xi_1^2] < \infty$, alors nous avons

$$\mathbf{E} \left[\left(\sum_{i=1}^{T(\omega)} (\xi_i - \mathbf{E} [\xi_1]) \right)^2 \right] = \mathbf{E} [(\xi_1 - \mathbf{E} [\xi_1])^2] \mathbf{E} [T].$$

Démonstration. On sait que

$$X_n = \sum_{i=1}^n (\xi_i - \mathbf{E} [\xi_1])$$

est une martingale. On considère la suite de t.d.a presque sûrement bornés $T_k = T \wedge k$ et on applique le théorème d'arrêt qui nous donne

$$\mathbf{E} \left[\sum_{i=1}^{T \wedge k} \xi_i \right] = \mathbf{E} [\xi_1] \mathbf{E} [T \wedge k].$$

Par le théorème de la limite monotone, $\mathbf{E} [T \wedge k]$ tend vers $\mathbf{E} [T]$ lorsque k tend vers $+\infty$.

Si ξ_1 est une variable aléatoire positive, le théorème de la limite monotone permet de conclure que $\mathbf{E} \left[\sum_{i=1}^{T \wedge k} \xi_i \right]$ tend vers $\mathbf{E} \left[\sum_{i=1}^T \xi_i \right]$ et on a donc le résultat voulu.

Dans le cas général, on remarque $\sum_{i=1}^{T \wedge k} \xi_i$ tend p.s. vers $\sum_{i=1}^T \xi_i$ et est majorée par $\sum_{i=1}^T |\xi_i|$ dont on sait par l'étude du cas positif que c'est une variable aléatoire intégrable. Par conséquent, le théorème de convergence dominée permet de conclure. ■

Applications aux marches aléatoires

On considère $(\eta_i, i \geq 1)$ une suite de variables aléatoires IID telles quel

$$\mathbf{P}(\eta_i = 1) = p = 1 - \mathbf{P}(\eta_i = -1).$$

On pose $S_0 = 0$ et $S_n = \sum_{i \leq n} \zeta_i$.

La loi forte des grands nombres stipule que S_n/n tend presque-sûrement vers $p - q$ quand n tend vers l'infini, donc

$$S_n \xrightarrow{n \rightarrow \infty} \begin{cases} +\infty & \text{si } p > q \\ -\infty & \text{si } p < q \\ 0 & \text{si } p = q = \frac{1}{2} \end{cases}$$

Quand $p > q$, on est donc sûr d'atteindre n'importe quel niveau $B > 0$ par contre rien ne garantit qu'on atteindra tout niveau négatif.

Supposons que $p > q$ et que l'on s'intéresse au temps d'atteinte de $B > 0$, i.e.

$$T = \inf\{n > 0, S_n = B\}.$$

Remarquons alors qu'en notant

$$\phi(t) = \mathbf{E} \left[e^{t\zeta_1} \right] = pe^t + qe^{-t}$$

la suite $Z_n = \exp(tS_n - n \ln(\phi(t)))$ est une martingale. En effet,

$$\begin{aligned} \mathbf{E} [Z_{n+1} | \mathcal{F}_n] &= \mathbf{E} \left[e^{tS_n} e^{t\zeta_{n+1}} | \mathcal{F}_n \right] e^{-(n+1) \ln(\phi(t))} \\ &= e^{tS_n} \mathbf{E} \left[e^{t\zeta_{n+1}} \right] e^{-(n+1) \ln(\phi(t))} \\ &= e^{tS_n} e^{-n \ln(\phi(t))} = Z_n \end{aligned}$$

En appliquant le théorème d'arrêt pour le temps d'arrêt $T \wedge n$, on obtient

$$\mathbf{E} \left[\exp(tS_{T \wedge n} - T \wedge n \ln(\phi(t))) \right] = 1.$$

D'après la convexité de la fonction exponentielle,

$$pe^t + qe^{-t} \geq \exp(pt - qt) = \exp((2p - 1)t).$$

Par conséquent, pour tout réel positif t , $\phi(t) \geq 1$. Par ailleurs, la loi forte des grands nombres implique que S_n/n a une limite strictement positive presque sûrement donc que S_n tend vers $+\infty$ p.s., donc T est p.s. fini et $tS_{T \wedge n} - T \wedge n \ln(\phi(t))$ tend vers $tB - T \ln(\phi(t))$ pour presque tout ω . Enfin, remarquons que $S_{T \wedge n} \leq B$ donc d'après le théorème de convergence dominée,

$$\lim_{n \rightarrow +\infty} \mathbf{E} \left[\exp(tS_{T \wedge n} - T \wedge n \ln(\phi(t))) \right] = e^{tB} \mathbf{E} \left[\exp(-T \ln(\phi(t))) \right] = 1,$$

pour tout réel t positif. Soit ψ l'inverse de la fonction $t \mapsto \ln(\phi(t))$, on a donc

$$\mathbf{E} \left[e^{-\lambda T} \right] = e^{-B\psi(\lambda)}.$$

Tous calculs faits, on trouve

$$\psi(\lambda) = \ln \left(\frac{\lambda + \sqrt{\lambda^2 - 4p + 4p^2}}{2p} \right).$$

8

Correction des exercices

Exercice 1.1 ▷

- 1 Si le théorème de Fubini s'applique, on peut intégrer dans l'ordre que l'on veut. Comme l'intégrande est anti-symétrique en x et y , on trouvera une valeur et son opposée donc l'intégrale est nulle.
- 2 Par intégration directe on a

$$\int_0^1 \frac{x^2 - y^2}{(x^2 + y^2)^2} dx = \left[-\frac{x}{x^2 + y^2} \right]_0^1 = -\frac{1}{1 + y^2}.$$

D'autre part,

$$\int_0^1 -\frac{1}{1 + y^2} = -[\arctan(y)]_0^1 = -\frac{\pi}{4} \neq 0.$$

3

Exercice 1.6 ▷

- a) On doit vérifier les trois axiomes qui définissent une tribu.
 - $T^{-1}(\emptyset) = \emptyset$ donc $\emptyset \in \mathcal{I}$.
 - Rappelons que $T^{-1}(A^c) = (T^{-1}(A))^c$ donc si $A \in \mathcal{I}$, $T^{-1}(A^c) = A^c$ soit $A^c \in \mathcal{I}$.
 - De même, pour une famille dénombrable d'ensembles $(A_n, n \geq 1)$,

$$T^{-1}\left(\bigcup_{n \geq 1} A_n\right) = \bigcup_{n \geq 1} T^{-1}(A_n).$$

Par conséquent, si les A_n appartiennent tous à \mathcal{I} ,

$$T^{-1}\left(\bigcup_{n \geq 1} A_n\right) = \bigcup_{n \geq 1} A_n \iff \bigcup_{n \geq 1} A_n \in \mathcal{I}.$$

- b) Nous devons montrer que pour tout $A \in \mathcal{B}(\mathbf{R})$, $f^{-1}(A)$ appartient à \mathcal{I} . Ce qui revient à montrer que $T^{-1} \circ f^{-1}(A) = f^{-1}(A)$. Or par définition d'une fonction invariante $f \circ T = f$ et par définition des images réciproques, $(f \circ T)^{-1} = T^{-1} \circ f^{-1}$.

- 3 Supposons que le système soit ergodique. Soit f une fonction invariante, pour tout $\alpha \in \mathbf{R}$,

$$\mathbf{P}(f^{-1}(] - \infty, \alpha]) = 0 \text{ ou } 1.$$

D'autre part, la fonction

$$G_f : \alpha \mapsto \mathbf{P}(f^{-1}(] - \infty, \alpha])$$

est croissante et continue à droite. Donc il existe α_0 tel que

$$\forall \beta < \alpha, G_f(\beta) = 0 \text{ et } G_f(\beta) = 1, \text{ pour } \beta \geq \alpha.$$

Ceci signifie que $\mathbf{P}(f^{-1}(\{\alpha_0\})) = 1$ donc que $f = \alpha$ \mathbf{P} -presque partout.

Réciproquement, supposons que toutes les fonctions invariantes f sont presque sûrement constantes. Si le système n'est pas ergodique, il existe un ensemble invariant de mesure non nulle :

$$\exists A \in \mathcal{E}, T^{-1}(A) = A \text{ et } \mathbf{P}(A) > 0.$$

Considérons alors la fonction $f = \mathbf{1}_A$. Elle est invariante par T puisque

$$f \circ T = \mathbf{1}_A \circ T = \mathbf{1}_{T^{-1}(A)} = \mathbf{1}_A = f,$$

et prend les valeurs 0 et 1 sur des ensembles de mesure non nulle, donc elle n'est pas constante, d'où une contradiction. Par conséquent, tous les ensembles T -invariants sont de mesure nulle.

- 4 Soit A un ensemble invariant. Appliquons (1.5) à $f = g = \mathbf{1}_A$. Comme $f \circ T = f$, il vient

$$\lim_{n \rightarrow \infty} \int_E \mathbf{1}_A^2 d\mathbf{P} = \left(\int_E \mathbf{1}_A d\mathbf{P} \right)^2,$$

soit $\mathbf{P}(A) = \mathbf{P}(A)^2$ donc $\mathbf{P}(A) = 0$ ou $\mathbf{P}(A) = 1$. Ce qui signifie que le système est ergodique.

