UNIVERSITÉ DE LORRAINE

Olivier GARET

Probabilités et Processus Stochastiques

VERSION DE TRAVAIL DU 11 janvier 2016

Table des matières

Chapitre 0

La première gorgée de processus : les variables de Bernoulli

Le premier processus que nous allons étudiée est celui formé par une suite de variables de Bernoulli indépendantes. C'est un modèle simple, mais suffisamment riche pour permettre de voir, ou revoir, un certain nombre de questions importantes de la théorie des probabilités.

0.1 La question de l'existence : de $[0,1]$ à $\{0,1\}^{\mathbb{N}}$

La première question est la question de l'existence. Est-on capable de fabriquer un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ sur lequel vivent une suite de variables indépendantes ? La réponse, positive, est donnée par le théorème suivant :

Théorème 1. *Soit $g \geq 2$ un entier naturel fixé.*

On considère l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1[, \mathcal{B}([0,1[), \lambda_{|[0,1[})$.

Soit $g \geq 2$ un entier. On pose $X_0^g(\omega) = \omega$. On définit les variables A_i^g et X_i^g par les récurrences $X_i^g = \{gX_{i-1}^g\}$ et $A_i^g = \lfloor gX_i \rfloor$. Alors, pour tout $\omega \in [0,1[$, on a

$$\omega = X_0(\omega) = \sum_{i=0}^{+\infty} \frac{A_i^g(\omega)}{g^{i+1}} \text{ avec } A_i^g \in \{0,1,\dots,g-1\}.$$

La suite $A_i^g(\omega)$ contient une infinité de termes différents de $g-1$: c'est le développement g-adique de ω. La suite $(A_i^g)_{i\geq 0}$ est une suite de variables aléatoires indépendantes suivant la loi uniforme sur $\{0,\dots,g-1\}$. En particulier, pour $g=2$, $(A_i^g)_{i\geq 0}$ est une suite de variables aléatoires indépendantes de Bernoulli de paramètre $1/2$.

Démonstration. Par définition de la partie fractionnaire, il est immédiat que la suite des X_i^g prend ses valeurs dans $[0,1[$. Comme $0 \le gX_i^g < g$, il est également clair que A_i^g prend ses valeurs dans $\{0,\dots,g-1\}$. On a $gX_i = \lfloor gX_i \rfloor + \{gX_i\}$, soit $gX_i = A_i + X_{i+1}$, ou encore $\frac{X_i}{g^i} = \frac{A_i}{g^{i+1}} + \frac{X_{i+1}}{g^{i+1}}$. Ainsi

$$\sum_{i=j}^{n} \frac{A_i}{g^{i+1}} = \sum_{i=j}^{n} \left(\frac{X_i}{g^i} - \frac{X_{i+1}}{g^{i+1}} \right) = \frac{X_j}{g^j} - \frac{X_{n+1}}{g^{n+1}}.$$

Soit en faisant tendre n vers l'infini : $\frac{X_j}{g^j} = \sum_{i=j}^{+\infty} \frac{A_i}{g^{i+1}}$. En particulier, pour $j = 0$, on obtient l'écriture voulue. Reste à voir que A_i ne peut être constamment égal à $j-1$ à partir d'un certain rang. En effet, si on avait $A_i = g-1$ pour $i > j$, on aurait $X_j = g^j \sum_{i=j}^{+\infty} \frac{g-1}{g^{i+1}} = 1$, ce qui est exclu, car $X_j \in [0,1[$.

On va montrer par récurrence que pour tout n, on a H_n :

— $(A_0,\dots,A_n)$ et X_{n+1} sont indépendants

— $(A_0,\dots,A_n)$ suit la loi uniforme sur $\{0,\dots,g-1\}^{n+1}$

— X_{n+1} suit la loi uniforme sur $[0,1]$.

Notons d'abord que pour tout $n \ge 1$, on a

$$\begin{aligned}\{A_{n-1} = b_n, X_n \in J\} &= \{\lfloor gX_{n-1} \rfloor = b_n, \{gX_{n-1}\} \in J\} \\ &= \{gX_{n-1} \in J + b_n\} = \left\{ X_{n-1} \in \frac{J+b_n}{g} \right\}\end{aligned}$$

Ainsi pour $n = 1$, on a

$$\mathbb{P}(A_0 = g_0, X_1 \in J) = \mathbb{P}(X_{n-1} \in \frac{J+b_n}{g}) = \lambda(\frac{J+b_n}{g}) = \frac{1}{g}\lambda(J),$$

ce qui montre que H_0 est vraie. Ensuite, on procède par récurrence : Comme

$$\begin{aligned}&\{A_0 = b_0, \dots, A_n = b_n, A_{n+1} = b_{n+1}, X_{n+1} \in J\} \\ &= \{A_0 = b_0, \dots, A_n = b_n, X_n \in \frac{J + b_{n+1}}{g}\},\end{aligned}$$

l'hypothèse de récurrence nous donne

$$\begin{aligned}&\mathbb{P}(A_0 = b_0, \dots, A_n = b_n, A_{n+1} = b_{n+1}, X_{n+1} \in J) \\ &= \mathbb{P}(A_0 = b_0, \dots, A_n = b_n)\mathbb{P}(X_n \in \frac{J+b_n}{g}) \\ &= \frac{1}{g^{n+1}}\lambda(\frac{J+b_n}{g}) = \frac{1}{g^n}\frac{1}{g}\lambda(J) \\ &= \frac{1}{g^{n+2}}\lambda(J)\end{aligned}$$

ce qui montre que l'hypothèse est vérifiée au rang $n+1$.

Ainsi pour tout $n \geq 1$ $(A_0, \dots A_{n-1})$ suit la loi uniforme sur $\{0, \dots, g-1\}^n$. Cependant la loi uniforme sur un produit d'ensembles finis, c'est le produit des lois uniformes sur les ensembles fini : les variables $(A_0, \dots A_{n-1})$ sont des variables aléatoires indépendantes suivant la loi uniforme sur $\{0, \dots, g-1\}$. Comme c'est vrai pour tout n, la suite $(A_n)_{n\geq 0}$ est une suite de variables aléatoires suivant la loi uniforme sur $\{0, \dots, g-1\}$. □

0.2 De $\{0,1\}^{\mathbb{N}}$ à $[0,1]$: où l'on a envie des processus

La section précédente nous a démontré que si un espace était suffisamment riche pour porter une variable aléatoire suivant la loi uniforme sur $[0,1]$, alors il supportait une suite de pile ou face indépendantes. Autrement dit, si on sait simuler une variable aléatoire suivant la loi uniforme sur $[0,1]$, on sait simuler une suite de pile ou face indépendantes. Le théorème qui suit montre que la réciproque est vraie.

Théorème 2. *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace probabilisé et* $(Y_n)_{n\geq 0}$ *une suite de variables de Bernoulli de paramètre* $1/2$ *sur cet espace. Alors, la variable aléatoire* V *définie par* $V = \sum_{i=0}^{+\infty} \frac{Y_i}{2^{i+1}}$ *suit la loi uniforme sur* $[0,1]$

Démonstration. La convergence de la série est évidente. Il suffit donc de caractériser la loi de V. Notons $\Psi_n(x_0, \dots, x_n) = \sum_{i=0}^{n-1} \frac{x_i}{2^{i+1}}$ et $V_n = \sum_{i=0}^{n-1} \frac{Y_i}{2^{i+1}}$. Comme $V_n = \psi_n((X_0, \dots, X_{n-1}))$, si on note $\gamma_n = \mathrm{ber}(1/2)^{\otimes n}$ et μ_n la loi de V_n, comme γ_n est la loi de $(X_0, \dots, X_{n-1})$, on peut dire que μ_n est la loi image de γ_n par ψ_n.

Revenons à la suite de la section précédente : si l'on pose $S_n = \sum_{i=0}^{n-1} \frac{A_i^2}{2^{i+1}}$, de sorte que $S_n = \psi_n(A_0^2, \dots, A_g^n)$. Comme γ_n est la loi de $(X_0, \dots, X_{n-1})$, la loi de S_n sous $\lambda_{|[0,1[}$ est également μ_n : pour toute fonction continue bornée

$$\int_{[0,1]} f(S_n(\omega)) \; d\lambda(\omega) = \int_{[0,1]} f(x) \; d\mu_n(x).$$

Nous savons que $S_n(\omega)$ tend vers ω pour tout $\omega \in [0,1[$. La convergence presque sûre entraîne la convergence en loi, donc

$$\lim_{n\to+\infty} \int_{[0,1]} f(S_n(\omega)) \; d\lambda(\omega) = \int_{[0,1]} f(x) \; d\lambda(x),$$

soit

$$\lim_{n\to+\infty} \int_{[0,1]} f(\omega) \; d\mu_n(\omega) = \int_{[0,1]} f(x) \; d\lambda(x),$$

Mais comme μ_n est la loi de V_n, on vient de montrer que V_n converge en loi vers $U[0,1]$. Comme V_n converge presque sûrement vers V, V_n converge en loi vers V, donc la loi de V est la loi uniforme sur $[0,1]$. □

La preuve est un peu compliquée. On aurait envie de faire plus court et de dire : d'après le Théorème 1 la loi image de $\mathrm{ber}(1/2)^{\otimes\mathbb{N}}$ par

$$x \mapsto \psi(x) = \sum_{i=0}^{+\infty} \frac{x_i}{2^{i+1}}$$

est $U[0,1]$. Or, la loi de $(Y_n)_{n\geq 1}$ est $\mathrm{ber}(1/2)^{\otimes\mathbb{N}}$, donc $V = \psi((X_n)_{n\geq 0}$ est $U[0,1]$. Ce genre de raisonnement sera facile avec la notion de loi d'un processus, que nous verrons au chapitre 8.

Mais dès à présent, nous pouvons en donner comme conséquence le résultat suivant :

Théorème 3. *Sur* $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1[, \mathcal{B}([0,1[), \lambda_{|[0,1[})$*, on peut faire vivre une suite de variables indépendantes suivant la loi uniforme sur* $[0,1]$*.*

Démonstration. Il suffit de poser $Z_j = \sum_{i=0}^{+\infty} \frac{A^2_{(2j+1)2^i}}{2^i}$ et d'appliquer consécutivement les deux théorèmes précédents. □

0.3 Inégalités ; lois des grands nombres

Les variables de Bernoulli de paramètre 1/2 ont leurs soeurs jumelles : les variables de Rademacher qui valent 1 et -1 avec probabilité 1/2. Ainsi X suit une loi de Bernoulli de paramètre 1/2 si et seulement si $Y = 2X - 1$ est une variable de Rademacher.

Les sommes $B_n = X_1 + \dots X_n$ et $S_n = Y_1 + \dots + Y_n$ sont liées par $S_n = 2B_n - n$. Ainsi, les (S_n), qui forment une marche aléatoire sur $\mathbb{Z}$, sont liées à la loi binomiale. On transfère ainsi couramment et facilement les résultats de l'un vers l'autre.

On sait par exemple que si les (Y_i) sont des Rademacher indépendantes, la loi des grands nombres dit que $S_n = Y_1 + \dots + Y_n$ vérifie S_n/n tend vers 0 presque sûrement. Peut-on faire mieux ? Oui.

Théorème 4. *Soient* $(Y_i)_{i\geq 1}$ *des variables de Rademacher indépendantes. Pour tout* $\alpha > 1/2$*,*

$$\lim_{n\to+\infty} \frac{Y_1 + \dots + Y_n}{n^\alpha} = 0.$$

On va s'appuyer sur un lemme :

Lemme 1. *Soient $(Y_i)_{i\geq 1}$ des variables de Rademacher indépendantes. On pose $S_n = Y_1 + \cdots + Y_n$. Alors pour tout $x > 0$ et tout $n \geq 0$:*

$$\mathbb{P}(|S_n - \mathbb{E}S_n| > x) \leq 2\exp\left\{-\frac{x^2}{2n}\right\}.$$

Démonstration. Rappelons l'inégalité de Hoeffding (voir par exemple Garet–Kurtzmann, page 346)

Proposition 1. *Soit $(X_n)_n$ une suite de variables aléatoires réelles indépendantes. Supposons qu'il existe deux suites de réels $(a_n)_{n\geq 0}$ et $(b_n)_{n\geq 0}$ telles que, pour tout $n \geq 0$, $a_n < b_n$ et*

$$\mathbb{P}(a_n \leq X_n \leq b_n) = 1.$$

Posons $S_n = X_1 + \ldots + X_n$. On a alors pour tout $x > 0$ et tout $n \geq 0$:

$$\mathbb{P}(|S_n - \mathbb{E}S_n| > x) \leq 2\exp\left\{-\frac{2x^2}{\sum_{i=1}^n (b_i - a_i)^2}\right\}.$$

Ici, on peut prendre les bornes $a_n = -1$ et $b_n = 1$, ce qui nous donne

$$\mathbb{P}(|S_n| > x) \leq 2\exp(-\frac{2x^2}{4n}).$$

On verra plus loin au chapitre 5 une forme plus générale de l'inégalité de Hoeffding, qui s'applique à certaines variables dépendantes.

Mais pour l'heure, si on ne veut pas admettre l'inégalité de Hoeffding, il est possible, dans ce cas particulier, de donner une preuve plus simple : on a, pour tout $\alpha > 0$

$$\mathbb{P}(|S_n| > x) \leq 2\mathbb{P}(S_n > x) \leq 2\mathbb{P}(e^{\alpha S_n} > e^{\alpha x}) \leq 2\frac{\mathbb{E}e^{\alpha S_n}}{e^{\alpha x}} = \frac{(\mathbb{E}e^{\alpha c_1})^n}{e^{\alpha x}}.$$

On a

$$\mathbb{E}e^{\alpha c_1} = \cosh\alpha = \sum_{k=0}^{+\infty} \frac{\alpha^{2k}}{(2k)!} \leq \sum_{k=0}^{+\infty} \frac{\alpha^{2k}}{2^k k!} = \exp(\frac{\alpha^2}{2}),$$

d'où

$$\mathbb{P}(|S_n| > x) \leq 2\exp(-f_{n,x}(\alpha)) \text{ avec } f_{n,x}(\alpha) = \alpha x - n\frac{\alpha^2}{2}.$$

Il faut évidemment rendre $f_{n,x}(\alpha)$ maximal, ce qui est facile puisque c'est un polynôme du second degré : on prend $\alpha = x/n$, d'où

$$\mathbb{P}(|S_n| > x) \leq 2\exp(-\frac{x^2}{2n}).$$

□

On peut maintenant passer à la preuve du théorème :

Démonstration. D'après le lemme,

$$\mathbb{P}(|S_n| > n^{1/2+\varepsilon}) \leq 2\exp(-\frac{n^{2\varepsilon}}{2}).$$

La série de terme général $\exp(-\frac{n^{2\varepsilon}}{2})$ converge (par exemple parce que $\frac{n^{2\varepsilon}}{2} \geq 2\log n$ pour n assez grand), donc d'après le lemme de Borel-Cantelli, presque sûrement $|S_n| \leq n^{1/2+\varepsilon}$ pour n assez grand, ce qui donne le résultat voulu. □

0.4 Variables de Rademacher et séries de Dirichlet

0.4.1 Une série de Dirichlet aléatoire

Théorème 5. *Soit $(c_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes identiquement distribuées avec $\mathbb{P}(c_n = 1) = \mathbb{P}(c_n = -1) = \frac{1}{2}$. Alors la série de terme général $\frac{c_n}{n^s}$ converge presque sûrement pour $s > \frac{1}{2}$.*

Démonstration. Il s'agit de mettre ensemble deux résultats :
— un résultat d'analyse sur les séries de Dirichlet : si les sommes partielles $s_n = \sum_{k=1}^n c_k$ vérifient $s_n = O(n^\alpha)$ pour un certain $\alpha \geq 0$, alors la série de Dirichet $\sum \frac{c_n}{n^s}$ converge pour $s > \alpha$. [1]
— un résultat de probabilité : dans le cas qui nous intéresse, on a presque sûrement $s_n = O(n^{1/2+\varepsilon})$ pour tout $\varepsilon > 0$.

Le résultat de probabilité a été montré plus haut. Il n'y a plus qu'à montrer le résultat d'analyse : On va montrer que la série de terme général $\frac{c_k}{k^{\alpha+\varepsilon}}$ converge dès que $M = \sup_{n\geq 1} |s_n n^{-\alpha}| < +\infty$.

$$\begin{aligned}
\sum_{k=1}^n \frac{c_k}{k^{\alpha+\varepsilon}} &= \sum_{k=1}^n (s_k - s_{k-1})\frac{1}{k^{\alpha+\varepsilon}} \\
&= \sum_{k=1}^n s_k \frac{1}{k^{\alpha+\varepsilon}} - \sum_{k=0}^{n-1} s_k \frac{1}{(k+1)^{\alpha+\varepsilon}} \\
&= \frac{s_n}{n^{\alpha+\varepsilon}} + \sum_{k=1}^{n-1} s_k(\frac{1}{k^{\alpha+\varepsilon}} - \frac{1}{(k+1)^{\alpha+\varepsilon}})
\end{aligned}$$

1. Vous avez sans doute déjà rencontré le cas où $\alpha = 0$ et $c_n = e^{in\theta}$ avec θ non congru à 0 modulo 2π.

Comme $\lim_{n\to+\infty}\frac{s_n}{n^{\alpha+\varepsilon}}=0$, la série de terme général $\frac{c_k}{k^{\alpha+\varepsilon}}$ est de même nature que la série de terme général $s_k(\frac{1}{k^{\alpha+\varepsilon}}-\frac{1}{(k+1)^{\alpha+\varepsilon}})$. Montrons que cette dernière converge absolument. D'après l'inégalité des accroissements finis, on a

$$|\frac{1}{k^{\alpha+\varepsilon}}-\frac{1}{(k+1)^{\alpha+\varepsilon}}|\le\frac{\alpha+\varepsilon}{k^{\alpha+\varepsilon+1}}.$$

Comme $|s_k|\le Mk^\alpha$, on a alors

$$|s_k(\frac{1}{k^{\alpha+\varepsilon}}-\frac{1}{(k+1)^{\alpha+\varepsilon}})|\le\frac{M(\alpha+\varepsilon)}{k^{1+\varepsilon}},$$

ce qui assure la convergence voulue. □

Remarquons que si on connait la théorie des séries de variables aléatoires indépendantes, on a une preuve beaucoup plus rapide, qui n'utilise pas la transformation d'Abel : les variables $\frac{c_n}{n^s}$ sont centrées, et le terme général de la série des variances est $\frac{1}{4n^{2s}}$, qui converge pour $s>1/2$, donc la série converge presque sûrement (et aussi dans L^2). Voir par exemple Garet–Kurtzmann, page 344.

Remarque culturelle : notons $\mu(n)$ la fonction de Moebius, définie par $\mu(n)=(-1)^k$ si n est le produit de k nombre premiers distincts, 0 sinon. si vous parvenez à démontrer que pour la suite $c_n=\mu(n)$, on a $s_n=O(n^{1/2+\varepsilon})$ pour tout $\varepsilon>0$, et donc que la série des $\frac{\mu(n)}{n^s}$ converge pour $s>1/2$, alors vous avez démontré un résultat équivalent à la fameuse conjecture de Riemann : les zéros non-triviaux de la fonction ζ de Riemann sont tous de partie réelle 1/2. Pour voir que cette propriété entraîne la conjecture de Riemann, voir par exemple Colmez, pages 319-320.

0.4.2 Comportement au bord

Théorème 6. *Soit $(c_n)_{n\ge1}$ une suite de variables aléatoires indépendantes identiquement distribuées avec $\mathbb{P}(c_n=1)=\mathbb{P}(c_n=-1)=\frac12$ On pose, pour $s>1/2$, $\zeta^*(s)=\sum_{n=0}^{+\infty}\frac{c_n}{n^s}$. Alors, on a la convergence en loi*

$$\sqrt{2h}\zeta^*(\frac12+h)\Longrightarrow\mathcal{N}(0,1)\ \text{quand } h \text{ tend vers } 0 \text{ par valeurs positives}$$

Démonstration. Posons $S_N(s)=\sum_{n=0}^{N}\frac{c_n}{n^s}$. La fonction caractéristique de $S_N(s)$ vaut $t\mapsto\prod_{n=0}^{N}\cos(\frac{t}{n^s})$. Comme $e^{itS_n(s)}$ converge presque sûrement vers $e^{it\zeta^*(s)}$, le théorème de convergence dominée nous donne

$$\varphi_{\zeta^*(s)}(t)=\prod_{n=0}^{+\infty}\cos(\frac{t}{n^s}),$$

d'où

$$\varphi_{\frac{\zeta^*(1/2+h)}{\sqrt{\zeta(1+2h)}}}(t) = \prod_{n=0}^{+\infty} \cos(\frac{t\zeta(1+2h)^{-1/2}}{n^{1/2+h}})$$

Par ailleurs

$$\exp(-\frac{t^2}{2}) = \prod_{n=0}^{+\infty} \exp(-\frac{t^2}{2}\frac{\zeta(1+2h)^{-1}}{n^{1+2h}})$$

On en déduit

$$|\varphi_{\frac{\zeta^*(1/2+h)}{\sqrt{\zeta(1+2h)}}}(t) - \exp(-\frac{t^2}{2})| \leq \sum_{n=0}^{+\infty} |\cos(\frac{t\zeta(1+2h)^{-1/2}}{n^{1/2+h}}) - \exp(-\frac{t^2}{2}\frac{\zeta(1+2h)^{-1}}{n^{1+2h}})|.$$

Mais il existe A tel que pour tout réel x $|\cos(x) - \exp(-\frac{x^2}{2})| \leq Ax^4$. On en déduit pour $h \in]0,1]$:

$$|\varphi_{\frac{\zeta^*(1/2+h)}{\sqrt{\zeta(1+2h)}}}(t) - \exp(-\frac{t^2}{2})| \leq \frac{At^2}{\zeta(2+2h)}\zeta(1+2h)^{-2} \leq \frac{At^2}{\zeta(4)}\zeta(1+2h)^{-2}.$$

Il est bien connu que $\zeta(1+h) \sim h^{-1}$ au voisinage de 0 : on en déduit la convergence de $\varphi_{\frac{\zeta^*(1/2+h)}{\sqrt{\zeta(1+2h)}}}(t)$ vers $\exp(-\frac{t^2}{2})$, et donc, d'après le théorème de Levy, la convergence en loi de $\frac{\zeta^*(1/2+h)}{\sqrt{\zeta(1+2h)}}$ vers $\mathcal{N}(0,1)$, puis, avec l'équivalent $\zeta(1+2h) \sim (2h)^{-1}$, la convergence en loi voulue.

Donnons une preuve courte de l'équivalent $\zeta(1+h) \sim h^{-1}$ au voisinage de 0 : avec l'inégalité des accoissements finis, on a, pour $h > 0$

$$|\sum_{k=1}^{+\infty} \frac{1}{k^{1+h}} - \sum_{k=1}^{+\infty} \int_k^{k+1} \frac{1}{t^{1+h}}\, dt| \leq \sum_{k=1}^{+\infty} \frac{1}{k^{2+h}} \leq \zeta(2),$$

d'où

$$\zeta(1+h) = \frac{1}{h} + O(1).$$

□

0.5 Exercices sur les variables de Bernoulli

0.5.1 Exercices corrigés

Exercice 1. Considérons une suite $(\omega_n)_{n\geq 1}$ de variables aléatoires indépendantes suivant la loi de Bernoulli de paramètre $p \in]0,1[$. Soit X la variable aléatoire définie par :

$$X = \sum_{n=1}^{\infty} \frac{\omega_n}{2^n}.$$

On note ν_p la loi de X. Pour quelle(s) valeur(s) de p la variable X admet-elle une densité par-rapport à la mesure de Lebesgue ? lien vers l'indication lien vers la solution

Chapitre 1

Équi-intégrabilité

Définition. *On dit qu'une famille $\mathcal{A}$ de variables aléatoires définies sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ est équi-intégrable (ou uniformément intégrable) si*

$$\lim_{M\to+\infty}\sup_{X\in\mathcal{A}}\mathbb{E}[|X|\ \mathbb{1}_{\{|X|\geq M\}}] = 0.$$

1.1 Premières propriétés

Remarque 1.
- *Une famille constituée d'une seule variable intégrable est équi-intégrable.*
- *La réunion de deux familles équi-intégrables est équi-intégrable. Par suite une famille finie de variables intégrables est équi-intégrable.*
- *Une famille équi-intégrable est toujours bornée dans L^1.*
  En effet, si M est choisi tel que $\sup_{X\in\mathcal{A}}\mathbb{E}[|X|\ \mathbb{1}_{\{|X|\geq M\}}] \leq 1$, alors comme $|X| \leq M + |X|\ \mathbb{1}_{\{|X|\geq M\}}$, on a $\mathbb{E}[|X|] \leq M + 1$ pour tout $X \in \mathcal{A}$.
- *Si la famille $\mathcal{A}$ est équi-intégrable et que pour tout $Y \in \mathcal{B}$, il existe $X \in \mathcal{A}$ avec $|Y| \leq |X|$, alors la famille $\mathcal{B}$ est équi-intégrable.*
- *Si la famille $\mathcal{A}$ est équi-intégrable, la famille $(\max(|X|, |Y|))_{(X,Y)\in\mathcal{A}^2}$ est équi-intégrable.*
  En effet, $\max(|X|, |Y|)\ \mathbb{1}_{\{\max(|X|,|Y|\})\geq M\}} \leq |X|\ \mathbb{1}_{\{|X|\geq M\}} + |Y|\ \mathbb{1}_{\{|Y|\geq M\}}$ *entraîne*

  $$\mathbb{E}[\max(|X|, |Y|)\ \mathbb{1}_{\{\max(|X|,|Y|\})\geq M\}}] \leq \mathbb{E}[|X|\ \mathbb{1}_{\{|X|\geq M\}}] + \mathbb{E}[|Y|\ \mathbb{1}_{\{|Y|\geq M\}}].$$

- *Par suite, si la famille $\mathcal{A}$ est équi-intégrable, la famille $(X+Y)_{(X,Y)\in\mathcal{A}^2}$ est équi-intégrable.*
  En effet, il suffit de remarquer que $|X + Y| \leq 2\max(|X|, |Y|)$ et d'appliquer les remarques précédentes.

Le résultat principal est le suivant.

Théorème 7. *Soit $(X_n)_{n\geq 1}$ une suite équi-intégrable de variables aléatoires. On suppose que X_n converge en loi vers X lorsque n tend vers l'infini. Alors X est intégrable et la suite $(\mathbb{E}X_n)_{n\geq 1}$ converge vers $\mathbb{E}X$.*

Pour ce résultat, on va avoir besoin d'un lemme intermédiaire.

Lemme 2. *Si $(X_n)_{n\geq 1}$ converge en loi vers X, alors $\mathbb{E}|X| \leq \underline{\lim}_{n\to+\infty}\mathbb{E}|X_n|$.*

Démonstration. Comme $|X_n|$ converge en loi vers $|X|$, on peut se ramener au cas où les variables X_n sont positives. On a pour tout n,

$$\mathbb{E}X_n = \int_{[0,+\infty[} \mathbb{P}(X_n > t)\ d\lambda(t).$$

On sait que $\mathbb{P}(X_n > t)$ converge vers $\mathbb{P}(X > t)$ en tous les points de continuité de F_X. Or les points de discontinuité de F_X sont au plus dénombrables, donc $\mathbb{P}(X_n > t)$ converge λ-presque partout vers $\mathbb{P}(X > t)$. On peut donc appliquer le lemme de Fatou

$$\begin{aligned}
\mathbb{E}X = \int_{[0,+\infty[} \mathbb{P}(X > t)\ d\lambda(t) &= \int_{[0,+\infty[} \underline{\lim}_{n\to+\infty}\mathbb{P}(X_n > t)\ d\lambda(t) \\
&\leq \underline{\lim}_{n\to+\infty} \int_{[0,+\infty[} \mathbb{P}(X_n > t)\ d\lambda(t) \\
&= \underline{\lim}_{n\to+\infty}\mathbb{E}X_n.
\end{aligned}$$

□

Remarque 2. *Certains auteurs invoquent ce théorème sous le nom de "lemme de Fatou".*

On peut maintenant passer à la démonstration du théorème.

Démonstration du théorème 7. D'après le lemme X est intégrable, donc $\{X_n; n \geq 1\} \cup \{X\}$ est équi-intégrable. ε étant fixé, on peut trouver M tel que

$$\sup_{n\geq 1}\mathbb{E}[X_n\ \mathbb{1}_{\{X_n\geq M\}}] \leq \varepsilon \text{ et } \mathbb{E}[X\mathbb{1}_{\{X\geq M\}}] \leq \varepsilon. \tag{1.1}$$

Notons que

$$\begin{aligned}
&\mathbb{E}[X_n] = \mathbb{E}[X_n \wedge M] + \mathbb{E}[(X_n - M)\mathbb{1}_{\{X_n\geq M\}}] \\
\text{et}\quad &\mathbb{E}[X] = \mathbb{E}[X \wedge M] + \mathbb{E}[(X - M)\mathbb{1}_{\{X\geq M\}}].
\end{aligned}$$

Comme $0 \leq (X_n - M)\mathbb{1}_{\{X_n\geq M\}} \leq X_n\mathbb{1}_{\{X_n\geq M\}}$ et
$0 \leq (X - M)\mathbb{1}_{\{X\geq M\}} \leq X\mathbb{1}_{\{X\geq M\}}$, on en déduit que

$$|\mathbb{E}[X_n] - \mathbb{E}[X]| \leq |\mathbb{E}[X_n \wedge M] - \mathbb{E}[X \wedge M]| + \varepsilon.$$

Mais la fonction $x \mapsto x \wedge M$ est continue bornée sur $\mathbb{R}_+$, donc, par définition de la convergence en loi $\lim_{n\to+\infty}\mathbb{E}[X_n \wedge M] = \mathbb{E}[X \wedge M]$, d'où

$$\overline{\lim}_{n\to+\infty}|\mathbb{E}[X_n] - \mathbb{E}[X]| \leq \varepsilon,$$

et comme ε est quelconque, $\overline{\lim}_{n\to+\infty}|\mathbb{E}[X_n] - \mathbb{E}[X]| = 0$, ce qui donne le résultat voulu. $\square$

Remarque 3. *On pourrait développer la notion d'équi-intégrabilité sur un espace mesuré (pas nécessairement un espace probabilisé) en disant que* (f_n) *est équi-intégrable si*

$$\lim_{M\to+\infty}\sup_{n\geq 1}\int |f_n| \ \mathbb{1}_{\{|f_n|\geq M\}} \ d\mu = 0.$$

Dans ce cas, le theorème 7 s'appelle théorème de Vitali. Mais ce cadre est beaucoup moins intéressant car la convergence presque partout jointe à l'équi-intégrabilité n'implique pas la convergence des intégrales.

Exemple: On prend pour μ la mesure de Lebesgue et $f_n = \frac{1}{n}\mathbb{1}_{[n,2n]}$. Pour $M > 1$, on a $\sup_{n\geq 1}\int |f_n| \ \mathbb{1}_{\{|f_n|\geq M\}} \ d\lambda = 0$ et f_n converge partout vers 0. Cependant, l'intégrale de f_n est constante égale à 1.

1.2 Application à la convergence dans L^p

Corollaire 1. *Soit* $(X_n)_{n\geq 1}$ *une suite de variables aléatoires. On suppose que* X_n *converge en probabilité vers* X *lorsque* n *tend vers l'infini. Si la famille* $(|X_n|^p)_{n\geq 1}$ *est équi-intégrable, alors* $X \in L^p$ *et* X_n *converge dans* L^p *vers* X.

Démonstration. Si X_n converge en probabilité vers X, alors la suite (Y_n) définie par $Y_n = |X_n - X|^p$ converge en probabilité, donc en loi, vers 0. La suite $|X_n|^p$ est équi-intégrable, donc $|X|^p$ est intégrable, soit $X \in L^p$.
Il reste à voir que (Y_n) est équi-intégrable, ce qui découle de l'inégalité $Y_n \leq 2^p \max(|X_n|^p, |X|^p)$ et des remarques faites plus haut. Il suffit alors d'appliquer le théorème à la suite $(Y_n)_{n\geq 1}$. $\square$

1.3 Une condition suffisante d'équi-intégrabilité

La manière la plus simple de montrer l'équi-intégrabilité d'une famille est de montrer sa bornitude dans L^p pour un certain $p > 1$.
En effet si $\mathbb{E}[|X|^p] \leq C$ pour tout $X \in \mathcal{A}$, on a pour tout $X \in \mathcal{A}$,

$$\mathbb{E}[|X|\mathbb{1}_{\{|X|\geq M\}}] \leq \mathbb{E}[\frac{|X|^p}{M^{p-1}}\mathbb{1}_{\{X\geq M\}}] \leq \frac{\mathbb{E}[|X|^p]}{M^{p-1}} \leq \frac{C}{M^{p-1}}.$$

Ainsi, si $(X_n)_{n\geq 1}$ converge en probabilité vers X et que la suite $(X_n)_{n\geq 1}$ est bornée dans L^q, on sait que X_n converge vers X dans L^p pour $p < q$.

1.4 Une version du lemme de Scheffé

On dit parfois que l'équi-intégrabilité est ce qui manque à la convergence presque sûre pour avoir la convergence L^1. Le théorème suivant précise (et renforce) cet énoncé.

Théorème 8. *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires positives, intégrables, convergeant en loi vers une variable aléatoire X intégrable. On suppose que $\mathbb{E}X_n$ tend vers $\mathbb{E}X$. Alors les variables aléatoires $(X_n)_{n\geq 1}$ sont uniformément intégrables.*

Démonstration. Comme

$$\mathbb{E}[X\ \mathbb{1}_{\{X\geq M\}}] = \mathbb{E}[X] - \int_{[0,M]} \mathbb{P}(t < X < M)\ d\lambda(t)$$

est vraie pour toute variable aléatoire positive X, on a la convergence de $\mathbb{E}[X_n\ \mathbb{1}_{\{X_n\geq M\}}]$ vers $\mathbb{E}[X\ \mathbb{1}_{\{X\geq M\}}]$ si M est un point de continuité de X. On peut trouver M tel que $\mathbb{E}[X\ \mathbb{1}_{\{X\geq M\}}] < \varepsilon$. Si M n'est pas un point de continuité de F_X, on le remplace par un point de continuité M' de F_X tel que $M' \geq M$. Pour n_0 assez grand, on a $\mathbb{E}[X_n\ \mathbb{1}_{\{X_n\geq M'\}}] < \varepsilon$ pour $n \geq n_0$. Comme la famille finie $X_1, \dots, X_{n_0-1}$ est équi-intégrable, il existe M'' tel que $\mathbb{E}[X_n\ \mathbb{1}_{\{X_n\geq M''\}}] < \varepsilon$ pour tout $n < n_0$.
Ainsi si on prend $M_1 = \max(M', M'')$, on a $\mathbb{E}[X_n\ \mathbb{1}_{\{X_n\geq M_1\}}] < \varepsilon$ pour tout $n \geq 1$. □

1.5 Caractérisation par l'équi-continuité

Théorème 9. *La famille $\mathcal{A}$ de variables aléatoires définies sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ est équi-intégrable si et seulement si elle est bornée dans L^1 et vérifie*

$$\lim_{\eta\to 0} \sup_{\substack{X\in\mathcal{A} \\ A\in\mathcal{F}:\mathbb{P}(A)\leq\eta}} \mathbb{E}[|X|\ \mathbb{1}_A] = 0.$$

Démonstration. On a déjà vu qu'une famille équi-intégrable est bornée. Maintenant Pour tout $A \in \mathcal{F}$, $X \in \mathcal{A}$ et $M > 0$, on a

$$|X|\mathbb{1}_A \leq |X|\mathbb{1}_{\{|X|\geq M\}} + M\mathbb{1}_A.$$

Fixons $\varepsilon > 0$. L'hypothèse d'équi-intégrabilité nous dit qu'on peut trouver M tel que pour tout $X \in \mathcal{A}$ $\mathbb{E}[|X|\mathbb{1}_{\{|X|\geq M\}}] \leq \varepsilon/2$. Maintenant, pour $\eta \leq \frac{\varepsilon}{2M}$, pour tout $X \in \mathcal{A}$, $\mathbb{P}(A) \leq \eta$ entraîne

$$\mathbb{E}[|X|\mathbb{1}_A] \leq \mathbb{E}[|X|\mathbb{1}_{\{|X|\geq M\}}] + M\mathbb{P}(A) \leq \varepsilon.$$

Réciproquement, supposons que la famille $\mathcal{A}$ est bornée dans L^1 et vérifie

$$\lim_{\eta\to 0}\sup_{\substack{X\in\mathcal{A}\\ A\in\mathcal{F}:\mathbb{P}(A)\leq\eta}} \mathbb{E}[|X|\ \mathbb{1}_A] = 0.$$

Soit $\varepsilon > 0$. On peut trouver $\eta > 0$ tel que pour tout $X \in \mathcal{A}$, $\mathbb{P}(A) \leq \eta$ entraîne $\mathbb{E}[|X|\mathbb{1}_A] \leq \varepsilon$. Maintenant, si $M \geq \frac{\sup_{X\in\mathcal{A}}\mathbb{E}[|X|]}{\eta}$, si je pose $A = \{|X| \geq M\}$, on a $\mathbb{P}(A) \leq \frac{\mathbb{E}[|X|]}{M} \leq \eta$, donc

$$\mathbb{E}[|X|\mathbb{1}_{\{|X|\geq M\}}] = \mathbb{E}[|X|\mathbb{1}_A] \leq \varepsilon.$$

□

1.6 Équi-intégrabilité d'une famille de lois

Dans ce qui précède, l'équi-intégrabilité a été présenté comme une propriété d'une famille de variables aléatoires définies sur un même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. Cependant, si on convient de dire qu'une famille $\mathcal{M}$ de mesures de probabilités sur $\mathbb{R}$ est équi-intégrable si et seulement

$$\lim_{M\to+\infty}\sup_{\mu\in\mathcal{A}} \int_{\mathbb{R}\backslash[-M,M]} |x|\ d\mu = 0,$$

on voit sans peine qu'une famille $\mathcal{A}$ de variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$ est équi-intégrable si et seulement la famille $\mathcal{M} = \{\mathbb{P}_X; X \in \mathcal{A}\}$.

Ainsi, l'équi-intégrabilité est essentiellement une notion qui concerne les lois. En particulier, l'équi-intégrabilité d'une famille de variables aléatoires ne dépend que de la famille des lois individuelles, pas des lois jointes.

Un grand nombre de remarques faites pour les familles de variables équi-intégrables s'adaptent donc sans douleur aux familles de lois équi-intégrables. En particulier le théorème 7 admet la variante suivante :

Théorème 10. *Soit $(\mu_n)_{n\geq 1}$ une suite équi-intégrable de mesures de probabilité sur $\mathbb{R}$. On suppose que μ_n converge en loi vers μ lorsque n tend vers l'infini. Alors $\int_{\mathbb{R}} |x|\ d\mu(x) < +\infty$ et la suite $(\int_{\mathbb{R}} x\ d\mu_n(x))_{n\geq 1}$ converge vers $\int_{\mathbb{R}} x\ d\mu(x)$.*

On applique parfois ce théorème à des suites de variables aléatoires qui ne sont pas toutes définies sur le même espace de probabilité.

1.7 Exercices sur l'équi-intégrabilité

1.7.1 Exercices corrigés

Exercice 2. Soit (X_n) une suite de variables aléatoires positives avec

$$\sup_{n\geq 1} \mathbb{E}[X_n \log(1+X_n)] < +\infty.$$

Montrer que cette famille est équi-intégrable. lien vers l'indication lien vers la solution

Exercice 3. Soient $(X_1, X_2, \dots, X_n)$ une suite de v-a i.i.d $\mathcal{U}[0,1]$. Soit $Z_n = \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n X_i^2}$. Le but de l'exercice est de montrer que Z_n converge presque sûrement et dans L^1 vers $\frac{3}{2}$.

1. Montrer la convergence presque sûre.
2. Soit N un entier naturel. On pose $Q_N = (\sum_{i=1}^N X_i^2)^{-2}$. Montrer que pour tout $n \geq N$, on a
$$Z_n^2 \leq 144 + n^2 Q_N \mathbb{1}_{\{N_n < n/3\}},$$
où $N_n = \sum_{k=1}^n \mathbb{1}_{\{X_k \geq 1/2\}}$.
3. Montrer que Q_9 est dans L^2.
4. Montrer que $(Z_n)_{n\geq 1}$ est équi-intégrable. Conclure.

lien vers l'indication lien vers la solution

Exercice 4. Soit X_n suivant la loi binomiale $\mathcal{B}(n, \frac{\lambda}{n})$. Montrer que (X_n) est bornée dans L^3. Retrouver alors simplement sans calcul l'espérance et la variance de la loi de Poisson. lien vers l'indication lien vers la solution

Exercice 5. Donner un équivalent en l'infini de la suite (u_n) définie par $u_n = \sum_{k=1}^n k \binom{2n}{n+k}$. lien vers l'indication lien vers la solution

1.7.2 Exercices non corrigés

Exercice 6. *Loi des grands nombres L^1.*
Soit $(X_n)_{n\geq 1}$ des variables aléatoires indépendantes identiquement distribuées avec un moment d'ordre 1. La loi faible des grands nombres nous dit que la suite $\frac{X_1+\cdots+X_n}{n}$ converge en probabilité vers $\mathbb{E}[X_1]$. Montrer que $(X_n)_{n\geq 1}$ converge dans L^1 vers $\mathbb{E}[X_1]$. lien vers l'indication

Exercice 7. Soient p et q positifs des exposants conjugués, c'est à dire que $1/p + 1/q = 1$. Montrer que si $(X_n)_{n\geq 1}$ est bornée dans L^p et que $(|Y_n|^q)_{n\geq 1}$ est équi-intégrable, alors la suite $(X_nY_n)_{n\geq 1}$ est équi-intégrable. lien vers l'indication

Exercice 8. Soient $(X_n)_{n\geq 1}$ une suite équi-intégrable et (U_n) une suite de variables aléatoires indépendantes suivant la loi uniforme sur $[0, 1]$. On suppose de plus que $(U_n)_{n\geq 1}$ est indépendante de $(X_n)_{n\geq 1}$. Pour $\theta \in]0, 1]$, on pose $N_\theta = \inf\{n \geq 1; U_n \leq \theta\}$, puis $Z_\theta = X_{N_\theta}$. Montrer que la famille $(Z_\theta)_{\theta\in]0,1]}$ est équi-intégrable. lien vers l'indication

Exercice 9. Soit (X_n) une suite de variables aléatoires positives. On suppose qu'il existe une variable aléatoire intégrable X telle que

$$\forall n \geq 1 \quad \forall t > 0 \quad \mathbb{P}(X_n > t) \leq \mathbb{P}(X > t).$$

Montrer que $(X_n)_{n\geq 1}$ est équi-intégrable. lien vers l'indication

Chapitre 2

Espérance conditionnelle

2.1 Motivation

Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé ; $A_1, \dots, A_N$ une partition de Ω et X une variable aléatoire intégrable sur $\Omega, \mathcal{F}, \mathbb{P})$. Soit $\mathcal{A}$ la tribu engendrée par la partition $\{A_1, \dots, A_N\}$.

On s'intéresse aux expressions de la forme $\mathbb{E}X\mathbb{1}_A$, où $A \in \mathcal{A}$.

Tout d'abord, on va remarquer qu'il existe une correspondance entre $\mathcal{A}$ et $\mathcal{P}(\{1, \dots, N\}$: tout élément $A \in \mathcal{A}$ peut s'écrire

$$A = \cup_{i \in B} A_i,$$

pour un certain $B \in \mathcal{P}(\{1, \dots, N\})$.

On a alors

$$\mathbb{E}(X\mathbb{1}_A) = \mathbb{E}\left(X \sum_{i=1}^{N} \mathbb{1}_{i \in B} \mathbb{1}_{A_i}\right) = \sum_{i=1}^{N} \mathbb{E}(\mathbb{1}_{i \in B} X \mathbb{1}_{A_i}) = \sum_{i=1}^{N} \mathbb{1}_{i \in B} \mathbb{E}X\mathbb{1}_{A_i}$$

Maintenant posons

$$X' = \sum_{j=1}^{N} \mathbb{1}_{A_i} \frac{\mathbb{E}X\mathbb{1}_{A_j}}{\mathbb{P}(A_j)}.$$

En remplaçant dans la formule précédente, on obtient

$$\mathbb{E}X'\mathbb{1}_A = \sum_{i=1}^{N} \mathbb{1}_{i \in B} \mathbb{E}X'\mathbb{1}_{A_i}$$

Mais

$$\mathbb{E}X'\mathbb{1}_{A_i} = \sum_{j=1}^{N} \frac{\mathbb{E}X\mathbb{1}_{A_j}}{\mathbb{P}(A_j)} \mathbb{E}\mathbb{1}_{A_i}\mathbb{1}_{A_j} = \mathbb{E}X\mathbb{1}_{A_i}.$$

Il s'ensuit que pour tout $A \in \mathcal{A}$, on a

$$\mathbb{E}X\mathbb{1}_A = \mathbb{E}X'\mathbb{1}_A \tag{2.1}$$

Ce qui est intéressant ici, c'est que X' a une propriété que X n'a pas en général : en effet, X' est $\mathcal{A}$-mesurable (car c'est une combinaison linéaire d'indicatrices d'éléments de $\mathcal{A}$).

Si X' est une variable aléatoire $\mathcal{A}$-mesurable et telle que (2.1) est vérifiée, on dit que X' est une espérance conditionnelle de X par rapport à la tribu $\mathcal{A}$.

Nous savons donc construire des espérances conditionnelles par rapport à des tribus finies. Le but de ce chapitre est de traiter le cas général et de donner les premières propriétés de ces objets.

2.2 construction

Lemme 3. *Soient X' et Y' des variables aléatoires intégrables et mesurables par rapport à une tribu $\mathcal{A}$. On suppose que pour tout $A \in \mathcal{A}$, on a*

$$\mathbb{E}X'\mathbb{1}_A \le \mathbb{E}Y'\mathbb{1}_A \tag{2.2}$$

alors $X' \le Y'$ $\mathbb{P}$ presque sûrement.

Démonstration. On pose $A = \{X' > Y'\}$ On a

$$\mathbb{E}X'\mathbb{1}_A \le \mathbb{E}Y'\mathbb{1}_A.$$

Ainsi $\mathbb{E}(X'-Y')\mathbb{1}_A \le 0$. Mais $(X'-Y')\mathbb{1}_A$ est positive, donc $(X'-Y')\mathbb{1}_A = 0$ presque sûrement. Ainsi $P(\{X' = Y'\} \cup A^c) = 1$, d'où $P(A^c) = 1$. □

Ainsi, si X' et Y' sont des espérances conditionnelles de X et Y par rapport à la même tribu, on voit que $X \le Y$ presque sûrement entraîne que $X' \le Y'$ presque sûrement.

Cela a deux conséquences faciles, mais importantes : d'une part, on voit que l'espérance conditionnelle est unique, à un négligeable près. D'autre part, on voit que l'espérance conditionnelle préserve l'ordre, en particulier l'espérance conditionnelle d'une variable (presque sûrement) positive est (presque sûrement) positive.

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé ; $\mathcal{A}$ une sous-tribu de $\mathcal{F}$. Notons $V_1 = L^1(\Omega, \mathcal{F}, \mathbb{P})$, $V_2 = L^2(\Omega, \mathcal{F}, \mathbb{P})$ et $H = L^2(\Omega, \mathcal{A}, \mathbb{P})$.

Ici, il convient de noter que les éléments V_1 et V_2 sont des classes de fonctions : $L^p(\Omega, \mathcal{F}, \mathbb{P})$ est le quotient de $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ par la relation d'égalité presque sûre.

Ainsi V_2 est un espace de Hilbert dont H est un sous-espace fermé.

On a

$$\forall x \in V_2 \quad \mathbb{E}x = \langle x, 1 \rangle,$$

où 1 représente la classe de la fonction constante égale à 1.

Notons P la projection orthogonale de V_2 sur H : par définition on a

$$\forall f \in V_2 \forall g \in H \quad \langle f - Pf, g \rangle = 0.$$

En particulier si $A \in \mathcal{A}$, $\mathbb{1}_A \in H$, et donc

$$\forall f \in V_2 \quad \langle f - Pf, \mathbb{1}_A \rangle = 0, \tag{2.3}$$

soit $\langle f, \mathbb{1}_A \rangle = \langle Pf, \mathbb{1}_A \rangle$, soit

$$\mathbb{E}f\mathbb{1}_A = \mathbb{E}Pf\mathbb{1}_A.$$

En particulier

$$\mathbb{E}f = \mathbb{E}Pf. \tag{2.4}$$

L'équation (2.3) dit que Pf est un bon candidat pour être l'espérance conditionnelle. Les propriétés de positivité évoquées plus haut sont également vérifiées, mais il faut être un peu soigneux car l'on travaille ici avec des classes de fonctions égales presque partout, non avec des fonctions.

Rappelons quelques propriétés simples : si F et G sont deux fonctions mesurables qui sont égales presque partout, alors pour tout borélien A les ensembles $F^{-1}(A) = \{F \in A\}$ et $G^{-1}(A) = \{G \in A\}$ sont égaux à un négligeable près : cela signifie que

$$\mathbb{P}(F^{-1}(A)\Delta G^{-1}(A)) = 0.$$

En effet

$$\{F \in A\}\Delta\{G \in A\} \subset \{F \neq G\},$$

donc $P(\{F \in A\}\Delta\{G \in A\}) \leq \mathbb{P}(F \neq G) = 0$. Ainsi

$$|\mathbb{1}_{\{F\in A\}} - \mathbb{1}_{\{G\in A\}}| = \mathbb{1}_{\{F\in A\}\Delta\{G\in A\}} = 0 \quad \mathbb{P} \text{ p.s.}$$

ce qui signifie que $\mathbb{1}_{\{F\in A\}}$ et $\mathbb{1}_{\{G\in A\}}$ ont la même classe dans L^p (avec p quelconque).

Ainsi, si $f \in L^p$, il est licite de noter $\mathbb{1}_{\{f\in A\}}$ la classe de l'indicatrice de $\{F \in A\}$, où F est un représentant quelconque de la classe f.

On peut ainsi parler des éléments positifs de L^p : ce sont les éléments f qui sont la classe d'une fonction positive.

Pour tout $f \in L^p$, on a $f = f.1 = f(\mathbb{1}_{f>0}+\mathbb{1}_{f=0}+\mathbb{1}_{f<0}) = f(\mathbb{1}_{f>0}+\mathbb{1}_{f<0}) = f^+ - f^-$, où $f^+ = f\mathbb{1}_{f>0}$ et $f^- = -f\mathbb{1}_{f>0}$. Il est facile de voir que $f^+ = f\mathbb{1}_{f>0}$ et $f^+ = f\mathbb{1}_{f>0}$

Démontrons maintenant l'analogue du lemme 2.1 : si f est un élément positif de L^2, on a

$$\mathbb{E}f\mathbb{1}_{Pf<0} = \mathbb{E}Pf\mathbb{1}_{Pf<0}$$

Mais $f\mathbb{1}_{Pf<0}$ est une classe de fonctions positives, donc $\mathbb{E}f\mathbb{1}_{Pf<0} \geq 0$, tandis que $Pf\mathbb{1}_{Pf<0}$ est une classe de fonctions négatives, donc $\mathbb{E}Pf\mathbb{1}_{Pf<0} \leq 0$, finalement $Pf\mathbb{1}_{Pf<0} = 0$, soit $(Pf)^- = 0$, ce qui signifie que Pf est un élément positif.

Soit maintenant $f \in L^2$:

Par linéarité, on a $Pf = Pf^+ - Pf^-$, d'où

$$\|Pf\|_1 \quad \leq \quad \|Pf^+\|_1 + \|Pf^-\|_1$$

Mais Pf^+ est positive, donc $\|Pf^+\|_1 = \mathbb{E}Pf^+$. En utilisant (2.4), il vient que $\|Pf^+\|_1 = \mathbb{E}f^+$. De la même manière $\|Pf^-\|_1 = \mathbb{E}f^-$. Finalement, on a

$$\|Pf\|_1 \quad \leq \quad \mathbb{E}f^+ + \mathbb{E}f^- = \mathbb{E}|f| = \|f\|_1$$

Ainsi, P prend ses valeurs dans $H \cap V_1 = L^1(\Omega, \mathcal{A}, \mathbb{P})$ et définit une application linéaire continue de $(V_2, \|.\|_1)$ dans $(L^1(\Omega, \mathcal{A}, \mathbb{P}), \|.\|_1)$. Comme V_2 est une partie dense de V_1, l'application P admet un unique prolongement linéaire continu $\tilde{P}$ de $(V_1, \|.\|_1)$ dans $(L^1(\Omega, \mathcal{A}, \mathbb{P}), \|.\|_1)$.

Ce prolongement a la même norme d'opérateur :

$$\forall f \in L^1 \quad \|\tilde{P}f\|_1 \quad \leq \|f\|_1$$

Par ailleurs, l'ensemble des (classes de) fonctions positives est fermé dans L^1, donc la propriété de positivité est conservée par $\tilde{P}$:

$$\forall f \in L_1 \quad f \geq 0 \Longrightarrow \tilde{P}f \geq 0$$

et

$$\forall f, g \in L_1 \quad f \geq g \Longrightarrow \tilde{P}f \geq \tilde{P}g.$$

Cet opérateur $\tilde{P}$ est l'opérateur d'espérance conditionnelle sachant la tribu $\mathcal{A}$.

Vérifions d'abord que $\tilde{P}$ vérifie les propriétés fondamentales requises : $\tilde{P}f$ est $\mathcal{A}$-mesurable et pour toute fonction f $\mathcal{F}$-mesurable dans L^1, pour tout g $\mathcal{A}$-mesurable dans L^∞,

$$\mathbb{E}gf = \mathbb{E}g\tilde{P}f \tag{2.5}$$

Preuve : pour f dans L^∞, posons $\varphi_g(f) = \mathbb{E}gf$ et $\psi_g(f) = \mathbb{E}g\tilde{P}f$. On a $|gf| \le \|g\|_\infty |f|$, d'où $|\varphi_g(f)| \le \|f\|_\infty \|f\|_1$. Par ailleurs

$$|\psi_g(f)| = \varphi_g(\tilde{P}f)| \le \|f\|_\infty \|\tilde{P}f\|_1 \le \|f\|_\infty \|f\|_1.$$

Ainsi φ_g et ψ_g sont deux formes linéaires continues sur V_1. Comme elles coïncident sur V_2 qui est dense dans V_1, elles coïncident.

Pour tout $f \in V_1$, on notera désormais $\mathbb{E}[f|\mathcal{A}] = \tilde{P}f$.

On remarquera que telle que nous l'avons défini, l'objet $\mathbb{E}[f|\mathcal{A}]$ n'est pas une variable aléatoire, mais une classe de variable aléatoires. En réalité, faire la confusion entre les deux objets est sans risque puisqu'on ne s'intéresse jamais aux valeurs en un ω particulier, mais seulement aux intégrales sur un ensemble mesurable, intégrales qui ne dépendent pas du représentant choisi.

2.2.1 Propriétés

D'après le lemme 3, on voit donc que les propriétés :Y $\mathcal{A}$-mesurable et

$$\forall A \in \mathcal{A}\ \mathbb{E}X\mathbb{1}_A = \mathbb{E}Y\mathbb{1}_A \tag{2.6}$$

entraînent que $Y = \mathbb{E}[X|\mathcal{A}]$.

Ainsi, si X est $\mathcal{A}$-mesurable, on a évidemment $\mathbb{E}[X|\mathcal{A}] = X$. En particulier, les fonctions constantes étant mesurables par rapport à n'importe quelle tribu, on a pour tout réel c $\mathbb{E}[c|\mathcal{A}] = c$.

Rappelons avec les nouvelles écritures quelques propriétés établies : on suppose que X et Y sont intégrables et que a et b sont des constantes réelles :

— $\mathbb{E}[aX + bY|\mathcal{A}] = a\mathbb{E}[X|\mathcal{A}] + b\mathbb{E}[Y|\mathcal{A}]$

— $X \le Y$ p.s. $\Longrightarrow \mathbb{E}[X|\mathcal{A}] \le \mathbb{E}[Y|\mathcal{A}]$ p.s.

— $X \ge 0$ p.s. $\Longrightarrow \mathbb{E}[X|\mathcal{A}] \ge 0$ p.s.

— Si Y est bornée et $\mathcal{A}$-mesurable, alors $\mathbb{E}XY = \mathbb{E}[\mathbb{E}[X|\mathcal{A}]Y]$.

Et voici quelques conséquences dont la preuve ne devrait pas laisser de grandes difficultés au lecteur :

— $\mathbb{E}\mathbb{E}[X|\mathcal{A}] = \mathbb{E}X$

— $\mathbb{E}[\mathbb{E}[X|\mathcal{A}]|\mathcal{A}] = \mathbb{E}[X|\mathcal{A}]$

— $|\mathbb{E}[X|\mathcal{A}]| \le \mathbb{E}[|X|\ |\mathcal{A}]$

Un peu plus compliqué, le théorème de convergence dominée conditionnel :

Théorème 11. *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires convergeant presque sûrement vers X. On suppose que*

$$\forall n \geq 1 |X_n| \leq Y p.s.,$$

où Y est une variable aléatoire intégrable. Alors

$$\mathbb{E}[X|\mathcal{A}] = \lim_{n\to+\infty} \mathbb{E}[X_n|\mathcal{A}] p.s.$$

Démonstration. Posons $Z_n = \sup_{k\geq n} |X_n - X|$. $(Z_n)_{n\geq 1}$ est une suite décroissante de variables aléatoires qui tend presque sûrement vers zéro. D'après la positivité de l'espérance conditionnelle, la suite $\mathbb{E}[Z_n|\mathcal{A}]$ est donc une suite décroissante de variables aléatoires positives : elle converge donc vers une variable aléatoire positive Z. Pour tout $n \geq 1$, on a $Z \leq Z_n$, donc $\mathbb{E}Z \leq \mathbb{E}Z_n$. Ainsi $\mathbb{E}Z \leq \inf_{n\geq 1} \mathbb{E}Z_n$, mais pour tout n $|Z_n| \leq 2Y$, donc d'après le théorème de convergence dominée $\mathbb{E}Z_n$ tend vers 0, ce qui montre que $\mathbb{E}Z = 0$, donc Z est presque sûrement nulle. Ainsi $\mathbb{E}[Z_n|\mathcal{A}]$ tend presque sûrement vers 0. Finalement, pour tout n, on a

$$\begin{aligned} |\mathbb{E}[X|\mathcal{A}] - \mathbb{E}[X_n|\mathcal{A}]| &= |\mathbb{E}[X - X_n|\mathcal{A}]| \\ &\leq \mathbb{E}[|X - X_n| \,|\mathcal{A}] \\ &\leq \mathbb{E}[Z_n|\mathcal{A}], \end{aligned}$$

ce qui entraîne donc le résultat voulu. □

On en déduit en particulier le résultat très important suivant :

Théorème 12. *Si Y est $\mathcal{A}$-mesurable, que X et XY sont intégrables, alors* $\mathbb{E}[XY|\mathcal{A}] = Y\mathbb{E}[X|\mathcal{A}]$

Démonstration. On va d'abord s'intéresser au cas où Y est borné. Pour montrer que $Y\mathbb{E}[X|\mathcal{A}]$ est une version de l'espérance conditionnelle de XY sachant $\mathcal{A}$, il suffit de montrer

— que $Y\mathbb{E}[X|\mathcal{A}]$ est $\mathcal{A}$-mesurable

— que pour toute fonction Z $\mathcal{A}$-mesurable bornée, on a $\mathbb{E}ZXY = \mathbb{E}ZY\mathbb{E}[X|\mathcal{A}]$.

Le premier point est clair, car $Y\mathbb{E}[X|\mathcal{A}]$ est le produit de deux fonctions $\mathcal{A}$-mesurables. Le deuxième point provient du fait que ZY est une fonction $\mathcal{A}$-mesurable bornée et de la propriété fondamentale de l'espérance conditionnelle.

Pour passer au cas général, il suffit de poser $Y_n = Y\mathbb{1}_{|Y|\leq n}$. D'après ce qui précède, on sait que pour tout n, on a

$$\mathbb{E}[XY_n|\mathcal{A}] = Y_n\mathbb{E}[X|\mathcal{A}].$$

XY_n converge presque sûrement vers XY et $|XY_n| \leq |XY|$. D'après le théorème de convergence dominée conditionnel, $\mathbb{E}[XY_n|\mathcal{A}]$ converge presque sûrement vers $\mathbb{E}[XY|\mathcal{A}]$. Comme Y_n converge presque sûrement vers Y, on a finalement

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

□

On va terminer cette section par quelques propriétés simples, mais utiles liées à la mesurabilité.

Théorème 13. *Soit X une variable aléatoire intégrable, $\mathcal{A}$ et $\mathcal{B}$ deux tribus avec $\mathcal{A} \subset \mathcal{B}$. Alors*

$$\mathbb{E}[\mathbb{E}[X|\mathcal{B}]|\mathcal{A}] = \mathbb{E}[X|\mathcal{A}].$$

Démonstration. Posons $X' = \mathbb{E}[\mathbb{E}[X|\mathcal{B}]|\mathcal{A}]$. Soit Y une variable aléatoire bornée $\mathcal{A}$-mesurable. Y est $\mathcal{A}$-mesurable, donc $Y\mathbb{E}[\mathbb{E}[X|\mathcal{B}]|\mathcal{A}] = \mathbb{E}[Y\mathbb{E}[X|\mathcal{B}]|\mathcal{A}]$. Par ailleurs, Y est $\mathcal{B}$-mesurable, donc $Y\mathbb{E}[X|\mathcal{B}] = \mathbb{E}XY|\mathcal{B}$, d'où $\mathbb{E}YX' = \mathbb{E}[\mathbb{E}XY|\mathcal{B}] = \mathbb{E}XY$. Ainsi, pour tout Y $\mathcal{A}$-mesurable bornée, on a $\mathbb{E}YX' = \mathbb{E}YX$. Comme X' est $\mathcal{A}$-mesurable, cela montre bien que $X' = \mathbb{E}[X|\mathcal{A}]$. □

Théorème 14. *On suppose que $\mathcal{A}$ et $\mathcal{B}$ sont deux tribus indépendantes sous $\mathbb{P}$. Alors, pour toute variable aléatoire intégrable X $\mathcal{B}$-mesurable, on a*

$$\mathbb{E}[X|\mathcal{A}] = \mathbb{E}X.$$

Démonstration. Soit Y une variable aléatoire bornée $\mathcal{A}$-mesurable.

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

Comme la variable constante $\mathbb{E}X$ est à l'évidence $\mathcal{A}$-mesurable, cela achève la preuve. □

Le résultat suivant peut également parfois être utile :

Théorème 15. *Si X est une variable intégrable, $\mathcal{A}$ et $\mathcal{B}$ des tribus telles que $\mathcal{B}$ est indépendante de $\sigma(\sigma(X), \mathcal{A})$, alors*

$$\mathbb{E}[X|\sigma(\mathcal{A}, \mathcal{B})] = \mathbb{E}[X|\mathcal{A}].$$

Démonstration. Par linéarité, en écrivant $X = X^+ - X^-$, on peut se ramener au cas où X est positive. Comme $Y = \mathbb{E}[X|\mathcal{A}]$ est $\sigma(\mathcal{A}, \mathcal{B})$-mesurable, il suffit de montrer que l'on a $\mathbb{E}[\mathbb{1}_C X] = \mathbb{E}[\mathbb{1}_C Y]$ pour tout $C \in \sigma(\mathcal{A}, \mathcal{B})$. Comme $C \mapsto \mathbb{E}[\mathbb{1}_C X]$ et $C \mapsto \mathbb{E}[\mathbb{1}_C Y]$ sont des mesures finies, il suffit de les identifier sur un π-système qui engendre $\sigma(\mathcal{A}, \mathcal{B})$: on prend naturellement les ensemble de la forme $C = A \cap B$, avec $A \in \mathcal{A}$ et $B \in \mathcal{B}$. Et, en effet, on a

$$\begin{aligned} \mathbb{E}[\mathbb{1}_C Y] &= \mathbb{E}[\mathbb{1}_A \mathbb{1}_B Y] \\ &= \mathbb{E}[\mathbb{1}_B]\mathbb{E}[\mathbb{1}_A Y] \\ &= \mathbb{E}[\mathbb{1}_B]\mathbb{E}[\mathbb{1}_A X] \\ &= \mathbb{E}[\mathbb{1}_B \mathbb{1}_A X] = \mathbb{E}[\mathbb{1}_C X], \end{aligned}$$

ce qui conclut la preuve. □

2.2.2 Inégalité de Jensen

Théorème 16. *Soit X une variable aléatoire intégrable à valeurs dans l'intervalle I. Soit f une fonction continue convexe de I dans $\mathbb{R}$. On suppose que $f(X)$ est intégrable. Alors, pour toute tribu $\mathcal{A}$, on a*

$$f(\mathbb{E}[X|\mathcal{A}]) \le \mathbb{E}[f(X)|\mathcal{A}].$$

Démonstration. Pour donner du sens à l'inégalité, remarquons d'abord que $\mathbb{E}[X|\mathcal{A}] \in I$ presque sûrement. En effet, il est bien connu si $X \ge a$ presque sûrement, on a alors $\mathbb{E}[X|\mathcal{A}] \ge a$ presque sûrement. Cependant, on va voir que si $X > a$ presque sûrement, on a encore $\mathbb{E}[X|\mathcal{A}] > a$ presque sûrement.

Par linéarité,on se ramène au cas où $a = 0$. On suppose donc que $X > 0$ et on pose $Y = \mathbb{E}[X|\mathcal{A}]$. On sait déjà que $Y \ge 0$ presque sûrement. Reste à montrer $\mathbb{P}(Y = 0) = 0$. Comme l'événement $\{Y = 0\}$ est $\mathcal{A}$-mesurable, on a $\mathbb{E}[\mathbb{1}_{\{Y=0\}} X] = \mathbb{E}[\mathbb{1}_{\{Y=0\}} Y] = 0$, donc $\mathbb{1}_{\{Y=0\}} X = 0$ presque sûrement. Comme $X > 0$ presque sûrement, on a $\mathbb{1}_{\{Y=0\}} = 0$ presque sûrement, donc $\mathbb{P}(Y = 0) = 0$.

Soit Φ une famille dénombrable de fonctions affines φ telles que

$$\forall x \in I \quad \varphi(x) \le f(x).$$

Soit $\varphi \in \Phi$. On a presque sûrement

$$\varphi(X) \le f(X).$$

On a donc $\mathbb{E}[\varphi(X)|\mathcal{A}] \le \mathbb{E}[f(X)|\mathcal{A}]$. Mais comme φ est une fonction affine, $\mathbb{E}[\varphi(X)|\mathcal{A}] = \varphi(\mathbb{E}[X|\mathcal{A}])$. Ainsi,

$$\forall \varphi \in \Phi \quad \varphi(\mathbb{E}[X|\mathcal{A}]) \le \mathbb{E}[f(X)|\mathcal{A}],$$

Comme Φ est dénombrable, on a alors

$$\sup_{\varphi \in \Phi} \varphi(\mathbb{E}[X|\mathcal{A}]) \le \mathbb{E}[f(X)|\mathcal{A}].$$

Il reste donc à démontrer que la famille peut être choisie de telle sorte que, presque sûrement, au point $x = \mathbb{E}[X|\mathcal{A}]$, on ait $\sup\{\varphi(x); \varphi \in \Phi\} = f(x)$

En tout point rationnel q de I, on considère l'application affine φ_q tangente à droite (ou à gauche) au point q à la courbe représentative de f. On prend alors $\Phi = (\varphi_q)_{q \in I \cap \mathbb{Q}}$. L'application $x \mapsto \sup_{\varphi \in \Phi} \varphi(x)$ est convexe, continue, et coïncide avec f sur $I \cap \mathbb{Q}$, donc, par densité et continuité, sur I. □

Pour retenir quel est le sens de l'égalité, prendre la fonction convexe $\varphi(x) = |x|$.

2.2.3 Espérance conditionnelle sachant une variable (ou un vecteur) aléatoire

Le cauchemar des conventions d'écriture

Soient A est un événement,$\mathcal{A}$ une tribu.

On note fréquemment $\mathbb{P}(A|\mathcal{A}) = \mathbb{E}[\mathbb{1}_A|\mathcal{A}]$.

Si Y,X, $X_1, \dots, X_n$ sont des variables aléatoires (ou des vecteurs aléatoires), on note fréquemment

$\mathbb{E}[Y|X]$ pour $\mathbb{E}[Y|\sigma(X)]$, et encore $\mathbb{E}[Y|X_1, \dots, X_n]$ pour $\mathbb{E}[Y|\sigma(X_1, \dots, X_n)]$.

Par ailleurs, si X est une variable aléatoire (ou un vecteur aléatoire), il est bien évident que $\mathbb{E}[Y|X]$ est $\sigma(X)$-mesurable. Alors, d'après le lemme de Doob, il existe une application mesurable f telle que

$$\mathbb{E}[Y|X] = f(X) \tag{2.7}$$

On écrit souvent $\mathbb{E}[Y|X = x] = f(x)$ pour signifier que l'application f vérifie l'équation (2.7). Bien sûr, il s'agit d'un abus d'écriture car la propriété est une propriété globale de la fonction, qui n'a pas de sens point par point, puisqu'en général (par exemple si la loi de X est à densité) l'événement $\{X = x\}$ est de probabilité nulle.

Il est plus facile de donner un sens à cette espérance conditionnelle – et aussi de la calculer – dans le cas où X est une variable aléatoire discrète.

Théorème 17. *Soit $Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, X une variable aléatoire sur $(\Omega, \mathcal{F}, \mathbb{P})$, et D un ensemble fini ou dénombrable avec $\mathbb{P}(X \in D) = 1$ et $P(X = n) > 0$ pour tout n dans D. Alors, $\mathbb{E}[Y|X] = g(X)$ avec*

$$\forall n \in D \quad g(n) = \frac{\mathbb{E}[\mathbf{1}_{\{X=n\}}Y]}{\mathbb{P}(X = n)},$$

ce qui s'écrit encore

$$\forall n \in D \quad \mathbb{E}[Y|X=n] = \frac{\mathbb{E}[\mathbf{1}_{\{X=n\}}Y]}{\mathbb{P}(X=n)}.$$

Démonstration. Pour tout n dans A, posons $g(n) = \frac{\mathbb{E}[\mathbf{1}_{\{X=n\}}Y]}{\mathbb{P}(X=n)}$. Soit $A \subset D$. Pour tout n dans A, on a

$$\mathbb{E}[Y\mathbb{1}_{\{X=n\}}] = g(n)\mathbb{P}(X=n) = \mathbb{E}[g(n)\mathbb{1}_{\{X=n\}}] = \mathbb{E}[g(X)\mathbb{1}_{\{X=n\}}]$$

En faisant la somme sur tous les $n \in A$, on obtient

$$\mathbb{E}[Y\mathbb{1}_{\{X\in A\}}] = \mathbb{E}[g(X)\mathbb{1}_{\{X\in A\}}]$$

Comme tous les éléments de $\sigma(X)$ peuvent s'écrire sous la forme $\{X \in A\}$ avec $A \subset D$ et que $g(X)$ est évidemment $\sigma(X)$-mesurable, on a bien l'identification voulue. □

Corollaire 2. *Soit X une variable aléatoire sur $(\Omega, \mathcal{F}, \mathbb{P})$, et D un ensemble fini ou dénombrable avec $\mathbb{P}(X \in D) = 1$ et $P(X = n) > 0$ pour tout n dans D. Soit A un événement. Pour tout $n \in D$, on a*

$$\mathbb{E}[\mathbb{1}_A|X=n] = \frac{\mathbb{E}[\mathbb{1}_{\{X=n\}}\mathbb{1}_A]}{\mathbb{P}(X=n)} = \frac{\mathbb{P}(\{X=n\}\cap A)}{\mathbb{P}(X=n)} = \mathbb{P}(A|X=n).$$

Ceci justifie que l'on écrive $\mathbb{P}(A|X)$ pour $\mathbb{E}(\mathbb{1}_A|X)$ sans qu'il y ait conflit entre les différentes conventions d'écriture.

Des techniques de calculs utiles

La première technique n'est que la particularisation de la formule précédente au cas d'un vecteur aléatoire.

Corollaire 3. *Soit $Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, $X_1, \dots, X_n$ des variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans S dénombrable, avec pour tous $x_1, \dots, x_n$ dans S $\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) > 0$. Alors, $\mathbb{E}[Y|X_1, \dots, X_n] = g(X_1, \dots, X_n)$ avec*

$$\forall n \in D \quad g(n) = \frac{\mathbb{E}[\mathbf{1}_{\{X_1=x_1,\dots,X_n=x_n\}}Y]}{\mathbb{P}(X_1=x_1,\dots,X_n=x_n)},$$

ce qui s'écrit encore

$$\forall (x_1,\dots,x_n) \in S^n \quad \mathbb{E}[Y|X_1=x_1,\dots,X_n=x_n] = \frac{\mathbb{E}[\mathbf{1}_{\{X_1=x_1,\dots,X_n=x_n\}}Y]}{\mathbb{P}(X_1=x_1,\dots,X_n=x_n)}.$$

On laisse au lecteur le soin de particulariser l'énoncé dans le cas où Y est l'indicatrice d'un événement.

Les théorèmes et corollaires précédents permettent, dans le cas de variables aléatoires discrètes, de calculer des espérances conditionnelles à l'aide d'opérations "classiques" sur les probabilités. En retour, les propriétés des espérances conditionnelles permettent souvent de simplifier des calculs de probabilités. Car exemple, on peut noter que

$$\begin{aligned}
\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) &= \mathbb{E}(\mathbb{1}_{\{X_1=x_1,\dots,X_n\}}) \\
&= \mathbb{E}(\mathbb{1}_{\{X_1=x_1,\dots,X_{n-1}=x_{n-1}\}}\mathbb{1}_{\{X_n=x_n\}}) \\
&= \mathbb{E}(\mathbb{E}(\mathbb{1}_{\{X_1=x_1,\dots,X_{n-1}=x_{n-1}\}}\mathbb{1}_{\{X_n=x_n\}}|X_1, \dots, X_{n-1})) \\
&= \mathbb{E}(\mathbb{1}_{\{X_1=x_1,\dots,X_{n-1}=x_{n-1}\}}\mathbb{E}(\mathbb{1}_{\{X_n=x_n\}}|X_1, \dots, X_{n-1})) \\
&= \mathbb{E}(\mathbb{1}_{\{X_1=x_1,\dots,X_{n-1}=x_{n-1}\}}\mathbb{P}(X_n = x_n|X_1, \dots, X_{n-1}))
\end{aligned}$$

Ce genre de manipulations sera très utile dans le cadre de l'étude des chaînes de Markov.

Un autre cas pratique très important est celui où la variable conditionnée est une fonction de deux variables indépendantes.

Théorème 18. *Soit X et Y deux vecteurs aléatoires indépendants, respectivement à valeurs dans $\mathbb{R}^n$ et $\mathbb{R}^p$. Soit g une application de $\mathbb{R}^n \times \mathbb{R}^p$ dans $\mathbb{R}$. On suppose que $g(X,Y)$ est une variable aléatoire intégrable. Alors $\mathbb{E}[g(X,Y)|X] = G(X)$, avec*

$$G(x) = \int g(x,y)d\mathbb{P}_Y(y) = \mathbb{E}g(x,Y).$$

Autrement dit, $\mathbb{E}[g(X,Y)|X = x] = \mathbb{E}[g(x,Y)]$.