- 5 Il suffit de le montrer pour $n = 1$, le résultat en découle par récurrence. Pour $f = \mathbf{1}_A$, c'est l'hypothèse d'invariance de \mathbf{P} par T qui permet d'écrire

$$\int_E \mathbf{1}_A \circ T d\mathbf{P} = \int_E \mathbf{1}_{T^{-1}(A)} d\mathbf{P} = \mathbf{P}(T^{-1}(A)) = \mathbf{P}(A) = \int_E \mathbf{1}_A d\mathbf{P}.$$

Par linéarité, l'équation (1.6) est vraie pour toutes les fonctions étagées positives puis par passage à la limite pour toutes les fonctions mesurables positives. Comme $L^2(\mathbf{P}) \subset L^1(\mathbf{P})$, toute fonction f de L^2 s'écrit $f = f^+ - f^-$ avec f^+ et f^- mesurables, positives et d'intégrale finie. Par différence, l'équation (1.6) est vraie pour tous les éléments de $L^2(\mathbf{P})$.

- 6 Soit D l'ensemble sur lequel (1.5) est vérifiée. Soit f et g des éléments de $L^2(\mathbf{P})$. Pour tout $\epsilon > 0$, il existe f_ϵ et g_ϵ appartenant à D tels que $\|f - f_\epsilon\|_2 < \epsilon$ et $\|g - g_\epsilon\|_2 < \epsilon$.

$$\begin{aligned} \int_E f \circ T^n g \, d\mathbf{P} &= \int_E (f \circ T^n - f_\epsilon \circ T^n) g \, d\mathbf{P} \\ &+ \int_E f_\epsilon \circ T^n (g - g_\epsilon) \, d\mathbf{P} + \int_E f_\epsilon \circ T^n g_\epsilon \, d\mathbf{P} := A_n + B_n + C_n. \end{aligned}$$

Par Cauchy-Schwarz et la question précédente,

$$|A_n| \leq \left(\int_E |(f - f_\epsilon) \circ T^n|^2 \, d\mathbf{P} \right)^{1/2} \|g\|_2 = \|f - f_n\|_2 \|g\|_2 \leq \epsilon \|g\|_2.$$

De même,

$$|B_n| \leq \|f \circ T^n\|_2 \|g - g_\epsilon\|_2 = \|f\|_2 \epsilon.$$

Enfin,

$$\begin{aligned} \left| C_n - \int_E f \, d\mathbf{P} \int_E g \, d\mathbf{P} \right| &\leq \left| \int_E (f_\epsilon - f) \, d\mathbf{P} \int_E g_\epsilon \, d\mathbf{P} \right| + \left| \int_E f \, d\mathbf{P} \int_E (g - g_\epsilon) \, d\mathbf{P} \right| \\ &\leq \|f - f_\epsilon\|_2 \|g_\epsilon\|_2 + \|f\|_2 \|g - g_\epsilon\|_2. \end{aligned}$$

D'après l'inégalité triangulaire, $\|g_\epsilon\|_2 \leq \|g\|_2 + \epsilon$ donc

$$\left| C_n - \int_E f \, d\mathbf{P} \int_E g \, d\mathbf{P} \right| \leq \epsilon(\|g\|_2 + \epsilon + \|f\|_2).$$

Si l'on pose

$$I_n(f, g) = \int_E f \circ T^n g \, d\mathbf{P} - \int_E f \, d\mathbf{P} \int_E g \, d\mathbf{P},$$

on a donc montré que pour tout $n \geq 1$, qu'il existe $c > 0$ tel que

$$|I_n(f, g) - I_n(f_\epsilon, g_\epsilon)| \leq c\epsilon.$$

On choisit n tel que $|I_n(f_\epsilon, g_\epsilon)| \leq \epsilon$ pour $m \geq n$, et alors dans ces conditions, $|I_m(f, g)| \leq (c + 1)\epsilon$.

- 7 La fonction T_1 est appelée *fonction tente* en raison de la forme de son graphe.

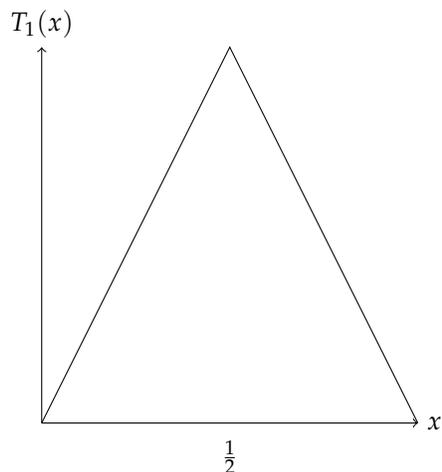
Pour démontrer que λ est invariante par T_1 , il suffit que l'on démontre que

$$\lambda(T_1^{-1}([a, b])) = \lambda([a, b]) = b - a,$$

pour tout a et b . Or

$$T_1^{-1}([a, b]) = [a/2, b/2] \cup [1 - b/2, 1 - a/2],$$

dont la mesure de Lebesgue est bien $b - a$.

FIGURE 8.1: Fonction T_1 .

8 Par définition,

$$e_k(T^n(x)) = \exp(2^n i \pi k x) \mathbf{1}_{[0,1/2]}(x) + \exp(-2^n i \pi k x) \mathbf{1}_{[1/2,1]}(x).$$

Par conséquent,

$$\int_0^1 e_k(T^n(x)) e_l(x) dx = \int_0^{1/2} e_{2^n-1k}(x) e_l(x) dx + \int_{1/2}^1 e_{-2^n-1k}(x) e_l(x) dx.$$

D'après le lemme de Riemman, ces deux intégrales tendent vers 0 quand n tend vers l'infini. On a donc bien

$$\int_0^1 e_k(T^n(x)) e_l(x) dx = 0 = \int_0^1 e_l(x) dx \int_0^1 e_k(x) dx.$$

9 Pour identifier la mesure μ , on procède en utilisant le théorème [??], soit ψ une fonction test, on a

$$\int_0^1 \psi(x) d\mu(x) = \int_0^1 \psi(T_1(x)) dx.$$

L'application Θ est bijective de $[0, 1]$ dans lui-même, on peut donc faire le changement de variable $y = \Theta(x)$ donc

$$dy = \pi \sin(\pi x/2) \cos(\pi x/2) dx = \pi \sqrt{y} \sqrt{1-y} dy.$$

Par conséquent,

$$\int_0^1 \psi(T_1(x)) dx = \int_0^1 \psi(y) \frac{1}{\pi \sqrt{y} \sqrt{1-y}} dy.$$

D'où $d\mu(y) = (\pi \sqrt{y(1-y)})^{-1} dy$.

10 On a le diagramme suivant Supposons que l'on ait démontré que

$$T \circ \Theta = \Theta \circ T_1.$$

Soit A invariant par T , i.e. $T^{-1}(A) = A$, et soit $B = \Theta^{-1}(A)$. On a alors

$$\begin{aligned} T_1^{-1}(B) &= T_1^{-1} \circ \Theta^{-1}(A) = (\Theta \circ T_1)^{-1}(A) = (T \circ \Theta)^{-1}(A) \\ &= \Theta^{-1}(T^{-1}(A)) = \Theta^{-1}(A) = B. \end{aligned}$$

Par conséquent, B est invariant par T_1 donc $\lambda(B) = 0$ ou $\lambda(B) = 1$. Comme μ est la mesure image de λ par Θ ,

$$\lambda(B) = \lambda(\Theta^{-1}(A)) = \mu(A),$$

on a bien $\mu(A) = 0$ ou $\mu(A) = 1$, soit (E, T, μ) ergodique. En vertu de théorème de Birkhoff, on a bien

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mathbf{1}_{[a,b]}(x_n) = \int_a^b \frac{1}{\pi \sqrt{x(1-x)}} dx.$$

Exercice 1.8 ▷

1. La première partie est l'inégalité de Markov. Ensuite

$$\int_{|f| > \|f\|_\infty - \varepsilon} |f|^p d\mu \leq \int |f|^p d\mu = \int |f|^{p-1} |f| d\mu \leq \|f\|_\infty^{p-1} \|f\|_1.$$

2. On passe à la puissance $1/p$ -ième et on prend les limites inf et sup.

$$\begin{aligned} (\|f\|_\infty - \varepsilon) \liminf_p \mu(|f| > \|f\|_\infty - \varepsilon)^{1/p} &\leq \liminf \|f\|_p \\ &\leq \limsup \|f\|_p \leq \limsup \|f\|_\infty^{1-1/p} \|f\|_1^{1/p}. \end{aligned}$$

On sait de la première question que $\mu(|f| > \|f\|_\infty - \varepsilon)$ est fini donc sa puissance $1/p$ -ième tend vers 1. De même pour $(\|f\|_1 / \|f\|_\infty)^{1/p}$.

Exercice 1.9 ▷

1. On fait le changement de variable $s = t/x$.

2.

$$\begin{aligned} \left(\int_0^\infty \left| \frac{1}{x} \int_0^x f(t) dt \right|^p \right)^{1/p} &\leq \left(\int_0^\infty \left| \int_0^1 f(sx) ds \right|^p \right)^{1/p} \\ &\leq \left(\int_0^\infty \left(\int_0^1 |f(sx)| ds \right)^p \right)^{1/p} \\ &\leq \int_0^1 \left(\int_0^\infty |f(sx)|^p dx \right)^{1/p} ds, \end{aligned}$$

d'après l'inégalité intégrale de Minkowski. Maintenant, on refait le changement de variables à l'envers,

$$\left(\int_0^\infty |f(sx)|^p dx \right)^{1/p} = \left(\int_0^\infty |f(u)|^p du \right)^{1/p} s^{-1/p}.$$

Comme

$$\int_0^1 s^{-1/p} ds = \frac{p}{p-1},$$

le résultat en découle.

Exercice 1.10 ▷

1. Par homogénéité, on peut se ramener à le prouver pour $y = 1$. On doit donc montrer que

$$1 + t^p \leq (1 + t^2)^{p/2} \iff f(t) \geq 0$$

pour $t \geq 0$. Or

$$f'(t) = p(t^2 + 1)^{p/2-1}t - pt^{p-1} \geq pt^{2(p/2-1)+1} - pt^{p-1} = 0.$$

Comme $f(0) = 0$, la fonction est positive sur \mathbf{R}^+ d'où le résultat.

2. On applique le résultat précédent à $x = (a + b)/2$ et $y = (a - b)/2$, on obtient

$$\begin{aligned} \left| \frac{a+b}{2} \right|^p + \left| \frac{a-b}{2} \right|^p &\leq \left(\left| \frac{a+b}{2} \right|^2 + \left| \frac{a-b}{2} \right|^2 \right)^{p/2} \\ &= \left(\frac{2a^2 + 2b^2}{4} \right)^{p/2} \\ &= \left(\frac{1}{2}a^2 + \frac{1}{2}b^2 \right)^{p/2} \\ &\leq \frac{1}{2}a^p + \frac{1}{2}b^p, \end{aligned}$$

d'après la convexité de la fonction $t \mapsto t^{p/2}$.

Exercice 2.1 ▷ D'une part,

$$\mathbf{E}[X] = 1/2$$

$$\mathbf{E}[Z] = 0$$

$$\mathbf{E}[Y] = \mathbf{E}[X] \mathbf{E}[Z] = 0$$

$$\mathbf{E}[XY] - \mathbf{E}[X] \mathbf{E}[Y] = \mathbf{E}[XY] = \mathbf{E}[ZX^2] = \mathbf{E}[Z] \mathbf{E}[X^2] = 0,$$

donc X et Y sont décorréélées. En revanche, on remarque que $|Y| = |X|$ donc $\mathbf{P}(X = 1, Y = 0) = 0$ alors que

$$\mathbf{P}(Y = 0) = \mathbf{P}(X = 0) = 1/2 \text{ donc } \mathbf{P}(X = 1)\mathbf{P}(Y = 0) = 1/4.$$

Les deux variables X et Y ne sont donc pas indépendantes.

Exercice 2.2 ▷

1) Dire que \mathbf{P} est la loi uniforme sur $\{0, 1\}^n$ signifie que

$$\mathbf{P}(\{x_1, \dots, x_n\}) = 2^{-n}$$

pour tout $(x_1, \dots, x_n) \in \{0, 1\}^n$. L'événement $(X_i = 1)$ s'écrit aussi

$$(X_i = 1) = \{(x_1, \dots, x_n) \in \{0, 1\}^n, x_i = 1\}.$$

Une seule coordonnée est fixée (en l'occurrence la i -ième) donc le cardinal de cet ensemble est 2^{n-1} . Par définition de la probabilité uniforme sur un ensemble fini (cf. ??)

$$\mathbf{P}(X_i = 1) = \frac{\text{card}(X_i = 1)}{2^n} = \frac{2^{n-1}}{2^n} = \frac{1}{2}.$$

Par passage au complémentaire, on en déduit que $\mathbf{P}(X_i = 0) = 1/2$. Maintenant, soit $\{i_1, \dots, i_k\} \subset \{1, \dots, n\}$ un ensemble d'indices (distincts). On doit prouver que pour tout $(x_1, \dots, x_k) \in \{0, 1\}^k$,

$$\mathbf{P}(X_{i_1} = x_1, \dots, X_{i_k} = x_k) = \mathbf{P}(X_{i_1} = x_1) \dots \mathbf{P}(X_{i_k} = x_k).$$

Or dans l'ensemble $(X_{i_1} = x_1, \dots, X_{i_k} = x_k)$ on a fixé k des n coordonnées donc son cardinal est 2^{n-k} . Par conséquent,

$$\mathbf{P}(X_{i_1} = x_1, \dots, X_{i_k} = x_k) = \frac{2^{n-k}}{2^n} = 2^{-k} = \mathbf{P}(X_{i_1} = x_1) \dots \mathbf{P}(X_{i_k} = x_k),$$

puisque chacun des termes de ce produit vaut $1/2$.

- 2) D'après sa définition, la longueur d'un mot est égale au nombre de ses bits à 1. En l'occurrence, cela revient à compter le nombre d'apparition de 1 dans une suite de n tirages iid où la probabilité d'apparition d'un 1 à chaque place est de $1/2$. Cela correspond à une distribution binomiale de paramètres n et $1/2$ (cf. ??) dont on sait que l'espérance vaut $n/2$.
- 3) On sait aussi que la variance d'une loi binomiale $B(n, p)$ est donnée par $np(1-p)$, soit ici $n/4$.
- 4) Par définition de la distance de Hamming,

$$d(X, Y) = \sum_{i=1}^n \mathbf{1}_{X_i \neq Y_i}.$$

Les variables aléatoires $Z_i = \mathbf{1}_{X_i \neq Y_i}$ sont des variables aléatoires de Bernoulli telles que (d'après la formule des probabilités totales cf. théorème [??])

$$\begin{aligned} \mathbf{P}(Z_i = 0) &= \mathbf{P}(X_i = Y_i) = \mathbf{P}(X_i = Y_i, X_i = 0) + \mathbf{P}(X_i = Y_i, X_i = 1) \\ &= \mathbf{P}(X_i = 0, Y_i = 0) + \mathbf{P}(X_i = 1, Y_i = 1) = \mathbf{P}(X_i = 0)\mathbf{P}(Y_i = 0) + \mathbf{P}(X_i = 1)\mathbf{P}(Y_i = 1), \end{aligned}$$

puisque X_i et Y_i sont indépendantes. Enfin, comme X_i et Y_i sont des Bernoulli de paramètre $1/2$, on obtient

$$\mathbf{P}(Z_i = 0) = 1/2 \cdot 1/2 + 1/2 \cdot 1/2 = 1/2.$$

D'après le lemme [??], les Z_i sont indépendantes. On sait que la variance de la somme de variables aléatoires indépendantes est la somme de leur variance donc

$$\text{var}(d(X, Y)) = \sum_{i=1}^n \text{var}(Z_i) = n/4.$$

Comme $\mathbf{E}[d(X, Y)] = \sum_{i=1}^n \mathbf{E}[Z_i] = n/2$, on tire

$$\mathbf{E}[d(X, Y)^2] = \text{var}(d(X, Y)) + \mathbf{E}[d(X, Y)]^2 = n/4 + n^2/4 = \frac{n(n+1)}{4}.$$

Exercice 2.3 ▷

- a) L'espace probabilisé est $\Omega = \{1, \dots, 6\} \times \{1, \dots, 6\}$ muni de la probabilité uniforme. On note de façon naturelle D_1 et D_2 les deux composantes ou si l'on préfère D_1 représente la valeur du dé numéro 1, D_2 celle du dé numéro 2. On doit calculer $\mathbf{P}(D_1 + D_2 = k)$ pour $k \in \{2, \dots, 12\}$. Par définition de la probabilité uniforme, pour k fixé, il faut calculer le nombre de couples de Ω dont la somme fait k . La probabilité d'obtenir k s'obtient en divisant ce cardinal par le cardinal de Ω qui est 36. Pour $k = 2$, un seul couple convient : $(1, 1)$. Pour $k = 3$, il y en a 2 : $(1, 2)$ et $(2, 1)$. On notera qu'ici les dés sont distingués donc ces deux couples diffèrent. Au final, on obtient

k	2	3	4	5	6	7	8	9	10	11	12
$\mathbf{P}(D_1 + D_2 = k)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Par ailleurs,

$$\mathbf{E}[D_1 + D_2] = 2\mathbf{E}[D_1] = 2 \cdot (1 + 2 + 3 + 4 + 5 + 6)/6 = 7.$$

- b) La variable aléatoire τ représente un premier temps de succès (cf. ??) si l'on considère que « succès » signifie l'obtention d'un 7. D'après la question précédente, la probabilité d'avoir 7 est de $1/6$, donc τ suit une loi géométrique de paramètre $1/6$. D'après l'exercice ??, sa moyenne vaut 6.

- c) Un initié connaît τ donc s'arrête au coup d'avant.
d) Le gain de l'initié s'exprime comme

$$G = \sum_{i=1}^{\tau-1} S_i = \sum_{i=1}^{\tau} S_i - 7, \quad (8.1)$$

puisque par définition de τ , $S_\tau = 7$. Pour calculer son espérance, on ne peut pas utiliser la linéarité puisque le nombre de termes dans la somme est aléatoire. Il faut d'abord décomposer l'espace en sous-ensembles où l'on connaît τ . D'après la formule des probabilités totale (cf. théorème [??])

$$\begin{aligned} & \mathbf{E} \left[\sum_{i=1}^{\tau} S_i \right] \\ &= \sum_{k=1}^{\infty} \mathbf{E} \left[\sum_{i=1}^{\tau} S_i \mathbf{1}_{\tau=k} \right] \\ &= \mathbf{E} [S_1 \mathbf{1}_{S_1=7}] + \sum_{k=2}^{\infty} \mathbf{E} \left[\sum_{i=1}^k S_i \mathbf{1}_{S_1 \neq 7, \dots, S_{k-1} \neq 7, S_k=7} \right] \\ &= \mathbf{E} [S_1 \mathbf{1}_{S_1=7}] \\ &+ \sum_{k=2}^{\infty} \left(\sum_{i=1}^{k-1} \mathbf{E} \left[(S_i \mathbf{1}_{S_i \neq 7}) \mathbf{1}_{S_1 \neq 7} \dots \mathbf{1}_{S_{k-1} \neq 7} \mathbf{1}_{S_k=7} \right] + \mathbf{E} \left[(S_k \mathbf{1}_{S_k=7}) \mathbf{1}_{S_1 \neq 7} \dots \mathbf{1}_{S_{k-1} \neq 7} \right] \right). \end{aligned}$$