Démonstration. D'abord, il faut vérifier que G est défini $\mathbb{P}_X$ presque partout. Pour cela, il faut montrer que pour $\mathbb{P}_X$ presque tout x : $\int |g(x,y)|d\mathbb{P}_Y(y) < +\infty$. Pour cela, il suffit de montrer que

$$\int \left(\int |g(x,y)|d\mathbb{P}_Y(y)\right) d\mathbb{P}_X(x) < +\infty,$$

ce qui découle facilement du théorème de Tonelli et de l'intégrabilité de $g(x,y)$ sous $\mathbb{P}_X \otimes \mathbb{P}_Y$.

Soit maintenant A un borélien de $\mathbb{R}^n$:

$$\begin{aligned}
\mathbb{E}\mathbb{1}_A(X)G(X) &= \int \mathbb{1}_A(x)G(x)d\mathbb{P}_X(x) \\
&= \int \mathbb{1}_A(x)\left(\int g(x,y)d\mathbb{P}_Y(y)\right)d\mathbb{P}_X(x) \\
&= \int \mathbb{1}_A(x)g(x,y)d(\mathbb{P}_X \otimes \mathbb{P}_Y)(x,y) \\
&= \int \mathbb{1}_A(x)g(x,y)d(\mathbb{P}_{X,Y})(x,y) \\
&= \mathbb{E}\mathbb{1}_A(X)g(X,Y)
\end{aligned}$$

Bien sûr $G(X)$ est $\sigma(X)$-mesurable, ce qui achève la preuve. □

Dans le même ordre d'idée, le résultat suivant peut également être utile :

Théorème 19. *On suppose que les vecteurs (X,Y) et (X',Y') à valeurs dans $\mathbb{R}^n \times \mathbb{R}^p$ ont même loi, que Y est intégrable avec $\mathbb{E}[Y|X] = f(X)$. Alors*

$$\mathbb{E}[Y'|X'] = f(X').$$

Démonstration. Bien sûr, $f(X')$ est $\sigma(X')$-mesurable. Il suffit donc de faire la vérification :

$$\begin{aligned}
\mathbb{E}[\mathbb{1}_A(X')Y'] &= \mathbb{E}[\mathbb{1}_A(X)Y] \text{ car } (X,Y) \text{ et } (X',Y') \text{ ont même loi} \\
&= \mathbb{E}[\mathbb{1}_A(X)f(X)] \text{ par définition de l'espérance conditionnelle} \\
&= \mathbb{E}[\mathbb{1}_A(X')f(X')] \text{ car } X \text{ et } X' \text{ ont même loi,}
\end{aligned}$$

ce qui donne le résultat voulu. □

2.3 Exercices sur l'espérance conditionnelle

2.3.1 Exercices corrigés

Exercice 10. 1. Soit X une variable aléatoire intégrable, N une variable aléatoire à valeurs dans $\mathbb{N}$. Soit Y une variable aléatoire intégrable $\sigma(N)$-mesurable.

Montrer que Y est une version de $\mathbb{E}[X|N]$ si et seulement si pour tout entier $n \in \mathbb{N}$ $\mathbb{E}[\mathbb{1}_{\{N=n\}}Y] = \mathbb{E}[\mathbb{1}_{\{N=n\}}X]$

2. Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires indépendantes suivant la loi de Bernoulli de paramètre q. On suppose que $1+T$ suit la loi géométrique de paramètre p et que T est indépendante de $\sigma((X_k)_{k\geq 1})$. On pose $U = X_1 + X_2 + \cdots + X_T$ (avec $U = 0$ si $T = 0$). Calculer $\mathbb{E}[U|T]$.

lien vers l'indication lien vers la solution

Exercice 11. On reprends l'énoncé du (b) de l'exercice précédent. Déterminer des réels α et β tels que $\mathbb{E}[T|U] = \alpha U + \beta$. lien vers l'indication lien vers la solution

Exercice 12. On note $\Omega_{n,b} = \{(x_1, \dots, x_{n+b}) \in \{0,1\}^{n+b}; \sum_{i=1}^{n+b} x_i = n\}$. On note $\mu_{n,b}$ la loi uniforme sur $\Omega_{n,b}$. On note $\mathbb{Q}_{n,b}$ la mesure de probabilités définie par

$$\mathbb{Q}_{n,b}(A) = \mu_{n,b}(A|x_1 = 1) = \frac{\mu_{n,b}(A \cap \{x_1 = 1\})}{\mu_{n,b}(x_1 = 1)}.$$

1. On note $X_1, \dots, X_{n+b}$ les projections canoniques de $\Omega_{n,b}$ dans $\mathbb{R}$. Montrer que la loi de $(X_2, \dots, X_{n+b})$ sous $\mathbb{Q}_{n,b}$ est $\mu_{n-1,b}$.
2. On suppose que $(X_1, \dots, X_{n+b})$ suit la loi $\mu_{n,b}$. On note $T = \inf\{k \geq 0; X_{k+1} = 0\}$. T représente le nombre de boules noires tirées dans une urne contenant n boules noires et b boules blanches, où l'on effectue une série de tirages en s'arrêtant au tirage de la première boule blanche. Montrer que $\mathbb{E}[T] = \frac{n}{b+1}$ (on pourra procéder par récurrence sur n, en conditionnant par le premier tirage).

lien vers l'indication lien vers la solution

2.3.2 Exercices non corrigés

Exercice 13. Soient X et Y deux variables aléatoires intégrables indépendantes. Calculer $\mathbb{E}[(1+X)(1+Y)|X]$. lien vers l'indication

Exercice 14. Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes identiquement distribuées admettant un moment d'ordre 1. Calculer

$$\mathbb{E}[X_1 | (X_1 + X_2 + \dots X_n)].$$

lien vers l'indication

Exercice 15. Soient X et Y indépendantes, avec $X \sim \mathcal{B}(n,p)$ et $Y \sim \mathcal{B}(n',p)$. Calculer $\mathbb{E}[X|X+Y]$

— par un calcul direct

— en utilisant le résultat de l'exercice précédent.

lien vers l'indication

Exercice 16. Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. On note $X(1), ..., X(n)$ les abscisses réordonnées de la plus petite à la plus grande. On pose

$$M_n = \max(X(i+1) - X(i); 1 \le i \le n-1).$$

Le but de l'exercice est de montrer que $\mathbb{E}M_n = O(\frac{\ln n}{n})$.

1. Montrer que pour toute variable aléatoire réelle X à valeurs dans $[0,1]$ et tout h réel positif, on a
$$\mathbb{E}X \le h + \mathbb{P}(X \ge h).$$
2. Pour $i \in \{1, \dots, n\}$, on note
$$A_i(h) = \{\exists j \in \{1, \dots n\} X_j \ge X_i + h\} \cap \cap_{1 \le j \le n, i \ne j} \{X_j \notin]X_i, X_i + h[\}.$$
Montrer que
$$\{M_n \ge h\} = \cup_{i=1}^n A_i(h).$$
3. On pose
$$Y_i(h) = \prod_{j \ne i} \mathbb{1}_{]X_i, X_i+h[^c}(X_j).$$
Montrer que $\mathbb{1}_{A_i}(h) \le \mathbb{1}_{\{X_i \le 1-h\}} Y_i(h)$.
4. Montrer que
$$\mathbb{E}[Y_i(h)|X_i] = \max(X_i, 1-h)^{n-1}$$
5. En déduire que $\mathbb{P}(A_i(h)) \le (1-h)^n$.
6. Conclure.

lien vers l'indication

Exercice 17. On suppose que la loi du couple (X, Y) sous $\mathbb{P}$ admet la densité $(x, y) \mapsto f(x, y)$ par rapport à la mesure $\mu \otimes \nu$. Montrer que

$$\mathbb{E}[X|Y = y] = \frac{\int x f(x, y) d\mu(x)}{\int f(x, y) d\mu(x)}.$$

(On commencera par élucider les abus de langage de l'énoncé).
Montrer que pour toute fonction φ mesurable telle que $\varphi(X)$ est intégrable, on a

$$\mathbb{E}[\varphi(X)|Y = y] = \frac{\int \varphi(x) f(x, y) d\mu(x)}{\int f(x, y) d\mu(x)}.$$

lien vers l'indication

Exercice 18. Soient X et Y deux variables aléatoires à valeurs dans un ensemble dénombrable D. Montrer que pour toute fonction φ bornée, et pour tout $i \in D$ avec $\mathbb{P}(X = i) > 0$, on a

$$\mathbb{E}[\varphi(Y)|X = i] = \sum_{j \in D} \mathbb{P}(Y = j|X = i)\varphi(j).$$

lien vers l'indication

Exercice 19. Soit

$$M = \begin{pmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

1. Montrer qu'on peut construire un vecteur gaussien centré (X, Y, Z) admettant M comme matrice de covariance.
2. Calculer $\mathbb{E}[X|Y, Z]$.

lien vers l'indication

Exercice 20. Soit

$$M = \begin{pmatrix} 3 & 0 & 2 \\ 0 & 2 & 1 \\ 2 & 1 & 2 \end{pmatrix}$$

1. Montrer qu'on peut construire un vecteur gaussien centré (X, Y, Z) admettant M comme matrice de covariance.
2. Calculer $\mathbb{E}[X|Y, Z]$.
3. Soit φ une fonction mesurable bornée telle que $\varphi(X)$ est intégrable. Montrer que

$$\mathbb{E}[\varphi(X)|Y = y, Z = z] = \int_{\mathbb{R}} \varphi(x) f_{y,z}(x) \; d\lambda(x),$$

où $f_{y,z}$ est la densité de la loi normale $\mathcal{N}(-\frac{2}{3}y + \frac{4}{3}z, \sigma^2)$, où

$$\sigma^2 = \langle Mv, v\rangle \text{ avec } v = (1, \frac{2}{3}, -\frac{4}{3})'.$$

lien vers l'indication

Exercice 21. Soit $(Y_n)_{n\geq 1}$ une suite de variables aléatoires positives intégrables, $\mathcal{A}$ une tribu quelconque. On pose $Y = \underline{\lim}_{n\to+\infty} Y_n$ et $Z = \underline{\lim}_{n\to+\infty}\mathbb{E}[Y_n|\mathcal{A}]$. Le but de l'exercice est de montrer que si Z est fini presque sûrement, alors Y est également fini presque sûrement.

1. Soit $M > 0$. Montrer que $\mathbb{E}[\mathbb{1}_{\{Y>M\}}|\mathcal{A}] = \lim_{n\to+\infty}\mathbb{E}[\mathbb{1}_{\{Y_n>M\}}|\mathcal{A}]$.
2. Montrer que pour tout $n \geq 1$, on a $M\mathbb{E}[\mathbb{1}_{\{Y_n>M\}}|\mathcal{A}] \leq \mathbb{E}[Y_n|\mathcal{A}]$.
3. En déduire que $\mathbb{E}[\mathbb{1}_{\{Y>M\}}|\mathcal{A}]$ tend presque sûrement vers 0 lorsque M tend vers l'infini.
4. Conclure.

lien vers l'indication

Exercice 22. Soit $(R_n)_{n\geq 0}$ une suite décroissante d'entiers naturels, avec $R_0 = N$. On pose, pour $n \geq 0$, $S_{n+1} = R_n - R_{n+1}$ et $\mathcal{F}_n = \sigma(S_1, \dots, S_n)$. On pose $T = \inf\{n \geq 0; R_n = 0\}$. On suppose en outre que

$$\forall k \geq 1 \quad \mathbb{E}[1_{T>k-1}(S_k - 1)|\mathcal{F}_{k-1}] = 0.$$

1. Montrer que pour tout entier naturel non nul k, on a

   $$\mathbb{P}(S_k = 0, T \geq k) = \mathbb{E}(1_{T\geq k}(S_k - 1)^+).$$

   En déduire que $\mathbb{P}(T < +\infty) = 1$.
2. Que peut-on dire de la suite $(M_i)_{i\geq 1}$ définie par

   $$M_i = \sum_{k=1}^{i} 1_{T\geq k}(S_k - 1) \quad ?$$

3. En déduire la valeur de $\mathbb{E}[T]$.
4. Application : On considère n personnes. On met leur nom dans une urne. Chacun tire un nom. Ceux qui ont tiré leur nom se retirent et on recommence avec les autres. On demande le nombre moyen de tirages qu'il faut faire pour "éliminer" tout le monde.

lien vers l'indication

Chapitre 3

Martingales

3.1 Définitions

3.1.1 Filtrations et martingales

Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé.

Définition: On appelle filtration toute suite croissante $(\mathcal{F}_n)_{n\geq 0}$ de sous-tribus de $\mathcal{F}$.

Définition: Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires et $(\mathcal{F}_n)_{n\geq 0}$ une filtration. On dit que la suite $(X_n)_{n\geq 0}$ est $(\mathcal{F}_n)_{n\geq 0}$ adaptée si pour tout n, X_n est $\mathcal{F}_n$-mesurable.

Définition: Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires. On appelle filtration naturelle adaptée à la suite $(X_n)_{n\geq 0}$ la filtration définie par $\mathcal{F}_n = \sigma(X_0, X_1, \dots X_n)$.

Définition: Soit $(\mathcal{F}_n)_{n\geq 0}$ une filtration et $(X_n)_{n\geq 0}$ une suite de variables aléatoires. On dit que la suite $(X_n)_{n\geq 0}$ est une martingale adaptée à la filtration $(\mathcal{F}_n)_{n\geq 0}$ si

1. la suite $(X_n)_{n\geq 0}$ est $(\mathcal{F}_n)_{n\geq 0}$ adaptée
2. Pour tout n, X_n est intégrable.
3. Pour tout n, $X_n = \mathbb{E}[X_{n+1}|\mathcal{F}_n]$.

Exemples:

1. Soit $(\mathcal{F}_n)_{n\geq 0}$ une filtration, X une variable aléatoire intégrable. La suite définie par $X_n = \mathbb{E}[X|\mathcal{F}_n]$ est une martingale
2. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes centrées. On pose pour tout $n \geq 1 : \mathcal{F}_n = \sigma(X_1, \dots X_n)$ et $S_n = X_1 + X_2 + \dots X_n$. Alors, $(S_n)_{n\geq 1}$ est une martingale adaptée à la filtration $(\mathcal{F}_n)_{n\geq 1}$.

Remarque: Une martingale est toujours adaptée à sa filtration naturelle.

3.1.2 Différences de martingales

Si X_n est une martingale adaptée à la filtration $(\mathcal{F}_n)_{n\geq 0}$, la suite Y_n définie par $Y_n = X_n - X_{n-1}$ vérifie

$$\forall n \geq 1 \quad \mathbb{E}[Y_n|\mathcal{F}_{n-1}] = 0. \tag{3.1}$$

Réciproquement, si la suite $(Y_n)_{n\geq 1}$ est adaptée à la filtration $(\mathcal{F}_n)_{n\geq 0}$ et vérifie 3.1, la suite des sommes partielles définie par $X_n = Y_1 + \cdots + Y_n$ est une martingale adaptée à la filtration $(\mathcal{F}_n)_{n\geq 0}$.

Définition: On appelle une telle suite (Y_n) une différence de martingale.

Remarque: Si la suite de différences de martingale $(Y_n)_{n\geq 1}$ est dans L^2, la variable Y_n est orthogonale à toutes les variables Z de carré intégrable qui sont $\mathcal{F}_{n-1}$-mesurables. En particulier, les $(Y_n)_{n\geq 1}$ forment une suite orthogonale dans L^2.

Comme on le verra dans les exercices, les martingales sont souvent utiles pour éudier des suites de variables aléatoires qui ne sont pas elles-mêmes des martingales. Mettre en évidence une martingale à partir d'une suite n'est pas toujours facile. Une bonne idée est de commencer par exhiber une différence de martingale. Par exemple, si $(X_n)_{n\geq 0}$ est une suite intégrable quelconque, $(X_{n+1} - \mathbb{E}[X_{n+1}|X_0, \dots X_n])_{n\geq 0}$ est toujours une différence de martingales.

3.1.3 Sous-martingales, sur-martingales

Définition: Soit $(\mathcal{F}_n)_{n\geq 0}$ une filtration et $(X_n)_{n\geq 0}$ une suite de variables aléatoires. On dit que la suite $(X_n)_{n\geq 0}$ est une sous-martingale adaptée à la filtration $(\mathcal{F}_n)_{n\geq 0}$ si

1. la suite $(X_n)_{n\geq 0}$ est $(\mathcal{F}_n)_{n\geq 0}$ adaptée
2. Pour tout n, X_n est intégrable.
3. Pour tout n, $X_n \leq \mathbb{E}[X_{n+1}|\mathcal{F}_n]$.

Définition: Soit $(\mathcal{F}_n)_{n\geq 0}$ une filtration et $(X_n)_{n\geq 0}$ une suite de variables aléatoires. On dit que la suite $(X_n)_{n\geq 0}$ est une surmartingale adaptée à la filtration $(\mathcal{F}_n)_{n\geq 0}$ si

1. la suite $(X_n)_{n\geq 0}$ est $(\mathcal{F}_n)_{n\geq 0}$ adaptée
2. Pour tout n, X_n est intégrable.
3. Pour tout n, $X_n \geq \mathbb{E}[X_{n+1}|\mathcal{F}_n]$.

Proposition 2. *Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires intégrables. La suite $(\mathbb{E}X_n)_{n\geq 0}$ est*
- *décroissante si la suite $(X_n)_{n\geq 0}$ est une surmartingale.*
- *croissante si la suite $(X_n)_{n\geq 0}$ est une sous-martingale.*
- *constante si la suite $(X_n)_{n\geq 0}$ est une martingale.*

Démonstration. On va juste prouver la première assertion. Pour tout n, on a $X_n \geq \mathbb{E}[X_{n+1}|\mathcal{F}_n]$.
En prenant l'espérance, on a $\mathbb{E}[X_n] \geq \mathbb{E}[\mathbb{E}[X_{n+1}|\mathcal{F}_n]] = \mathbb{E}[X_{n+1}]$. □

3.2 Premières inégalités

3.2.1 Martingales et fonctions convexes

Théorème 20. *Soit $(\mathcal{F}_n)_{n\geq 0}$ une filtration et φ une fonction convexe.*
- *Si la suite $(X_n)_{n\geq 0}$ est une martingale adaptée à la filtration $(\mathcal{F}_n)_{n\geq 0}$ et que les $(\varphi(X_n))_{n\geq 0}$ sont intégrables, alors la suite la suite $(\varphi(X_n))_{n\geq 0}$ est une sous-martingale adaptée à la filtration $(\mathcal{F}_n)_{n\geq 0}$.*
- *Si la suite $(X_n)_{n\geq 0}$ est une sous-martingale adaptée à la filtration $(\mathcal{F}_n)_{n\geq 0}$, que φ est croissante et que les $(\varphi(X_n))_{n\geq 0}$ sont intégrables, alors la suite $(\varphi(X_n))_{n\geq 0}$ est une sous-martingale adaptée à la filtration $(\mathcal{F}_n)_{n\geq 0}$.*

Démonstration.
- Comme $X_n = \mathbb{E}[X_{n+1}|\mathcal{F}_n]$, on a $\varphi(X_n) = \varphi(\mathbb{E}[X_{n+1}|\mathcal{F}_n]) \leq \mathbb{E}[\varphi(X_{n+1})|\mathcal{F}_n]$, d'après l'inégalité de Jensen conditionnelle.
- $X_n \leq \mathbb{E}[X_{n+1}|\mathcal{F}_n]$ entraîne, avec l'hypothèse de croissance $\varphi(X_n) \leq \varphi(\mathbb{E}[X_{n+1}|\mathcal{F}_n])$ et on conclut comme précédemment avec l'inégalité de Jensen conditionnelle.

□

Exemple: En particulier, si une suite $(X_n)_{n\geq 0}$ de variables aléatoires positives est une sous-martingale de carré intégrable, alors la suite $(X_n^2)_{n\geq 0}$ est une sous-martingale.

3.2.2 Inégalité de Kolmogorov

Théorème 21. *Soit $(X_n)_{n\geq 0}$ une sous-martingale. Pour tout $\alpha > 0$, on a*

$$\mathbb{P}(\max_{1\leq i\leq n} X_i \geq \alpha) \leq \frac{1}{\alpha}\mathbb{E}|X_n|.$$

Démonstration. Notons $\tau = \inf\{i \geq 1; X_i \geq \alpha\}$. Il est clair que

$$\{\max_{1 \leq i \leq n} X_i \geq \alpha\} = \{\tau \leq n\}.$$

Soit k entre 1 et n : l'événement $\tau = k$ est $\mathcal{F}_k$-mesurable, donc d'après la propriété de sous-martingale, on a

$$\mathbb{E}[X_n \mathbb{1}_{\{\tau=k\}}] = \mathbb{E}\ \mathbb{E}[X_n \mathbb{1}_{\{\tau=k\}} | \mathcal{F}_k] = \mathbb{E}\ \mathbb{1}_{\{\tau=k\}} \mathbb{E}[X_n | \mathcal{F}_k] \geq \mathbb{E}\ \mathbb{1}_{\{\tau=k\}} X_k.$$

Mais $\mathbb{1}_{\{\tau=k\}} X_k \geq \alpha \mathbb{1}_{\{\tau=k\}}$. Ainsi, en intégrant, on obtient

$$\mathbb{E}[X_n \mathbb{1}_{\{\tau=k\}}] \geq \alpha \mathbb{P}(\tau = k).$$

En faisant la somme pour k variant de 1 à n, on obtient

$$\mathbb{E}[X_n \mathbb{1}_{\{\tau \leq n\}}] \geq \alpha \mathbb{P}(\tau \leq n).$$

Si $(X_n)_{n \geq 1}$ est une sous-martingale positive, on a fini, car alors

$$\mathbb{E} X_n \geq \mathbb{E}[X_n \mathbb{1}_{\{\tau \leq n\}}] \geq \alpha \mathbb{P}(\tau \leq n).$$

Sinon, comme $(X_n^+)_{n \geq 1}$ est une sous-martingale positive, on peut lui appliquer le résultat que l'on vient de démontrer, et l'on a

$$\mathbb{P}(\max_{1 \leq i \leq n} X_i \geq \alpha) = \mathbb{P}(\max_{1 \leq i \leq n} X_i^+ \geq \alpha) \leq \frac{1}{\alpha} \mathbb{E} X_n^+ \leq \frac{1}{\alpha} \mathbb{E}|X_n|,$$

ce qui donne le résultat voulu. □

3.3 Convergence des martingales de carré intégrable

Théorème 22. *Soit $(X_n)_{n \geq 0}$ une martingale adaptée à la filtration $(\mathcal{F}_n)_{n \geq 1}$ telle que*

$$\sup_{n \geq 1} \mathbb{E} X_n^2 < +\infty.$$

Alors $(X_n)_{n \geq 0}$ converge presque sûrement et dans L^2 vers une variable X_∞ de carré intégrable.

Démonstration. La convergence quadratique s'obtient par des méthodes hilbertiennes classiques : comme L^2 est complet, il suffit en effet de montrer que la suite $(X_n)_{n \geq 0}$ est de Cauchy. Soit $p < n$ entiers. Comme X_n est dans L^2, on sait que $\mathbb{E}[X_n|\mathcal{F}_p] = X_p$ est le projeté orthogonal de X_n sur le sous-espace des variables $\mathcal{F}_p$-mesurables. Ainsi, on peut écrire l'identité de Pythagore :

$$\|X_n\|_2^2 = \|X_p\|_2^2 + \|X_n - X_p\|_2^2,$$

ou encore

$$\mathbb{E}X_n^2 = \mathbb{E}X_p^2 + \mathbb{E}(X_n - X_p)^2.$$

Il est alors clair que la suite $(\mathbb{E}X_n^2)_{n\geq 1}$ est croissante : elle converge donc vers $\alpha = \sup_{n\geq 1} \mathbb{E}X_n^2$ que nous avons supposé fini. Soit $\varepsilon > 0$ et N tel que $\alpha \geq \mathbb{E}X_n^2 \geq \alpha - \varepsilon^2$ pour $n \geq N$. Alors, pour $n, p \geq N$, on a $\|X_n - X_p\|_2 \leq \varepsilon$, ce qui contre bien que la suite est de Cauchy, et donc convergente.

On va maintenant montrer la convergence presque sûre. Pour cela, on va montrer que la suite $(X_n)_{n\geq 1}$ est presque sûrement de Cauchy, c'est à dire que $R_n = \sup_{i,j\geq n} |X_i - X_j|$ tend presque sûrement vers 0. Comme la suite $(R_n)_{n\geq}$ est monotone décroissante, il suffit de montrer qu'elle admet une sous-suite qui converge presque sûrement vers 0. Pour démontrer que $(R_n)_{n\geq}$ admet une sous-suite qui converge presque sûrement vers zéro, il suffit (voir le cours de licence) de montrer que R_n converge en probabilité vers 0.

Soit donc $\varepsilon > 0$. On a

$$\{R_n > \varepsilon\} \subset \cup_{i\geq n}\{|X_n - X_i| > \varepsilon/2\}.$$

Ainsi

$$\begin{aligned} \mathbb{P}(R_n > \varepsilon) &\leq \mathbb{P}(\sup_{i\geq n} |X_n - X_i| \geq \varepsilon/2) \\ &\leq \mathbb{P}(\sup_{i\geq n} |X_n - X_i| > \varepsilon/3). \end{aligned}$$

D'après le théorème de continuité séquentielle croissante, on a

$$\mathbb{P}(\sup_{i\geq n} |X_n - X_i| > \varepsilon/3) = \sup_{N\geq n} \mathbb{P}(\sup_{n\leq i\leq N} |X_n - X_i| > \varepsilon/3).$$

La suite $(X_n - X_i)_{n\geq i}$ est une martingale, donc la suite $((X_n - X_i)^2)_{n\geq i}$ est une sous-martingale positive : on a donc

$$\mathbb{P}(\sup_{n\leq i\leq N} |X_n - X_i|^2 > \varepsilon^2/9) \leq \frac{9}{\varepsilon^2}\mathbb{E}(X_n - X_N)^2.$$

Ainsi

$$\mathbb{P}(R_n > \varepsilon) \leq \frac{9}{\varepsilon^2} \sup_{N\geq n} \mathbb{E}(X_n - X_N)^2,$$

cette dernière suite tend bien vers 0, puisque $(X_n)_{n\geq 1}$ converge en moyenne quadratique.

□

3.4 Temps d'arrêts

On dit que variable aléatoire T à valeurs dans $\mathbb{N} \cup \{+\infty\}$ est un *temps d'arrêt* adapté à la filtration $(\mathcal{F}_n)_{n\geq 0}$ si pour tout $n \in \mathbb{N}$, l'événement $T \leq n$ est $\mathcal{F}_n$-mesurable.

Comme $\{T = n\} = \{T \leq n\} \backslash \{T \leq n-1\}$, il s'ensuit que $T = n$ est également $\mathcal{F}_n$-mesurable.

Exemple: Toute constante est un temps d'arrêt adapté à toute filtration.

Démonstration. Si T est constant, $\{T \leq n\}$ ne peut valoir que Ω ou $\varnothing$, et est donc toujours dans $\mathcal{F}_n$. □

Exemple: Si $(X_n)_{n\geq 1}$ est une suite de variables aléatoires $(\mathcal{F}_n)_{n\geq 1}$-adaptée à valeurs dans S et A un borélien de S, alors

$$T_A = \inf\{n \geq 1; X_n \in A\}$$

est un temps d'arrêt $(\mathcal{F}_n)_{n\geq 1}$-adapté.

Preuve : $\{T_A \leq n\} = \cup_{k=1}^n \{X_k \in A\}$.

Définition: On dit qu'un événement A se produit avant T si pour tout n, l'événement $A \cap \{T \leq n\}$ est $\mathcal{F}_n$-mesurable.

Remarque: Si deux temps d'arrêts S et T adaptés à la filtration $(\mathcal{F}_n)_{n\geq 0}$ vérifient $S \leq T$, alors tout événement qui se produit avant S se produit avant T.

Démonstration. Soit A se produisant avant S. Comme $S \leq T$, on a $\{T \leq n\} \cap A = (A \cap \{S \leq n\}) \cap \{T \leq n\}$. Comme A se produit avant S, $A \cap \{S \leq n\} \in \mathcal{F}_n$. Par définition d'un temps d'arrêt, $\{T \leq n\} \in \mathcal{F}_n$. On en déduit que $\{T \leq n\} \cap A \in \mathcal{F}_n$. Comme n est quelconque, A se produit avant T. □

Proposition 3. *Soit T un temps d'arrêt adapté à une filtration $(\mathcal{F}_n)_{n\geq 0}$. L'ensemble $\mathcal{F}_T$ des événements qui se produisent avant T forme une tribu.*

Démonstration.
— Montrons que $\varnothing \in \mathcal{F}_T$. Pour tout n, on a $\varnothing \cap \{T \leq n\} = \varnothing \in \mathcal{F}_n$, car toute tribu contient $\varnothing$ donc on a bien $\varnothing \in \mathcal{F}_T$.
— Soit $A \in \mathcal{F}_T$. Montrons que $A^c \in \mathcal{F}_T$. Soit n entier. On a $A^c \cap \{T \leq n\} = \{T \leq n\} \backslash (A \cap \{T \leq n\})$. Les événements $\{T \leq n\}$ et $A \cap \{T \leq n\}$ sont tous deux $\mathcal{F}_n$-mesurables, donc $A^c \cap \{T \leq n\}$ est bien dans $\mathcal{F}_n$. Comme n est quelconque, $A^c \in \mathcal{F}_T$.
— Soit $(A_p)_{p\geq 1}$ une suite d'éléments de $\mathcal{F}_T$. Il faut montrer que $A = \cup_{p\geq 1} A_p \in \mathcal{F}_T$ Soit n entier. On a

$$A \cap \{T \leq n\} = \cup_{p\geq 1}(A_p \cap \{T \leq n\}),$$

A est réunion dénombrable d'éléments de $\mathcal{F}_n$, donc $A \in \mathcal{F}_n$, et finalement $A \in \mathcal{F}_T$.

□

Remarque: T est $\mathcal{F}_T$ mesurable.

Démonstration. Il suffit de montrer que pour tout $t \in \mathbb{R}$ $\{T \le t\} \in \mathcal{F}_T$. Soit $n \in \mathbb{N}$: Si on note i la partie entière de t, on a

$$\{T \le t\} \cap \{T \le n\} = \{T \le i\} \cap \{T \le n\} = \{T \le i \wedge n\} \in \mathcal{F}_{i \wedge n} \subset \mathcal{F}_n.$$

□

Théorème 23 (Théorème de Hunt)**.** *Soit $(\mathcal{F}_n)_{n\ge 0}$ une filtration et $(X_n)_{n\ge 0}$ une sous-martingale adaptée à la filtration $(\mathcal{F}_n)_{n\ge 0}$.*

Soient S et T deux temps d'arrêts $(\mathcal{F}_n)_{n\ge 0}$-adaptés bornés avec $S \le T$. Alors

$$E[X_T|\mathcal{F}_S] \ge X_S.$$

Démonstration. Il s'agit de montrer que pour tout événement A $\mathcal{F}_S$-mesurable, on a $\mathbb{E}X_T\mathbb{1}_A \ge \mathbb{E}X_S\mathbb{1}_A$. Soit M un entier déterministe tel que l'on ait $S \le T \le M$. Posons $\Delta_k = X_k - X_{k-1}$, avec $X_{-1} = 0$. Notons, que comme (X_n) est une sous-martingale $\mathbb{E}\mathbb{1}_B\Delta_k$ est positif pour tout ensemble B $\mathcal{F}_{k-1}$-mesurable. On a

$$\begin{aligned}\mathbb{E}((X_T - X_S)\mathbb{1}_A) &= \mathbb{E}(\sum_{k=0}^{M} \Delta_k \mathbb{1}_{S<k\le T}\mathbb{1}_A) \\ &= \sum_{k=0}^{M} \mathbb{E}\Delta_k \mathbb{1}_{\{S<k\le T\}\cap A}\end{aligned}$$

Pour conclure, il reste donc à voir que $\{S < k \le T\} \cap A$ est $\mathcal{F}_{k-1}$-mesurable, ce qui vient de l'identité

$$\{S < k \le T\} \cap A = (\{S \le k-1\} \cap A) \cap (T \le k-1)^c.$$

□

On en déduit facilement les deux résultats suivants :

Corollaire 4. *Soit $(\mathcal{F}_n)_{n\ge 0}$ une filtration et $(X_n)_{n\ge 0}$ une surmartingale adaptée à la filtration $(\mathcal{F}_n)_{n\ge 0}$.*

Soient S et T deux temps d'arrêts $(\mathcal{F}_n)_{n\ge 0}$-adaptés bornés avec $S \le T$. Alors

$$E[X_T|\mathcal{F}_S] \le X_S.$$

Corollaire 5. *Soit $(\mathcal{F}_n)_{n\geq 0}$ une filtration et $(X_n)_{n\geq 0}$ une martingale adaptée à la filtration $(\mathcal{F}_n)_{n\geq 0}$.*

Soient S et T deux temps d'arrêts $(\mathcal{F}_n)_{n\geq 0}$-adaptés bornés avec $S \leq T$. Alors
$$E[X_T|\mathcal{F}_S] = X_S.$$

Théorème 24 (Théorème d'arrêt)**.** *Soit $(\mathcal{F}_n)_{n\geq 0}$ une filtration et $(X_n)_{n\geq 0}$ une sous-martingale adaptée à la filtration $(\mathcal{F}_n)_{n\geq 0}$. Soit T un temps d'arrêt adapté à la filtration $(\mathcal{F}_n)_{n\geq 0}$. Alors la suite $(X_{n\wedge T})_{n\geq 0}$ est une sous-martingale adaptée à la filtration $(\mathcal{F}_n)_{n\geq 0}$.*

Démonstration. Soit A un événement $\mathcal{F}_n$-mesurable et n un entier. On a
$$\begin{aligned} \mathbb{E}\mathbb{1}_A X_{(n+1)\wedge T} &= \mathbb{E}\mathbb{1}_A X_{(n+1)\wedge T}\mathbb{1}_{\{T\leq n\}} + \mathbb{E}\mathbb{1}_A X_{n+1\wedge T}\mathbb{1}_{\{T>n\}} \\ &= \mathbb{E}X_{n\wedge T}\mathbb{1}_A\mathbb{1}_{\{T\leq n\}} + \mathbb{E}X_{(n+1)\wedge T}\mathbb{1}_A\mathbb{1}_{\{T>n\}} \end{aligned}$$

Pour l'instant, on a juste utilisé que quand $T \leq n$, on a $(n+1)\wedge T = T = n \wedge T$.

Montrons que $A \cap \{T > n\}$ est $\mathcal{F}_{n\wedge T}$-mesurable : soit p un entier ; on doit montrer que $A \cap \{T > n\} \cap \{n \wedge T \leq p\}$ est dans $\mathcal{F}_p$. Si $n > p$, l'intersection est l'ensemble vide, donc est $\mathcal{F}_p$-mesurable. Si $n \leq p$, alors $\{n \wedge T \leq p\} = \Omega$, donc
$$A \cap \{T > n\} \cap \{n \wedge T \leq p\} = A \cap \{T > n\} \in \mathcal{F}_n \subset \mathcal{F}_p.$$

Ainsi, en appliquant le théorème de Hunt aux temps d'arrêts $(n+1)\wedge T$ et $n \wedge T$, on a
$$\mathbb{E}X_{(n+1)\wedge T}\mathbb{1}_A\mathbb{1}_{\{T>n\}} \geq \mathbb{E}X_{n\wedge T}\mathbb{1}_A\mathbb{1}_{\{T>n\}}$$
D'où
$$\begin{aligned} \mathbb{E}\mathbb{1}_A X_{(n+1)\wedge T} &= \mathbb{E}X_{n\wedge T}\mathbb{1}_A\mathbb{1}_{\{T\leq n\}} + \mathbb{E}X_{(n+1)\wedge T}\mathbb{1}_A\mathbb{1}_{\{T>n\}} \\ &\geq \mathbb{E}X_{n\wedge T}\mathbb{1}_A\mathbb{1}_{\{T\leq n\}} + \mathbb{E}X_{n\wedge T}\mathbb{1}_A\mathbb{1}_{\{T>n\}} \\ &\geq \mathbb{E}X_{n\wedge T}\mathbb{1}_A \end{aligned}$$
ce qui achève la preuve.

□

On en déduit facilement les deux résultats suivants :

Corollaire 6. *Soit $(\mathcal{F}_n)_{n\geq 0}$ une filtration et $(X_n)_{n\geq 0}$ une surmartingale adaptée à la filtration $(\mathcal{F}_n)_{n\geq 0}$. Soit T un temps d'arrêt adapté à la filtration $(\mathcal{F}_n)_{n\geq 0}$. Alors la suite $(X_{n\wedge T})_{n\geq 0}$ est une surmartingale adaptée à la filtration $(\mathcal{F}_n)_{n\geq 0}$.*

Corollaire 7. *Soit* $(\mathcal{F}_n)_{n\geq 0}$ *une filtration et* $(X_n)_{n\geq 0}$ *une martingale adaptée à la filtration* $(\mathcal{F}_n)_{n\geq 0}$*. Soit* T *un temps d'arrêt adapté à la filtration* $(\mathcal{F}_n)_{n\geq 0}$*. Alors la suite* $(X_{n\wedge T})_{n\geq 0}$ *est une martingale adaptée à la filtration* $(\mathcal{F}_n)_{n\geq 0}$*.*

Remarque: Certains auteurs appellent « théorème d'arrêt »ce que nous avons appelé « théorème de Hunt ». En fait, ces deux résultats sont équivalents. Ici, nous avons choisi de démontrer le théorème de Hunt et d'en déduire le théorème d'arrêt, tandis que d'autres auteurs (par exemple Baldi, Mazliak et Priouret) font le choix inverse.

Les deux lemmes suivants seront utiles par la suite. Leur preuve relève des méthodes classiques.