Dans chaque terme de la forme

$$\mathbf{E} \left[(S_i \mathbf{1}_{S_i \neq 7}) \mathbf{1}_{S_1 \neq 7} \dots \mathbf{1}_{S_{k-1} \neq 7} \mathbf{1}_{S_k=7} \right],$$

on a un produit de fonctions de chacune des variables S_i donc en vertu de l'indépendance des variable aléatoire r . S_k et du théorème [??], on peut séparer les espérances :

$$\begin{aligned} & \mathbf{E} \left[(S_i \mathbf{1}_{S_i \neq 7}) \mathbf{1}_{S_1 \neq 7} \dots \mathbf{1}_{S_{k-1} \neq 7} \mathbf{1}_{S_k=7} \right] \\ &= \mathbf{E} [(S_i \mathbf{1}_{S_i \neq 7})] \mathbf{P}(S_1 \neq 7) \dots \mathbf{P}(S_{k-1} \neq 7) \mathbf{P}(S_k = 7). \end{aligned}$$

D'une part,

$$\mathbf{E} [S_k \mathbf{1}_{S_k=7}] = \mathbf{E} [7 \mathbf{1}_{S_k=7}] = 7 \mathbf{P}(S_k = 7) = \frac{7}{6}.$$

D'autre part, calculons maintenant $\mathbf{E} [S_i \mathbf{1}_{S_i \neq 7}]$.

$$\mathbf{E} [S_i \mathbf{1}_{S_i \neq 7}] = \mathbf{E} [S_i] - \mathbf{E} [S_i \mathbf{1}_{S_i=7}] = 7 - 7 \mathbf{P}(S_i = 7) = 7 \left(1 - \frac{1}{6}\right) = \frac{35}{6}.$$

On en tire que

$$\mathbf{E} \left[(S_i \mathbf{1}_{S_i \neq 7}) \mathbf{1}_{S_1 \neq 7} \dots \mathbf{1}_{S_{k-1} \neq 7} \mathbf{1}_{S_k=7} \right] = \frac{35}{6} \left(\frac{5}{6}\right)^{k-2} \frac{1}{6}.$$

$$\begin{aligned}
\mathbf{E} \left[\sum_{i=1}^{\tau} S_i \right] &= \frac{7}{6} + \sum_{k=2}^{\infty} \left(\sum_{i=1}^{k-1} \left(\frac{5}{6} \right)^{k-2} \frac{1}{6} \frac{35}{6} + \left(\frac{5}{6} \right)^{k-1} \frac{7}{6} \right) \\
&= \frac{7}{6} + \frac{35}{36} \sum_{k=2}^{\infty} (k-1) \left(\frac{5}{6} \right)^{k-2} + \frac{7}{6} \sum_{k=2}^{\infty} \left(\frac{5}{6} \right)^{k-1} \\
&= \frac{7}{6} + \frac{35}{36} \sum_{k=1}^{\infty} k \left(\frac{5}{6} \right)^{k-1} + \frac{7}{6} \sum_{k=0}^{\infty} \left(\frac{5}{6} \right)^k.
\end{aligned}$$

On reconnaît dans les expressions de la forme

$$\sum_{k=1}^{\infty} k \rho^{k-1}$$

la dérivée de la série géométrique donc

$$\sum_{k=1}^{\infty} k \rho^{k-1} = \frac{d}{d\rho} \sum_{k=0}^{\infty} \rho^k = \frac{d}{d\rho} \frac{1}{1-\rho} = \frac{1}{(1-\rho)^2}.$$

On obtient donc

$$\mathbf{E} \left[\sum_{i=1}^{\tau} S_i \right] = \frac{7}{6} + \frac{35}{36} \frac{1}{(1-5/6)^2} + \frac{35}{36} \frac{1}{1-5/6} = \frac{7}{6} + 35 + \frac{35}{6} = 35 + 7.$$

Au vu de l'équation (8.3), on obtient $\mathbf{E}[G] = 35$.

- e) Conditionnellement à $(X_n = i)$, les valeurs possibles de X_{n+1} sont 0 si $i = 0$ et

$$\{0\} \cup \{i+1, i+2, \dots, i+6, 0, i+8, \dots, i+12\} = \{0\} \cup \{i+B\}$$

où $B = \{2, 3, 4, 5, 6, 8, 9, 10, 11, 12\}$, si $i > 0$. Au vu des rappels ou du corollaire ??,

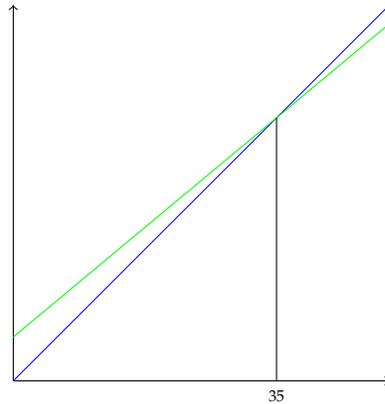
$$\begin{aligned}
\mathbf{E}[X_{n+1} | X_n = i] &= \sum_{j \in i+B} j \mathbf{P}(X_{n+1} = j | X_n = i) \\
&= \sum_{k \in B} (i+k) \mathbf{P}(S_{n+1} = k | X_n = i) \\
&= \sum_{k \in B} (i+k) \mathbf{P}(S_{n+1} = k),
\end{aligned}$$

car X_n qui est une fonction de S_1, \dots, S_n est indépendant de S_{n+1} d'après le lemme [??]. On peut transformer cette quantité en remarquant que B contient toutes les valeurs possibles de S_{n+1} sauf 7

$$\mathbf{E}[X_{n+1} | X_n = i] = i \mathbf{P}(S_{n+1} \neq 7) + \mathbf{E}[S_{n+1} \mathbf{1}_{S_{n+1} \neq 7}] = \frac{5}{6} i + \frac{35}{6},$$

d'après les questions précédentes.

- f) La droite d'équation $y = \frac{5}{6}x + \frac{35}{6}$ croise la première bissectrice en $(35, 35)$. Avant elle est au dessus, après elle est en dessous. Cela signifie que si le score courant est inférieur (respectivement supérieur) à 35, on peut « espérer » s'enrichir (respectivement s'appauvrir). La stratégie consiste donc à jouer tant que l'on n'a pas atteint 35 et à s'arrêter dès que ce palier est atteint.

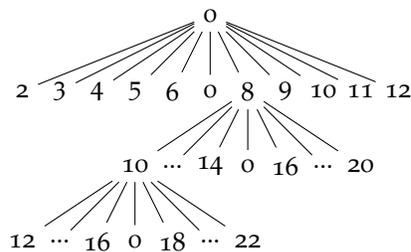


- g) La méthode de la question d) ne permet pas de calculer l'espérance du gain explicitement sous la stratégie optimale définie dans la question précédente. Nous présentons ici deux méthodes pour résoudre numériquement ce calcul.

- Par simulation (cf. ??). On simule un grand nombre de trajectoires du jeu sous cette stratégie : on joue jusqu'à perdre ou atteindre 35 ou plus. On fait la moyenne des gains obtenus et l'intervalle de confiance. Pour $N = 1\,000\,000$ de tirages, on obtient comme intervalle de confiance

$$[14,1378267459162; 14,3732532540876].$$

- La deuxième solution consiste à utiliser l'arbre de toutes les trajectoires possibles.



Chaque branche est pondérée par sa probabilité : ainsi la branche de 0 vers 2 a un poids de $\frac{1}{36}$. On parcourt l'arbre de manière récursive jusqu'à atteindre toutes les feuilles de valeur supérieure

ou égale à 35. Chaque chemin entre la racine et une feuille est pondérée par le produit des poids rencontrés sur le chemin et on fait la somme de tous ces poids sur toutes les feuilles. Evidemment, on n'examine pas les branches qui correspondent à un 7. Cela donne le résultat exact.

```
def score(init,proba,seuil,distrib):
    c=0.
    if (init>seuil):
        return init*proba
    else:
        for j,l in distrib.iteritems():
            c+=score(init+j,proba*l,seuil,distrib)
        return c
```

On notera qu'à travers le paramètre `distrib`, on peut changer la distribution des tirages (par exemple considérer un jeu à deux à 4 faces avec comme nombre interdit 5). L'appel à la fonction se fait donc par le code

```
distrib={2:1/36.,3:2/36.,4:3/36.,5:4/36.,6:5/36.,
8:5/36.,9:4/36.,10:3/36.,11:2/36.,12:1/36.}
scoremoyen=score(0.,1.,35,distrib)
```

Ce qui donne comme résultat

Gain moyen = ??.