Lemme 4. *Soit* $(\mathcal{F}_n)_{n\geq 0}$ *une filtration et* $(X_n)_{n\geq 0}$ *une martingale adaptée à la filtration* $(\mathcal{F}_n)_{n\geq 0}$*. Soit* T *un temps d'arrêt adapté à la filtration* $(\mathcal{F}_n)_{n\geq 0}$*. Alors, la variable aléatoire* $\mathbb{1}_{\{T<+\infty\}}X_T$ *est* $\mathcal{F}_T$*-mesurable.*

Démonstration. Il suffit de voir que pour tout borélien de $\mathbb{R}$, l'événement $\{\mathbb{1}_{\{T<+\infty\}}X_T \in A\}$ est $\mathcal{F}_T$-mesurable. Soit n un entier : on a

$$\{\mathbb{1}_{\{T<+\infty\}}X_T \in A\} \cap \{T \leq n\} = \cup_{i\leq n}\{X_i \in A; T = i\},$$

qui est bien $\mathcal{F}_n$-mesurable. □

Lemme 5. *Soit* $(\mathcal{F}_n)_{n\geq 0}$ *une filtration et* $(A_n)_{n\geq 0}$ *une suite d'événements adaptée à la filtration* $(\mathcal{F}_n)_{n\geq 0}$*. Soit* T *un temps d'arrêt adapté à la filtration* $(\mathcal{F}_n)_{n\geq 0}$*. Alors, la variable aléatoire* S *définie par*

$$S(\omega) = \inf\{n \geq T(\omega); \omega \in A_n\}.$$

est un temps d'arrêt.

Démonstration. Il suffit de remarquer que pour tout n

$$\{S \leq n\} = \cup_{i\leq n} A_i \cap \{T \leq i\},$$

qui est clairement $\mathcal{F}_n$-mesurable. □

3.5 Convergence des martingales bornées dans L^1

3.5.1 Théorème des traversées montantes

Théorème 25. *Soit* $(X_n)_{n\geq 1}$ *une sous-martingale* a, b *deux réels avec* $a < b$*. On note* $U_n^{[a,b]}$ *le nombre de traversées montantes de* a *à* b *entre les instants* 1 *et* n*. Alors, on a*

$$\mathbb{E}U_n^{[a,b]} \leq \frac{\mathbb{E}(X_n - a)^+}{b - a}$$

Démonstration. Posons pour $n \geq 1$: $Y_n = \frac{(X_n-a)^+}{b-a}$: comme $|Y_n| \leq \frac{1}{b-a}(|X_n|+|a|)$ et que $\varphi(x) = \frac{(x-a)^+}{b-a}$ est croissante et convexe, $(Y_n)_{n\geq 1}$ est une sous-martingale.

Posons $\tau_0 = 0$, et pour $k \geq 1$

$$\sigma_k = \inf\{n \geq \tau_{k-1}; X_n \leq a\}$$

et

$$\tau_k = \inf\{n \geq \sigma_k; X_n \geq b\}.$$

En utilisant le lemme 5, on voit facilement par récurrence que les (σ_k) et les (τ_k) sont des temps d'arrêt. D'autre part, par construction, on a $\tau_0 \leq \sigma_1 \leq \tau_1 \leq \sigma_2 \leq \cdots \leq \sigma_n \leq \tau_n$.

On a

$$U_n^{[a,b]} = \sum_{k=1}^n \mathbb{1}_{\{\tau_k \leq n\}}.$$

Montrons que $\mathbb{1}_{\{\tau_k \leq n\}} \leq Y_{n\wedge\tau_k} - Y_{n\wedge\sigma_k}$: il y a trois cas possibles

— si $\sigma_k > n$, alors $\tau_k > n$: on a

$$\mathbb{1}_{\{\tau_k \leq n\}} = 0 = Y_n - Y_n = Y_{n\wedge\tau_k} - Y_{n\wedge\sigma_k}.$$

— si $\sigma_k \leq \tau_k \leq n$, alors on a $Y_{n\wedge\tau_k} = Y_{\tau_k} \geq 1$ tandis que $Y_{n\wedge\sigma_k} = Y_{\sigma_k} = 0$, de sorte que

$$\mathbb{1}_{\{\tau_k \leq n\}} = 1 \leq Y_{n\wedge\tau_k} - Y_{n\wedge\sigma_k}.$$

— si $\sigma_k \leq n < \tau_k$, alors $Y_{n\wedge\tau_k} = Y_n \geq 0$ tandis que $Y_{n\wedge\sigma_k} = Y_{\sigma_k} = 0$, de sorte que

$$\mathbb{1}_{\{\tau_k \leq n\}} = 0 \leq Y_{n\wedge\tau_k} - Y_{n\wedge\sigma_k}.$$

On a donc

$$U_n^{[a,b]} \leq \sum_{k=1}^n Y_{n\wedge\tau_k} - Y_{n\wedge\sigma_k}.$$

On peut écrire

$$\begin{aligned}
Y_n - Y_0 &= Y_{\tau_n\wedge n} - Y_{\tau_0\wedge n} \\
&= \sum_{k=1}^n Y_{\tau_k\wedge n} - Y_{\tau_{k-1}\wedge n} \\
&= \sum_{k=1}^n (Y_{\tau_k\wedge n} - Y_{\sigma_k\wedge n}) + (Y_{\sigma_k\wedge n} - Y_{\tau_{k-1}\wedge n})
\end{aligned}$$

Il s'ensuit que

$$\begin{aligned} U_n^{[a,b]} &\leq Y_n - Y_0 - \sum_{k=1}^{n} Y_{\sigma_k \wedge n} - Y_{\tau_{k-1} \wedge n} \\ &\leq Y_n - \sum_{k=1}^{n} Y_{\sigma_k \wedge n} - Y_{\tau_{k-1} \wedge n}, \end{aligned}$$

soit, en prenant l'espérance

$$\mathbb{E}U_n^{[a,b]} \leq \mathbb{E}Y_n - \sum_{k=1}^{n} \mathbb{E}[Y_{\sigma_k \wedge n} - Y_{\tau_{k-1} \wedge n}].$$

Mais $\sigma_k \wedge n$ et $\tau_{k-1} \wedge n$ sont des temps d'arrêts bornés avec $\tau_{k-1} \leq \sigma_k$. D'après le théorème de Hunt, on a donc, en réintégrant, $\mathbb{E}[Y_{\sigma_k \wedge n} - Y_{\tau_{k-1} \wedge n}] \geq 0$, d'où

$$\mathbb{E}U_n^{[a,b]} \leq \mathbb{E}Y_n,$$

ce qui était le résultat voulu. □

3.5.2 Le théorème de convergence de Doob

Théorème 26. *Soit $(X_n)_{n\geq 1}$ une sous-martingale. Les quantités*

$$K = \sup_{n\geq 1} \mathbb{E}[X_n^+] \text{ et } L = \sup_{n\geq 1} \mathbb{E}[|X_n|]$$

sont simultanément finies ou infinies. Si elles sont finies, la suite $(X_n)_{n\geq 1}$ converge presque surement vers une variable aléatoire X intégrable, avec $\mathbb{E}|X| \leq L$.

Démonstration. On a pour tout n,

$$\mathbb{E}(X_n^+) \leq E(|X_n|) = \mathbb{E}(2X_n^+ - X_n) = 2\mathbb{E}(X_n^+) - \mathbb{E}(X_n) \leq 2\mathbb{E}(X_n^+) - \mathbb{E}(X_0),$$

d'où $K \leq L \leq 2K - \mathbb{E}(X_0)$. Passons à la preuve de la convergence.

Faisons d'abord une remarque d'analyse : si une suite $(x_n)_{n\geq 1}$ ne converge pas dans $\overline{\mathbb{R}}$, alors il existe deux rationnels a et b tels que la suite $(x_n)_{n\geq 1}$ traverse une infinité de fois l'intervalle $[a, b]$ de bas en haut : pour cela, il suffit de considérer deux rationnels a et b vérifiant $\underline{\lim}_{n\to+\infty} x_n < a < b < \overline{\lim}_{n\to+\infty} x_n$. Ainsi, si l'on note $U^{[a,b]}$ le nombre de traversées de l'intervalle $[a, b]$, on a

$$\{X_n \text{ ne converge pas dans } \overline{\mathbb{R}}\} \subset \cup_{a<b,(a,b)\in\mathbb{Q}^2} \{U^{[a,b]} = +\infty\}.$$

Cette réunion est dénombrable. Pour achever la preuve, il suffit donc de montrer que pour tout couple (a, b) avec $a < b$, on a

$$\mathbb{P}(U^{[a,b]} = +\infty) = 0.$$

Cependant on a par convergence monotone $\mathbb{E}U^{[a,b]} = \lim_{n\to+\infty}\mathbb{E}[U_n^{[a,b]}] \leq \frac{K+|a|}{b-a}$ d'après le théorème des traversées montantes et la définition de K. Comme $U^{[a,b]}$ est intégrable, elle est presque sûrement finie.

Ainsi $(X_n)_{n\geq 1}$ converge presque sûrement vers un élément X de $\overline{\mathbb{R}}$. D'après le lemme de Fatou, on a $\mathbb{E}|X| = \mathbb{E}\underline{\lim}_{n\to+\infty}|X_n| \leq \underline{\lim}_{n\to+\infty}\mathbb{E}|X_n| \leq L$. Ainsi, $|X|$ est intégrable, ce qui implique qu'elle prend presque sûrement ses valeurs dans $\mathbb{R}$.

□

Corollaire 8. *Soit $(X_n)_{n\geq 0}$ une surmartingale positive intégrable adaptée à la filtration $(\mathcal{F}_n)_{n\geq 1}$. Alors $(X_n)_{n\geq 0}$ converge presque sûrement vers une variable X_∞ telle que $\mathbb{E}[X_\infty] \leq \mathbb{E}[X_0]$.*

Démonstration. $(-X_n)_{n\geq 0}$ est une sous-martingale. Comme $\mathbb{E}(|-X_n|) = \mathbb{E}(X_n) \leq \mathbb{E}(X_0)$, on applique le théorème avec $L = \mathbb{E}[X_0]$. □

3.5.3 Martingales inverses

Soit $(\mathcal{F}_n)_{n\geq 0}$ une suite décroissante de sous-tribus de $\mathcal{F}$. On dit qu'une suite $(X_n)_{n\geq 0}$ de variables aléatoires intégrables est une surmartingale renversée adaptée à la filtration $(\mathcal{F}_n)_{n\geq 0}$ si

— Pour tout $n \geq 0$, X_n est $\mathcal{F}_n$-mesurable.

— Pour tout $n \geq 0$, $\mathbb{E}(X_n|\mathcal{F}_{n+1}) \leq X_{n+1}$.

Théorème 27. *Une surmartingale renversée (X_n) adaptée à la filtration $(\mathcal{F}_n)_{n\geq 0}$ converge presque sûrement vers une variable aléatoire X_∞ mesurable par rapport à la tribu $\mathcal{F}_\infty = \cap_{n\geq 0}\mathcal{F}_n$.*

Démonstration. On procède comme dans le théorème de Doob : si l'on note $U^{[a,b]}$ (resp. $U_n^{[a,b]}$) le nombre de traversées de traversées de l'intervalle $[a, b]$ (resp. le nombre de traversées entre 0 et n), on a par convergence monotone $\mathbb{E}U^{[a,b]} = \lim_{n\to+\infty}\mathbb{E}[U_n^{[a,b]}]$. Cependant, la suite $(X_n, X_{n-1}, \dots, X_0)$ est une surmartingale, donc d'après le théorème des traversées montantes, on a pour tout n : $\mathbb{E}[U_n^{[a,b]}] \leq \frac{\mathbb{E}(X_0-a)^+}{b-a}$. Ainsi $\mathbb{E}U^{[a,b]} \leq \frac{\mathbb{E}(X_0-a)^+}{b-a} < +\infty$. On conclut comme dans la preuve du théorème de Doob. □

3.6 Approximation L^1 par des martingales

Le but de cette section est de démontrer les deux théorèmes suivants :

Théorème 28 (Convergence L^1 des martingales)**.** *Soit $(\mathcal{F}_n)_{n\geq 1}$ une filtration de l'espace $(\Omega, \mathcal{F}, \mathbb{P})$. On note $\mathcal{F}_\infty = \sigma(\cup_{n\geq 1}\mathcal{F}_n))$. Soit Z une variable aléatoire possédant un moment d'ordre 1. On pose $Y_n = \mathbb{E}[Z|\mathcal{F}_n]$. La suite Y_n converge presque sûrement et dans L^1 vers $\mathbb{E}[Z|\mathcal{F}_\infty]$*

Théorème 29 (Convergence L^1 des martingales inverses)**.** *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, $(\mathcal{G}_n)_{n\geq 1}$ une suite décroissante de sous-tribu de $\mathcal{F}$. On note $\mathcal{G}_\infty = \cap_{n\geq 1}\mathcal{G}_n$. Soit Z une variable aléatoire possédant un moment d'ordre 1. On pose $Z_n = \mathbb{E}[Z|\mathcal{G}_n]$. La suite Y_n converge presque sûrement et dans L^1 vers $\mathbb{E}[Z|\mathcal{G}_\infty]$*

On va s'appuyer sur un lemme utile dans les deux preuves.

Lemme 6. *Soient $\mathcal{C}$ et $\mathcal{D}$ deux tribus, $a > 0$. Pour Z variable positive intégrable, on a*

$$\begin{aligned}\|\mathbb{E}[Z|\mathcal{C}] - \mathbb{E}[Z|\mathcal{D}]\|_1 \leq \|&\mathbb{E}[Z \wedge a|\mathcal{C}] - \mathbb{E}[Z \wedge a|\mathcal{D}]\|_1 \\ &+ 2\mathbb{E}[Z\mathbb{1}_{\{Z>a\}}]\end{aligned}$$

Démonstration. On a

$$\mathbb{E}(Z|\mathcal{C}) - \mathbb{E}(Z \wedge a|\mathcal{C}) = \mathbb{E}((Z-a)\mathbb{1}_{\{Z>a\}}|\mathcal{C})$$

Comme l'espérance conditionnelle est une contraction dans L^1, on a

$$\|\mathbb{E}(Z|\mathcal{C}) - \mathbb{E}(Z \wedge a|\mathcal{C})\|_1 \leq \mathbb{E}((Z-a)\mathbb{1}_{\{Z>a\}}).$$

De même,

$$\|\mathbb{E}(Z|\mathcal{D}) - \mathbb{E}(Z \wedge a|\mathcal{D})\|_1 \leq \mathbb{E}((Z-a)\mathbb{1}_{\{Z>a\}}).$$

En utilisant l'inégalité triangulaire dans L^1, on obtient l'inégalité voulue. □

Passons aux preuves des théorèmes.

Démonstration. Par linéarité, avec $Z = Z^+ - Z^-$, il suffit de traiter le cas où $Z \geq 0$. On peut également supposer sans restriction que $\mathbb{E}Z > 0$. D'après le théorème de convergence des martingales de Doob, on sait que (Y_n) converge presque sûrement. Montrons que (Y_n) est de Cauchy dans L^1. Soit $\varepsilon > 0$. Par convergence dominée, on peut se donner a tel que $\mathbb{E}(Z\mathbb{1}_{\{Z>a\}}) \leq \varepsilon/3$. La suite $\mathbb{E}(Z \wedge a|\mathcal{F}_n)$ est une martingale, bornée dans L^2 (par a) , donc qui converge dans L^2, et à plus forte raison dans L^1.

Ainsi, la suite $\mathbb{E}(Z \wedge a|\mathcal{F}_n)$ est de Cauchy dans L^1 : on peut trouver N tels que pour $n,p \geq N$, $\|\mathbb{E}(Z \wedge a|\mathcal{F}_n) - \mathbb{E}(Z \wedge a|\mathcal{F}_p)\|_1 \leq \varepsilon/3$. Avec le lemme, on a $\|\mathbb{E}(Z|\mathcal{F}_n) - \mathbb{E}(Z \wedge a|\mathcal{F}_p)\|_1 \leq \varepsilon$ pour $n,p \geq N$, ce qui montre que la suite (Y_n) est de Cauchy dans L^1, donc convergente dans L^1. Notons Y' la limite. Il reste à voir que $Y' = \mathbb{E}[Z|\mathcal{F}_\infty]$. Pour tout n, Y_n est $\mathcal{F}_n$-mesurable, donc $\mathcal{F}_\infty$-mesurable, et la limite Y' est $\mathcal{F}_\infty$-mesurable.

Fixons un entier naturel n. Soit $A \in \mathcal{F}_n$. Pour $p \geq n$, comme $A \in \mathcal{F}_p$, on a $\mathbb{E}[Z\mathbb{1}_A] = \mathbb{E}[Y_p\mathbb{1}_A]$, donc

$$\begin{aligned}|\mathbb{E}[Z\mathbb{1}_A] - \mathbb{E}[Y'\mathbb{1}_A]| &= |\mathbb{E}[Y_p\mathbb{1}_A] - \mathbb{E}[Y'\mathbb{1}_A]| \\ &\leq \mathbb{E}(|Y_p - Y']1_A) \leq \mathbb{E}|Y_p - Y'|\end{aligned}$$

Finalement $|\mathbb{E}[Z\mathbb{1}_A] - \mathbb{E}[Y'\mathbb{1}_A]| \leq \lim \mathbb{E}|Y_p - Y'| = 0$, d'où $\mathbb{E}[Z\mathbb{1}_A] = \mathbb{E}[Y'\mathbb{1}_A]$. En particulier, prenant $A = \Omega$, on a $\mathbb{E}(Z) = E(Y')$. Les applications $A \mapsto \frac{\mathbb{E}[Z\mathbb{1}_A]}{\mathbb{E}Z}$ et $A \mapsto \frac{\mathbb{E}[Y'\mathbb{1}_A]}{\mathbb{E}Z}$ sont des mesures – ce sont en fait des mesures à densité par rapport à $\mathbb{P}$. Ce sont même des probabilités. D'après ce qui précèdent, ces deux probabilités coïncident sur $\cup_{n\geq 1}\mathcal{F}_n)$ qui est un Π-système engendrant $\mathcal{F}_\infty$: elles coïncident donc sur $\mathcal{F}_\infty$, ce qui montre que Y' est bien une version de l'espérance conditionnelle de Z sachant $\mathcal{F}_\infty$.

□

Les deux preuves sont assez semblables, la deuxième étant peut-être un peu plus facile. On suggère donc à la lectrice de lire la première preuve, puis d'essayer de faire la seconde preuve seule avant de lire la preuve proposée.

Démonstration. Par linéarité, avec $Z = Z^+ - Z^-$, il suffit de traiter le cas où $Z \geq 0$. D'après les théorèmes de convergence des surmartingales renversées, on sait que (Z_n) converge presque sûrement. Montrons qu'elle est de Cauchy dans L^1. Soit $\varepsilon > 0$ et prenons a tel que $\mathbb{E}(Z\mathbb{1}_{\{Z>a\}}) \leq \varepsilon/3$.

La suite $\mathbb{E}(Z \wedge a|\mathcal{G}_n)$ est une martingale renversée : elle converge presque sûrement. Mais une suite de variables aléatoires qui converge presque sûrement et est uniformément bornée par une constante converge dans L^1. Ainsi, il existe N tel que pour $n,p \geq N$, $\|\mathbb{E}(Z \wedge a|\mathcal{G}_n) - \mathbb{E}(Z \wedge a|\mathcal{G}_p)\|_1 \leq \varepsilon/3$, ce qui entraîne avec le lemme que $\|\mathbb{E}(Z|\mathcal{G}_n) - \mathbb{E}(Z \wedge a|\mathcal{G}_p)\|_1 \leq \varepsilon$ pour $n,p \geq N$: la suite (Z_n) est de Cauchy dans L^1, donc convergentes dans L^1. Notons Z' la limite. Il reste à voir que $Z' = \mathbb{E}[Z|\mathcal{G}_\infty]$.

Fixons n. Pour $p \geq n$, Z_p est $\mathcal{G}_p$-mesurable, donc $\mathcal{G}_n$-mesurable, et la limite Z' est $\mathcal{G}_n$-mesurable. Mais si Z' est $\mathcal{G}_n$-mesurable pour tout n, elle est $\mathcal{G}_\infty$-mesurable. Soit $A \in \mathcal{G}_\infty$. Comme $A \in \mathcal{G}_n$, on a $\mathbb{E}[Z\mathbb{1}_A] = \mathbb{E}[Z_n\mathbb{1}_A]$, donc

$$\begin{aligned}|\mathbb{E}[Z\mathbb{1}_A] - \mathbb{E}[Z'\mathbb{1}_A]| &= |\mathbb{E}[Z_n\mathbb{1}_A] - \mathbb{E}[Z'\mathbb{1}_A]| \\ &\leq \mathbb{E}(|Z_n - Z']1_A) \leq \mathbb{E}|Z_n - Z'|\end{aligned}$$

Finalement $|\mathbb{E}[Z\mathbb{1}_A]-\mathbb{E}[Z'\mathbb{1}_A]| \leq \lim \mathbb{E}|Z_n-Z'| = 0$, d'où $\mathbb{E}[Z\mathbb{1}_A] = \mathbb{E}[Z'\mathbb{1}_A]$: Z' est bien une version de l'espérance conditionnelle de Z sachant $\mathcal{G}_\infty$. □

3.7 Décomposition de Doob (*)

Définition: On dit qu'un processus $(\mathcal{F}_n)_{n\geq 0}$- adapté $(C_n)_{n\geq 0}$ est un processus croissant prévisible si $C_0 = 0$, $C_n \leq C_{n+1}$ et si C_{n+1} est $\mathcal{F}_n$-mesurable.

Théorème 30. *Toute sous-martingale $(X_n)_{ge0}$ s'écrit de manière unique comme somme d'une martingale $(M_n)_{n\geq 0}$ et d'un processus croissant prévisible intégrable $(C_n)_{n\geq 0}$.*

Démonstration. Supposons qu'une telle décomposition existe : on a alors

$$\begin{aligned} \mathbb{E}[X_{n+1} - X_n|\mathcal{F}_n] &= \mathbb{E}[M_{n+1} - M_n|\mathcal{F}_n] + \mathbb{E}[C_{n+1} - C_n|\mathcal{F}_n] \\ &= 0 + (C_{n+1} - C_n) \\ &= C_{n+1} - C_n, \end{aligned}$$

car $(M_n)_{n\geq 0}$ est une martingale et $C_{n+1} - C_n$ est $\mathcal{F}_n$-mesurable. Comme $C_0 = 0$, on doit nécessairement avoir

$$C_n = \sum_{i<n} \mathbb{E}[X_{i+1} - X_i|\mathcal{F}_i].$$

Chaque terme de la somme est bien dans $\mathcal{F}_{n-1}$ et est positif, car $(X_n)_{n\geq 0}$ est une sous-martingale, donc C_n est bien croissant prévisible. □

On présente maintenant une jolie application de la Décomposition de Doob. On va donner une nouvelle preuve du théorème de Hunt, basée sur son corollaire relatif aux martingales (lequel peut bien sûr se montrer sans utiliser la version sous-martingale du théorème de Hunt, voir par exemple Stroock).

Corollaire 9 (Théorème de Hunt, version 2)**.** *Soit $(\mathcal{F}_n)_{n\geq 0}$ une filtration et $(X_n)_{n\geq 0}$ une sous-martingale adaptée à la filtration $(\mathcal{F}_n)_{n\geq 0}$.*

Soient S et T deux temps d'arrêts $(\mathcal{F}_n)_{n\geq 0}$-adaptés bornés avec $S \leq T$. Alors

$$E[X_T|\mathcal{F}_S] \geq X_S.$$

Démonstration. On écrit la décomposition de Doob $X_n = M_n + A_n$, avec la martingale $(M_n)_{n\geq 0}$ et le processus croissant prévisible intégrable $(C_n)_{n\geq 0}$. $X_T = M_T + C_T \geq M_T + C_S$, car $(C_n)_{n\geq 0}$ est croissant. Donc $\mathbb{E}[X_T|\mathcal{F}_S] \geq \mathbb{E}[M_T|\mathcal{F}_S] + \mathbb{E}[C_S|\mathcal{F}_S]$. D'après le théorème de Hunt sur les martingales,

$\mathbb{E}[M_T|\mathcal{F}_S] = M_S$. D'autre part , le processus $(C_n)_{n\geq 0}$ est adapté, donc d'après le lemme 4, C_S est $\mathcal{F}_S$-mesurable et $\mathbb{E}[C_S|\mathcal{F}_S] = C_S$. Finalement, $\mathbb{E}[X_T|\mathcal{F}_S] \geq M_S + C_S = X_S$. $\square$

3.8 Exercices sur les martingales

3.8.1 Exercices corrigés

Exercice 23. Soit $a \in \mathbb{R}$. Soit $(U_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. On définit par récurrence une suite X_n par $X_0 = a$ et $X_{n+1} = U_{n+1} + (1-U_{n+1})X_n^2$.

1. Montrer que la suite $(X_n)_{n\geq 0}$ est une sous-martingale.
2. On suppose maintenant que $a \in [0,1]$.
   (a) Montrer que $(X_n)_{n\geq 0}$ converge presque sûrement vers une variable aléatoire X^*.
   (b) Donner la valeur de X^*.

lien vers l'indication lien vers la solution

Exercice 24. *Urne de Pólya*
Une urne contient des boules de m couleurs différentes. On suit le procédé suivant : on prend au hasard (avec équiprobabilité) une boule dans l'urne, puis on la replace dans l'urne en même temps qu'on y ajoute S boules de la couleur tirée. On s'intéresse au comportement asymptotique de la répartition des couleurs dans l'urne.

On modélise le problème comme suit : on suppose que l'urne contient initialement d_1 boules de couleur 1, ..., d_m boules de couleur m. On pose $d = d_1 + \dots + d_m$. On se donne un vecteur déterministe $(B_1, \dots, B_d)$ à valeurs dans $\{1, \dots, m\}$ de telle sorte que d_i est le nombre d'apparitions de i dans la suite $B_1, \dots, B_d$. On prend maintenant $(U_n)_{n\geq 1}$, une suite de variables aléatoires indépendantes telles que pour tout n, U_n suive la loi uniforme sur $\{1, \dots, d+(n-1)S\}$. On définit alors par récurrence les suites $(B_i)_{i>d}$ et $(T_n)_{n\geq 1}$ en posant, pour $n \geq 1$:

$$T_n = B_{U_n} \text{ puis, pour } i \text{ entre } 1 \text{ et } S : B_{d+i+(n-1)S} = T_n.$$

$(B_i)_{i\geq 1}$ représente la suite des couleurs des boules successivement ajoutées dans l'urne, tandis que $(T_i)_{i\geq 1}$ représente la suite des tirages.

On note $V_n = \sum_{k=1}^{d+nS} e_{B_k}$: V_n représente le vecteur des effectifs des différentes couleurs avant le $n+1$-ième tirage. On a donc $V_0 = (d_1, \dots, d_m)$. Notons que par construction $V_{n+1} = V_n + Se_{T_{n+1}}$. On pose $V_n^i = \langle V_n, e_i \rangle$: c'est le nombre de boules de couleur i dans l'urne avant le n-ième tirage.

Pour $n \geq 1$, n note $\mathcal{F}_n$ la tribu engendrée par $U_1, \dots, U_n$; on pose également $\mathcal{F}_0 = \{\varnothing, \Omega\}$.

1. (a) Montrer que pour tout $n \geq 1$ et pour tout i entre 1 et S, $B_{d+(n-1)S+i}$ est $\mathcal{F}_n$-mesurable.

(b) Montrer que $V_{n+1}^i = V_n^i + S\sum_{k=1}^{d+nS} \mathbb{1}_{\{U_{n+1}=k\}}\mathbb{1}_{\{B_k=i\}}$

(c) En déduire que $\mathbb{E}[V_{n+1}^i|\mathcal{F}_n] = V_n^i + \frac{SV_n^i}{d+Sn}$.

(d) Montrer que $(\frac{V_n^i}{Sn+d})_{n\geq 0}$ est une martingale.

(e) En déduire que la suite de vecteurs aléatoires $V_n/(Sn+d)$ converge presque sûrement vers un vecteur aléatoire W.

(f) Montrer que pour tout $n \geq 1$ et tout $i \in \{1,\ldots,m\}$, on a $\mathbb{P}(T_n = i) = \frac{d_i}{d}$.

2. Dans la suite de l'exercice, on va chercher à déterminer la loi de W. On a besoin à cet effet de quelques rappels sur les lois de Dirichlet. Par définition, la loi de Dirichlet de paramètre $(a_1,\ldots,a_m)$ est la loi du vecteur $(Y_1,\ldots,Y_m) = (\frac{X_1}{X_1+\cdots+X_m},\ldots,\frac{X_m}{X_1+\cdots+X_m})$, où $X_1,\ldots,X_m$ sont des variables aléatoires indépendantes avec pour tout i entre 1 et m : $X_i \sim \Gamma(a_i,1)$. On peut démontrer (et c'est en fait la seule propriété des lois de Dirichlet qui sera utile ici) que pour toute suite d'entiers $(b_1,\ldots,b_m)$, on a

$$\mathbb{E}[\prod_{i=1}^m Y_i^{b_i}] = \frac{\tilde{B}(a+b)}{\tilde{B}(a)}, \text{ avec } \tilde{B}(a) = \frac{\prod_{i=1}^m \Gamma(a_i)}{\Gamma(\sum_{i=1}^m a_i)}.$$

(a) Soit $n \geq 1$, $(t_1,\ldots,t_n) \in \{1,\ldots,m\}^n$. Montrer que

$$\mathbb{P}(T_1 = t_1,\ldots T_n = t_n) = \frac{\prod_{i=1}^m d_i(d_i+S)\ldots(d_i+(a_i-1)S)}{d(d+S)\ldots(d+(n-1)S)}, \tag{3.2}$$

où a_i est le nombre d'apparitions de i dans la suite $t_1,\ldots,t_n$, puis que

$$\mathbb{P}(T_1 = t_1,\ldots T_n = t_n) = \mathbb{E}\left(\prod_{i=1}^m Y_i^{a_i}\right), \tag{3.3}$$

où le vecteur $(Y_1,\ldots,Y_m)$ suit la loi de Dirichlet de paramètre $\frac{d}{S}$. Remarque : pour k entre 1 et n, on pourra avoir intérêt à noter a_i^k le nombre d'apparitions de i dans la suite $t_1,\ldots,t_k$.

(b) Montrer que pour toute suite d'entiers $a_1,\ldots,a_m$ avec $a_1+\ldots a_m = n$, on a

$$\mathbb{P}((V_n-d)/S = (a_1,\ldots,a_m)) = \binom{n}{a_1,\ldots,a_m}\mathbb{E}[\prod_{i=1}^m Y_i^{a_i}].$$

On rappelle que le coefficient multinomial $\binom{n}{a_1,a_2,\dots,a_m}$ est, par définition, le nombre d'applications de $\{1,\dots,n\}$ dans $\{1,\dots,m\}$ prenant a_i fois la valeur i. On a la formule du multinôme :

$$\left(\sum_{j=1}^m X_j\right)^n = \sum_{(a_1,\dots,a_m)} \binom{n}{a_1,a_2,\dots,a_m} \prod_{k=1}^m X_k^{a_k},$$

où la sommation a lieu sur les m-uplets d'entiers naturels de somme n.

(c) On note $\varphi_n(u)$ la fonction caractéristique du vecteur $(V_n - d)/S$. Montrer que

$$\varphi_n(u) = \mathbb{E}\left(\sum_{k=1}^m Y_k e^{iu_k}\right)^n.$$

(d) Montrer que $V_n/(Sn+d)$ converge en loi vers Y.

(e) Identifier la loi de W.

Toute la preuve semble reposer sur l'identité miraculeuse (3.3). En réalité, l'équation (3.2) entraîne que les T_i sont échangeables, c'est à dire que leur loi est invariante par permutation d'un nombre fini de coordonnées. Dans ce cas, un théorème de De Finetti–Hewitt–Savage entraîne que conditionnellement à une certaine tribu $\mathcal{T}$, les T_i sont indépendants et de même loi. Ici, par exemple on peut démontrer que

$$\mathbb{P}(T_1 = t_1, \dots T_n = t_n | W) = \prod_{i=1}^m W_{t_i},$$

ce qui entraîne (3.3).

lien vers l'indication lien vers la solution

3.8.2 Exercices non corrigés

Exercice 25. Soient τ_1 et τ_2 des temps d'arrêts adaptés à la filtration $(\mathcal{F}_t)_{t\ge 0}$. Montrer que $\tau = \max(\tau_1, \tau_2)$ est un temps d'arrêt et que $\mathcal{F}_\tau = \sigma(\mathcal{F}_{\tau_1}, \mathcal{F}_{\tau_2})$.
lien vers l'indication

Exercice 26. Soit $(X_n)_{n\ge 1}$ une suite de variables aléatoires indépendantes identiquement distribuées admettant un moment d'ordre 3 et vérifiant $\mathbb{E}X_1 = \mathbb{E}X_1^3 = 0$ et $\mathbb{E}X_1^2 = 1$. On note $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ et on pose

$$S_n = \sum_{k=1}^n X_i.$$

1. On pose $Y_n = S_n^3 - 3nS_n$. Montrer que $(Y_n)_{n\geq 1}$ est une martingale par rapport à la filtration $(\mathcal{F}_n)_{n\geq 1}$.
2. Soient a, b, c, d des réels. On pose

$$Q(x,t) = x^2 + axt + bx + ct + d.$$

   Pour quelles valeurs du quadruplet (a, b, c, d) la suite $Z_n = Q(S_n, n)$ forme-t-elle une martingale rapport à la filtration $(\mathcal{F}_n)_{n\geq 1}$?

lien vers l'indication

Exercice 27. Soit $(X_n)_{n\geq 1}$ une surmartingale $(\mathcal{F}_n)_{n\geq 1}$-adaptée. On suppose que les $(X_n)_{n\geq 1}$ ont toutes la même loi.

1. Montrer que $(X_n)_{n\geq 1}$ est une martingale.
2. Montrer que pour tout réel a les suites $(X_n - a)^+_{n\geq 1}$ et $(X_n - a)^-_{n\geq 1}$ sont des martingales.
3. En déduire que pour $n > p \geq 1$, X_n est presque sûrement supérieur ou égal à a sur l'événement $\{X_p \geq a\}$.
4. En déduire que $(X_n)_{n\geq 1}$ est presque sûrement constante.

lien vers l'indication

Exercice 28. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes identiquement distribuées dont la loi commune est non dégénérée et à support compact. On pose $S_n = X_1 + \cdots + X_n$, $\varphi(t) = \log \mathbb{E}e^{tX_1}$ et

$$Y_n^t = e^{tS_n - n\varphi(t)}.$$

1. Montrer que $(Y_n^t)_{n\geq 1}$ est une martingale par rapport à la filtration naturelle associée aux $(X_n)_{n\geq 1}$.
2. On suppose désormais que t est non-nul. Montrer que $\varphi(t/2) < \varphi(t)/2$.
3. En déduire que $(Y_n^t)_{n\geq 1}$ converge presque sûrement vers 0.
4. Retrouver ce résultat à partir de la loi forte des grands nombres.

lien vers l'indication

Exercice 29. *Retour sur les théorèmes d'approximation L^1 par des martingales*
Lors de la preuve des théorèmes de convergence d'approximation L^1 par des martingales (théorèmes 28 et 29) , on a vu qu'une étape importante de la preuve était de passer de la convergence presque sûre à la convergence L^1. Pour cela, on s'est appuyé sur le lemme 6. Une solution un peu plus longue,

mais éclairante, est de passer par l'équi-intégrabilité.

Soit X une variable aléatoire intégrable sur $(\Omega, \mathcal{F}, \mathbb{P})$. Montrer que la famille de variables $\mathbb{E}(X|\mathcal{G})$, où $\mathcal{G}$ décrit l'ensemble des sous-tribus de $\mathcal{F}$ est équi-intégrable. Conclure. lien vers l'indication

Exercice 30. Soit $a \in [0, \pi/2]$. Soit $(U_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant la loi uniforme sur $[0, 1]$. On définit par récurrence une suite X_n par $X_0 = a$ et $X_{n+1} = U_{n+1} \sin X_n$. Montrer que $(X_n)_{n\geq 1}$ prend des valeurs positives, puis que la suite $(2^n X_n)_{n\geq 0}$ est une surmartingale. lien vers l'indication

Exercice 31. *Arrêt optimal pour une marche aléatoire.*
On considère le jeu suivant. J'ai un pion qui se déplace sur les entiers compris entre 0 et n. A chaque point $i \in \{1, \dots, n-1\}$ est associé un somme $f(i)$ strictement positive (on la prolonge par $f(0) = f(n) = 0$). On prolonge encore f en une fonction continue par morceau définie sur $[0, n]$ en posant $f(\theta k + (1-\theta)(k+1)) = \theta f(k) + (1-\theta) f(k+1)$ pour tout $k \in \{0, \dots, n-1\}$ et tout $\theta \in]0, 1[$.

Mon pion part d'un point $i_0 \in \{1, \dots, n-1\}$; à chaque étape, je peux décider de partir avec le gain correspondant à ma position actuelle, ou alors lancer une pièce équilibrée qui me donnera ma position suivante (juste à droite si 'pile', juste à gauche si 'face'). Si je touche 0 ou n je ne gagne rien et je suis éliminé. Quelle stratégie adopter ?

Ce problème revient à trouver un temps d'arrêt T optimal pour la marche aléatoire. Notons X_n la position de mon pion à l'instant n, et $\mathcal{F}_n$ la tribu engendrée par $X_0, \dots, X_n$.