Exercice 2.4 ▷

- 1) Pour obtenir une configuration avec k poissons marqués, il faut choisir k poissons parmi les r marqués et $s - k$ poissons parmi les $N - r$ pas marqués. Lorsque l'on tire s poissons dans un lot de N , il y a exactement $\binom{N}{s}$ tirages possibles. On a donc

$$\mathbf{P}(X = k) = \frac{\binom{r}{k} \binom{N-r}{s-k}}{\binom{N}{s}}.$$

- 2) Calculons le rapport des deux termes en deux parties.

$$\begin{aligned} \frac{p_k}{p_{k-1}} &= \frac{r!(N-r)!}{k!(r-k)!(s-k)!(N-r-s+k)!} \\ &\times \frac{(k-1)!(r-(k-1))!(s-(k-1))!(N-r-s+k-1)!}{r!(N-r)!} \\ &= \frac{k(N-r-s+k)}{(r-k+1)(s-k+1)}. \end{aligned}$$

D'autre part,

$$\begin{aligned} \frac{p_k}{p_{k+1}} &= \frac{r!(N-r)!}{k!(r-k)!(s-k)!(N-r-s+k)!} \\ &\quad \times \frac{(k+1)!(r-(k+1))!(s-(k+1))!(N-r-s+k+1)!}{r!(N-r)!} \\ &= \frac{(r-k)(s-k)}{(k+1)(N-r-s+k+1)}. \end{aligned}$$

Par conséquent,

$$\frac{p_k}{p_{k-1}p_{k+1}} = \frac{k}{k+1} \frac{N-r-s+k}{N-r-s+k+1} \frac{s-k}{s-k+1} \frac{r-k}{r-k+1} < 1.$$

3) En passant au logarithme on obtient

$$2 \ln p_k < \ln p_{k-1} + \ln p_{k+1},$$

ce qui signifie que la suite $(\ln p_k, k \geq 0)$ est (strictement) concave (autrement dit la suite d'origine est *log-concave*). D'après les rappels sur les fonctions convexes (cf. théorème [??]), cela signifie que la suite $(\ln p_k, k \geq 0)$ admet un unique maximum. Comme logarithme est une fonction monotone, il en va de même pour la suite $(p_k, k \geq 0)$.

4) Pour que p_k/p_{k-1} soit inférieur à 1, il faut et il suffit que

$$\begin{aligned} k(N-r-s+k) &\leq (r-k+1)(s-k+1) \\ \iff k(N'-r'-s'+k) &\leq (r'-k)(s'-k) \\ \iff N'k &\leq r's'. \end{aligned}$$

Pour que p_k/p_{k+1} soit inférieur à 1, il faut et il suffit que

$$\begin{aligned} (r-k)(s-k) &\leq (k+1)(N-r-s+k+1) \\ \iff (r'-(k+1))(s'-(k+1)) &\leq (k+1)(N'-r'-s'+k+1) \\ \iff N'(k+1) &\geq r's'. \end{aligned}$$

En combinant les deux inégalités, on obtient

$$\frac{r's'}{N'} - 1 \leq k \leq \frac{r's'}{N'}.$$

On en déduit que k_0 est la partie entière de $r's'/N'$.

5) On construit un estimateur \hat{N} selon le raisonnement suivant. Si N est connu, l'effectif que l'on a le plus de chance d'observer est $r's'/(N+2)$. On part alors du principe que si l'on a observé X poissons marqués, c'est que $X = r's'/(N+2)$ d'où $N = r's'/X - 2$.

Exercice 2.5 ▷

- 1) Cet encadrement est connu sous le nom *d'inégalités de Bonferonni* et constitue un raffinement de la formule de Poincaré (ou principe d'inclusion-exclusion).

On remarque que $X = 0$ si et seulement si $X_i = 0$ pour tout i , en d'autres termes si et seulement si $\prod_{i=1}^n (1 - X_i) = 1$: par ailleurs, ce produit, s'il ne vaut pas 1 vaut 0. D'où

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

La quantité *aléatoire* $\prod_{i=1}^n (1 - X_i)$ prend des valeurs du type $\prod_{i=1}^n (1 - x_i)$ où chaque x_i vaut 0 ou 1. Chacune de ces valeurs est donc de la forme $f(1)$ où

$$f : t \mapsto \prod_{i=1}^n (1 - tx_i).$$

Une des formules de Taylor nous dit que

$$f(1) = f(0) + f'(0) + f''(0) + \dots + f^{(r)}(0) + f^{(r+1)}(c) \quad (8.2)$$

où $c \in]0, 1[$. Or comme chaque x_i est dans $[0, 1]$, on constate que le signe de la quantité $f^{(r+1)}(c)$ ne dépend que de la parité de r . Passer aux espérances dans (8.2) donne les inégalités demandées.

- b) On a $E[X_i] = (1 - p)^{n-1} \sim e^{-c}/n$. D'où $E[X] = nE[X_1] \sim e^{-c}$.
 c) On a $F^{(k)} = \binom{n}{k} E[X_1 \cdots X_k]$. De plus,

$$\begin{aligned} E[X_1 \cdots X_k] &= P[\text{les sommets } 1, \dots, k \text{ sont isolés}] \\ &= (1 - p)^{k(n-1) - \binom{k}{2}} \\ &= E[X_1]^k (1 - p)^{-\binom{k}{2}} \end{aligned}$$

Donc, pour k fixé, $F^{(k)} \rightarrow \binom{n}{k} E[X_1]^k \rightarrow \frac{e^{-ck}}{k!}$.

- d) Il s'agit d'utiliser la question préliminaire : on se donne $\varepsilon > 0$ et on choisit r tel que

$$\left| \sum_{k=0}^{2r} (-1)^k \frac{e^{-ck}}{k!} - e^{-e^{-c}} \right| < \frac{\varepsilon}{2}$$

puis un n_0 tel que pour tout $n > n_0$ et $0 \leq k \leq 2r$

$$\left| F^{(k)} - \frac{e^{-ck}}{k!} \right| < \frac{\varepsilon}{2(2k+1)}.$$

On en déduit que pour tout n suffisamment grand

$$\mathbf{P}[X = 0] < e^{-e^{-c}} + \varepsilon.$$

On obtient une borne inférieure sur $P[X = 0]$ de manière similaire. Comme les ε sont arbitraires, on en déduit le résultat.

Commentaire : on pourrait montrer de la même manière que

$$\mathbf{P}(X = j) \rightarrow e^{-e^{-c}} e^{-cj} / j!.$$

La loi de X se rapproche d'une loi de Poisson, ce qui veut dire que les X_i se comportent de manière « de plus en plus indépendantes ».

- e) Il y a $\binom{n}{2}$ paires de sommets. la probabilité qu'une paire de sommets donnée constitue une composante connexe vaut $p(1-p)^{2(n-2)}$. L'espérance du nombre de composantes connexes à deux sommets vaut donc

$$\binom{n}{2} p(1-p)^{2(n-2)} \sim \frac{p}{2} (ne^{-pn})^2 = \frac{p}{2} e^{-2c} \rightarrow 0$$

car p tend vers 0 quand $n \rightarrow \infty$.

- f) On en déduit qu'avec probabilité tendant vers 1 le nombre de composantes connexes à t éléments avec $2 \leq t \leq n/2$ tend vers 0.
- g) Or $G_{n,p}$ n'est pas connexe si et seulement s'il existe une composante connexe à t sommets pour $1 \leq t \leq n/2$. La probabilité d'être non connexe se comporte donc comme la probabilité d'avoir (au moins) un point isolé. Autrement dit, la probabilité que $G_{n,p}$ soit connexe tend vers $e^{-e^{-c}}$. En particulier on en déduit que si p grandit moins vite que $\ln n/n + c/n$ pour tout c , alors $G_{n,p}$ n'est pas connexe avec probabilité tendant vers 1. Par contre si p grandit plus vite que $\ln n/n + c/n$ pour tout c , alors $G_{n,p}$ est connexe avec probabilité tendant vers 1.

Exercice 2.6 ▷ D'après la formule de Markov pour $p = 2$ (cf. proposition ??),

$$\begin{aligned} \mathbf{P}\left(\left|\frac{X_n - \mu_n}{b_n}\right| > \eta\right) &\leq \frac{1}{\eta} \mathbf{E}\left[\left|\frac{X_n - \mu_n}{b_n}\right|^2\right] \\ &= \frac{\sigma_n^2}{\eta b_n^2} \xrightarrow{n \rightarrow \infty} 0, \end{aligned}$$

d'après les hypothèses. D'après la définition de la convergence en probabilités, cela signifie bien que $\frac{X_n - \mu_n}{b_n}$ tend vers 0 en probabilité.

Exercice 2.7 ▷

- 1 Pour $\theta > 0$, la fonction ($x \mapsto \exp(\theta x)$) est strictement croissante donc

$$(X \geq \eta) = (e^{\theta X} \geq e^{\theta \eta}).$$

- 2 D'après l'inégalité de Markov pour $p = 1$ (cf. proposition ??),

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

3 Pour une loi de Poisson de paramètre λ ,

$$\mathbf{E} \left[e^{\theta X} \right] = \sum_{k=0}^{\infty} e^{\theta k} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^{\theta} \lambda)^k}{k!} = \exp(-\lambda + \lambda e^{\theta}).$$

4 Pour η fixé, on doit donc calculer

$$\operatorname{argmin}_{\theta > 0} \exp(-\theta \eta - \lambda + \lambda e^{\theta}) = \operatorname{argmin}_{\theta > 0} (-\theta \eta + \lambda e^{\theta}).$$

Comme

$$\frac{d}{d\theta} (-\theta \eta + \lambda e^{\theta}) = -\eta + \lambda e^{\theta},$$

le point critique est obtenu pour $\theta_0 = \ln(\eta/\lambda)$. La dérivée seconde étant positive, c'est bien un minimum.

5 On obtient donc

$$\ln \mathbf{P}(X \geq \eta) \leq -\theta_0 \eta + \eta - \lambda.$$

Pour $\eta = K\lambda$, $\theta_0 = \ln K$, on a donc

$$\ln \mathbf{P}(X \geq \eta) \leq -\lambda(K \ln K - K + 1)$$

On doit résoudre l'équation

$$K \ln K - K + 1 = -\frac{\ln(0.001)}{\lambda}.$$

Avec le code SAGE suivant, on la résout pour $\lambda = 5, 10, 15, 20, 25$ (la variable `lambda` est un mot réservé en SAGE...)