1. Soit g une fonction concave supérieure à f. Montrer que $(g(X_n))_{n\in\mathbb{N}}$ est une surmartingale.
2. Soit T un temps d'arrêt fini p.s. Montrer que
$$\mathbb{E}(f(X_T)) \leq \mathbb{E}(g(X_T)) \leq g(i_0).$$
3. Notons Ψ l'enveloppe concave de f, c'est à dire
$$\Psi(x) = \inf\{g(x); g \in S(f)\},$$
où $S(f)$ est l'ensemble des fonctions concaves de $[0, n]$ dans $\mathbb{R}$ qui sont supérieures à f. Montrer que
$$\mathbb{E}(f(X_T)) \leq \Psi(i_0).$$

4. Montrer que Ψ est une fonction concave. Soit $]s,t[\subset [0,n]$ tel que $\{x \in [s,t]; f(x) = \Psi(x)\} = \{s,t\}$. Montrer que Ψ est affine sur $[s,t]$.
5. On définit $T_{\text{opt}} = \min\{n \in \mathbb{N}, f(X_n) = \Psi(X_n)\}$, ainsi que
$$A = \min\{j \geq i_0, f(j) = \Psi(j)\} \text{ et } B = \max\{j \leq i_0, f(j) = \Psi(j)\}.$$
Calculer $\mathbb{E}(f(X_{T_{\text{opt}}}))$ en fonction de $f(A)$ et $f(B)$, et en déduire que
$$\mathbb{E}(f(X_{T_{\text{opt}}})) = \Psi(i_0).$$

lien vers l'indication

Exercice 32. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes centrées, de variance 1. Montrer que la série de terme général $Y_n = a_n X_1 \dots X_n$ converge presque sûrement et dans L^2. lien vers l'indication

Chapitre 4

Compléments de théorie de la mesure

4.1 Rappels de topologie

4.1.1 Topologie produit

Si les (E_i, d_i) sont des espaces métriques, on définit une distance sur $\prod_{i=1}^{+\infty} E_i$ par

$$\forall (x,y) \in \prod_{i=1}^{+\infty} E_i \quad d(x,y) = \sum_{i=1}^{+\infty} 2^{-i} \arctan d_i(x_i, y_i).$$

Notons π_i la projection sur la i-ième coordonnée : $\pi_i(x) = x_i$.

La suite $(x^{(n)})_{n\geq 1}$ converge vers x si et seulement si pour tout i $(\pi_i(x^{(n)})_{n\geq 1} = (x_i^{(n)})_{n\geq 1}$ converge vers $x_i = \pi_i(x)$.

Démonstration. En effet, on a d'un côté l'inégalité

$$d_i(x_i^n, x_i) = \tan(\arctan d_i(x_i^n, x_i)) \leq \tan(2^i d(x^n, x)).$$

Ainsi, si $d(x^n, x)$ tend vers 0, $d_i(x_i^n, x_i)$ tend vers 0. Réciproquement, si $d_i(x_i^n, x_i)$ tend vers 0 pour tout i, on a

— $\forall i \geq 1 \quad \lim_{n\to+\infty} 2^{-i} \arctan d_i(x_i, x_i^n) = 0$

— $\forall i \geq 1 \quad \forall n \geq 1 \quad |2^{-i} \arctan d_i(x_i, x_i^n)| \leq 2^{-i}\pi/2$

— $\sum_{i\geq 1} 2^{-i}\pi/2 < +\infty$

Ainsi, le résultat annoncé découle du théorème de convergence dominée pour la mesure de comptage. □

Corollaire 10. *Un produit dénombrable d'espaces métriques complets est complet.*

4.1.2 Espaces polonais

Définition. *On introduit les deux définitions suivantes :*
- *Un espace métrique est séparable s'il contient une partie dénombrable dense*
- *Un espace polonais est un espace métrique complet et séparable.*

Par exemple, $\mathbb{R}$ muni de la métrique usuelle est un espace polonais : $\mathbb{R}$ est complet et $\mathbb{Q}$ est dense dans $\mathbb{R}$.

Théorème 31 (Lemme de Doob)**.** *Soient Ω_1, Ω_2 des ensembles. On suppose que X est une application de Ω_1 dans Ω_2, que $\mathcal{F}_2$ est une tribu sur Ω_2. On note alors $\mathcal{F}_1$ la tribu engendrée par X, c'est à dire la plus petite tribu $\mathcal{F}_1$ qui rende l'application X $(\Omega_1, \mathcal{F}_1) - (\Omega_2, \mathcal{F}_2)$ mesurable. Alors, si $(E, \mathcal{B})$ est un espace polonais muni de sa tribu borélienne, les fonctions $(\Omega_1, \mathcal{F}_1) - (E, \mathcal{B})$ mesurables sont celles qui s'écrivent sous la forme*

$$Y = f(X),$$

où f est une application $(\Omega_2, \mathcal{F}_2) - (E, \mathcal{B})$ mesurable.

Démonstration. Prenons Y une variable $(\Omega_1, \mathcal{F}_1) - (E, \mathcal{B})$ mesurable et montrons qu'elle s'écrit sous la forme demandée (le sens inverse est immédiat). On commence par traiter le cas où Y ne prend qu'un nombre fini de valeurs $y_1, \ldots, y_n$. Pour tout k entre 1 et n, $\{y_k\} \in \mathcal{E}$, donc $Y^{-1}(\{y_k\}) \in \mathcal{F}_1$. Mais les éléments $\mathcal{F}_1$ sont exactement les images réciproques par X des éléments de $\mathcal{F}_2$, donc il existe $A_k \in \mathcal{F}_2$, avec $Y^{-1}(\{y_k\}) = X^{-1}(A_k)$. Si l'on pose $f(x) = \sum_{i=1}^k y_k \mathbb{1}_{A_k}$, il est alors clair que f est une application $(\Omega_2, \mathcal{F}_2) - (E, \mathcal{B})$ mesurable et que $Y = f(X)$.

Pour traiter le cas général, on a besoin d'une technique d'approximation : considérons une suite $(y_n)_{n \geq 1}$ d'éléments de E qui est dense dans E. On pose alors pour tout $y \in E$:

$$\varphi_n(y) = \sum_{k=1}^n \prod_{i<k} \mathbb{1}_{\{d(y,y_i) > d(y,y_k)\}} \prod_{k<i\leq n} \mathbb{1}_{\{d(y,y_i) \geq d(y,y_k)\}} y_k.$$

Comme les fonctions $y \mapsto d(y, y_i)$ sont continues, elles sont $(E, \mathcal{B}) - (\mathbb{R}, \mathcal{B}(\mathbb{R}))$-mesurable. Ainsi, les fonctions $y \mapsto d(y, y_i) - d(y, y_k)$ sont $(E, \mathcal{B}) - (\mathbb{R}, \mathcal{B}(\mathbb{R}))$-mesurables : on en déduit que les ensembles $\{d(y, y_i) > d(y, y_k)\}$ et $\{d(y, y_i) \geq d(y, y_k)\}$ sont dans $\mathcal{B}$, puis que φ_n est $(E, \mathcal{B}) - (E, \mathcal{B})$-mesurable. $\varphi_n(y)$ est l'élément le plus proche de y parmi $y_1, \ldots, y_k$ (on prend celui de plus petit index en cas d'égalité). Comme la suite (y_n) est dense dans E, $\varphi_n(y)$ converge vers y quand n tend vers l'infini. [1]

1. Si $E = \mathbb{R}$ ou $E = \mathbb{R}^d$, on peut prendre plus simplement $\varphi_n(x) = \frac{\lfloor nx \rfloor}{n} \mathbb{1}_{[0,n]}(\|x\|)$.

Prenons maintenant Y générale. Si on pose $Y_n = \varphi_n(Y)$, Y_n est une fonction $(\Omega_1, \mathcal{F}_1)-(E,\mathcal{B})$ mesurable qui ne prend qu'un nombre fini de valeurs : on peut écrire $Y_n = f_n(X)$, où f_n est une application $(\Omega_2, \mathcal{F}_2)-(E,\mathcal{B})$ mesurable. Comme (E,d) est complet, si on pose $C = \{x \in \Omega_2; (f_n(x))_{n\geq 1} \text{ converge }\}$, on a

$$C = \cap_{k\geq 1} \cup_{N\geq 1} \cap_{n,p\geq N} \{d(f_n(x), f_p(x)) \leq 1/k\},$$

donc $C \in \mathcal{F}_2$. Soit $z_0 \in E$ quelconque. On pose $g_n(x) = \mathbb{1}_C(x) f_n(x) + z_0 \mathbb{1}_{C'}$. Il est aisé de voir que g_n est encore $(\Omega_2, \mathcal{F}_2) - (E,\mathcal{B})$ mesurable. Par construction, $g_n(x)$ converge pour tout x. La limite g est $(\Omega_2, \mathcal{F}_2) - (E,\mathcal{B})$ mesurable comme limite simple d'applications $(\Omega_2, \mathcal{F}_2)-(E,\mathcal{B})$ mesurables. [2] Pour tout $\omega \in \Omega_1$, on a $f_n(X(\omega)) = \varphi_n(Y(\omega))$ qui converge vers $Y(\omega)$, donc $X(\omega) \in C$, et $g_n(X(\omega)) = f_n(X(\omega))$ converge vers $Y(\omega)$. Comme $g_n(X(\omega))$ converge vers $g(X(\omega))$, on a $Y(\omega) = g(X(\omega))$, ce qui est le résultat voulu. $\square$

Théorème 32. *Un produit dénombrable d'espaces polonais est polonais.*

Démonstration. Soient (E_i, d_i) des espaces polonais. Vu le corollaire précédent, il suffit de montrer que $\prod_{i=1}^{+\infty} E_i$ a une partie dénombrable dense. Notons D_i une partie dénombrable dense dans E_i et x^* un élément quelconque de $\prod_{i=1}^{+\infty} E_i$. Posons $D = \cup_{n\geq 1} E_n$, avec

$$E_n = \prod_{i=1}^{n} D_i \times \prod_{i=n}^{+\infty} \{x_i^*\}.$$

E_n est dénombrable pour tout n, donc D est dénombrable. Montrons que D est dense dans $\prod_{i=1}^{+\infty} E_i$. Soit x dans E Soit $\varepsilon > 0$. Il existe n tel que $\pi 2^{-n} \leq \varepsilon$. Par densité, on peut trouver $y \in E_n$ tel que pour tout i entre 1 et n, $d(y_i, x_i) \leq \frac{\varepsilon}{2}$. On a alors

$$\begin{aligned}
d(x,y) &= \sum_{i=1}^{+\infty} 2^{-i} \arctan d_i(x_i, y_i) \\
&\leq \sum_{i=1}^{n} 2^{-i} \arctan d_i(x_i, y_i) + \sum_{i=n+1}^{+\infty} 2^{-i} \arctan d_i(x_i, y_i) \\
&\leq \sum_{i=1}^{n} 2^{-i} d_i(x_i, y_i) + \sum_{i=n+1}^{+\infty} 2^{-i} \pi/2 \\
&\leq \sum_{i=1}^{n} 2^{-i} \varepsilon/2 + \sum_{i=n+1}^{+\infty} 2^{-i} \pi/2 \\
&\leq \varepsilon/2 + 2^{-n}\pi/2 \leq \varepsilon
\end{aligned}$$

2. Cette propriété est bien connue pour des variables aléatoires réelles. Le cas des variables à valeur dans un espace métrique sera traité en exercice.

□

En particulier $\mathbb{R}^{\mathbb{N}^*}$ est un espace polonais. C'est important en probabilités car $\mathbb{R}^{\mathbb{N}^*}$ est typiquement l'espace sur lequel on peut faire vivre une suite de variables aléatoires.

Théorème 33. *Soient (E, d) un espace métrique séparable et $\mathcal{B}$ sa tribu borélienne. $\mathcal{B}$ est engendrée par une famille dénombrable de boules ouvertes.*

Démonstration. Soit O un ouvert de E. E est un espace métrique qui admet une partie dense D, donc O s'écrit comme réunion de boules ayant leur centre dans $O \cap D$ et un rayon rationnel.

En effet, pour $x \in O$, on peut trouver $n_x \in \mathbb{N}^*$ tel que $B(x, 2/n_x) \subset O$. Si on prend c_x quelconque avec $c_x \in B(x, 1/n_x) \cap D$, alors $x \in B(c_x, 1/n_x) \subset B(x, 2/n_x) \subset O$. On a alors $O = \cup_{x \in O} B(c_x, 1/n_x)$.

Cette réunion est nécessairement dénombrable, puisque les centres et les rayons sont dénombrables, donc O est dans la tribu engendrée par les boules dont le centre est dans D et le rayon rationnel. □

Théorème 34. *La tribu $\mathcal{B}$ engendrée par les ouverts de $\mathbb{R}^{\mathbb{N}^*}$ coïncide avec la tribu engendrée par les applications projection π_i*

Démonstration. On a vu que l'application π est continue de $(\mathbb{R}^{\mathbb{N}^*}, d)$ dans $(\mathbb{R}, |\cdot|)$. Elle est donc mesurable entre les tribus boréliennes associées. On a donc $\sigma((\pi_i)_{i \geq 1}) \subset \mathcal{B}$. Pour y dans $\mathbb{R}^{\mathbb{N}^*}$, la fonction

$$d_y(x) = d(x, y) = \sum_{i=1}^{+\infty} \frac{\arctan |x_i - y_i|}{2^i} = \lim_{N \to +\infty} \sum_{i=1}^{N} \arctan \frac{|\pi_i(x) - y_i|}{2^i}$$

est une limite d'applications $\sigma((\pi_i)_{i \geq 1})$-mesurables donc est $\sigma(\pi_i)$-mesurable. Il s'ensuit que les boules de $(\mathbb{R}^{\mathbb{N}^*}, d)$ sont des éléments de $\sigma((\pi_i)_{i \geq 1})$. Or d'après le théorème précédent, les boules ouvertes engendrent la tribu borélienne, donc $\mathcal{B}$ est incluse dans $\sigma((\pi_i)_{i \geq 1})$, ce qui achève la preuve. □

Théorème 35. *Soient (E, d) un espace métrique séparable. Il existe une famille dénombrable $(O_i)_{i \in I}$ d'ouverts de (E, d) telles que deux mesures boréliennes qui coïncident sur les O_i sont égales.*

Démonstration. Comme précédemment, on a une famille dénombrable de boules $(B_d)_{d \in D}$ telles que tout ouvert s'écrive comme réunion dénombrable de boules $(B_{d_n})_{n \geq 1}$ avec $d_n \in D$ pour tout n. Soit I l'ensemble des parties finies de D. I est dénombrable. Posons $O_A = \cap_{i \in A} B_i$.

Soit O un ouvert, avec $O = \cup_{n\geq 1} B_{d_n}$. On a pour toute mesure μ :

$$\begin{aligned}\mu(O) &= \lim_{n\to+\infty} \mu(\cup_{k=1}^n B_{d_k}) \\ &= \lim_{n\to+\infty} \sum_{A\subset\{d_1,\dots,d_n\}, A\neq\varnothing} (-1)^{|A|+1}\mu(O_A)\end{aligned}$$

Ainsi, si deux mesures coïncident sur les O_A, elles coïncident sur les ouverts, donc sur la tribu borélienne. □

4.2 Notion de loi conditionnelle

Soit Ω un espace polonais. On note $M_1(\Omega)$ l'ensemble des mesures de probabilités sur $(\Omega, \mathcal{B}(\Omega))$. On munit $M_1(\Omega)$ de la tribu $\mathcal{T}$ engendrée par les applications $\mu \mapsto \int f\ d\mu$, où f décrit l'ensemble des fonctions continues bornées sur Ω.

4.2.1 Le théorème général

Théorème 36. *Soit Ω un espace polonais. On pose $\mathcal{F} = \mathcal{B}(\Omega)$. Soit P une probabilité sur $(\Omega, \mathcal{F})$. Pour toute sous-tribu $\mathcal{S}$ de $\mathcal{F}$, il existe une application*

$$\begin{aligned}\Omega &\to M_1(\Omega) \\ \omega &\mapsto \mathbb{P}_\omega^{\mathcal{S}}\end{aligned}$$

qui est $(\Omega, \mathcal{S})$-$(M_1(\Omega), \mathcal{T})$ mesurable et telle que pour tout $A \in \mathcal{S}$ et $B \in \mathcal{F}$,

$$P(A\cap B) = \int_A \mathbb{P}_\omega^{\mathcal{S}}(B)\ dP(\omega). \tag{4.1}$$

Cette application est P presque sûrement unique. Pour toute variable aléatoire X à valeurs dans $[0, +\infty]$, l'application

$$\omega \mapsto \int_\Omega X\ d\mathbb{P}_\omega^{\mathcal{S}}$$

est une version de l'espérance conditionnelle de X sachant $\mathcal{S}$.

Avant de faire la preuve, on rappelle (?) un résultat d'analyse fonctionnelle :

Proposition 4 (Théorème de représentation de Riesz)**.** *Pour toute forme linéaire positive L sur $(\mathcal{C}([0,1]^d), \|\cdot\|_\infty)$, il existe une mesure μ sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ telle que*

$$\forall f \in \mathcal{C}([0,1]^d) \quad L(f) = \int_{[0,1]^d} f\ d\mu.$$

Démonstration. Il n'y a pas de preuve simple du théorème de représentation de Riesz. On devine vite qu'une bonne définition de μ - si elle existe est

$$\mu(A) = \sup\{L(f); 0 \le f \le \mathbb{1}_A.\}$$

Toute la difficulté consiste à montrer que l'objet μ ainsi défini est bien une mesure sur la tribu borélienne. Les preuves reposent de manière plus ou moins explicite sur la notion de mesure extérieure, de manière très explicite dans Briane–Pagès [1] (chapitre 10), de manière plus cachée dans Rudin [4] (chapitre 1) ou Hirsch–Lacombe [3] (chapitre 2).

En fait, le théorème de représentation de Riesz est la pierre de base de la théorie de l'intégration par rapport à une mesure de Radon. Cette théorie part de la notion d'intégrale (une forme linéaire positive sur les fonctions continues à support compact) pour aller à la notion de mesure de Radon. À l'inverse, la théorie de l'intégrale par rapport à une mesure abstraite, telle que développée dans Briane–Pagès [1] ou Garet–Kurtzmann [2] part d'une mesure (pas nécessairement de Radon) pour construire l'intégrale. La mesure de Lebesgue étant une mesure de Radon, la théorie de l'intégrale par rapport à une mesure de Radon est suffisante pour la plupart des besoins en analyse, en revanche, elle est incomplète aux yeux des probabilistes, qui ont besoin d'intégrer sur des espaces plus généraux. □

Démonstration. Montrons l'unicité. Soit $\omega \mapsto \mathbb{P}^{\mathcal{S}}$ et $\omega \mapsto \mathbb{Q}^{\mathcal{S}}$ deux solutions. Comme Ω est polonais, il existe une famille dénombrable d'ouverts $(O_n)_{n\ge1}$ telle que décrite dans le théorème 35. Pour tout n, $\omega \mapsto \mathbb{P}^{\mathcal{S}}_\omega(O_n)$ et $\omega \mapsto \mathbb{Q}^{\mathcal{S}}_\omega(O_n)$ sont des versions de l'espérance conditionnelle de $\mathbb{1}_{O_n}$ sachant $\mathcal{S}$. Ainsi, l'ensemble

$$B = \cap_{n\ge1}\{\omega : \mathbb{P}^{\mathcal{S}}_\omega(O_n) = \mathbb{Q}^{\mathcal{S}}_\omega(O_n)\}$$

est de mesure 1. D'après le théorème 35, pour tout $\omega \in B$, les mesures $\mathbb{P}^{\mathcal{S}}_\omega$ et $\mathbb{Q}^{\mathcal{S}}_\omega$ coïncident : c'est ce que l'on voulait démontrer.

Passons à l'existence. Afin d'éviter les détails techniques d'analyse fonctionnelle, on va se contenter de donner la preuve dans le cas où $\Omega = [0,1]^d$. [3] On renvoie le lecteur à l'ouvrage de Stroock [5] pour une preuve dans le cas général. On sait que les polynômes forment une famille dense de $C([0,1]^d, \|\cdot\|_\infty)$. On en déduit aisément que l'ensemble $\mathcal{P}$ des polynômes à coefficient rationnel est une famille dénombrable dense de $C([0,1]^d, \|\cdot\|_\infty)$.

3. C'est une restriction minime. On verra dans la preuve du théorème de Kolmogorov (théorème 61, voir aussi la note de bas de page suivant la preuve) que dès qu'un espace permet de faire vivre une variable aléatoire suivant la loi uniforme sur $[0,1]$, on peut y faire vivre à peu près tout ce qu'on veut.

Posons $g_{0,\dots,0} = 1$, puis, pour $n \in \mathbb{N}^d \backslash \{(0,\dots,0)\}$, notons g_n une version de l'espérance conditionnelle de $\omega \mapsto \prod_{i=1}^d \omega_i^{n_i}$ sachant $\mathcal{S}$. À ω fixé, il existe une application linéaire $\Lambda_\omega : \mathbb{R}[X_1,\dots,X_d] \to \mathbb{R}$ telle que $\Lambda_\omega(\prod_{i=1}^d X_i^{n_i}) = g_n(\omega)$ pour tout n. Il est aisé de constater que pour tout $f \in \mathbb{R}[X_1,\dots,X_d]$, $\omega \mapsto \Lambda_\omega(f)$ est une version de l'espérance conditionnelle de f sachant $\mathcal{S}$, c'est à dire que $\Lambda_\omega(f)$ est $\mathcal{S}$-mesurable et que

$$\forall A \in \mathcal{S} \quad \int_A f \; d\mathbb{P} = \int_A \Lambda_\omega(f) \; d\mathbb{P}(\omega) \tag{4.2}$$

Posons

$$\mathcal{N} = \cup_{f \in \mathbb{Q}[X_1,\dots,X_d]; f \geq \alpha} \{\omega \in \Omega; \Lambda_\omega(f) < 0\}.$$

Comme réunion dénombrable d'éléments de $\mathcal{S}$ de probabilité nulle, $\mathcal{N}$ est un élément de $\mathcal{S}$ de probabilité nulle. Pour $\omega \in \mathcal{N}^c$ et $f \in \mathbb{Q}[X_1,\dots,X_d]$, on a pour tout $\alpha > \|f\|_\infty$ avec α rationnel : $\alpha + f \geq 0$ et $\alpha - f \geq 0$, donc $\Lambda_\omega(\alpha + f) \geq 0$, soit $\alpha\Lambda_\omega(1) + \Lambda_\omega(f) \geq 0$ et $\alpha\Lambda_\omega(1) - \Lambda_\omega(f) \geq 0$. Comme $\Lambda_\omega(1) = 1$, on en déduit $|\Lambda_\omega(f)| \leq \alpha$. Comme α peut être pris aussi proche de $\|f\|_\infty$ que l'on veut, on a $|\Lambda_\omega(f)| \leq \|f\|_\infty$. On a donc montré

$$\forall \omega \in \mathcal{N}^c \quad \forall f \in \mathcal{P} \quad |\Lambda_\omega(f)| \leq \|f\|_\infty.$$

Comme $\mathcal{P}$ est dense dans $C(\Omega, \|\cdot\|_\infty)$, pour tout $\omega \in \mathcal{N}^c$, Λ_ω se prolonge en une application linéaire de norme 1 de $C(\Omega, \|\cdot\|_\infty)$ dans $\mathbb{R}$.

Soit $f \in C(\Omega)$ une fonction positive. On peut trouver $\tilde{f} \in \mathcal{P}$ avec $\|f - \tilde{f}\|_\infty \leq 1/n$. On a

$$\begin{aligned}
\Lambda_\omega(f) &= \Lambda_\omega(\tilde{f} + 1/n) + \Lambda_\omega(f - \tilde{f} - 1/n) \\
&\geq \Lambda_\omega(\tilde{f} + 1/n) - \|f - \tilde{f} - 1/n\|_\infty \\
&\geq \Lambda_\omega(\tilde{f} + 1/n) - \frac{2}{n}
\end{aligned}$$

Mais $\tilde{f}+1/n$ est une fonction positive de $\mathcal{P}$. Comme $\omega \in \mathcal{N}^c$, $\Lambda_\omega(\tilde{f}+1/n) \geq 0$, d'où $\Lambda_\omega(f) \geq -2/n$. En faisant tendre n vers l'infini, on obtient $\Lambda_\omega(f) \geq 0$. Posons pour $f \in C(\Omega)$:

$$\tilde{\Lambda}_\omega(f) = \mathbb{E}_\mathbb{P}(f)\mathbb{1}_\mathcal{N}(\omega) + \Lambda_\omega(f)\mathbb{1}_{\mathcal{N}^c}(\omega).$$

Comme $\mathcal{N} \in \mathcal{S}$, $\tilde{\Lambda}_\omega(f)$ pour $f \in \mathcal{P}$. Mais pour $f \in C(\Omega$, il existe une suite f_n d'éléments de $\mathcal{P}$ telle que f_n converge uniformément vers f et $\tilde{\Lambda}_\omega(f_n)$ converge partout (pour tout ω) vers $\tilde{\Lambda}_\omega(f)$: $\omega \mapsto \tilde{\Lambda}_\omega(f)$ est donc $\mathcal{S}$-mesurable comme limite ponctuelle d'applications $\mathcal{S}$-mesurables.

Comme $\mathcal{N}$ est de mesure nulle, on a encore pour tout $f \in \mathcal{P}$:

$$\forall A \in \mathcal{S} \quad \int_A f \, d\mathbb{P} = \int_A \tilde{\Lambda}_\omega(f) \, d\mathbb{P}(\omega) \tag{4.3}$$

Pour tout A, les formes linéaires $f \mapsto \int_A f \, d\mathbb{P}$ et $f \mapsto \int_A \tilde{\Lambda}_\omega(f) \, d\mathbb{P}(\omega)$ sont continues sur $C(\Omega)$. Comme elles coïncident sur $\mathcal{P}$, elles coïncident sur $C(\Omega)$.

D'après le théorème de représentation de Riesz, pour tout $\omega \in \Omega$, il existe $\mathbb{P}_\omega^{\mathcal{S}}$ telle que pour tout $f \in C(\Omega)$, $\tilde{\Lambda}_\omega(f) = \int_\Omega f \, d\mathbb{P}_\omega^{\mathcal{S}}$.

Comme $\omega \mapsto \int_\Omega f \, d\mathbb{P}_\omega^{\mathcal{S}}$ est $\mathcal{S}$-mesurable pour tout $f \in C(\Omega)$, la définition de $\mathcal{T}$ entraîne, avec le théorème fondamental de la mesurabilité, que $\omega \mapsto \mathbb{P}_\omega^{\mathcal{S}}$ est $\mathcal{S}$-mesurable[4].

Soit maintenant K un compact de Ω : on peut construire une suite (f_n) bornée de fonctions continues telle que f_n converge simplement vers $\mathbb{1}_K$. Par convergence dominée, pour tout ω, $\int_\Omega f_n \, d\mathbb{P}_\omega^{\mathcal{S}}$ converge vers $\mathbb{P}_\omega^{\mathcal{S}}(K)$, ce qui entraîne en particulier que $\omega \mapsto \mathbb{P}_\omega^{\mathcal{S}}(K)$ est $\mathcal{S}$-mesurable. Comme

$$\forall A \in \mathcal{S} \quad \int_A f_n \, d\mathbb{P} = \int_A \int_\Omega f_n \, d\mathbb{P}_\omega^{\mathcal{S}} \, d\mathbb{P}(\omega),$$

En appliquant encore une fois le théorème de convergence dominée, on a

$$\forall A \in \mathcal{S} \quad \int_A \mathbb{1}_K \, d\mathbb{P} = \int_A \int_\Omega \mathbb{1}_K \, d\mathbb{P}_\omega^{\mathcal{S}} \, d\mathbb{P}(\omega).$$

Passons maintenant au cas d'un borélien quelconque : comme P est une mesure de probabilités sur $[0,1]^d \subset \mathbb{R}^d$, P est régulière, c'est à dire que pour B borélien de Ω, on peut trouver une suite (K_n) de compacts et une suite (O_n) d'ouverts, avec $K_n \subset B \subset O_n$ et $\lim \mathbb{P}(O_n \backslash K_n) = 0$. Pour une preuve de cette propriété, on pourra par exemple se reporter à l'annexe B. de Garet-Kurtzmann [2].

Avec le lemme de Fatou

$$\begin{aligned} \int \underline{\lim}_{n\to+\infty} (\mathbb{P}_\omega^{\mathcal{S}}(O_n) - \mathbb{P}_\omega^{\mathcal{S}}(K_n)) \, d\mathbb{P} &\le \underline{\lim}_{n\to+\infty} \int (\mathbb{P}_\omega^{\mathcal{S}}(O_n) - \mathbb{P}_\omega^{\mathcal{S}}(K_n)) \, d\mathbb{P} \\ &= \underline{\lim}_{n\to+\infty} \mathbb{P}(O_n \backslash K_n) = 0 \end{aligned}$$

Comme la suite $(\mathbb{P}_\omega^{\mathcal{S}}(O_n) - \mathbb{P}_\omega^{\mathcal{S}}(K_n))_{n\ge 0}$ est décroissante, il s'ensuit que pour $\mathbb{P}$-presque tout ω, $\mathbb{P}_\omega^{\mathcal{S}}(O_n) - \mathbb{P}_\omega^{\mathcal{S}}(K_n)$ tend vers 0. Ainsi, $\mathbb{P}$-presque tout ω, $\mathbb{P}_\omega^{\mathcal{S}}(K_n)$ tend vers $\mathbb{P}_\omega^{\mathcal{S}}(A)$. Ainsi, à un $\mathbb{P}$-négligeable près $\omega \mapsto \mathbb{P}_\omega^{\mathcal{S}}(A)$ est $\mathcal{S}$-mesurable. L'identité $\int_A \mathbb{1}_{K_n} \, d\mathbb{P} = \int_A \mathbb{P}_\omega^{\mathcal{S}}(K_n) \, d\mathbb{P}$ entraîne à la limite $\int_A \mathbb{1}_B \, d\mathbb{P} = \int_A \mathbb{P}_\omega^{\mathcal{S}}(B) \, d\mathbb{P}$. Ainsi, $\mathbb{P}_\omega^{\mathcal{S}}(B)$ est bien une version de la probabilité conditionnelle de B sachant $\mathcal{S}$.

Le passage au fonctions étagées, puis aux fonctions mesurables positives, est classique et laissé au lecteur.

□

4. voir l'exercice 36

4.2.2 Loi d'un vecteur sachant un autre

Un cas particulier important est le cas où $\mathbb{P}$ est une loi sur le produit $E \times F$ de deux espaces polonais [5] et que $\mathcal{S}$ est la famille des ensembles de la forme $A \times F$, où A décrit l'ensemble des boréliens de E. Autrement dit, c'est le cas où $\mathcal{S}$ est la tribu engendrée par la variable X, projection sur la première coordonnée. Notons x un élément de E, y un élément de F.

Dans ce cas, comme $(x,y) \mapsto P^{\mathcal{S}}_{(x,y)}$ est $\mathcal{S}$-mesurable, $\mathbb{P}^{\mathcal{S}}_{(x,y)}$ ne dépend que de x [6]. On se fixe $y_0 \in F$ quelconque et l'on peut poser

$$\tilde{\mathbb{P}}_x(B) = \mathbb{P}^{\mathcal{S}}_{(x,y_0)}(E \times B).$$

On peut vérifier que $(x \mapsto \tilde{\mathbb{P}}_x)$ est $(E, \mathcal{B}(E))$-$(M_1(F), \mathcal{T})$-mesurable. Pour tous x, y, on a $\mathbb{P}^{\mathcal{S}}_{(x,y)}(E \times B) = \tilde{\mathbb{P}}_x(B)$.

Cette mesure $B \mapsto \tilde{\mathbb{P}}_x(B)$ est couramment appelée “loi de Y sachant $X = x$.” On a ainsi

$$\begin{aligned}\mathbb{P}(A \times B) &= \int_{E\times F} \mathbb{1}_A(x)\mathbb{P}^{\mathcal{S}}_{(x,y)}(E \times B)\; d\mathbb{P}(x,y) &(4.4)\\ &= \int_{E\times F} \mathbb{1}_A(x)\tilde{\mathbb{P}}_x(B)\; d\mathbb{P}(x,y) = \int_E \mathbb{1}_A(x)\tilde{\mathbb{P}}_x(B)\; d\mathbb{P}_X(x). &(4.5)\end{aligned}$$

Si $\mathbb{P}$ admet une densité $f(x,y)$ par rapport à la mesure produit $\mu \otimes \nu$, alors $\tilde{\mathbb{P}}_x$ est presque sûrement la mesure μ_x de densité

$$y \mapsto \frac{f(x,y)}{f_X(x)} = \frac{f(x,y)}{\int_F f(x,y')\; d\nu(y')}$$

par rapport à ν.

En effet, on sait que X admet par rapport à μ la densité $x \mapsto f_X(x) = \int_F f(x,y)\; d\nu(y)$. f_X est $\mathbb{P}_X$-presque partout strictement positif, donc μ_x est bien défini pour $\mathbb{P}_X$ presque tout x. Ainsi, pour $\mathbb{P}_X$ presque tout x, on peut écrire $\mu_x(B) = \frac{\int_F \mathbb{1}_B(y)f(x,y)\; d\nu(y)}{\int_F f(x,y)\; d\nu(y)} = \frac{\int_F \mathbb{1}_B(y)f(x,y)\; d\nu(y)}{f_X(x)}$. On a alors

$$\begin{aligned}\int_E \mathbb{1}_A(x)\mu_x(B)\; d\mathbb{P}_X(x) &= \int_E \left(\mathbb{1}_A(x)\frac{\int_F \mathbb{1}_B(y)f(x,y)\; d\nu(y)}{f_X(x)} f_X(x)\right) d\mu(x)\\ &= \int_E\int_F \mathbb{1}_A(x)\mathbb{1}_B(y)f(x,y)\; d\nu(y)\; d\mu(x)\\ &= \int_{E\times F} \mathbb{1}_{A\times B}(x,y)\; d\mathbb{P}(x,y) = \mathbb{P}(A\times B),\end{aligned}$$

5. Typiquement $E = \mathbb{R}^n$ et $F = \mathbb{R}^p$.

6. On laisse ce point à préciser en exercice. On notera qu'on n'a pas besoin ici du lemme de Doob.

ce qui donne bien le résultat voulu.
Exemple: supposons que (X,Y) est gaussien de centrage $(m_X, m_Y) = (\mathbb{E}X, \mathbb{E}Y)$ et de matrice de covariance $M = \begin{pmatrix} A & C \\ C^* & B \end{pmatrix}$ inversible.
En posant $M^{-1} = \begin{pmatrix} A' & C' \\ C'^* & B' \end{pmatrix}$, on a

$$f(x,y) = \frac{(\det M)^{-1/2}}{(2\pi)^{(n+p)/2}} \exp\left(-\frac{1}{2}\begin{pmatrix} \langle A'(x-m_X),(x-m_X)\rangle \\ +2\langle C'^*(x-m_X),(y-m_Y)\rangle \\ +\langle B'(y-m_Y),(y-m_Y)\rangle \end{pmatrix}\right)$$

d'où la densité de μ_x :

$$y \mapsto k_x \exp\left(-\frac{1}{2}\left(2\langle C'^*(x-m_X),(y-m_Y)\rangle + \langle B'(y-m_Y),(y-m_Y)\rangle\right)\right).$$

Posant $\overline{x} = x - m_X$, $\overline{y} = x - m_Y$, et $[x,y] = \langle Bx, y\rangle$, on note que

$$\begin{aligned} 2\langle C'^*\overline{x},\overline{y}\rangle + \langle B'\overline{y},\overline{y}\rangle &= \langle B'\overline{y},\overline{y}\rangle + 2\langle BB^{-1}C'^*\overline{x},\overline{y}\rangle \\ &= [\overline{y},\overline{y}] + 2[B^{-1}C'^*\overline{x},\overline{y}] \\ &= [\overline{y}+B^{-1}C'^*\overline{x},\overline{y}+B^{-1}C'^*\overline{x}] - 2[B^{-1}C'^*\overline{x},B^{-1}C'^*\overline{x}] \\ &= \langle(\overline{y}+B'^{-1}C'^*\overline{x}),B'(\overline{y}+B'^{-1}C'^*\overline{x})\rangle + c_x \end{aligned}$$

On reconnait ici (à une constante près) la densité d'un vecteur gaussien : on a ainsi $\mu_x = \mathcal{N}(m_Y - B'^{-1}C'^*(x-m_X), B'^{-1})$. Cependant en identifiant par bloc les termes du produit $M^{-1}.M = \mathrm{I}_{n+p}$, on obtient $C'^*A + B'C^* = 0$, d'où $B'^{-1}(C'^*A + B'C^*)A^{-1} = 0$, soit $B'^{-1}C'^* + C^*A^{-1} = 0$. On a donc également

$$\mu_x = \mathcal{N}(m_Y + C^*A^{-1}(x-m_X), B'^{-1}).$$

En particulier, on note que $\mathbb{E}[Y|X] = m_Y + C^*A^{-1}(X-m_X)$: il n'est pas nécessaire d'inverser M pour obtenir le centrage.

On a également

$$\begin{aligned} C'^*C + B'B &= \mathrm{I}_p \\ C'^*AA^{-1}C + B'B &= \mathrm{I}_p \\ -B'C^*A^{-1}C + B'B &= \mathrm{I}_p \\ B'(B - C^*A^{-1}C) &= \mathrm{I}_p, \end{aligned}$$

Soit $B'^{-1} = B - C^*A^{-1}C$.

4.2.3 Échantillonneur de Gibbs

Le résultat qui suit est très important en simulation : il exprime que si on sait simuler (la loi de) X et les lois conditionnelles de Y sachant X, on sait simuler le couple (X,Y).

Théorème 37. *Soit μ une loi sur $E\times F$. On note μ_X la loi image de μ par la projection canonique de $E\times F$ sur E (la loi de la première marginale). On suppose qu'on a un espace probabilisé sur lequelle vivent des variables aléatoires X et U (respectivement à valeurs dans E et G) et qu'existe une fonction $\varphi : E\times G\to F$ telle que*

- $\mathbb{P}_X=\mu_X$
- *U est indépendante de X*
- *Pour tout $x\in E$, $\varphi(x,U)$ suit la loi $\tilde{\mu}_x$.*

Alors $(X,\varphi(X,U))$ suit la loi μ.