Listing 8.1 – Calcul de K

```
def f(x,l,epsilon):
    return x*log(x)-x+1+log(epsilon)/l
for i in range(5):
    l=(i+1)*5
    def g(x):
        return f(x,l,0.001)
    print l,find_root(g, l, 4*l, rtol=0.0001)
```

On trouve les résultats suivants :

λ	5	10	15	20	25
K	3,08	2,39	2,10	1,94	1,83

On voit que si l'on doit stocker une loi de Poisson en mémoire, i.e. stocker les valeurs des probabilités $\mathbf{P}(X = k)$, en ne gardant que les valeurs de k jusqu'à trois fois la moyenne (λ), on a une erreur inférieure à 0,1 pour cent. On peut évidemment raffiner en prenant des valeurs de seuil plus basses.

Exercice 2.8 ▷

- 1 La variable aléatoire $T_{i+1} - T_i$ représente le nombre de tablettes nécessaires pour obtenir une nouvelle image sachant que l'on en a déjà i . Soit S_i l'ensemble des i images obtenues, à chaque nouvelle tablette, on trouve une image de cet ensemble avec probabilité $|S_i|/N = i/N$. Le nombre de tablettes nécessaires pour en obtenir une nouvelle suit donc un loi géométrique (cf. ??) : on a un succès lorsque l'image obtenue n'est pas dans S_i , ce qui se produit avec probabilité $1 - i/N$. Au final,

$$\mathbf{P}(T_{i+1} - T_i = k) = \left(\frac{i}{N}\right)^{k-1} \left(1 - \frac{i}{N}\right).$$

- 2 Notons $(X_k, k \geq 1)$ le numéro de l'image obtenue dans la tablette k . Par hypothèse, les variable aléatoire $r. X_k$ sont indépendantes (et de loi uniforme de sur $\{1, \dots, N\}$). L'événement $(T_1 - T_0 = i_1, \dots, T_k - T_{k-1} = i_k)$ dépend des variables aléatoires $X_1, \dots, X_{i_1+\dots+i_k}$. Conditionnellement à cet événement, l'événement $(T_{k+1} - T_k = i_{k+1})$ dépend des variables aléatoires $X_{i_1+\dots+i_{k+1}}$ et suivantes. Ces deux ensembles de variables aléatoires sont disjoints donc en vertu du lemme [??], il y a indépendance conditionnelle :

$$\begin{aligned} \mathbf{P}(T_{k+1} - T_k = i_{k+1} \mid T_1 - T_0 = i_1, \dots, T_k - T_{k-1} = i_k) \\ = \mathbf{P}(T_{k+1} - T_k = i_{k+1}) \mathbf{P}(T_1 - T_0 = i_1, \dots, T_k - T_{k-1} = i_k). \end{aligned}$$

Par récurrence, on en déduit alors que

$$\mathbf{P}(T_1 - T_0 = i_1, \dots, T_N - T_{N-1} = i_N) = \prod_{k=0}^{N-1} \mathbf{P}(T_{k+1} - T_k = i_{k+1}).$$

Ce qui signifie que les variable aléatoire $r. T_0, T_1 - T_0, \dots, T_N - T_{N-1}$ sont indépendantes dans leur ensemble.

- 3 On a la relation évidente

$$T_N = (T_N - T_{N-1}) + (T_{N-1} - T_{N-2}) + \dots + (T_1 - T_0) + T_0.$$

Compte-tenu de la linéarité de l'espérance et de l'exercice ?? où l'on a calculé la moyenne et la variance d'une loi géométrique, on a pour

$$\mathbf{E}[T_{k+1} - T_k] = \frac{1}{1 - k/N} = \frac{N}{N - k},$$

d'où

$$\mathbf{E}[T_N] = N + \frac{N}{2} + \frac{N}{3} + \dots + \frac{N}{N-1} = N \sum_{k=1}^{N-1} \frac{1}{k}.$$

Il est alors supposé connu que

$$\sum_{k=1}^{N-1} \frac{1}{k} \sim \ln(N)$$

pour N tendant vers l'infini. D'où l'équivalent de $\mathbf{E}[T_N]$ qui est $N \ln(N)$. Pour la variance, d'après l'exercice ??,

$$\text{var}(T_{k+1} - T_k) = \frac{1 - (1 - k/N)}{(1 - k/N)^2} = \frac{kN^2}{N(N - k)^2} = N \frac{k}{(N - k)^2}.$$

Comme les variable aléatoire r . sont indépendantes, la variance de la somme est la somme des variance (cf. proposition ??) donc

$$\text{var}(T_N) = N \sum_{k=1}^{N-1} \frac{k}{(N - k)^2} \leq N^2 \sum_{k=1}^{N-1} \frac{1}{k^2},$$

en majorant k par N et en réindexant. Comme la série de terme général $1/k^2$ converge, on en déduit que $\text{var}(T_N) = O(N^2)$.

- 4 Pour pouvoir appliquer le résultat de l'exercice 3.1, il suffit de vérifier que

$$\frac{\text{var}(T_N)}{\mathbf{E}[T_N]^2} \xrightarrow{N \rightarrow \infty} 0.$$

Or d'après la question précédente,

$$\frac{\text{var}(T_N)}{\mathbf{E}[T_N]^2} = O\left(\frac{N^2}{N^2 \ln(N)^2}\right) = O\left(\frac{1}{\ln(N)^2}\right) \xrightarrow{N \rightarrow \infty} 0.$$

Exercice 2.11 ▷

- a) L'utilisation de la fonction génératrice permet de calculer simplement les moments par dérivation (cf. théorème ??). Pour X Bernoulli de paramètre p ,

$$\begin{aligned} \Phi_X(s) &= \mathbf{E}[s^X] = ps^1 + (1 - p)s^0 = 1 - p + ps \\ \mathbf{E}[X] &= \left. \frac{d}{ds} \Phi_X(s) \right|_{s=1} \\ &= p \\ \mathbf{E}[X(X - 1)] &= \left. \frac{d^2}{ds^2} \Phi_X(s) \right|_{s=1} = 0 \\ \text{var}(X) &= \mathbf{E}[X^2] - \mathbf{E}[X]^2 \\ &= \mathbf{E}[X(X - 1)] + \mathbf{E}[X] - \mathbf{E}[X]^2 = 0 + p - p^2 \\ &= p(1 - p). \end{aligned}$$

Pour X géométrique de paramètre p ,

$$\begin{aligned}\Phi_X(s) &= \sum_{k=1}^{\infty} s^k (1-p)^{k-1} p = ps \sum_{k=1}^{\infty} ((1-p)s)^{k-1} \\ &= ps \sum_{k=0}^{\infty} ((1-p)s)^k \\ &= \frac{ps}{1 - (1-p)s} \\ \mathbf{E}[X] &= \frac{p}{(1 - (1-p)s)^2} \Big|_{s=1} \\ &= \frac{1}{p} \\ \mathbf{E}[X(X-1)] &= \frac{2p(1-p)}{(1 - (1-p)s)^3} \Big|_{s=1} = 2 \frac{1-p}{p^2} \\ \text{var}(X) &= \frac{1-p}{p^2}.\end{aligned}$$

Pour la loi de Poisson de paramètre λ ,

$$\begin{aligned}\Phi_X(s) &= \sum_{k=0}^{\infty} s^k e^{-\lambda} \frac{\lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda s)^k}{k!} \\ &= e^{-\lambda} e^{\lambda s} \\ \mathbf{E}[X] &= \lambda e^{-\lambda} e^{\lambda s} \Big|_{s=1} \\ &= \lambda \\ \mathbf{E}[X(X-1)] &= \lambda^2 e^{-\lambda} e^{\lambda s} \Big|_{s=1} = \lambda^2 \\ \text{var}(X) &= \lambda^2 + \lambda - \lambda^2 \\ &= \lambda.\end{aligned}$$

- b) Calculons la fonction génératrice de $\sum_{k=1}^n X_k$. Comme ces variables aléatoires sont indépendantes, en vertu de théorème [??] et de la question précédente, on a

$$\mathbf{E} \left[s^{\sum_{k=1}^n X_k} \right] = \prod_{k=1}^n \mathbf{E} \left[s^{X_k} \right] = \prod_{k=1}^n e^{-\lambda_k(1-s)} = \exp \left(- \sum_{k=1}^n \lambda_k (1-s) \right).$$

On remarque que $\sum_{k=1}^n X_k$ a même fonction génératrice qu'une variable aléatoire de loi de Poisson de paramètre $\sum_{k=1}^n \lambda_k$. Comme la fonction génératrice caractérise la loi (cf. corollaire ??), on en déduit que $\sum_{k=1}^n X_k$ suit une loi de Poisson de paramètre $\sum_{k=1}^n \lambda_k$.

Exercice 3.1 ▷ D'après la formule de Markov pour $p = 2$ (cf. proposi-

tion ??),

$$\begin{aligned} \mathbf{P}\left(\left|\frac{X_n - \mu_n}{b_n}\right| > \eta\right) &\leq \frac{1}{\eta} \mathbf{E}\left[\left|\frac{X_n - \mu_n}{b_n}\right|^2\right] \\ &= \frac{\sigma_n^2}{\eta b_n^2} \xrightarrow{n \rightarrow \infty} 0, \end{aligned}$$

d'après les hypothèses. D'après la définition de la convergence en probabilités, cela signifie bien que $\frac{X_n - \mu_n}{b_n}$ tend vers 0 en probabilité.

Exercice 3.2 ▷

- 1 Pour $\theta > 0$, la fonction ($x \mapsto \exp(\theta x)$) est strictement croissante donc

$$(X \geq \eta) = (e^{\theta X} \geq e^{\theta \eta}).$$

- 2 D'après l'inégalité de Markov pour $p = 1$ (cf. proposition ??),

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

- 3 Pour une loi de Poisson de paramètre λ ,

$$\mathbf{E}[e^{\theta X}] = \sum_{k=0}^{\infty} e^{\theta k} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^{\theta} \lambda)^k}{k!} = \exp(-\lambda + \lambda e^{\theta}).$$

- 4 Pour η fixé, on doit donc calculer

$$\operatorname{argmin}_{\theta > 0} \exp(-\theta \eta - \lambda + \lambda e^{\theta}) = \operatorname{argmin}_{\theta > 0} (-\theta \eta + \lambda e^{\theta}).$$

Comme

$$\frac{d}{d\theta} (-\theta \eta + \lambda e^{\theta}) = -\eta + \lambda e^{\theta},$$

le point critique est obtenu pour $\theta_0 = \ln(\eta/\lambda)$. La dérivée seconde étant positive, c'est bien un minimum.

- 5 On obtient donc

$$\ln \mathbf{P}(X \geq \eta) \leq -\theta_0 \eta + \eta - \lambda.$$

Pour $\eta = K\lambda$, $\theta_0 = \ln K$, on a donc

$$\ln \mathbf{P}(X \geq \eta) \leq -\lambda(K \ln K - K + 1)$$

On doit résoudre l'équation

$$K \ln K - K + 1 = -\frac{\ln(0.001)}{\lambda}.$$

Avec le code SAGE suivant, on la résout pour $\lambda = 5, 10, 15, 20, 25$ (la variable `lambda` est un mot réservé en SAGE...)

Listing 8.2 – Calcul de K

```

def f(x, l, epsilon):
    return x*log(x)-x+1+log(epsilon)/l
for i in range(5):
    l=(i+1)*5
    def g(x):
        return f(x, l, 0.001)
    print l, find_root(g, l, 4*l, rtol=0.0001)

```

On trouve les résultats suivants :

λ	5	10	15	20	25
K	3,08	2,39	2,10	1,94	1,83

On voit que si l'on doit stocker une loi de Poisson en mémoire, i.e. stocker les valeurs des probabilités $\mathbf{P}(X = k)$, en ne gardant que les valeurs de k jusqu'à trois fois la moyenne (λ), on a une erreur inférieure à 0,1 pour cent. On peut évidemment raffiner en prenant des valeurs de seuil plus basses.

Exercice 3.3 ▷

- 1 La variable aléatoire $T_{i+1} - T_i$ représente le nombre de tablettes nécessaires pour obtenir une nouvelle image sachant que l'on en a déjà i . Soit S_i l'ensemble des i images obtenues, à chaque nouvelle tablette, on trouve une image de cet ensemble avec probabilité $|S_i|/N = i/N$. Le nombre de tablettes nécessaires pour en obtenir une nouvelle suit donc un loi géométrique (cf. ??) : on a un succès lorsque l'image obtenue n'est pas dans S_i , ce qui se produit avec probabilité $1 - i/N$. Au final,

$$\mathbf{P}(T_{i+1} - T_i = k) = \left(\frac{i}{N}\right)^{k-1} \left(1 - \frac{i}{N}\right).$$

- 2 Notons $(X_k, k \geq 1)$ le numéro de l'image obtenue dans la tablette k . Par hypothèse, les v.a.r. X_k sont indépendantes (et de loi uniforme de sur $\{1, \dots, N\}$). L'événement $(T_1 - T_0 = i_1, \dots, T_k - T_{k-1} = i_k)$ dépend des variables aléatoires $X_1, \dots, X_{i_1+\dots+i_k}$. Conditionnellement à cet événement, l'événement $(T_{k+1} - T_k = i_{k+1})$ dépend des variables aléatoires $X_{i_1+\dots+i_{k+1}}$ et suivantes. Ces deux ensembles de variables aléatoires sont disjoints donc en vertu du lemme [??], il y a indépendance conditionnelle :

$$\begin{aligned} \mathbf{P}(T_{k+1} - T_k = i_{k+1} \mid T_1 - T_0 = i_1, \dots, T_k - T_{k-1} = i_k) \\ = \mathbf{P}(T_{k+1} - T_k = i_{k+1})\mathbf{P}(T_1 - T_0 = i_1, \dots, T_k - T_{k-1} = i_k). \end{aligned}$$

Par récurrence, on en déduit alors que

$$\mathbf{P}(T_1 - T_0 = i_1, \dots, T_N - T_{N-1} = i_N) = \prod_{k=0}^{N-1} \mathbf{P}(T_{k+1} - T_k = i_{k+1}).$$

Ce qui signifie que les v.a.r. $T_0, T_1 - T_0, \dots, T_N - T_{N-1}$ sont indépendantes dans leur ensemble.

3 On a la relation évidente

$$T_N = (T_N - T_{N-1}) + (T_{N-1} - T_{N-2}) + \dots + (T_1 - T_0) + T_0.$$

Compte-tenu de la linéarité de l'espérance et de l'exercice ?? où l'on a calculé la moyenne et la variance d'une loi géométrique, on a pour

$$\mathbf{E}[T_{k+1} - T_k] = \frac{1}{1 - k/N} = \frac{N}{N - k},$$

d'où

$$\mathbf{E}[T_N] = N + \frac{N}{2} + \frac{N}{3} + \dots + \frac{N}{N-1} = N \sum_{k=1}^{N-1} \frac{1}{k}.$$

Il est alors supposé connu que

$$\sum_{k=1}^{N-1} \frac{1}{k} \sim \ln(N)$$

pour N tendant vers l'infini. D'où l'équivalent de $\mathbf{E}[T_N]$ qui est $N \ln(N)$. Pour la variance, d'après l'exercice ??,

$$\text{var}(T_{k+1} - T_k) = \frac{1 - (1 - k/N)}{(1 - k/N)^2} = \frac{kN^2}{N(N - k)^2} = N \frac{k}{(N - k)^2}.$$

Comme les v.a.r. sont indépendantes, la variance de la somme est la somme des variance (cf. proposition ??) donc

$$\text{var}(T_N) = N \sum_{k=1}^{N-1} \frac{k}{(N - k)^2} \leq N^2 \sum_{k=1}^{N-1} \frac{1}{k^2},$$

en majorant k par N et en réindexant. Comme la série de terme général $1/k^2$ converge, on en déduit que $\text{var}(T_N) = O(N^2)$.

4 Pour pouvoir appliquer le résultat de l'exercice 3.1, il suffit de vérifier que

$$\frac{\text{var}(T_N)}{\mathbf{E}[T_N]^2} \xrightarrow{N \rightarrow \infty} 0.$$

Or d'après la question précédente,

$$\frac{\text{var}(T_N)}{\mathbf{E}[T_N]^2} = O\left(\frac{N^2}{N^2 \ln(N)^2}\right) = O\left(\frac{1}{\ln(N)^2}\right) \xrightarrow{N \rightarrow \infty} 0.$$

Exercice 6.1 ▷

- a) L'espace probabilisé est $\Omega = \{1, \dots, 6\} \times \{1, \dots, 6\}$ muni de la probabilité uniforme. On note de façon naturelle D_1 et D_2 les deux composantes ou si l'on préfère D_1 représente la valeur du dé numéro 1, D_2 celle du dé numéro 2. On doit calculer $\mathbf{P}(D_1 + D_2 = k)$ pour $k \in \{2, \dots, 12\}$. Par définition de la probabilité uniforme, pour k fixé, il faut calculer le nombre de couples de Ω dont la somme fait k . La probabilité d'obtenir k s'obtient en divisant ce cardinal par le cardinal de Ω qui est 36. Pour $k = 2$, un seul couple convient : $(1, 1)$. Pour $k = 3$, il y en a 2 : $(1, 2)$ et $(2, 1)$. On notera qu'ici les dés sont distingués donc ces deux couples diffèrent. Au final, on obtient

k	2	3	4	5	6	7	8	9	10	11	12
$\mathbf{P}(D_1 + D_2 = k)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Par ailleurs,

$$\mathbf{E}[D_1 + D_2] = 2\mathbf{E}[D_1] = 2 \cdot (1 + 2 + 3 + 4 + 5 + 6)/6 = 7.$$

- b) La v.a.r. τ représente un premier temps de succès (cf. ??) si l'on considère que « succès » signifie l'obtention d'un 7. D'après la question précédente, la probabilité d'avoir 7 est de $1/6$, donc τ suit une loi géométrique de paramètre $1/6$. D'après l'exercice ??, sa moyenne vaut 6.
- c) Un initié connaît τ donc s'arrête au coup d'avant.
- d) Le gain de l'initié s'exprime comme

$$G = \sum_{i=1}^{\tau-1} S_i = \sum_{i=1}^{\tau} S_i - 7, \quad (8.3)$$

puisque par définition de τ , $S_\tau = 7$. Pour calculer son espérance, on ne peut pas utiliser la linéarité puisque le nombre de termes dans la somme est aléatoire. Il faut d'abord décomposer l'espace en sous-ensembles où l'on connaît τ . D'après la formule des probabilités totale (cf. théorème [??])