Démonstration. Soient A et B des boréliens de E et F On a

$$\mathbb{P}(X\in A,\varphi(X,U)\in B)=\mathbb{E}[\mathbb{1}_A(X)\mathbb{1}_B(\varphi(X,U))].$$

Comme X et U sont indépendants, on a $\mathbb{E}[\mathbb{1}_A(X)\mathbb{1}_B(\varphi(X,U))|X]=\psi(X)$, avec

$$\psi(x)=\mathbb{E}[\mathbb{1}_A(x)\mathbb{1}_B(\varphi(x,U))]=\mathbb{1}_A(x)\mathbb{P}(\varphi(x,U)\in B)=\mathbb{1}_A(x)\tilde{\mu}_x(B).$$

En réintégrant, on a maintenant

$$\begin{aligned}\mathbb{P}(X\in A,\varphi(X,U)\in B)=\mathbb{E}[\psi(X)]&=\int_E\psi(x)\ d\mathbb{P}_X\\&=\int_E\psi(x)\ d\mu_X=\mu(A\times B)\end{aligned}$$

avec (4.4). □

Les hypothèses d'existence d'une telle fonction φ ne sont pas extravagantes. Rappelons le résultat classique suivant :

Théorème 38. *Soit F une fonction de $\mathbb{R}$ dans $\mathbb{R}$, croissante, continue à droite, dont la limite est nulle en $-\infty$ et vaut 1 en $+\infty$. On suppose que sur $(\Omega,\mathcal{F},\mathbb{P})$, U est une variable aléatoire suivant la loi uniforme sur $[0,1]$. On pose*

$$\forall u\in]0,1[\quad Q^*(u)=\min\{x\in\mathbb{R}:\ 1-F(x)\le u\}.$$

Alors $Q^*(U)$ est une variable aléatoire réelle dont la fonction de répartition est F.

Démonstration. Posons $Q = 1-F$. On a, comme $1-F$ est continue à droite,

$$\{x : \ Q^*(U) > x\} = \{x : \ Q(x) > U\},$$

donc

$$\forall x \in \mathbb{R} \quad \mathbb{P}(Q^*(U) > x) = \mathbb{P}(Q(x) > U) = Q(x).$$

Il reste à vérifier que $Q^*(U)$ prend presque sûrement ses valeurs dans $\mathbb{R}$. Pour $u \in]0,1[$, $\{x \in \mathbb{R}; 1-F(x) \le u\}$ n'est ni l'ensemble vide, ni $\mathbb{R}$ tout entier, car F admet 0 comme limite en $-\infty$ et 1 en $+\infty$. Comme $\mathbb{P}(U \in]0,1[) = 1$, le résultat voulu s'ensuit. □

Ainsi, si on pose

$$\varphi(x,u) = \min\{y \in \mathbb{R} : \tilde{\mu}_x([y,+\infty[) \le x\},$$

pour U variable aléatoire suivant la loi uniforme sur $[0,1]$, la fonction φ répond aux hypothèses du théorème.

Ainsi, en admettant que l'on sache exprimer φ, on sait simuler (X,Y) dès lors que l'on sait simuler X et une variable aléatoire suivant la loi uniforme sur $[0,1]$ indépendante de X.

Mais en réalité, le principe de l'échantillonneur de Gibbs est plutôt utilisé pour exhiber des chaînes de Markov qui admettent une mesure donnée μ comme mesure invariante.

Avec un peu de chance, la dynamique ainsi formée convergera vers la mesure limité, ce qui donne un manière de simuler (ou d'approcher une simulation de la loi).

Corollaire 11. *Soit Λ un ensemble fini, μ une mesure sur $\mathbb{R}^\Lambda$. On note $X^i : \mathbb{R}^\Lambda \to \mathbb{R}$ la projection canonique sur la i-ème coordonnée. Pour tout $i \in \Lambda$ et $x \in \mathbb{R}^{\Lambda\backslash\{i\}}$, on note $\tilde{\mu}_x^i$ la loi conditionnelle de X_i sachant $X^j = x_j$ pour tout $j \in \Lambda\backslash\{i\}$. On suppose que pour tout $x \in \mathbb{R}^{\Lambda\backslash\{i\}}$, si U suit la loi $\mathcal{U}$, alors $\varphi^i(x,U)$ suit la loi $\tilde{\mu}_x^i$.*

Soit maintenant (U_n) une suite de variables aléatoires indépendantes de loi $\mathcal{U}$, (V_n) une suite de variables aléatoires indépendantes suivant la loi uniforme sur Λ. Alors, si X_0 est indépendantes de $\sigma(U_n, V_n, n \ge 1)$ la suite définie par $X_{n+1} = (X_n^{V_n}, \varphi^{V_n}(X_n, U_n))$ est une chaîne de Markov admettant μ comme mesure invariante.

Démonstration. Le caractère markovien ne fait pas de mystère vu la représentation par récurrence. Il faut montrer que si X_0 suit la loi μ, X_1 aussi. Soit Ψ une fonction mesurable bornée de $\mathbb{R}^\Lambda$ dans $\mathbb{R}$. Par indépendance, on a

$$\mathbb{E}(\Psi(X_1)|V_0) = g(V_0),$$

où $g(i) = \mathbb{E}\psi(X_0^i, \varphi^i(X_0, U_0))$. Or, d'après le théorème, $(X_0^i, \varphi^i(X_0, U_0))$ a même loi que X_0, donc $\mathbb{E}\psi(X_0^i, \varphi^i(X_0, U_0)) = \mathbb{E}\Psi(X_0)$. Ainsi, $\mathbb{E}(\Psi(X_1)|V_0) = \mathbb{E}\Psi(X_0)$, et en réintégrant $\mathbb{E}(\Psi(X_1)) = \mathbb{E}\Psi(X_0)$. Comme Ψ est prise mesurable bornée quelconque, X_0 et X_1 ont même loi. $\square$

Dans le cadre de ce cours, les chaînes de Markov sont à espace d'état fini et la situation est très favorable. Disons que l'espace d'états S est inclus dans E^Λ, où E et Λ sont des ensembles finis.

Voilà comment on passe de l'état au temps n à l'étape au temps $n+1$:

On note x la configuration au temps n.

1. Choisir uniformément $k \in \Lambda$.
2. Choisir $e \in E$ suivant la loi μ conditionnée par le fait que la configuration coincide avec x en tout point de $\Lambda \backslash k$.
3. Définir la configuration y au temps $n+1$ par

$$y_l = \begin{cases} x_l & \text{si } l \in \Lambda \backslash k \\ e & \text{si } l = k \end{cases}$$

Ainsi, si deux configurations x et y diffèrent en plus d'un site, la probabilité de passage $p(x, y)$ de x à y vaut zéro. En revanche, si x et y diffèrent en un unique site k, on a

$$p(x, y) = \frac{1}{|\Lambda|} \frac{\mu(y)}{\sum_{z \in S; z_{k^c} = x_{k^c}} \mu(z)}.$$

La chaîne est apériodique car

$$p(x, x) = \frac{\mu(x)}{|\Lambda|} \sum_{k \in \Lambda} \left(\sum\nolimits_{z \in S; z_{k^c} = x_{k^c}} \mu(z) \right)^{-1} > 0.$$

Le seul point un peu critique qu'il faut vérifier pour pouvoir appliquer le théorème de convergence des chaînes de Markov est celui de l'irréductibilité : il n'est pas automatique qu'on puisse passer de n'importe quel état de S en changeant un point à la fois.

Une condition suffisante est que les lois conditionnelles d'une coordonnées sachant tous les autres chargent tous les points de E : dans ce cas, on peut passer d'une configuration à l'autre en changeant une coordonnée à la fois – et $|\Lambda|$ étapes suffisent.

4.3 Théorème de Radon–Nikodým

Définition. *Soient μ et ν deux mesures sur $(\Omega, \mathcal{F})$. On rappelle qu'on dit qu'une mesure μ est une mesure absolument continue par rapport à la mesure ν, ce qui est noté $\mu \ll \nu$, si pour tout $A \in \mathcal{F}$, on a : $\nu(A) = 0$ implique $\mu(A) = 0$.*

Lemme 7. *Soient μ et ν deux mesures de probabilité sur $(\Omega, \mathcal{F})$. Si $\mu \ll \nu$, alors pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que $\nu(A) \leq \eta \Longrightarrow \mu(A) \leq \varepsilon$.*

Démonstration. On raisonne par l'absurde : on suppose qu'il existe $\varepsilon > 0$, tel que pour tout n, il existe A_n avec $\nu(A_n) \leq \frac{1}{n^2}$ et $\mu(A_n) > \varepsilon$. Si on pose $A = \overline{\lim}_{n\to+\infty} A_n = \cap_{n\geq 1} \cup_{k\geq 1} A_k$, le lemme de Borel-Cantelli nous dit que $\nu(A) = 0$. Cependant, d'après le lemme de Fatou, on a $\mu(A) \geq \overline{\lim} \mu(A_n) \geq \varepsilon$, ce qui contredit l'hypothèse. $\square$

Théorème 39 (Radon–Nikodým)**.** *Soit $(\Omega, \mathcal{F})$ un espace mesurable. $\mathbb{P}$ et $\mathbb{Q}$ deux mesures de probabilités sur $(\Omega, \mathcal{F})$. On suppose qu'il existe une suite $(A_n)_{n\geq 1}$ d'éléments de $\mathcal{F}$ tels que $\mathcal{F} = \sigma(A_n, n \geq 1)$.*

Alors $Q \ll \mathbb{P}$ si et seulement si $\mathbb{Q}$ admet une densité φ par rapport à $\mathbb{P}$, c'est à dire qu'on a la représentation,

$$\mathbb{Q}(A) = \int_A \varphi(x)\ d\mathbb{P}(x).$$

Démonstration. Bien sûr, si $\mathbb{Q} = \varphi\mathbb{P}$, on a $\mathbb{Q} \ll \mathbb{P}$. Voyons la réciproque. Notons $\mathcal{F}_n$ la tribu engendrée par les ensembles $A_1, \ldots, A_n$. $\mathcal{F}_n$ est engendrée par une partition finie mesurable $\mathcal{P}_n$, et l'on peut définir

$$X_n = \sum_{A\in\mathcal{P}_n} \frac{\mathbb{Q}(A)}{\mathbb{P}(A)} \mathbb{1}_A.$$

Notons qu'une fonction f $\mathcal{F}_n$-mesurable s'écrit $f = \sum_{A\in\mathcal{P}_n} \mathbb{1}_A \alpha(A)$.

Ainsi, on a $fX_n = \sum_{A\in\mathcal{P}_n} \frac{\mathbb{Q}(A)}{\mathbb{P}(A)} \alpha(A) \mathbb{1}_A$ et

$$\mathbb{E}[fX_n] = \sum_{A\in\mathcal{P}_n} \mathbb{Q}(A)\alpha(A) \qquad = \mathbb{E}_{\mathbb{Q}}[\sum_{A\in\mathcal{P}_n} \alpha(A)\mathbb{1}_A] = \mathbb{E}_{\mathbb{Q}}[f]. \tag{4.6}$$

Comme f est aussi $\mathcal{F}_{n+1}$-mesurable, on a $\mathbb{E}[fX_{n+1}] = \mathbb{E}_Q(f)$, d'où $\mathbb{E}[fX_n] = \mathbb{E}[fX_{n+1}]$. Ainsi, on a $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$ et $(X_n)_{n\geq 1}$ est une martingale positive.

L'équation (4.6) nous sera encore utile par la suite. Elle exprime que X_n est la densité de $\mathbb{Q}$ par rapport à $\mathbb{P}$ sur la tribu $\mathcal{F}_n$. Montrons que $(X_n)_{n\ge 1}$ est équi-intégrable. Soit $\varepsilon > 0$. D'après le lemme, il existe η tel que $\mathbb{P}(B) \le \eta \Longrightarrow \mathbb{Q}(B) \le \varepsilon$. Soit $c > 1/\eta$. Comme $\mathbb{1}_{\{X_n > c\}}$ est $\mathcal{F}_n$-mesurable, on a

$$\mathbb{E}[X_n \mathbb{1}_{\{X_n > c\}}] = \mathbb{E}_{\mathbb{Q}}[\mathbb{1}_{\{X_n > c\}}] = \mathbb{Q}(X_n > c).$$

Pour montrer que $\mathbb{Q}(X_n > c) \le \varepsilon$, il suffit de montrer que $\mathbb{P}(X_n > c) \le \eta$. Or, l'inégalité de Markov nous donne

$$\mathbb{P}(X_n > c) \le \frac{1}{c}\mathbb{E}[X_n] = \frac{1}{c}\mathbb{E}_{\mathbb{Q}}[1] = \frac{1}{c} \le \eta,$$

ce qui donne l'équi-intégrabilité de la suite (X_n). Comme (X_n) est une martingale positive, X_n converge presque sûrement vers une fonction positive φ, et on a $\mathbb{E}[\varphi] \le \mathbb{E}[X_1] = \mathbb{E}_{\mathbb{Q}}[1] = 1$. Soit $A \in \mathcal{F}_n$. Pour tout $p \ge n$, on a $A \in \mathcal{F}_p$, d'où

$$\mathbb{Q}(A) = \mathbb{E}_Q[\mathbb{1}_A] = \mathbb{E}X_p \mathbb{1}_A.$$

Comme $(X_p \mathbb{1}_A)_p$ est équi-intégrable, de limite $\mathbb{P}$-presque sûre $\varphi \mathbb{1}_A$, on a $\lim_{p\to+\infty} \mathbb{E}[X_p \mathbb{1}_A] = \mathbb{E}[\varphi \mathbb{1}_A]$, soit $\mathbb{Q}(A) = \mathbb{E}[\varphi \mathbb{1}_A]$. Ainsi, les mesures Q et $\varphi\mathbb{P}$ coïncident sur $\cup_{n\ge 1}\mathcal{F}_n$. Comme ce π-système engendre la tribu $\mathcal{F}$, les deux mesures coïncident et φ est bien la densité de $\mathbb{Q}$ par rapport à $\mathbb{P}$ sur $\mathcal{F}$. □

Le lemme suivant permet d'étendre de nombreux résultats des mesures de probabilité aux mesures σ-finies.

Lemme 8. *Si $(\Omega, \mathcal{F}, \mu)$ est un espace mesuré σ-fini, alors il existe une probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{F})$ et une fonction μ-presque sûrement positive f tels que f est la densité de $\mathbb{P}$ par rapport à μ et $1/f$ la densité de μ par rapport à $\mathbb{P}$.*

Démonstration. Soit $(A_n)_{n\ge 1}$ une suite avec $0 < \mu(A_n) < +\infty$ pour tout $n \ge 1$ et $\Omega = \cup_{n\ge 1} A_n$. On pose

$$f = \sum_{n=1}^{+\infty} \frac{1}{2^n \mu(A)} \mathbb{1}_A.$$

$\mathbb{P} = f\mu$ est une mesure. Avec Tonelli, on a

$$\begin{aligned}\mathbb{P}(\Omega) &= \int_\Omega f \, d\mu \\ &= \int_\Omega \sum_{n=1}^{+\infty} \frac{1}{2^n \mu(A)} \mathbb{1}_A \, d\mu \\ &= \sum_{n=1}^{+\infty} \frac{1}{2^n \mu(A)} \mu(A) = 1\end{aligned}$$

$\mathbb{P}$ est donc bien une mesure de probabilité, de densité f par rapport à μ. Maintenant, pour g mesurable positive, on a $\int g\ d\mu = \int g\ \frac{1}{f} f\ d\mu = \int g\ \frac{1}{f}\ d\mathbb{P}$, ce qui montre que $1/f$ est la densité de μ par rapport à $\mathbb{P}$. □

Corollaire 12 (Radon–Nikodým)**.** *Soient* $(\Omega, \mathcal{F})$ *un espace mesurable.* μ *et* ν *deux mesures de probabilités* σ*-finies sur* $(\Omega, \mathcal{F})$*. On suppose qu'il existe une suite* $(A_n)_{n\geq 1}$ *d'éléments de* $\mathcal{F}$ *tels que* $\mathcal{F} = \sigma(A_n, n \geq 1)$*.*

Alors $\mu \ll \nu$ *si et seulement si* μ *admet une densité* φ *par rapport à* ν*, c'est à dire qu'on a la représentation,*

$$\mu(A) = \int_A \varphi(x)\ d\nu(x).$$

Démonstration. Là encore, la condition est nécessaire. Ensuite, construisons $\mathbb{P}$ et $\mathbb{Q}$ telles que données par le lemme 8. On a $\mathbb{Q} \ll \mu$, $\mu \ll \nu$, $\nu \ll \mathbb{P}$, donc $\mathbb{Q} \ll \mathbb{P}$. Avec le théorème, $\mathbb{Q}$ admet une densité par rapport à ν. Considérons les densités respectives $\frac{d\mu}{d\mathbb{Q}}$, $\frac{d\mathbb{Q}}{d\mathbb{P}}$, $\frac{d\mathbb{P}}{d\nu}$ de μ par rapport à $\mathbb{Q}$, $\mathbb{Q}$ par rapport à $\mathbb{P}$, $\mathbb{P}$ par rapport à ν. En appliquant 3 fois le théorème de transfert, on voit que

$$\frac{d\mu}{d\mathbb{Q}} \cdot \frac{d\mathbb{Q}}{d\mathbb{P}} \cdot \frac{d\mathbb{P}}{d\nu}$$

est une densité de μ par rapport à ν. □

Remarque 4. *Dans les théorèmes de Radon–Nicodým, l'hypothèse d'existence d'une famille dénombrable engendrant la tribu est par exemple vérifiée dans le cas où* $\mathcal{F}$ *est la tribu borélienne d'un espace séparable. Toutefois, il faut noter que cette hypothèse n'est pas nécessaire, les théorèmes de Radon–Nicodým pouvant se démontrer directement par des techniques hilbertiennes.*

4.4 Exercices sur les compléments

4.4.1 Exercices corrigés

Exercice 33. Soit $(\Omega, \mathcal{F})$ un espace mesurable et $(E, \mathcal{B})$ un espace métrique muni de sa tribu borélienne. On suppose que $(X_n)_{n\geq 1}$ est une suite d'applications $(\Omega, \mathcal{F}) - (E, \mathcal{B})$ mesurables et que pour tout $\omega \in \Omega$, $X_n(\omega) \to X(\omega)$. Montrer que X est $(\Omega, \mathcal{F}) - (E, \mathcal{B})$ mesurables. lien vers l'indication lien vers la solution

Exercice 34. On considère les applications X et Y sur l'espace mesuré $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ définies par $X(x, y) = x$ et $Y(x, y) = y$. On suppose que $(\Omega, \mathcal{F})$ est un espace mesuré et que $\mathcal{F}$ contient les singletons. Montrer que pour toute application Z de $\mathbb{R}^2$ dans Ω qui est $(\mathbb{R}^2, \sigma(X)) - (\Omega, \mathcal{F})$ mesurable, il existe une application H $(\mathbb{R}, \mathcal{B}(\mathbb{R})) - (\Omega, \mathcal{F})$ mesurable avec $Z = H(X)$. lien vers l'indication lien vers la solution

Exercice 35. Soient X et Y deux variables aléatoires indépendantes suivant respectivement les lois $\mathcal{P}(\lambda)$ et $\mathcal{P}(\mu)$.

1. Calculer la loi conditionnelle de X sachant $X + Y$.
2. En déduire la valeur de $\mathbb{E}[X^2|X + Y]$.

lien vers l'indication lien vers la solution

4.4.2 Exercices non corrigés

Exercice 36. Soient $(\Omega_1, \mathcal{F}_1)$, $(\Omega_2, \mathcal{F}_2)$ et $(\Omega_3, \mathcal{F}_3)$ des espaces mesurés.

On suppose que $\mathcal{F}_2$ est la plus petite tribu qui rend mesurable les applications $(\pi_i)_{i\geq 1}$ de Ω_2 dans $(\Omega_3, \mathcal{F}_3)$. Montrer que $f : \Omega_1 \to \Omega_2$ est $(\Omega_1, \mathcal{F}_1) - (\Omega_2, \mathcal{F}_2)$ mesurable si et seulement pour tout i $\pi_i \circ f$ est $(\Omega_1, \mathcal{F}_1) - (\Omega_3, \mathcal{F}_3)$ mesurable. lien vers l'indication

Exercice 37. Soient μ, ν, ρ des mesures σ-finies sur $(\Omega, \mathcal{F})$

1. Montrer que si $\mu \ll \nu$ et $\nu \leq \rho$, alors $\mu \ll \rho$, et les densités relatives vérifient $\frac{d\mu}{d\rho} = \frac{d\mu}{d\nu} \cdot \frac{d\nu}{d\rho}$.
2. Montrer que $\nu \ll \mu$ si et seulement si $\frac{d\mu}{d\nu}$ est μ presque sûrement non nulle, et qu'alors $\frac{d\nu}{d\mu} = \left(\frac{d\mu}{d\nu}\right)^{-1}$.

lien vers l'indication

Exercice 38. Soient $\mathbb{P}$ et $\mathbb{Q}$ deux mesures de probabilité sur l'espace mesuré $(\Omega, \mathcal{F})$. On suppose qu'il existe une famille dénombrable $\mathcal{A} \subset \mathcal{F}$ engendrant $\mathcal{F}$ et une constante C telles que

$$\forall A \in \mathcal{A} \quad C^{-1}\mathbb{Q}(A) \leq \mathbb{P}(A) \leq \mathbb{Q}(A).$$

Montrer que $\mathbb{P} \ll \mathbb{Q} \ll \mathbb{P}$ et que, à la fois $\mathbb{P}$ presque sûrement et $\mathbb{Q}$ presque sûrement, on a

$$C^{-1} \leq \frac{d\mathbb{P}}{d\mathbb{Q}} \leq C.$$

lien vers l'indication

Chapitre 5

Inégalités

Le but de ce chapitre est de présenter, en application de la théorie des martingales, et, plus généralement, de l'espérance conditionnelle, quelques inégalités utiles en probabilités. Il ne s'agit pas ici d'applications scolaires : elles sont choisies pour leur utilité dans la pratique courante d'un chercheur en probabilités et en statistiques.

5.1 Inégalité d'Efron–Stein

On commence par une inégalité utile en statistiques.

Théorème 40. *Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes et Z une variable aléatoire de carré intégrable $\sigma(X_1, \dots, X_n)$ mesurable. Alors, si l'on pose $Z_i = \mathbb{E}[Z|\sigma(X_j)_{j \neq i}]$, on a*

$$\operatorname{Var} Z \le \sum_{i=1}^{n} \mathbb{E}[(Z - Z_i)^2].$$

Démonstration. On pose $Y_i = \mathbb{E}[Z|X_1, \dots, X_i] - \mathbb{E}[Z|X_1, \dots, X_{i-1}]$. La suite Y_i est une suite de différences de martingales, donc une suite orthogonale dans L^2. On a donc

$$\begin{aligned} \operatorname{Var} Z &= \mathbb{E}(Y_1 + \cdots + Y_n)^2 \\ &= \sum_{i=1}^{n} \mathbb{E}[Y_i^2] \end{aligned}$$

Comme $\sigma(X_1, \dots, X_{i-1})$ est une sous-tribu de $\sigma(X_j)_{j \neq i}$, on a

$$\mathbb{E}[Z_i|X_1, \dots X_{i-1}] = \mathbb{E}[Z|X_1, \dots X_{i-1}].$$

Cependant la tribu engendrée par Z_i et $X_1, \dots X_{i-1}$ est une sous-tribu de $\sigma(X_j)_{j\neq i}$ qui est indépendante de X_i, donc d'après le théorème 15, on a

$$\mathbb{E}[Z_i|X_1, \dots X_{i-1}] = \mathbb{E}[Z_i|X_1, \dots X_{i-1}, X_i].$$

Finalement $\mathbb{E}[Z|X_1, \dots X_{i-1}] = \mathbb{E}[Z_i|X_1, \dots X_i]$ et $Y_i = \mathbb{E}[(Z-Z_i)|X_1, \dots, X_i]$.

Avec l'inégalité de Jensen conditionnelle, on a

$$Y_i^2 = (\mathbb{E}[(Z-Z_i)|X_1, \dots, X_i])^2 \leq \mathbb{E}[(Z-Z_i)^2|X_1, \dots, X_i],$$

d'où $\mathbb{E}[Y_i^2] \leq \mathbb{E}[(Z-Z_i)^2]$ et on a le résultat voulu. □

5.2 L'inégalité de Hoeffding–Azuma

5.2.1 Le théorème

Théorème 41. *Soient $(\mathcal{F}_n)_{n\geq 0}$ une filtration et $(Y_n)_{n\geq 0}$ une martingale adaptée à la filtration $(\mathcal{F}_n)_{n\geq 0}$.*

On suppose qu'il existe des réels $(k_n)_{n\geq 1}$ tels que $|Y_n - Y_{n-1}| \leq k_n$ presque sûrement.

On suppose que la série de terme général k_n^2 est convergente, de somme σ^2. Alors, si on note Y_∞ la limite de Y_n, on a $\mathbb{E}[Y_\infty] = \mathbb{E}[Y_0]$ et pour tout $x > 0$, on a

$$\mathbb{P}(Y_\infty - Y_0 \geq x) \leq \exp\left(-\frac{x^2}{2\sigma^2}\right) \text{ et } \mathbb{P}(Y_\infty - Y_0 \leq -x) \leq \exp\left(-\frac{x^2}{2\sigma^2}\right),$$

de sorte que $\mathbb{P}(|Y_\infty - Y_0| \geq x) \leq 2\exp\left(-\frac{x^2}{2\sigma^2}\right)$.

Démonstration. On va s'appuyer sur un lemme :

Lemme 9. *Soient X une variable aléatoire à valeurs dans $[-1,1]$ et $\mathcal{A}$ une sous-tribu de $\mathcal{F}$. On suppose que $\mathbb{E}[X|\mathcal{A}] = 0$. Alors*

$$\mathbb{E}[e^{\alpha X}|\mathcal{A}] \leq \cosh \alpha \leq \exp(\frac{\alpha^2}{2}).$$

Démonstration. On pose $f(x) = e^{\alpha x}$. f est convexe. Or on a la combinaison convexe $\frac{1-x}{2}(-1) + \frac{1+x}{2}(1)$ (noter que $0 \leq \frac{1-x}{2} \leq 1$ et $\frac{1-x}{2} + \frac{1+x}{2} = 1$). Ainsi $f(x) \leq \frac{1-x}{2}f(-1) + \frac{1+x}{2}f(1)$, soit

$$e^{\alpha x} \leq \frac{1-x}{2}e^{-\alpha} + \frac{1+x}{2}e^{\alpha}.$$

En substituant dans l'inégalité précédente, on a

$$e^{\alpha X} \leq \cosh\alpha + (\sinh\alpha)X.$$

Par positivité de l'expérance conditionnelle,

$$\mathbb{E}[e^{\alpha X}|\mathcal{A}] \quad \leq \quad \cosh\alpha + (\sinh\alpha)\mathbb{E}[X|\mathcal{A}] = \cosh\alpha$$

On a

$$\cosh(\alpha) = 1 + \sum_{n=1}^{+\infty} \frac{\alpha^{2n}}{(2n)!}$$

et

$$\exp(\alpha^2/2) = 1 + \sum_{n=1}^{+\infty} \frac{\left(\frac{\alpha^2}{2}\right)^n}{n!} = 1 + \sum_{n=1}^{+\infty} \frac{\alpha^{2n}}{2^n n!}$$

Or pour tout $n \geq 1$, on a

$$\frac{\alpha^{2n}}{(2n)!} = \frac{\alpha^{2n}}{2^n n!} \frac{2 \quad . \; 2 \qquad \dots \quad 2}{(n+1).(n+2)\dots(2n)} \leq \frac{\alpha^{2n}}{2^n n!}$$

En sommant les inégalités, on obtient bien que $\cosh(\alpha) \leq \exp\frac{\alpha^2}{2}$. D'où finalement

$$\mathbb{E}[e^{\alpha X}|\mathcal{A}] \leq \exp(\frac{\alpha^2}{2}).$$

□

On note la suite des sommes partielles : $L_n = \sum_{i=1}^n k_i^2$.

$e^{\theta(Y_{n+p}-Y_p)} = e^{\theta(Y_{n+p}-Y_{n+p-1})}e^{\theta(Y_{n+p-1}-Y_p)}$. $e^{\theta(Y_{n+p-1}-Y_p)}$ est $\mathcal{F}_{n+p-1}$-mesurable, donc

$$\mathbb{E}[e^{\theta(Y_{n+p}-Y_p)}|\mathcal{F}_{n+p-1}] = \left(e^{\theta(Y_{n+p-1}-Y_p)}\right)\mathbb{E}[e^{\theta D_{n+p}}|\mathcal{F}_{n+p-1}],$$

où l'on a posé $D_n = Y_n - Y_{n-1}$. Cependant, comme $(Y_n)_{n\geq 1}$ est une martingale,

$$\mathbb{E}[Y_{n+p} - Y_{n+p-1}|\mathcal{F}_{n+p-1}] = \mathbb{E}[Y_{n+p}|\mathcal{F}_{n+p-1}] - Y_{n+p-1} = Y_{n+p-1} - Y_{n+p-1} = 0.$$

Ainsi si l'on pose $X = \frac{Y_{n+p}-Y_{n+p-1}}{k_{n+p}}$, $\alpha = \theta k_{n+p}$ et $\mathcal{A} = \mathcal{F}_{n+p-1}$, X est à valeurs dans $[-1,1]$ et $\mathbb{E}[X|\mathcal{A}] = 0$: d'après le lemme, on a donc

$$\mathbb{E}[e^{\alpha X}|\mathcal{F}_{n+p-1}] \leq \exp(\frac{\alpha^2}{2}) = \exp(\frac{\theta^2}{2}k_{n+p}^2),$$

soit

$$\mathbb{E}[e^{\theta D_{n+p}}|\mathcal{F}_{n+p-1}] \leq \exp(\frac{\theta^2}{2}k_{n+p}^2),$$

d'où
$$\mathbb{E}[e^{\theta(Y_{n+p}-Y_p)}|\mathcal{F}_{n+p-1}] \le e^{\theta(Y_{n+p-1}-Y_p)}\exp(\frac{1}{2}\theta^2 k_{n+p}^2).$$

En prenant l'espérance, on a

$$\mathbb{E}[e^{\theta(Y_{n+p}-Y_p)}] \le \mathbb{E}[e^{\theta(Y_{n+p-1}-Y_p)}]\exp(\frac{1}{2}\theta^2 k_{n+p}^2),$$

soit
$$\frac{\mathbb{E}[e^{\theta(Y_{n+p}-Y_p)}]}{\mathbb{E}[e^{\theta(Y_{n+p-1}-Y_p)}]} \le \exp(\frac{1}{2}\theta^2 k_{n+p}^2).$$

En faisant le produit pour n variant de 1 à ℓ, on obtient

$$\frac{\mathbb{E}[e^{\theta(Y_{\ell+p}-Y_p)}]}{\mathbb{E}[e^{\theta(Y_p-Y_p)}]} \le \prod_{n=1}^{\ell}\exp(\frac{1}{2}\theta^2 k_{n+p}^2),$$

soit
$$\mathbb{E}[e^{\theta(Y_{\ell+p}-Y_p)}] \le \exp(\frac{1}{2}\theta^2\sum_{n=1}^{\ell}k_{n+p}^2) = \exp\left(\frac{1}{2}\theta^2(L_{\ell+p}-L_p)\right).$$

Avec l'inégalité de Markov, on a

$$\begin{aligned}\mathbb{P}(Y_{\ell+p}-Y_p \ge x) &\le \mathbb{P}(e^{\theta(Y_{\ell+p}-Y_p)} \ge e^{\theta x})\\ &\le \frac{\mathbb{E}(e^{\theta(Y_{\ell+p}-Y_p)})}{e^{\theta x}}\\ &\le \exp\left(-\theta x+\frac{1}{2}\theta^2(L_{\ell+p}-L_p)\right).\end{aligned}$$

En prenant $\theta = \frac{x}{L_{\ell+p}-L_p}$, on obtient

$$\mathbb{P}(Y_{\ell+p}-Y_p \ge x) \le \exp\left(-\frac{1}{2}x^2\frac{1}{L_{\ell+p}-L_p})\right).$$

Cependant $(-Y_n)_{n\ge 0}$ est également une martingale, avec $|-Y_n-(-Y_{n-1})| \le k_n$, donc on a également

$$\mathbb{P}(-Y_{\ell+p}+Y_p \ge x) \le \exp\left(-\frac{1}{2}x^2\frac{1}{L_{\ell+p}-L_p}\right).$$

Finalement

$$\mathbb{P}(|Y_{\ell+p}-Y_p| \ge x) \le \mathbb{P}(Y_{\ell+p}-Y_p \ge x)+\mathbb{P}(-Y_{\ell+p}+Y_p \ge x) \le 2\exp\left(-\frac{x^2}{2(L_{\ell+p}-L_p)}\right).$$

On a pour tout n :

$$\mathbb{E}|Y_0 - Y_n|^2 = \int_0^{+\infty} 2t\mathbb{P}(|Y_0 - Y_n| > t)\ dt \leq \int_0^{+\infty} 4t \exp\left(-\frac{t^2}{2\sigma^2}\right)\ dt = 4\sigma^2.$$

$(Y_n - Y_0)_{n\geq 0}$ est une martingale centrée bornée dans L^2 : elle converge presque sûrement et dans L^2 vers une variable $Y_\infty - Y_0$ d'espérance nulle. Soient $x > 0$ et $\varepsilon \in]0, x/2[$

$$\begin{aligned}\mathbb{P}(Y_\infty - Y_0 \geq x) &\leq \mathbb{P}(|Y_\infty - Y_n| > \varepsilon) + \mathbb{P}(Y_n - Y_0 > x - 2\varepsilon) \\ &\leq \mathbb{P}(|Y_\infty - Y_n| > \varepsilon) + \exp\left(-\frac{(x-2\varepsilon)^2}{2\sigma^2}\right)\end{aligned}$$

En faisant tendre n vers $+\infty$, puis ε vers 0, on obtient une des deux inégalités voulue. L'autre s'obtient de la même manière. □

5.2.2 Principe de Maurey

Théorème 42. *Soient $(X_1, \dots, X_n)$ n des variables indépendantes à valeurs dans un ensemble E. Soit f une fonction telle que si $(x_1, \dots, x_n) \in E^n$ et $(y_1, \dots, y_n) \in E^n$ vérifient $x_j = y_j$ pour $j \neq i$, alors $|f(x) - f(y)| \leq k_i$. On suppose que $Z = f(X_1, \dots, X_n)$ est intégrable. Alors, si on pose $\sigma^2 = \sum_{i=1}^n k_i^2$, on a pour tout $x > 0$,*

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq x) \leq \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

et

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq -x) \leq \exp\left(-\frac{x^2}{2\sigma^2}\right),$$

d'où

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq x) \leq 2\exp\left(-\frac{x^2}{2\sigma^2}\right).$$

Démonstration. On pose $Y_0 = \mathbb{E}[Z]$, et pour tout entier $i \in \{1, \dots, n\}$:

$$Y_i = \mathbb{E}[Z|X_1, \dots, X_i].$$

Pour pouvoir appliquer l'inégalité d'Hoeffding-Azuma, il suffit de vérifier que $|Y_{i-1} - Y_i| \leq k_i$. Soient $(X'_1, \dots, X'_n)$ indépendant de $(X_1, \dots, X_n)$ et de même loi. En appliquant successivement les théorèmes 19 et 15, on a

$$\begin{aligned}& \mathbb{E}[f(X_1, \dots, X_{i-1}, X_i, X_{i+1}, \dots, X_n)|X_1, \dots, X_{i-1}] \\ = \ & \mathbb{E}[f(X_1, \dots, X_{i-1}, X'_i, X_{i+1}, \dots, X_n)|X_1, \dots, X_{i-1}] \\ = \ & \mathbb{E}[f(X_1, \dots, X_{i-1}, X'_i, X_{i+1}, \dots, X_n)|X_1, \dots, X_{i-1}, X_i]\end{aligned}$$

Ainsi

$$\begin{aligned}
& Y_{i-1} - Y_i \\
= \quad & \mathbb{E}[f(X_1, \ldots, X_{i-1}, X'_i, X_{i+1}, \ldots, X_n)|X_1, \ldots, X_{i-1}, X_i] \\
& -\mathbb{E}[f(X_1, \ldots, X_{i-1}, X_i, X_{i+1}, \ldots, X_n)|X_1, \ldots, X_{i-1}, X_i] \\
= \quad & \mathbb{E}\left[\begin{array}{cc} & f(X_1, \ldots, X_{i-1}, X'_i, X_{i+1}, \ldots, X_n) \\ - & f(X_1, \ldots, X_{i-1}, X_i, X_{i+1}, \ldots, X_n)\end{array}\middle| X_1, \ldots, X_{i-1}, X_i\right]
\end{aligned}$$

qui est presque sûrement majorée en norme par k_i puisque

$$f(X_1, \ldots, X_{i-1}, X'_i, X_{i+1}, \ldots, X_n) - f(X_1, \ldots, X_{i-1}, X_i, X_{i+1}, \ldots, X_n)$$

l'est. □

Étude d'un exemple

On jette n boules dans n urnes de manière indépendante et uniforme. Comment se concentre le nombre d'urnes vides ?