$$\begin{aligned} & \mathbf{E} \left[\sum_{i=1}^{\tau} S_i \right] \\ &= \sum_{k=1}^{\infty} \mathbf{E} \left[\sum_{i=1}^{\tau} S_i \mathbf{1}_{\tau=k} \right] \\ &= \mathbf{E} [S_1 \mathbf{1}_{S_1=7}] + \sum_{k=2}^{\infty} \mathbf{E} \left[\sum_{i=1}^k S_i \mathbf{1}_{S_1 \neq 7, \dots, S_{k-1} \neq 7, S_k=7} \right] \\ &= \mathbf{E} [S_1 \mathbf{1}_{S_1=7}] \\ &+ \sum_{k=2}^{\infty} \left(\sum_{i=1}^{k-1} \mathbf{E} \left[(S_i \mathbf{1}_{S_i \neq 7}) \mathbf{1}_{S_1 \neq 7} \dots \mathbf{1}_{S_{k-1} \neq 7} \mathbf{1}_{S_k=7} \right] + \mathbf{E} \left[(S_k \mathbf{1}_{S_k=7}) \mathbf{1}_{S_1 \neq 7} \dots \mathbf{1}_{S_{k-1} \neq 7} \right] \right). \end{aligned}$$

Dans chaque terme de la forme

$$\mathbf{E} \left[(S_i \mathbf{1}_{S_i \neq 7}) \mathbf{1}_{S_1 \neq 7} \dots \mathbf{1}_{S_{k-1} \neq 7} \mathbf{1}_{S_k = 7} \right],$$

on a un produit de fonctions de chacune des variables S_i donc en vertu de l'indépendance des v.a.r. S_k et du théorème [??], on peut séparer les espérances :

$$\begin{aligned} & \mathbf{E} \left[(S_i \mathbf{1}_{S_i \neq 7}) \mathbf{1}_{S_1 \neq 7} \dots \mathbf{1}_{S_{k-1} \neq 7} \mathbf{1}_{S_k = 7} \right] \\ &= \mathbf{E} \left[(S_i \mathbf{1}_{S_i \neq 7}) \right] \mathbf{P}(S_1 \neq 7) \dots \mathbf{P}(S_{k-1} \neq 7) \mathbf{P}(S_k = 7). \end{aligned}$$

D'une part,

$$\mathbf{E} [S_k \mathbf{1}_{S_k = 7}] = \mathbf{E} [7 \mathbf{1}_{S_k = 7}] = 7 \mathbf{P}(S_k = 7) = \frac{7}{6}.$$

D'autre part, calculons maintenant $\mathbf{E} [S_i \mathbf{1}_{S_i \neq 7}]$.

$$\mathbf{E} [S_i \mathbf{1}_{S_i \neq 7}] = \mathbf{E} [S_i] - \mathbf{E} [S_i \mathbf{1}_{S_i = 7}] = 7 - 7 \mathbf{P}(S_i = 7) = 7 \left(1 - \frac{1}{6}\right) = \frac{35}{6}.$$

On en tire que

$$\mathbf{E} \left[(S_i \mathbf{1}_{S_i \neq 7}) \mathbf{1}_{S_1 \neq 7} \dots \mathbf{1}_{S_{k-1} \neq 7} \mathbf{1}_{S_k = 7} \right] = \frac{35}{6} \left(\frac{5}{6}\right)^{k-2} \frac{1}{6}.$$

$$\begin{aligned} \mathbf{E} \left[\sum_{i=1}^{\tau} S_i \right] &= \frac{7}{6} + \sum_{k=2}^{\infty} \left(\sum_{i=1}^{k-1} \left(\frac{5}{6}\right)^{k-2} \frac{1}{6} \frac{35}{6} + \left(\frac{5}{6}\right)^{k-1} \frac{7}{6} \right) \\ &= \frac{7}{6} + \frac{35}{36} \sum_{k=2}^{\infty} (k-1) \left(\frac{5}{6}\right)^{k-2} + \frac{7}{6} \sum_{k=2}^{\infty} \left(\frac{5}{6}\right)^{k-1} \\ &= \frac{7}{6} + \frac{35}{36} \sum_{k=1}^{\infty} k \left(\frac{5}{6}\right)^{k-1} + \frac{7}{6} \sum_{k=0}^{\infty} \left(\frac{5}{6}\right)^k. \end{aligned}$$

On reconnaît dans les expressions de la forme

$$\sum_{k=1}^{\infty} k \rho^{k-1}$$

la dérivée de la série géométrique donc

$$\sum_{k=1}^{\infty} k \rho^{k-1} = \frac{d}{d\rho} \sum_{k=0}^{\infty} \rho^k = \frac{d}{d\rho} \frac{1}{1-\rho} = \frac{1}{(1-\rho)^2}.$$

On obtient donc

$$\mathbf{E} \left[\sum_{i=1}^{\tau} S_i \right] = \frac{7}{6} + \frac{35}{36} \frac{1}{(1-5/6)^2} + \frac{35}{36} \frac{1}{1-5/6} = \frac{7}{6} + 35 + \frac{35}{6} = 35 + 7.$$

Au vu de l'équation (8.3), on obtient $\mathbf{E} [G] = 35$.

- e) Conditionnellement à $(X_n = i)$, les valeurs possibles de X_{n+1} sont 0 si $i = 0$ et

$$\{0\} \cup \{i+1, i+2, \dots, i+6, 0, i+8, \dots, i+12\} = \{0\} \cup \{i+B\}$$

où $B = \{2, 3, 4, 5, 6, 8, 9, 10, 11, 12\}$, si $i > 0$. Au vu des rappels ou du corollaire ??,

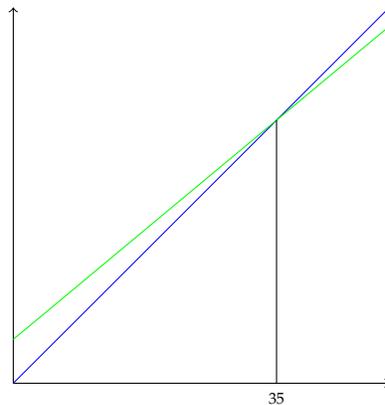
$$\begin{aligned} \mathbf{E}[X_{n+1} | X_n = i] &= \sum_{j \in i+B} j \mathbf{P}(X_{n+1} = j | X_n = i) \\ &= \sum_{k \in B} (i+k) \mathbf{P}(S_{n+1} = k | X_n = i) \\ &= \sum_{k \in B} (i+k) \mathbf{P}(S_{n+1} = k), \end{aligned}$$

car X_n qui est une fonction de S_1, \dots, S_n est indépendant de S_{n+1} d'après le lemme [??]. On peut transformer cette quantité en remarquant que B contient toutes les valeurs possibles de S_{n+1} sauf 7

$$\mathbf{E}[X_{n+1} | X_n = i] = i\mathbf{P}(S_{n+1} \neq 7) + \mathbf{E}[S_{n+1}\mathbf{1}_{S_{n+1} \neq 7}] = \frac{5}{6}i + \frac{35}{6},$$

d'après les questions précédentes.

- f) La droite d'équation $y = \frac{5}{6}x + \frac{35}{6}$ croise la première bissectrice en $(35, 35)$. Avant elle est au dessus, après elle est en dessous. Cela signifie que si le score courant est inférieur (respectivement supérieur) à 35, on peut « espérer » s'enrichir (respectivement s'appauvrir). La stratégie consiste donc à jouer tant que l'on n'a pas atteint 35 et à s'arrêter dès que ce palier est atteint.

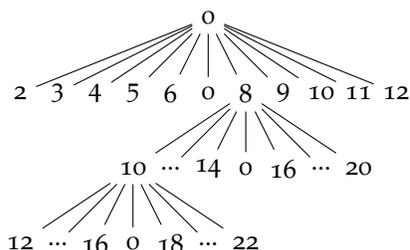


- g) La méthode de la question d) ne permet pas de calculer l'espérance du gain explicitement sous la stratégie optimale définie dans la question précédente. Nous présentons ici deux méthodes pour résoudre numériquement ce calcul.

- Par simulation (cf. ??). On simule un grand nombre de trajectoires du jeu sous cette stratégie : on joue jusqu'à perdre ou atteindre 35 ou plus. On fait la moyenne des gains obtenus et l'intervalle de confiance. Pour $N = 1\,000\,000$ de tirages, on obtient comme intervalle de confiance

[14,1378267459162; 14,3732532540876].

- La deuxième solution consiste à utiliser l'arbre de toutes les trajectoires possibles.



Chaque branche est pondérée par sa probabilité : ainsi la branche de 0 vers 2 a un poids de $1/36$. On parcourt l'arbre de manière récursive jusqu'à atteindre toutes les feuilles de valeur supérieure ou égale à 35. Chaque chemin entre la racine et une feuille est pondérée par le produit des poids rencontrés sur le chemin et on fait la somme de tous ces poids sur toutes les feuilles. Evidemment, on n'examine pas les branches qui correspondent à un 7. Cela donne le résultat exact.

```

def score(init,proba,seuil,distrib):
    c=0.
    if (init>seuil):
        return init*proba
    else:
        for j,l in distrib.iteritems():
            c+=score(init+j,proba*l,seuil,distrib)
        return c
  
```

On notera qu'à travers le paramètre `distrib`, on peut changer la distribution des tirages (par exemple considérer un jeu à deux à 4 faces avec comme nombre interdit 5). L'appel à la fonction se fait donc par le code

```

distrib={2:1/36.,3:2/36.,4:3/36.,5:4/36.,6:5/36.,
8:5/36.,9:4/36.,10:3/36.,11:2/36.,12:1/36.}
scoremoyen=score(0.,1.,35,distrib)
  
```

Ce qui donne comme résultat

$$\text{Gain moyen} = 12,3.$$