Notons $X_1, \ldots, X_n$ n variables aléatoires indépendantes suivant la loi uniforme sur $\{1, \ldots, n\}$. Le nombre d'urnes vides est $Z_n = f(X_1, \ldots, X_n)$, avec

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} \prod_{j=1}^{n} \mathbb{1}_{\{x_j \neq i\}}.$$

On a aisément $\mathbb{E}[Z_n] = n(1 - \frac{1}{n})^n \sim \frac{n}{e}$. Bien sûr, si on change une boule d'urne, on modifie au plus de 1 le nombre d'urnes vides, on peut donc appliquer le théorème 42, d'où

$$\mathbb{P}(|Z_n - \mathbb{E}Z_n| > x) \leq 2\exp(-\frac{x^2}{2n}).$$

5.3 Inégalité de Harris

Théorème 43. *Soient $(X_1, \ldots, X_n)$ des variables aléatoires indépendantes, f et g deux fonctions de $\mathbb{R}^n$ dans $\mathbb{R}$, croissantes en chacune de leurs coordonnées. Alors*

$$\mathbb{E}[f(X_1, \ldots, X_n)g(X_1, \ldots, X_n)] \geq \mathbb{E}[f(X_1, \ldots, X_n)]\mathbb{E}[g(X_1, \ldots, X_n)].$$

Démonstration. On va montrer le résultat par récurrence sur n.
On commence avec $n = 1$. Soit X'_1 indépendant de X_1 de même loi que X_1. On a $(f(X_1) - f(X'_1))(g(X_1) - g(X'_1)) \geq 0$, d'où en prenant l'espérance

$$\mathbb{E}[f(X_1)g(X_1)] + \mathbb{E}[f(X'_1)g(X'_1)] - \mathbb{E}[f(X_1)]\mathbb{E}[g(X'_1)] - \mathbb{E}[f(X'_1)]\mathbb{E}[g(X_1)] \geq 0,$$

soit $2(\mathbb{E}[f(X_1)g(X_1)]-\mathbb{E}f(X_1)\mathbb{E}g(X_1)) \geq 0$, ce qui est l'inégalité voulue pour $n = 1$.

Supposons le résultat acquis jusqu'à n et soient f, g deux fonctions de $\mathbb{R}^{n+1}$ dans $\mathbb{R}$. On pose

$$\begin{aligned}
H(x_1, \dots, x_n) &= \mathbb{E}[f(x_1, \dots, x_n, X_{n+1})g(x_1, \dots, x_n, X_{n+1})], \\
F(x_1, \dots, x_n) &= \mathbb{E}[f(x_1, \dots, x_n, X_{n+1})] \\
G(x_1, \dots, x_n) &= \mathbb{E}[g(x_1, \dots, x_n, X_{n+1})].
\end{aligned}$$

On peut noter que

- F et G sont des fonctions croissantes
- les fonctions H, F, G évaluées en $X_1, \dots, X_n$ sont des versions de l'espérance conditionnelle de $f(X_1, \dots, X_{n+1})g(X_1, \dots, X_{n+1})$, $f(X_1, \dots, X_{n+1})$ et $g(X_1, \dots, X_{n+1})$, sachant la tribu $\sigma(X_1, \dots X_n)$.

En appliquant l'inégalité de Harris pour $n = 1$, on obtient

$$\forall (x_1, \dots, x_n) \in \mathbb{R}^n \quad H(x_1, \dots, x_n) \geq F(x_1, \dots, x_n)G(x_1, \dots, x_n),$$

d'où

$$\mathbb{E}[H(X_1, \dots, X_n)] \geq \mathbb{E}[F(X_1, \dots, X_n)G(X_1, \dots, X_n)].$$

Mais F et G sont des fonctions croissantes, donc d'après l'hypothèse de récurrence

$$\mathbb{E}[F(X_1, \dots, X_n)G(X_1, \dots, X_n)] \geq \mathbb{E}[F(X_1, \dots, X_n)]\mathbb{E}[G(X_1, \dots, X_n)].$$

Comme $\mathbb{E}[H(X_1, \dots, X_n)] = \mathbb{E}[f(X_1, \dots, X_{n+1})g(X_1, \dots, X_{n+1})]$ tandis que $\mathbb{E}[F(X_1, \dots, X_n)] = \mathbb{E}[f(X_1, \dots, X_{n+1})]$ et $\mathbb{E}[G(X_1, \dots, X_n)] = \mathbb{E}[g(X_1, \dots, X_{n+1})]$, on a le résultat voulu. □

5.4 Exercices sur les inégalités

5.4.1 Exercices corrigés

Exercice 39. Soit

$$\Delta_n = \{(x_1, \dots, x_n) \in [0,1]^n ; \forall i,j \in \{1,\dots,n\} \quad i \neq j \Longrightarrow x_i \neq x_j\}.$$

Pour $x \in \Delta_n$ et $\sigma \in \mathfrak{S}_n$, on définit $\sigma.x$ par $(\sigma.x)_i = x_{\sigma(i)}$. On définit une application h de Δ_n dans $\mathfrak{S}_n$ par

$$\forall i \in \{1,\dots,n\} \quad [h(x)](i) = |\{j \in \{1,\dots,n\}, x_j \leq x_i\}|.$$

1. Montrer que $h(\sigma.x) = h(x) \circ \sigma$.
2. Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. Montrer que $h(X_1, \dots, X_n)$ suit la loi uniforme sur $\mathfrak{S}_n$.
3. Soit $n \geq 1$ un entier. On note $\gamma_{i,j}$ l'unique permutation γ de $\mathfrak{S}_n$ telle que $\gamma(i) = j$ et telle que la restriction de γ à $\{1,\dots,n\} \backslash \{i\}$ est croissante. Soit g une fonction de $\mathfrak{S}_n$ dans $\mathbb{R}$, telle que pour toute permutation σ et tous i,j entre 1 et n, on a $|f(\gamma_{i,j}\sigma) - f(\sigma)| \leq M$. Montrer que si σ_n est une variable aléatoire suivant la loi uniforme sur $\mathfrak{S}_n$, alors pour tout $x > 0$, on a
$$\mathbb{P}(|g(\sigma_n) - \mathbb{E}(g(\sigma_n))| > x) \leq 2\exp(-\frac{x^2}{2nM^2}).$$
4. *Application à la loi hypergéométrique.* Une urne contient r boules rouges et b boules bleues. On en tire k boules sans remises. Soit X le nombre de boules rouges tirées. Montrer que
$$\mathbb{P}\left(\left|X - \frac{kr}{b+r}\right| > x\right) \leq 2\exp(-\frac{x^2}{2(b+r)}).$$

lien vers l'indication lien vers la solution

Exercice 40. On a au temps zéro n urnes vides. Puis, à tout instant $t \in \mathbb{N}^*$ on place une boule dans l'une des urnes. On cherche l'espérance du nombre maximum d'urnes contenant exactement une boule.

On prend donc une suite $(Y_k)_{k\geq 1}$ de variables aléatoires indépendantes suivant la loi uniforme sur $\{1,\dots,n\}$. Y_k est le numéro de l'urne dans laquelle on dépose une boule au temps k. On pose $N_{k,p} = \sum_{i=1}^k \mathbb{1}_{\{Y_i=p\}}$: $N_{k,p}$ représente le nombre de boules dans l'urne p au temps k.

Soit X_k^n le nombre d'urnes contenant exactement une boule à l'instant k. Le nombre maximum d'urnes contenant exactement une boule est

$$X^n = \max_{k\geq 0} X_k^n.$$

1. Exprimer X_k^n en fonction des $N_{k,p}$. En déduire que
$$\mathbb{E}[X_k^n] = n\frac{k}{n}\Big(1-\frac{1}{n}\Big)^{k-1}.$$
2. On pose $v_n = \max_{0\leq k\leq \lfloor n^{3/2}\rfloor}\mathbb{E}[X_k^n]$. Montrer que $v_n \sim \frac{n}{e}$.
3. On pose $X_*^n = \max(X_k^n, 0\leq k\leq \lfloor n^{3/2}\rfloor)$. On fixe $\varepsilon>0$. Montrer qu'il existe N tel que
$$\forall n\geq N \quad \mathbb{P}(|X_*^n - n/e|\geq n\varepsilon) \leq \sum_{k=0}^{\lfloor n^{3/2}\rfloor}\mathbb{P}(|X_k^n - \mathbb{E}[X_k^n]| \geq \frac{\varepsilon}{2}n).$$
4. En déduire que X_n^*/n converge en probabilité vers $1/e$
5. Montrer l'inclusion
$$\{X^n \neq X_*^n\} \subset \cup_{p=1}^n \{N_{\lfloor n^{3/2}\rfloor,p} < 2\}.$$
En déduire que X^n/n converge en probabilité vers $1/e$.
6. Conclure.
7. Un étudiant dispose d'une boite de pilules pour améliorer ses chances de réussite aux examens. Il doit en prendre une demie chaque matin afin d'améliorer ses performances intellectuelles.
Chaque jour, il prend une pilule au hasard dans sa boite. Si c'est une pilule entière, il la coupe en deux avant d'en avaler une moitié et de remettre l'autre dans la boite. Si c'est une demi pilule, il l'avale. Il continue sa cure jusqu'au jour où la boite est vide. Si la boite contient initialement n pilules, montrer que l'espérance du nombre maximal de demi-pilules présentes dans la boite est équivalent à Kn quand n tend vers l'infini, K étant à déterminer.

lien vers l'indication lien vers la solution

5.4.2 Exercices non corrigés

Exercice 41. Lors d'un vote, les deux candidats en présence optiennent chacun n voix. On note E_n l'écart maximal observé dans les bulletins dépouillés au cours du scrutin. Montrer que pour tout $x>0$,

$$\mathbb{P}(|E_n - \mathbb{E}(E_n)| > x) \leq 2\exp(-\frac{x^2}{2n}).$$

lien vers l'indication

Exercice 42. *Un résultat sur les statistiques d'ordre, d'après Björnberg et Broman*
Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes, admettant un moment d'ordre deux. On note $X^{(1)} \le X^{(2)} \le \dots X^{(n)}$ les statistiques d'ordres associées. Soit $A \subset \{1, \dots, n\}$. On pose $S_A = \sum_{k \in A} X_{(k)}$. Montrer que

$$\operatorname{Var} S_A \le \operatorname{Var}(X_1 + \dots X_n).$$

lien vers l'indication

Exercice 43. Dans une assemblée de n personnes, chaque couple de deux personnes se connaît avec probabilité p, de manière indépendante des autres. On leur donne des chemises de couleur, de telle manière que deux personnes qui se connaissent ne portent pas des chemises de même couleur On note χ le nombre minimum de couleurs nécessaires. Montrer que pour tout $\lambda > 0$,

$$\mathbb{P}(|\chi - \mathbb{E}(\chi)| > \lambda\sqrt{n-1})| \le 2e^{-\lambda^2/2}.$$

lien vers l'indication

Chapitre 6

Statistiques exhaustives

On suppose ici que pour tout $\theta \in \Theta$, $\mathbb{P}_\theta$ est une mesure de probabilité sur $(\Omega, \mathcal{F})$. Le problème fondamentale de la statistique est d'obtenir des informations sur la valeur du paramètre θ, réputé inconnu, à partir d'une observation. Une telle collection $(\mathbb{P}_\theta)_{\theta\in\Theta}$ est appelée un modèle statistique. On note X l'observation générique, c'est à dire que l'on travaille directement sur l'espace observé : $X(\omega) = \omega$. Les applications mesurables sur $(\Omega, \mathcal{F})$ sont appelées des statistiques.

À un modèle statistique $(\mathbb{P}_\theta)_{\theta\in\Theta}$ sur $(\Omega, \mathcal{F})$ est naturellement associé un autre modèle sur $(\Omega^n, \mathcal{F}^n)$: la $(\mathbb{P}_\theta^{\otimes n})_{\theta\in\Theta}$: Dans ce cas, les projections canoniques $X_1, \ldots, X_n$ de Ω^n sur Ω constituent sous $\mathbb{P}_\theta^{\otimes n}$ un échantillon de la loi $\mathbb{P}_\theta$, c'est à dire n variables indépendantes suivant la loi $\mathbb{P}_\theta$.

Définition. *Une statistique $S(X)$ est dite exhaustive (pour le paramètre θ) si la probabilité conditionnelle d'observer X sachant $S(X)$ est indépendante de θ : il existe une famille de mesures (P_s) telles que, pour tout $\theta \in \Theta$:*

$$\mathbb{P}_\theta(S \in A_0, X \in A) = \mathbb{E}_\theta[\mathbb{1}_{\{S\in A_0\}} P_S(A)].$$

soit encore

$$\mathbb{P}_\theta[X \in A|S] = P_S(A) \quad \mathbb{P}_\theta\text{-}p.s \tag{6.1}$$

Plus généralement, on dira qu'une tribu $\mathcal{S}$ est exhaustive si on a des lois conditionnelles P_ω avec $\omega \mapsto P_\omega(A)$ $\mathcal{S}$-mesurable avec

$$\forall\theta \in \Theta \quad \mathbb{P}_\theta[X \in A|\mathcal{S}](\omega) = P_\omega(A) \quad \mathbb{P}_\theta\text{-}p.s \tag{6.2}$$

6.1 Hypothèse de domination – dominante privilégiée

Définition. *On dit qu'un modèle statistique* $(\mathbb{P}_\theta)_{\theta\in\Theta}$ *est dominé s'il existe une mesure* μ σ*-finie telle que* $\mathbb{P}_\theta \ll \mu$ *pour tout* $\theta \in \mu$.

Lemme 10 (lemme de Halmos et Savage)**.** *Si une famille de mesures* $(m_\theta)_{\theta\in\theta}$ *est telle que chaque* m_θ *est absolument continue par rapport à une même mesure* μ*, alors on peut construire une mesure* m *par rapport à laquelle les* m_θ *sont absolument continues et qui est un mélange dénombrable des* m_θ*, c'est à dire que* m *s'écrit*

$$m = \sum_{i\in D} \alpha_i m_{\theta_i}, \tag{6.3}$$

avec $0 < \alpha_i$ *pour tout* i*,* D *dénombrable et* $\sum_{i\in D} \alpha_i = 1$*. Une telle mesure est appelée* dominante privilégiée *de la famille* $(m_\theta)_{\theta\in\theta}$*.*

Démonstration. D'abord, on peut supposer sans perte de généralité que μ n'est pas seulement finie, mais que c'est une mesure finie : si $(A_n)_{n\geq 1}$ est une suite croissante d'ensembles de réunion Ω avec $\mu(A_n) < +\infty$, $A_0 = \varnothing$ et $\mu(A_{n+1}) > \mu(A_n)$ pour tout n, l'identité $\mu'(A) = \sum_{n\geq 1} 2^{-n} \frac{\mu(A\cap(A_n\backslash A_{n-1}))}{\mu(A_n\backslash A_{n-1})}$ définit une mesure de probabilité.

$\mu'(A) = 0$ si et seulement si $\frac{\mu(A\cap(A_n\backslash A_{n-1}))}{\mu(A_n\backslash A_{n-1})} = 0$ pour tout n, soit si et seulement si $\mu(A\cap(A_n\backslash A_{n-1})) = 0$ pour tout n, ce qui arrive si et seulement si $\mu(A) = 0$. Les mesures μ et μ' sont donc absolument continues l'une par rapport à l'autre.

Notons $\mathcal{P}$ l'ensemble des mesures de la forme (6.3), et regardons la classe

$$\mathcal{A} = \cup_{\gamma\in\mathcal{P}}\{A : \gamma(A) > 0, \mu(\cdot\cap A) \ll \gamma\}.$$

Formons une suite (A_n) telle $\mu(A_n)$ converge vers $\sup_{A\in\mathcal{A}}\mu(A)$. Soit γ_i une mesure telle que $\mu(\cdot\cap A_i) \ll \gamma_i$. On pose

$$m = \sum_{n=1}^{+\infty} \frac{1}{2^n}\gamma_i.$$

Notons que si on pose $A_\infty = \cup_{n\geq 1} A_i$, on a $\mu(\cdot\cap A_\infty) \ll m$. En effet, si $m(C\cap A_\infty) = 0$, on a $\gamma_i(C\cap A_\infty) = 0$ pour tout i, donc $\mu(C\cap A_i) = 0$ pour tout i, ce qui entraîne que $\mu(C\cap A_\infty) = 0$.

Comme $m \in \mathcal{P}$, on en déduit que

$$\sup_{A\in\mathcal{A}}\mu(A) = \mu(A_\infty).$$

On va montrer que tous les éléments de $\mathcal{P}$ sont absolument continus par rapport à m. Cela donnera le résultat voulu puisque les m_θ sont dans $\mathcal{P}$.

Supposons que $m(C) = 0$ et prenons $\gamma \in \mathcal{P}$. On va montrer que $\gamma(C) = 0$. On a

$$\gamma(C) = \gamma(C \cap A_\infty) + \gamma(C \cap A_\infty^c).$$

Comme $m(C) = 0$ et $\mu(\cdot \cap A_\infty) \ll m$, $\mu(C \cap A_\infty) = 0$, ce qui entraîne $\gamma(C \cap A_\infty) = 0$ puisque $\gamma \ll \mu$. Il n'y a plus qu'à montrer que $\gamma(C \cap A_\infty^c) = 0$.

Notons $p(x)$ une version positive de la densité de γ par rapport à μ. On pose $N = \{x : p(x) = 0\}$ et $S = \{x : p(x) > 0\}$. $\gamma(N) = \int_N p(x)d\ d\mu = 0$. On en déduit que $\gamma(C \cap A_\infty^c) = \gamma(C \cap A_\infty^c \cap S)$. On pose $R = C \cap A_\infty^c \cap S$. Montrons que $\mu(\cdot \cap R) \ll \gamma$. Posons $f(x) = p(x)^{-1}$ si $x \in S$, $f(x) = 0$ sinon Soit E un ensemble tel que $\gamma(R) = 0$.

$$\mu(E \cap R) = \int \mathbb{1}_E(x)\mathbb{1}_R f(x)p(x)\ d\mu(x) = \int \mathbb{1}_E(x)\mathbb{1}_R(x)p(x)\ d\gamma(x) = 0,$$

car $\mathbb{1}_R = 0$ γ-presque sûrement : on a bien $\mu(\cdot \cap R) \ll \gamma$.

Il est maintenant aisé de voir que $\mu(\cdot \cap (A_\infty \cup R)) \ll \frac{m+\gamma}{2}$, ce qui montre que $A_\infty \cup R \in \mathcal{A}$. On en déduit

$$\mu(A_\infty) = \sup_{A \in \mathcal{A}} \mu(A) \geq \mu(A_\infty \cup R) = \mu(A_\infty) + \mu(R),$$

ce qui entraîne que $\mu(R) = 0$. Comme $\gamma \ll \mu$, $\gamma(R) = 0$, ce qui achève la preuve.

□

Remarque 5. *Si les mesures m_θ ont le même support, on peut prendre tout simplement $m = m_{\theta_0}$ pour $\theta_0 \in \Theta$ quelconque !*

6.2 Théorème de factorisation de Neyman-Fisher

Le but des théorèmes de factorisation Neyman-Fisher est de caractériser simplement l'existence de statistique exhaustive. En 1922, Fisher a démontré que la condition de factorisation que nous allons présenter était suffisante. Plus tard, en 1935, Neyman a montré qu'elle était également suffisante, sous quelques hypothèses supplémentaires. La forme générale que l'on enseigne aujourd'hui est en réalité dûe à Halmos et Savage, et a été démontrée en 1949.

Théorème 44. *On suppose que la famille des lois $(P_\theta)_{\theta\in\Theta}$ est dominée par rapport à une même mesure μ σ-finie.*

Alors, une statistique S est exhaustive si et seulement si il existe une fonction h et des fonctions ψ_θ telles que la fonction de vraisemblance $f_\theta(x)$ s'écrive $\psi_\theta(S(x))h(x)$ presque partout.

et

Théorème 45. *On suppose que la famille des lois $(P_\theta)_{\theta\in\Theta}$ admet une densité par rapport à une même mesure μ σ-finie.*

Alors, une tribu $\mathcal{S}$ est exhaustive si et seulement si il existe une fonction h et des fonctions $\mathcal{S}$-mesurables ψ_θ telles que la fonction de vraisemblance f_θ s'écrive

$$f_\theta(x) = \psi_\theta(x)h(x) \quad \mu \text{ presque partout.}$$

Vu le lemme de Doob, il suffit de démontrer le théorème 45 : en l'appliquant à une tribu engendrée par une statistique, on obtient le théorème 44.

On va commencer par traiter le cas où la mesure de référence est une dominante privilégiée.

Théorème 46. *On suppose que la famille des lois $(P_\theta)_{\theta\in\Theta}$ est dominée et que m en est une dominante privilégié.*

Alors, une tribu $\mathcal{S}$ est exhaustive si et seulement pour tout $\theta\in\Theta$, il existe une fonction g_θ $\mathcal{S}$-mesurable telle que

g_θ soit une densité de $\mathbb{P}_\theta$ par rapport à m.

Démonstration. Supposons que la statistique est exhaustive et prenons $\theta\in\Theta$. On note r_θ la densité de $\mathbb{P}_\theta$ par rapport à m.

Rappelons qu'on sait que pour tout θ

$$\mathbb{P}_\theta[X\in A|\mathcal{S}](\omega) = P_\omega(A) \quad \mathbb{P}_\theta\text{-p.s} \tag{6.4}$$

Comme m est un mélange dénombrable des $\mathbb{P}_\theta$, on a encore

$$m[X\in A|\mathcal{S}](\omega) = P_\omega(A) \quad m\text{-p.s.}$$

On a

$$\begin{aligned}
\mathbb{P}_\theta(X\in A) &= \int \mathbb{P}_\theta(X\in A|\mathcal{S})\ d\mathbb{P}_\theta \\
&= \int P_\omega(A)\ d\mathbb{P}_\theta(\omega) \\
&= \int P_\omega(A)\ r_\theta(\omega)\ dm(\omega) \\
&= \int P_\omega(A)\ g_\theta(\omega)\ dm(\omega),
\end{aligned}$$

où l'on a posé $g_\theta(\omega) = \mathbb{E}_m[r_\theta|\mathcal{S}]$.

Mais $\mathbb{E}_m[\mathbb{1}_{X\in A} g_\theta|\mathcal{S}] = g_\theta \mathbb{E}_m[\mathbb{1}_{X\in A}|\mathcal{S}] = g_\theta \mathbb{P}_\omega(A)$ m-presque sûrement,donc

$$\int P_\omega(A) g_\theta(\omega)\ dm(\omega) = \mathbb{E}_m[\mathbb{E}_m[\mathbb{1}_{X\in A} g_\theta|\mathcal{S}]] = \mathbb{E}_m(\mathbb{1}_{X\in A} g_\theta).$$

On a donc montré

$$\mathbb{P}_\theta(X \in A) = \mathbb{E}_m(\mathbb{1}_{X\in A} g_\theta),$$

c'est à dire que g_θ est la densité de la loi $\mathbb{P}_\theta$ par rapport à m.

Réciproquement, supposons que pour tout θ, il existe une fonction g_θ $\mathcal{S}$-mesurable telle que pour tout A $\mathbb{P}_\theta(A) = \mathbb{E}_m(\mathbb{1}_A g_\theta)$ et montrons que

$$\mathbb{P}_\theta(A|\mathcal{S}) = \mathbb{P}_m(A|\mathcal{S}) \quad \mathbb{P}_\theta - \text{p.s.}.$$

Fixons θ et A, et considérons sur $\mathcal{S}$ la mesure $\gamma_{A,\theta} = \mathbb{P}_\theta(A \cap \cdot)$. On va calculer de deux manières différentes une densité de $\gamma_{A,\theta}(\cdot)$ par rapport à m.

En utilisant tout de suite la forme particulière de la densité de $\mathbb{P}_\theta$ par rapport à m, on a

$$\begin{aligned}\gamma_{A,\theta}(C) &= \mathbb{E}_m(\mathbb{1}_C \mathbb{1}_A g_\theta)\\ &= \mathbb{E}_m[\mathbb{E}_m[\mathbb{1}_C \mathbb{1}_A g_\theta|\mathcal{S}]]\\ &= \mathbb{E}_m[\mathbb{1}_C g_\theta \mathbb{P}_m[A|\mathcal{S}]\end{aligned}$$

donc la dérivée de Radon–Nicodým de $\gamma_{A,\theta}(\cdot)$ par rapport à m est $g_\theta(S)\mathbb{P}_m[A|\mathcal{S}]$. Cependant, en commençant par un calcul direct, on obtient

$$\begin{aligned}\gamma_{A,\theta}(C) &= \mathbb{E}_\theta[\mathbb{1}_A \mathbb{1}_C]\\ &= \mathbb{E}_\theta(\mathbb{E}_\theta[\mathbb{1}_A \mathbb{1}_C|\mathcal{S}])\\ &= \mathbb{E}_\theta(\mathbb{1}_C \mathbb{P}_\theta[A|\mathcal{S}])\\ &= \mathbb{E}_m(\mathbb{1}_C \mathbb{P}_\theta[A|\mathcal{S}] g_\theta),\end{aligned}$$

donc la dérivée de Radon–Nicodým de $\gamma_{A,\theta}(\cdot)$ par rapport à m est $g_\theta \mathbb{P}_\theta[A|\mathcal{S}]$.

On en déduit que

$$g_\theta \mathbb{P}_\theta[A|\mathcal{S}] = g_\theta(S)\mathbb{P}_m[A|\mathcal{S}] \quad m - \text{p.s.}$$

Comme $\mathbb{P}_\theta \ll m$, on a encore

$$g_\theta \mathbb{P}_\theta[A|\mathcal{S}] = g_\theta \mathbb{P}_m[A|\mathcal{S}] \quad \mathbb{P}_\theta - \text{p.s.}$$

Notons enfin $\mathbb{P}_\theta(g_\theta = 0) = \mathbb{E}_m[\mathbb{1}_{\{g_\theta=0\}} g_{\theta(S)}] = 0$. Ceci permet de diviser par $g_\theta(S)$ et on obtient

$$\mathbb{P}_\theta[A|\mathcal{S}] = \mathbb{P}_m[A|\mathcal{S}] \quad \mathbb{P}_\theta - \text{p.s.}$$

□

Preuve du théorème de Fisher 45. Si la statistique est exhaustive, le lemme précédent montre qu'on a une fonction g_θ $\mathcal{S}$-mesurable qui est la densité de $\mathbb{P}_\theta$ par rapport à la dominante privilégiée m. Alors, $g_\theta \frac{dm}{d\mu}$ est évidemment la densité de $\mathbb{P}_\theta$ par rapport à μ.

Réciproquement, si la densité de $\mathbb{P}_\theta$ par rapport à μ s'écrit $\psi_\theta(S)h$, on peut noter, comme on l'a déjà vu plusieurs fois, qu'une densité est presque partout non nulle par rapport à la loi de la densité.

En particulier $\mathbb{P}_\theta(h = 0) = 0$ pour tout θ, ce qui entraîne que $m(h = 0) = 0$ par définition de m.

Par ailleurs, on peut calculer explicitement la densité de m par rapport à μ. En effet, pour tout A on a

$$\begin{aligned} m(A) &= \sum_{i \in D} a_i \mathbb{P}_{\theta_i}(A) \\ &= \sum_{i \in D} a_i \int \mathbb{1}_A(x) \psi_{\theta_i} h(x)\, d\mu(x) \qquad = \int \mathbb{1}_A(x) r(x) h(x)\, d\mu(x), \end{aligned}$$

où on a posé $r(s) = \sum_{i \in D} a_i \psi_{\theta_i}(s)$. La fonction r peut éventuellement être infinie, cependant l'intégrale par rapport à μ de $r.h$ est 1, donc $\mu(r.h = +\infty) = 0$, ce qui entraîne $m(r.h = +\infty) = 0$. Comme $m(h = 0) = 0$, on a $m(r = +\infty) = 0$, ce qui entraîne que $\mathbb{P}_\theta(r = +\infty) = 0$ pour tout θ. $r.h$. Ainsi $r.h$ est la densité de m par rapport à μ. Comme $r.h$ est la densité de m par rapport à μ, $m(r.h = 0) = 0$. Comme $\mathbb{P}_\theta \ll m$, $\mathbb{P}_\theta(r.h = 0) = 0$. Notons $Z = \{r.h \neq 0\}$.

Pour tout A, on a

$$\begin{aligned} \mathbb{P}_\theta(A) &= \mathbb{P}_\theta(A \cap Z) \\ &= \int \mathbb{1}_A \mathbb{1}_Z \psi_\theta h\, d\mu \\ &= \int \mathbb{1}_A \mathbb{1}_Z \psi_\theta h \frac{rh}{rh}\, d\mu \\ &= \int \mathbb{1}_A \mathbb{1}_Z \frac{\psi_\theta}{r} rh\, d\mu \\ &= \int \mathbb{1}_A \mathbb{1}_Z \frac{\psi_\theta}{r}\, dm \\ &= \int \mathbb{1}_A \frac{\psi_\theta}{r}\, dm \end{aligned}$$

En posant $g_\theta(s) = \frac{\psi_\theta(s)}{r(s)}$, on peut alors appliquer le lemme.

□

6.3 Amélioration de Rao-Blackwell

Le théorème de Rao-Blackwell-Kolmogorov permet comment la connaissance d'une statistique exhaustive permet d'améliorer des estimateurs.

Théorème 47. *Soit $\hat{g}$ un estimateur de $g(\theta)$ et S une statistique exhaustive. Alors*

$$\mathbb{E}_\theta[\hat{g}|S]$$

est un estimateur de g qui est préférable g au sens où

$$\forall \theta \in \Theta \quad \mathbb{E}_\theta((\mathbb{E}_\theta[\hat{g}|S] - g(\theta))^2) \leq \mathbb{E}_\theta(\hat{g} - g(\theta))^2.$$

À θ fixé, l'égalité n'a lieu que si $\hat{g} = \mathbb{E}_\theta[\hat{g}|S]$ $\mathbb{P}_\theta$-presque sûrement. $\mathbb{E}_\theta[\hat{g}|S]$ à même biais que $\hat{g}$.

Démonstration. Comme S est une statistique exhaustive, on peut écrire

$$\mathbb{E}_\theta[g|S] = \psi(S), \text{ avec } \psi(s) = \int g \, dP_s.$$

Ainsi, $\mathbb{E}_\theta[g|S]$ est bien un estimateur. $\mathbb{E}_\theta[\hat{g}|S] - g(\theta)$ est l' espérance conditionnelle de $\hat{g} - g(\theta)$ sous $\mathbb{E}_\theta$ conditionnellement. Comme l'espérance conditionnelle est une contraction de L^2, le résultat s'ensuit. Le cas d'égalité découle du théorème de Pythagore pour l'espérance conditionnelle. Bien sûr $\mathbb{E}_\theta[\hat{g}|S]$ et $\hat{g}$ ont même espérance sous $\mathbb{E}_\theta$: c'est donc le même biais. □

Définition. *On dit qu'un estimateur sans biais $\hat{g}$ est uniformément de variance minimum parmi les estimateurs sans biais (UVMB) si pour tout estimateur sans biais $\tilde{g}$, on a*

$$\forall \theta \in \Theta \quad \operatorname{Var}_\theta \tilde{g} \geq \operatorname{Var}_\theta \hat{g}.$$

6.4 Statistiques exhaustives minimales

On dit qu'une tribu exhaustive $\mathcal{S}_0$ est minimimale si toute tribu exhaustive $\mathcal{S}$ vérifie $\mathcal{S}_0 \subset \mathcal{S}$.

Une statistique exhaustive est dite minimale si la tribu qu'il engendre est minimale.

Théorème 48. *Soit une famille de mesures de probabilité $(\mathbb{P}_\theta)_{\theta\in\Theta}$ telle que chaque m_θ est absolument continue par rapport à une même mesure μ. On note f_θ la densité de $\mathbb{P}_\theta$ par rapport à μ. Alors, si, avec les notations du lemme 10, on pose $f_m = \sum \alpha_{i\in D} f_{\theta_i}$ et $r_\theta = \frac{f_\theta}{f_m}$, alors la tribu $\mathcal{T} = \sigma(r_\theta)_{\theta\in D}$ est une tribu exhaustive minimale.*

Démonstration. On a $\mathbb{P}_\theta \ll m \ll \mu$, donc on peut écrire

$$\frac{d\mathbb{P}_\theta}{d\mu} = \frac{d\mathbb{P}_\theta}{dm}\frac{dm}{d\mu},$$

soit $f_\theta = \frac{d\mathbb{P}_\theta}{dm} f_m$. Ainsi r_θ est la densité de $\mathbb{P}_\theta$ par rapport à m. Évidemment, r_θ est $\mathcal{T}$-mesurable, donc d'après le théorème de Fisher, $\mathcal{T}$ est exhaustive. Soit $\mathcal{S}$ une tribu quelconque supposée exhaustive pour $(\mathbb{P}_\theta)$. D'après le théorème de Fisher, on a une écriture

$$f_\theta = \psi_\theta h,$$

où ψ_θ est $\mathcal{S}$-mesurable. On a alors

$$f_m = \sum_{i \in D} \alpha_i f_{\theta_i} = \sum_{i \in D} \alpha_i \psi_{\theta_i} h$$

et pour tout $\theta \in D$, on a μ-presque sûrement :

$$r_\theta = \frac{f_\theta}{f_m} = \frac{\psi_\theta h}{\sum_{i \in D} \alpha_i \psi_{\theta_i} h} = \frac{\psi_\theta}{\sum_{i \in D} \alpha_i \psi_{\theta_i}}$$

(En effet h est μ-presque sûrement non nulle.) Cette identité montre que r_θ est $\mathcal{S}$-mesurable, d'où $\mathcal{T} = \sigma((r_\theta)_{\theta \in D}) \subset \mathcal{S}$, ce qui montre bien que $\mathcal{T}$ est minimale. □

Corollaire 13. *On suppose que les mesures $\mathbb{P}_\theta$ ont toutes le même support et que les hypothèses du théorème de Factorisation sont vérifiées :*

$$f_\theta(x) = h(x)\psi_\theta(S(x)).$$

Si il existe θ_1, θ_2 avec $\theta_1 \neq \theta_2$ tels que l'application $x \mapsto \frac{\psi_{\theta_1}}{\psi_{\theta_2}}$ est bijective, alors S est une statistique exhaustive minimale.

Démonstration. Par définition, la tribu minimale vérifie $\mathcal{T} \subset \sigma(S)$. D'après le théorème précédent $\frac{\psi_{\theta_1}}{\psi_{\theta_2}}(S(x))$ est $\mathcal{T} \subset \sigma(T)$-mesurable. Si $q = \frac{\psi_{\theta_1}}{\psi_{\theta_2}}$ est bijective, alors $S = q^{-1}(\frac{\psi_{\theta_1}}{\psi_{\theta_2}}(S(x))$ est $\mathcal{T} \subset \sigma(T)$-mesurable, donc $\sigma(S) \subset \mathcal{T}$, ce qui donne l'égalité voulue. □

6.5 Statistiques complètes

Définition. *On dit qu'une statistique S est complète relativement à la famille $(\mathbb{P}_\theta)_{\theta \in \Theta}$ si pour tout fonction φ mesurable.*

$$(\forall\theta \in \Theta \quad \mathbb{E}_\theta[\varphi(S)] = 0) \Longrightarrow (\forall\theta \in \Theta \quad \varphi(S) = 0 \quad \mathbb{P}_\theta \text{ p.s.}).$$

La définition peut sembler étrange : en réalité elle exprime l'injectivité de l'application linéaire

$$X \mapsto (\mathbb{E}_\theta(X))_{\theta\in\Theta}$$

sur l'ensemble des statistiques $\sigma(S)$-mesurables qui sont dans $\cap_{\theta\in\Theta} L^1(\mathbb{P}_\theta)$.

Théorème 49 (théorème de Lehmann-Scheffé)**.** *Soit S une statistique exhaustive complète, et $\hat{g}$ un estimateur sans biais de $g(\theta)$. Alors*

— $\mathbb{E}_\theta(\hat{g}|S)$ *est un estimateur sans biais de $g(\theta)$.*

— $\mathbb{E}_\theta(\hat{g}|S)$ *ne dépend pas du choix de $\hat{g}$*

— $\mathbb{E}_\theta(\hat{g}|S)$ *est un estimateur sans biais UVMB : pour tout estimateur sans biais T, on a*

$$\forall\theta \in \Theta \operatorname{Var}_\theta(\mathbb{E}_\theta(\hat{g}|S)) \le \operatorname{Var}_\theta T.$$

Démonstration. D'après le théorème de Rao-Blackwell-Kolmogorov, $\mathbb{E}_\theta(\hat{g}|S)$ est bien un estimateur, qui plus est sans biais. Si T est un autre estimateur sans biais de $g(\theta)$, $\mathbb{E}_\theta(\hat{g}|S)$ et $\mathbb{E}_\theta(T|S)$ sont deux estimateurs $\sigma(S)$-mesurables qui ont même espérance : ils sont donc égaux car S est complète. Maintenant, Rao-Blackwell-Kolmogorov nous dit que

$$\operatorname{Var}_\theta(\mathbb{E}_\theta(\hat{g}|S)) = \operatorname{Var}_\theta(\mathbb{E}_\theta(T|S)) \le \operatorname{Var}_\theta T,$$

ce qui achève la preuve. □

On peut également noter le résultat suivant :

Théorème 50. *Une statistique exhaustive complète S est une statistique exhaustive minimale.*

Démonstration. Quitte à remplacer S par $\arctan S$, on peut supposer que S est bornée (en effet la tribu engendrée par S et celle engendrée par $\arctan S$ coïncident). Soit $\mathcal{T}$ une tribu exhaustive minimale. Bien sûr $\mathcal{T} \subset \sigma(S)$. Notons que comme $\mathcal{T}$ est exhaustive, on a

$$\mathbb{E}_\theta(S|\mathcal{T})(\omega) = E_\omega(S).$$

Posons $\varphi = S - E(S) = S - \mathbb{E}_\theta(S|\mathcal{T})$: par construction φ est $\sigma(S)$-mesurable. Les propriétés de l'espérance conditionnelle nous donne $\mathbb{E}_\theta\varphi = 0$. Comme φ est une statistique $\sigma(S)$-mesurable et que S est complète, on a $\varphi = 0$. Donc $S = \mathbb{E}_\theta(S|\mathcal{T})$ $\mathbb{P}_\theta$ presque sûrement pour tout θ, ce qui donne la mesurabilité de S par rapport à $\mathcal{T}$: S est donc complète.

□

Montrer qu'une statistique est complète est essentiellement un problème d'analyse, qui peut être difficile. Heureusement, on a un théorème générique qui peut être appliqué dans de nombreux cas.

6.6 Modèles exponentiels

Définition. *On dit qu'une famille* $(\mathbb{P}_\theta)$ *forme un modèle exponentiel si les densités s'écrivent*

$$f_\theta(x) = \beta(\theta)\xi(x)\exp(\langle\alpha(\theta), S(x)\rangle)$$

— *β et ξ sont à valeurs dans $\mathbb{R}_+$. α est appelé le paramètre naturel*
— *α et S sont à valeurs dans $\mathbb{R}^d$.*

Remarque 6. — *D'après le théorème de Neymann-Fisher, S est une statistique exhaustive.*
— *Si $(\mathbb{P}_\theta)_{\theta\in\Theta}$ est un modèle exponentiel, alors $(\mathbb{P}_\theta^{\otimes n})_{\theta\in\Theta}$ l'est aussi. En effet la densité de $(X_1, \dots, X_n)$ s'écrit*

$$\beta(\theta)^n \prod_{i=1}^n \zeta(x_i) \exp\left(\langle\alpha(\theta), \sum_{i=1}^n S(x_i)\rangle\right),$$

et $\sum_{i=1}^n S(X_i)$ est la statistique naturelle associée.

Théorème 51. *Dans un modèle exponentiel, si l'image de Θ par le paramètre naturel contient un ouvert de $\mathbb{R}^d$; alors la statistique naturelle du modèle est complète.*

Démonstration. Soit φ une application bornée telle que pour tout θ

$$\int f_\theta(x)\varphi(S(x))\, d\mu = 0.$$

Pour tout $\theta \in \Theta$, on a

$$\int \xi(x)(\varphi_+(S(x)) - \varphi_-(S(x)))\exp(\langle\alpha(\theta), S(x)\rangle)\, d\mu = 0.$$

Si on note $\tilde{\Theta}$ l'image de Θ par α, et $\nu = \xi\mu$, on a Pour tout $\theta \in \Theta$, on a

$$\int(\varphi_+(S(x)) - \varphi_-(S(x)))\exp(\langle\alpha, S(x)\rangle)\, d\nu = 0,$$

soit encore

$$\int \varphi_+(y)\exp(\langle\alpha, y\rangle)\, d\mu_S(y) = \int \varphi_+(y)\exp(\langle\alpha, y\rangle)\, d\mu_S(y)$$

Les mesures $\varphi_+\mu_S$ et $\varphi_-\mu_S$ ont même transformée de Laplace sur un ouvert : elles sont égales $\varphi_+ = \varphi_-$ μ_S presque sûrement : donc $\varphi_+^2 = \varphi_-^2 = \varphi_+\varphi_- = 0$, d'où $\varphi = 0$ ν_S presque sûrement : On a donc

$$\begin{aligned} 0 = &\int 1_{\{\varphi(y)\neq 0\}}\ d\nu_S \\ &\int 1_{\{\varphi(S(x))\neq 0\}}\ d\nu \\ &\int 1_{\{\varphi(S(x))\neq 0\}}\xi(x)\ d\mu \end{aligned}$$

Donc $1_{\{\varphi(S(x))\neq 0\}}\xi(x)$ est μ presque partout nulle ; comme $\mathbb{P}_\theta \ll \mu$, on a $1_{\{\varphi(S(x))\neq 0\}}\xi(x) = 0$ $\mathbb{P}_\theta$ presque partout. Comme ξ est $\mathbb{P}_\theta$ presque partout non nulle, donc $1_{\{\varphi(S(x))\neq 0\}} = 0$ $\mathbb{P}_\theta$ presque partout, soit $\mathbb{P}_\theta(\varphi(S(x) \neq 0)) = 0$, et on peut dire que S est complète. $\square$

6.7 Exercices sur les statistiques exhaustives

6.7.1 Exercices corrigés

Exercice 44. Soit $(\mathbb{P}_\theta)$ un modèle dominé de dominante privilégiée m, Z une statistique et S une statistique exhaustive.

1. Montrer que si Z et S sont indépendantes sous m, alors Z est libre.
2. *Théorème de Basu* : Montrer que si Z est libre et S une statistique exhaustive complète, alors Z et S sont indépendantes sous $\mathbb{P}_\theta$, quelque soit $\theta \in \Theta$.

lien vers l'indication

6.7.2 Exercices non corrigés

Exercice 45. On considèle le modèle $(\mathbb{R}, \mathcal{B}(\mathbb{R}^n), \mathbb{P}_\theta)_{\theta\in\Theta}$, avec $\Theta = \mathbb{R}\times]0, +\infty[$, $\mathbb{P}_{(m,\sigma^2)} = \mathcal{N}(m, \sigma^2)^{\otimes n}$. À l'aide du théorème de Neyman–Fisher, trouver une statistique exhaustive pour ce modèle. lien vers l'indication

Exercice 46. Soit n un entier naturel non nul. On considèle le modèle $(\mathbb{R}, \mathcal{B}(\mathbb{R}^n), \mathbb{P}_\theta)_{\theta\in\Theta}$, avec $\Theta =]0, +\infty[$, $\mathbb{P}_\lambda = \mathcal{P}(\lambda)^{\otimes n}$.

1. Montrer que $S_n = \sum_{i=1}^n X_i$ est une statistique exhaustive du modèle.
2. Déterminer la loi de S_n sous $\mathbb{P}_\theta$.
3. Soit f une fonction bornée. Montrer que
$$F(z) = \sum_{k=0}^{+\infty} \frac{(nz)^k}{k!} f(k)$$
définit une fonction holomorphe sur $\mathbb{C}$.
4. Montrer que S_n est une statistique exhaustive complète du modèle.

lien vers l'indication

Exercice 47. Soit $(X_1, \dots, X_n)$ un n-échantillon de la loi uniforme sur $[0, \theta]$, où θ décrit $]0, +\infty[$.

1. Calculer l'estimateur du maximum de vraisemblance pour θ. On le notera M_n dans la suite. Montrer que cet estimateur est une statistique exhaustive.
2. Déterminer la loi de M_n, puis montrer M_n est une statistique exhaustive complète.
3. Construire un estimateur sans biais de θ.

4. En déduire sans calcul la valeur de $\mathbb{E}_\theta[\overline{X}_n | \max(X_1, \dots, X_n)]$, où $\overline{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$.

lien vers l'indication

Exercice 48. Soit $(X_1, \dots, X_n)$ un n-échantillon de la loi uniforme sur $[0, \theta]$, où θ décrit $]0, +\infty[$. On veut estimer $\mathbb{P}_\theta(X_1 \leq t)$.

1. Construire une statistique exhaustive complète du modèle.
2. Construire un estimateur sans biais de $\mathbb{P}_\theta(X_1 \leq t)$.
3. Trouver le meilleur estimateur sans biais de $\mathbb{P}_\theta(X_1 \leq t)$.

lien vers l'indication

Exercice 49. On se fixe $\theta > 0$ connu, et on considère le modèle $(P_m)_{m \in \mathbb{R}}$, où $\mathbb{P}_m = \mathcal{N}(m, \sigma^2)^n$ et m décrit $\mathbb{R}$.

1. Donner une statistique exhaustive complète du modèle.
2. À l'aide du théorème de Basu (vu en exercice), montrer le théorème de Fisher pour les échantillons gaussiens : si $(X_1, \dots, X_n)$ est un n-échantillon de la loi $\mathcal{N}(m, \sigma^2)$, alors les variables

$$\overline{X}_n = \frac{1}{n}(X_1 + \cdots + X_n) \text{ et } S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

sont indépendantes.

lien vers l'indication

Chapitre 7

Information de Fisher

Soit Θ un ouvert de $\mathbb{R}$ Dans tout ce qui suit, on suppose que $(\mu_\theta)_{\theta\in\Theta}$ est une famille de lois sur $\Omega \subset \mathbb{R}^d$ et μ une loi sur Ω telles que :

— $\mathbb{P}_\theta$ admet une densité f_θ par rapport à μ avec $f_\theta > 0$.

— Pour μ presque tout $x \in \Omega$, $\theta \to f_\theta(x)$ est dérivable par rapport à θ.

7.1 Hypothèses

Si h est une fonction mesurable bornée, on a

$$\forall \theta \in \Theta \quad \mathbb{E}_\theta h(X) = \int_\Omega h(x) f_\theta x \; d\mu.$$

Si jamais

$$\frac{\partial}{\partial\theta} \int_\Omega h(x) f_\theta(x) \; d\mu = \int_\Omega h(x) \frac{\partial}{\partial\theta} f_\theta(x) \; d\mu,$$

alors on aura

$$\frac{\partial}{\partial\theta} \mathbb{E}_\theta[h(X)] = \mathbb{E}_\theta[W_\theta h(X)], \tag{7.1}$$

avec

$$W_\theta = (\frac{\partial}{\partial\theta} \log f_\theta)(X).$$

$$\forall h \in \mathcal{H} \quad \frac{\partial}{\partial\theta} \mathbb{E}_\theta[h(X)] = \mathbb{E}_\theta[W_\theta h(X)]. \tag{7.2}$$

Il existe plusieurs types d'espaces $\mathcal{H}$ et d'hypothèses qui permettent cette intervertion de la dérivée. Par exemple

Théorème 52. *Si l'on prend pour $\mathcal{H}$ l'ensemble des fonctions bornées, alors une condition suffisante pour (7.2) est la suivante :*

Pour tout $\theta \in \Theta$, il existe un voisinage V de θ tel que

$$\sup_{\theta \in V} \frac{\partial}{\partial \theta} f_\theta \in L^1(\mu).$$

Démonstration. C'est une simple application du théorème de dérivation sous le signe intégrale. □

Le théorème suivant est plus subtil :

Théorème 53. *Si l'on prend pour $\mathcal{H}$ l'ensemble des fonctions h telles que la fonction $\theta \mapsto \mathbb{E}_\theta[h^2(X)]$ est bornée au voisinage de tout point de Θ, une condition suffisante pour (7.2) est la suivante : pour μ-presque tout x, la fonction $\theta \mapsto g_x(\theta) = \sqrt{f_\theta(x)}$ est de classe C_1, et la fonction*

$$\theta \mapsto \mathcal{I}(\theta) = \int (g'_x(\theta))^2 \, d\mu(x) = \int \left(\frac{\partial}{\partial \theta} \sqrt{f_\theta(x)} \right)^2 \, d\mu(x)$$

est continue, à valeurs réelles.

Démonstration. La preuve sera vue en exercice. □

Remarquons que pour vérifier la continuité de $\mathcal{I}(\theta)$, il faudra dans la plupart des cas utiliser le théorème de dérivation sous le signe intégrale.

Remarques

— La quantité $(\frac{\partial}{\partial \theta} \log f_\theta)(X)$ est indépendante de la mesure de référence μ. En effet si les (μ_θ) ont des densités (f_θ) par rapport à μ et (g_θ) par rapport à μ, alors les (μ_θ) ont des densités (h_θ) par rapport à $\mu + \nu$ et on a

$$\frac{d\mu_\theta}{d(\mu+\nu)} = \frac{d\mu_\theta}{d\mu} \frac{d\mu}{d(\mu+\nu)} \text{ et } \frac{d\mu_\theta}{d(\mu+\nu)} = \frac{d\mu_\theta}{d\nu} \frac{d\nu}{d(\mu+\nu)}.$$

$\frac{d\mu_\theta}{d(\mu+\nu)}$ est $\mu + \nu$ presque partout non-nulle. Soit x un point de $\mathbb{R}^d$: on suppose que $f_{\theta_0}(x) > 0$ (ce qui est $\mu + \nu$ partout équivalent à $h_{\theta_0}(x) > 0$) : on a sur un voisinage V de θ_0 :

$$\log h_\theta(x) = \log f_\theta(x) + \log \frac{d\mu}{d(\mu+\nu)}(x),$$

d'où l'égalité des deux dérivées partielles par rapport à θ.

— Sous l'hypothèse (7.2), on a $\mathbb{E}_\theta(W_\theta) = 0$.

— Il peut exister au plus une collection (W_θ) vérifiant (7.2), car si (W_θ^1) et (W_θ^2) conviennent $W_\theta^1 - W_\theta^2$ est orthogonal à $L^2(\mathbb{P}_\theta)$.

Définition : information de Fischer

Définition. *On appelle score du modèle $(\mathbb{P}_\theta)_{\theta\in\Theta}$ la statistique W_θ définie par*

$$W_\theta = (\frac{\partial}{\partial\theta}\log f_\theta)(X),$$

avec la convention que $\frac{\partial}{\partial\theta}\log f_\theta$ est nulle en dehors des intervalles où f est strictement positive. On appelle information de Fisher la quantité $\mathcal{I}(\theta)$ définie par

$$\mathcal{I}(\theta) = \mathbb{E}_\theta W_\theta^2$$

Remarque 7. *L'hypothèse (7.2) est très importante. C'est elle qui donne du sens à l'information de Fisher. Néanmoins, il est intéressant de définir $\mathcal{I}(\theta)$ avant de savoir si (7.2) est réalisée, car on verra que certaines propriétés de $\mathcal{I}$ peuvent parfois être utiles pour démontrer que (7.2) est vérifiée.*

7.2 Inégalité de Cramer-Rao

Théorème 54 (Inégalité de Cramer-Rao)**.** *On suppose que l'hypothèse (7.2) est vérifiée et que $\hat{g} = h(X)$ est un estimateur sans biais de $g(\theta)$ avec $h \in \mathcal{H}$. Alors*

$$\mathrm{Var}_\theta\,\hat{g} \geq \frac{g'(\theta)^2}{\mathcal{I}(\theta)}.$$

Démonstration. Comme $\hat{g}$ est un estimateur sans biais de $g(\theta)$, on a $g(\theta) = \mathbb{E}_\theta[h(X)W_\theta]$. En appliquant l'hypothèse (7.2), on a

$$g'(\theta) = \mathbb{E}_\theta[h(X)W_\theta] = \mathbb{E}_\theta[\hat{g}W_\theta].$$

Mais W_θ est centrée, donc on a également

$$g'(\theta) = \mathbb{E}_\theta[(\hat{g} - \mathbb{E}_\theta\hat{g})W_\theta].$$

L'inégalité de Cauchy-Schwartz donne alors

$$g'(\theta)^2 \leq \mathbb{E}_\theta(\hat{g} - \mathbb{E}_\theta\hat{g})^2\mathbb{E}_\theta W_\theta^2 = (\mathrm{Var}_\theta\,\hat{g})\mathcal{I}(\theta).$$

□

Théorème 55 (Cas d'égalité)**.** *On suppose que l'hypothèse (7.2) est vérifiée et que $\hat{g} = h(X)$ est un estimateur sans biais de $g(\theta)$. Alors si*

$$\forall \theta \in \Theta \quad 0 < \operatorname{Var}_\theta \hat{g} = \frac{g'(\theta)^2}{\mathcal{I}(\theta)},$$

$\hat{g}$ est une statistique exhaustive et le modèle est exponentiel.

Réciproquement, un modèle exponentiel vérifiant l'hypothèse (7.2) atteint la borne de Cramer-Rao pour une certaine fonction g.

Démonstration. D'après le cas d'égalité dans l'inégalité de Cauchy-Schwartz, les variables $\hat{g} - \mathbb{E}_\theta \hat{g}$ et W_θ sont liées dans $L^2(\mathbb{P}_\theta)$. Comme $0 < \operatorname{Var}_\theta \hat{g}$, il existe $\alpha(\theta)$ avec

$$W_\theta = \alpha(\theta)(\hat{g} - \mathbb{E}_\theta \hat{g}) = \alpha(\theta)(h(X) - g(\theta)) \quad \mathbb{P}_\theta - p.s.$$

Ainsi, sous la dominante privilégiée m, on a

$$W_\theta = \alpha(\theta)(h(X) - g(\theta)) \quad m - p.s.$$

En intégrant l'égalité, on voit que la densité f_θ sous m vérifie

$$\ln f_\theta(X) = A(\theta)h(X) + B(\theta) + c(X),$$

et donc

$$f_\theta(x) = \exp(A(\theta)h(X) + B(\theta) + c(X))$$

On a donc bien un modèle exponentiel de paramètre naturel $A(\theta)$ et h est une statistique exhaustive du modèle.

Réciproquement, considérons un modèle exponentiel : W_θ s'écrit $W_\theta = \alpha'(\theta)S + \beta'(\theta)$. Comme $\mathbb{E}_\theta[W_\theta] = 0$, on a $\alpha'(\theta)\mathbb{E}_\theta[S] + \beta'(\theta) = 0$. Ainsi $W_\theta = \alpha'(\theta)(S - \mathbb{E}_\theta(S))$, et on est dans le cas d'égalité de l'inégalité de Cauchy-Schwartz : S est un estimateur optimal pour la fonction $g(\theta) = \mathbb{E}_\theta(S)$. □

7.3 Quelques propriétés

7.3.1 Information de Fisher d'un produit

Théorème 56. *On suppose que $(\mathbb{P}^1_\theta)$ et $(\mathbb{P}^2_\theta)$ ont même support pour tout θ et que les modèles $(\mathbb{P}^1_\theta)$ et $(\mathbb{P}^2_\theta)$ vérifient les hypothèses du théorème 53. Alors, le modèle $\mathbb{P}_\theta = \mathbb{P}^1_\theta \otimes \mathbb{P}^2_\theta$ vérifie les hypothèses du théorème 53, on a*

$$W_{(\theta_1,\theta_2)}(X) = W_1(X_1) + W_2(X) \quad \textit{et} \quad \mathcal{I}(\theta) = \mathcal{I}_1(\theta) + \mathcal{I}_2(\theta).$$

Démonstration. Le modèle $\mathbb{P}_\theta$ admet la densité $h_{(\theta_1,\theta_2)}(x) = f_{\theta_1}(x_1)g_{\theta_2}(x)$, et on a sur leur support commun

$$\frac{\partial}{\partial\theta}\log h_{(\theta_1,\theta_2)}(x) = \frac{\partial}{\partial\theta}\log f_{\theta_1}(x_1) + \frac{\partial}{\partial\theta}\log h_{\theta_2}(x_2),$$

ce qui donne la première identité L'indépendance de X_1 et X_1 donne $\mathcal{I}(\theta) = \mathrm{Var}_\theta\, W(X) = \mathrm{Var}_\theta\, W_1(X_1) + \mathrm{Var}_\theta(X_2) = I_1(\theta) + I_2(\theta)$, donc I est continue comme somme de deux fonctions continues. De même, $\sqrt{h}$ est C^1 comme produit de deux fonctions C^1 □

Corollaire 14. *Si le modèles $(\mathbb{P}_\theta)$ vérifient les hypothèses du théorème 53 , le modèle $(\mathbb{P}_\theta^{\otimes n})$ également, et son information de Fisher $I_n(\theta)$ vérifie $I_n(\theta) = nI\theta$.*

7.3.2 Information de Fisher d'une statistique

Théorème 57. *Soit $(\Omega, \mathcal{F}, \mathbb{P}_\theta)$ un modèle statistique vérifiant l'hypothèse (7.2), avec $\mathcal{H} \subset \cap_{\theta\in\Theta} L^2(\mathbb{P}_\theta)$. Soit $T : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ une statistique. Alors le modèle $(\Omega', \mathcal{F}', (\mathbb{P}_\theta)_S)$ vérifie l'hypothèse (7.2) et son information de Fisher $I^T(\theta)$ vérifie*

$$\forall\theta \in \Theta \quad I^S(\theta) \leq \mathcal{I}(\theta).$$

Si on suppose de plus que l'hypothèse (7.2) est vérifiée, alors il y a égalité si et seulement si S est une statistique exhaustive pour $(\Omega', \mathcal{F}', (\mathbb{P}_\theta)_S)$.

Démonstration. D'après le théorème de transfert, on a pour tout $\theta \in \Theta$ et tout $h \in L^2((\mathbb{P}_\theta)_S)$,

$$\int h(x)\ d(\mathbb{P}_\theta)_S(x) = \int h(S(x))\ d\mathbb{P}_\theta(x) = \mathbb{E}_\theta[(h(S(x))]$$

D'après l'hypothèse (7.2), on a donc

$$\begin{aligned}\frac{\partial}{\partial\theta}\int h(x)\ d(\mathbb{P}_\theta)_S(x) &= \mathbb{E}_\theta[W_\theta h(S(x))] \\ &= \mathbb{E}_\theta[\mathbb{E}_\theta[W_\theta|S]h(S(x))] \qquad = \mathbb{E}_\theta[\psi_\theta(S)h(S(x))]\end{aligned}$$

où ψ_θ est telle que $\mathbb{E}_\theta[W_\theta|S] = \psi_\theta(S)$. Ainsi, avec le théorème de transfert ψ_θ est un score pour $(\Omega', \mathcal{F}', (\mathbb{P}_\theta)_S)$ et l'on a

$$\begin{aligned}I^S(\theta) &= \int \psi_\theta^2(x)\ d(\mathbb{P}_\theta)_S(x) \\ &= \mathbb{E}_\theta[\psi_\theta^2(x)] = \mathbb{E}_\theta[\mathbb{E}_\theta[W_\theta|S]^2]\end{aligned}$$

Comme l'espérance conditionnelle est la projection dans L^2. On a

$$I^S(\theta) = \mathbb{E}_\theta[\mathbb{E}_\theta[W_\theta|S]^2] \leq \mathbb{E}_\theta(W_\theta)^2 = \mathcal{I}(\theta),$$

avec égalité si et seulement si

$$\mathbb{E}_\theta[W_\theta|S] = W_\theta,$$

autrement dit W_θ est $\mathcal{S}$ mesurable. Supposons qu'il y ait égalité : on a $(\log f_\theta)' = A_\theta(S(x))$, d'où en intégrant par rapport à θ : $\log f_\theta(x) = B_\theta(S(x)) + c(x)$, soit

$$f_\theta(x) = \exp(B_\theta(S(x)) \exp(c(X)),$$

ce qui montre que S est une statistique exhaustive. La réciproque est facile.

$\square$

7.4 Exercices sur l'information de Fisher

7.4.1 Exercices corrigés

Exercice 50. On appelle loi de Pareto $P_{a,\alpha}$ de paramètre (a, α) la loi sur $\mathbb{R}$ de densité $\frac{\alpha a^\alpha}{x^{\alpha+1}} \mathbb{1}_{]a,+\infty[}(x)$. On considère le modèle $(P_{1,\alpha})_{\alpha>0}$.

1. Calculer le score W_α, puis $\mathcal{I}(\alpha)$.
2. En déduire que $T = \log X$ est un estimateur sans biais de $g(\alpha) = \frac{1}{\alpha}$.
3. Calculer $\operatorname{Var}_\alpha T$. Comparer avec la borne de Cramer-Rao $\frac{g'(\alpha)^2}{\mathcal{I}(\alpha)}$. Expliquer
4. Existe-t'il un estimateur sans biais de α ?

lien vers l'indication lien vers la solution

Exercice 51. Soit $(X_1, \ldots, X_d)$ un d-échantillon sous la loi $\mathcal{P}(\lambda)$ dans le modèle $(\mathbb{P}_\lambda)_{\lambda>0}$.

1. Calculer $\mathbb{E}_\lambda X_1(X_1 - 1)$. En déduire que $E_d = \frac{1}{d}\sum_{i=1}^d X_i(X_i - 1)$ est un estimateur sans biais de λ^2.
2. Montrer que $S_d = X_1 + \cdots + X_d$ est une statistique exhaustive complète.
3. En déduire que $\frac{S_d(S_d-1)}{d^2}$ est le meilleur estimateur quadratique de λ^2.
4. Montrer simplement que $\mathbb{E}_\lambda[X_1^2 + \ldots X_d^2 | S_d] = \frac{S_d(S_d+d-1)}{d}$.
5. Comparer la variance de $\frac{S_d(S_d-1)}{d^2}$ avec la borne de Cramer-Rao. $\frac{S_d(S_d-1)}{d^2}$ est-il un estimateur efficace ?

lien vers l'indication lien vers la solution

7.4.2 Exercices non corrigés

Exercice 52. 1. calculer l'information de Fischer lorsque $\Theta =]0, +\infty[$ et μ_θ est la loi de Poisson de paramètre θ.

2. Même question lorsque $\mu_\theta = \mathcal{P}(\theta)^{\otimes n}$

lien vers l'indication

Exercice 53. Le but de cet exercice est de démontrer le théorème suivant annoncé en cours :

Soit $(\mathbb{P}_\theta)_{\theta\in\Theta}$ un modèle dominé par μ, de densité f_θ. On suppose que pour μ-presque tout x, la fonction $\theta \mapsto g_x(\theta) = \sqrt{f_\theta(x)}$ est de classe C_1, et la fonction

$$\theta \mapsto \mathcal{I}(\theta) = \int (g_x'(\theta))^2 \, d\mu(x) = \int \left(\frac{\partial}{\partial\theta} \sqrt{f_\theta(x)} \right)^2 \, d\mu(x)$$

est continue, à valeurs réelles.

Alors pour toute fonction h telle que la fonction $\theta \mapsto \mathbb{E}_\theta[h^2(X)]$ est bornée au voisinage de tout point de Θ, on a :

$$\frac{\partial}{\partial\theta}\mathbb{E}_\theta[h(X)] = \mathbb{E}_\theta[W_\theta h(X)] \text{ avec } W_\theta = (\frac{\partial}{\partial\theta}\log f_\theta)(X).$$

On convient que $(\frac{\partial}{\partial\theta}\log f_\theta)$ désigne la fonction nulle sur les parties où f_θ est nulle.

On peut supposer sans perte de généralité que μ est une mesure de probabilité (on peut prendre par exemple la dominante privilégiée).

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé sur lesquelles vivent des variables aléatoires X et U indépendantes, avec $\mathbb{P}_X = \mu$ et U suit la loi uniforme sur $[0, 1]$.

Soit $\theta \in \Theta$ et $(\varepsilon_n)_{n\geq 0}$ une suite quelconque de limite nulle, avec $\theta \in \varepsilon_n \in \Theta$ pour tout n.

On pose

$$X_n = 2h(X)\sqrt{f_{\theta+U\varepsilon_n}(x)} \quad Y_n = \frac{\partial}{\partial\theta}(\sqrt{f_{\theta+U\varepsilon_n}})(X), \qquad \text{et } Z_n = X_n Y_n.$$

1. Montrer que $\mathbb{E}[Y_n^2|U] = \mathcal{I}(\theta + U\varepsilon_n)$.
2. En déduire que $\lim_{n\to+\infty}\mathbb{E}[Y_n^2] = \mathcal{I}(\theta)$.
3. Montrer que $(Y_n^2)_{n\geq 1}$ est équi-intégrable.
4. Montrer que $(X_n)_{n\geq 1}$ est bornée dans L^2.
5. Montrer que $(Z_n)_{n\geq 1}$ est équi-intégrable.
6. Montrer que $\lim_{n\to+\infty}\mathbb{E}[Z_n] = \int h(x)\frac{\partial}{\partial\theta}f_\theta(x)\ d\mu(x)$.
7. Montrer que $\mathbb{E}Z_n = \frac{\mathbb{E}_{\theta+\varepsilon_n}h(X) - \mathbb{E}_\theta h(X)}{\varepsilon_n}$.
8. Conclure.

lien vers l'indication

Chapitre 8

Loi d'un processus

8.1 Loi d'un processus

Définition: Un processus stochastique est une famille infinie $(X_t)_{t\in\mathbb{T}}$ de variables aléatoires définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$

Le plus souvent, T est un ensemble ordonné qui joue le rôle du temps, par exemple $T = \mathbb{N}, \mathbb{Z}, \mathbb{R}$. Le cas $T = \mathbb{Z}^d$, qui évoque plutôt une structure spatiale est également intéressant.

Définition: On appelle trajectoire de X tout élément $X(\omega) = (X_n(\omega), n \in \mathbb{T}), \omega \in \Omega$.

Définition: On définit la tribu borélienne sur $\mathbb{R}^{\mathbb{T}}$, notée $\mathcal{B}(\mathbb{R}^{\mathbb{T}})$, comme étant la plus petite tribu qui rend mesurable les projections $\Pi_i : \mathbb{R}^{\mathbb{T}} \longrightarrow \mathbb{R}, \omega = (\omega_n)_{n\in\mathbb{T}} \longmapsto \omega_i$.

On a vu au chapitre 4 que dans le cas où $\mathbb{T}$ est dénombrable, $\mathcal{B}(\mathbb{R}^{\mathbb{T}})$ coïncide avec la tribu borélienne de $\mathbb{R}^{\mathbb{T}}$. C'est encore vrai lorsque $\mathbb{T}$ est infini dénombrable, mais dans ce cas la topologie qui doit être mise sur $\mathbb{R}^{\mathbb{T}}$ (la topologie produit) n'est pas métrisable.

Définition: Pour toute parties non vides S et S' de $\mathbb{T}$ telle que $S \supseteq S'$, on appelle projection de $\mathbb{R}^S$ sur $\mathbb{R}^{S'}$ la fonction $\Pi^S_{S'} : \mathbb{R}^S \longrightarrow \mathbb{R}^{S'}$ définie par

$$\forall (x_s, s \in S) \in \mathbb{R}^S, \Pi^S_{S'}(x_s, s \in S) = (x_s, s \in S').$$

Définition: - théorème : loi d'un processus

Théorème 58. *Soit $(X_n)_{n\in\mathbb{T}}$ une suite de variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$. L'application $X : \omega \mapsto (X_n(\omega))_{n\in\mathbb{T}}$ est une application mesurable de $(\Omega, \mathcal{F})$ dans l'espace des trajectoires $(\mathbb{R}^{\mathbb{T}}, \mathcal{B}(\mathbb{R}^{\mathbb{T}}))$. La loi image $\mathbb{P}_X$ de $\mathbb{P}$ par X est appelée loi de la suite (ou du processus) $(X_n)_{n\in\mathbb{T}}$*

Démonstration. Comme les ensembles de la forme $\Pi_i^{-1}(A)$, avec A borélien, engendrent $\mathcal{B}$, il suffit de montrer que pour B de la forme $B = \Pi_i^{-1}(A)$, avec A borélien, on a $X^{-1}(B) \in \mathcal{F}$. Or

$$X^{-1}(B) = X^{-1}(\Pi_i^{-1}(A)) = (\Pi_i \circ X)^{-1}(A) = X_i^{-1}(A),$$

qui est bien dans $\mathcal{F}$ puisque X_i est une variable aléatoire. □

Notation : $\mathcal{P}^*(\mathbb{T}), \mathcal{F}(\mathbb{T}), \mathcal{D}(\mathbb{T})$ désignent respectivement l'ensemble des parties non vides, l'ensemble des parties finies non vides et l'ensemble des parties dénombrables non vides de $\mathbb{T}$.

Il est clair que $\mathcal{F}(\mathbb{T}) \subseteq \mathcal{D}(\mathbb{T}) \subseteq \mathcal{P}^*(\mathbb{T})$.

Définition: On appelle loi de dimension finie d'un processus $X = (X_n)_{n\in\mathbb{T}}$ la loi de tout vecteur extrait $(X_n)_{n\in S}$, où $S \in \mathcal{F}(\mathbb{T})$.

Proposition 5. *Les lois de dimension finie d'un processus $X = (X_n)_{n\in\mathbb{T}}$ où*

$$\forall n \in \mathbb{T}, X_n : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

sont les images de la loi $\mathbb{P}_X$ de X par les projections de $\mathbb{R}^{\mathbb{T}}$ sur les espaces-produits finis $\mathbb{R}^S, S \in \mathcal{F}(\mathbb{T})$.

Théorème 59. *Deux processus stochastiques $X = (X_n)_{n\in\mathbb{T}}$ et $X' = (X'_n)_{n\in\mathbb{T}}$ où*

$$\forall n \in \mathbb{T}, X_n : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

et

$$X'_n : (\Omega', \mathcal{F}', \mathbb{P}') \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

ont la même loi si et seulement s'ils ont les mêmes lois de dimension finie.

Démonstration. La condition nécessaire est évidente, d'après la proposition précédente. Réciproquement, si X et X' ont les mêmes lois de dimension finie, d'après la proposition précédente, $\mathbb{P}_X$ et $\mathbb{P}'_{X'}$ ont les mêmes images par projection sur les espaces-produits de dimension finie $(\mathbb{R}^S, \mathcal{B}(\mathbb{R}^S))$. On considère

$$\mathcal{C} = \{\prod_{n\in\mathbb{T}} A_n, \forall n, A_n \in \mathcal{B}(\mathbb{R}), \text{ et } \exists N, \forall n \geq N A_n = \mathbb{R}\}$$

$\mathbb{P}_X$ et $\mathbb{P}'_{X'}$ coïncident sur $\mathcal{C}$, c'est-à-dire X et X' ont même loi sur $\mathcal{C}$. $\mathcal{C}$ est un Π-système qui engendre $\mathcal{B}(\mathbb{R}^{\mathbb{T}})$, donc X et X' ont la même loi. □

Définition: On appelle processus canonique associé à un processus stochastique $X = (X_n, n \in \mathbb{T})$ où $\forall n \in \mathbb{T}, X_n : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ le processus $\Pi = (\Pi_n, n \in \mathbb{T})$ formé par les projections Π_n de $\mathbb{R}^{\mathbb{T}}$ sur $\mathbb{R}$, qui à $X = (X_k, k \in \mathbb{T})$ associe X_n, sa n-ième composante.

Théorème 60. *Tout processus stochastique a même loi que son processus canonique associé, quand on munit l'espace de ses trajectoires de la loi $\mathbb{P}_X$.*

Démonstration. $\Pi : (X_n)_{n\in\mathbb{T}} \longmapsto (X_n(\omega), \omega \in \Omega)_{n\in\mathbb{T}}$ est l'application identité de $\mathbb{R}^{\mathbb{T}}$. Donc l'image de $\mathbb{P}_X$ par Π est $\mathbb{P}_X$. □

8.2 Théorème d'existence de Kolmogorov

Définition: Système projectif

On considère une famille $(Q_S, S \in \mathcal{F}(\mathbb{T}))$ où pour tout S, Q_S désigne une probabilité sur $(\mathbb{R}^S, \mathcal{B}(\mathbb{R}^S))$. On dit que $(Q_S, S \in \mathcal{F}(\mathbb{T}))$ est un système projectif de lois si pour tous S, S' de $\mathcal{F}(\mathbb{T})$ tels que $S \supseteq S'$, $Q_{S'}$ est l'image de Q_S par $\Pi^S_{S'}$

Remarque: Si on a seulement défini Q_S pour des ensembles S de la forme $S = \{1, \dots, n\}$ et que l'on sait que pour tout $n \geq 1$, $Q_{\{1,\dots,n\}}$ est la mesure image de $Q_{\{1,\dots,n+1\}}$ par $\Pi^{\{1,\dots,n+1\}}_{\{1,\dots,n\}}$, alors on peut définir pour S partie finie de $\mathbb{N}$ une mesure Q_S comme étant la mesure image de $Q_{\{1,\dots,\max S\}}$ par $\Pi^{\{1,\dots,n\}}_S$. Il n'est alors pas difficile de vérifier que (Q_S) est un système projectif de lois.

Exemple: Si pour tout $i \in \mathbb{T}$, μ_i est une mesure de probabilité sur $\mathbb{R}$, la famille $(Q_S)_{S\in\mathcal{F}(T)}$ définie par $Q_S = \otimes_{i\in S}\mu_i$ est un système projectif de lois.

Théorème 61. *Théorème d'existence de Kolmogorov.*
On se place sur $(\mathbb{R}^{\mathbb{T}}, \mathcal{B}(\mathbb{R}^{\mathbb{T}}))$. Pour toute partie $S \in \mathcal{F}(\mathbb{T})$, soit Q_S une mesure de probabilité sur $(\mathbb{R}^S, \mathcal{B}(\mathbb{R}^S))$. Les trois conditions suivantes sont équivalentes :

1. *Il existe un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et pour tout $n \in \mathbb{T}$ une variable aléatoire $X_n : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que $(Q_S, S \in \mathcal{F}(\mathbb{T}))$ soit l'ensemble des lois de dimension finie du processus $X = (X_n, n \in \mathbb{T})$*
2. *Il existe une mesure de probabilité Q sur $(\mathbb{R}^{\mathbb{T}}, \mathcal{B}(\mathbb{R}^{\mathbb{T}}))$ dont l'image par Π_S soit Q_S quelle que soit la partie finie, non vide S de $\mathbb{T}$,*
3. *$(Q_S, S \in \mathcal{F}(\mathbb{T}))$ est un système projectif de lois.*

Démonstration. On va seulement donner la preuve dans le cas où $\mathbb{T} = \mathbb{N}$. Le cas où $\mathbb{T}$ est dénombrable s'en déduit immédiatement ; en revanche la preuve dans le cas général demanderait un argument supplémentaire.

(1)$\Longrightarrow$(2) : il suffit de prendre $Q = \mathbb{P}_X$.

(2)$\Longrightarrow$(3) : si $S' \subset S$, on a $\Pi'_S = \Pi^S_{S'} \circ \Pi_S$. $Q'_{S'}$ est la mesure image de Q par Π'_S, mais c'est aussi la mesure image par $\Pi^S_{S'}$ de la mesure image de Q par Π_S, soit donc la mesure image de Q_S par $\Pi^S_{S'}$, ce qui montre bien que le système (Q_S) est projectif

(3)$\Longrightarrow$(1) : c'est évidemment le gros morceau. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité sur lequel vit une suite $(U_n)_{n\geq 0}$ de variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. D'après le théorème 3, l'espace $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1[, \mathcal{B}([0,1[), \lambda|_{[0,1[})$ convient. Notons F^0 la fonction de répartition de Q_0, puis, pour $x \in \mathbb{R}^n$, notons F^n_x la fonction de répartition de la loi de Π_n sachant $\Pi_{\{0,\dots,n-1\}} = x$ sous $Q_{\{0,\dots,n\}}$. Ainsi, on a pour tout u réel

$$\mathbb{E}_{Q_{\{0,\dots,n\}}}[\Pi_n \leq u | \Pi_{\{0,\dots,n-1\}}] = F^n_{\Pi_{\{0,\dots,n-1\}}}(u).$$

On pose encore $Q_0^*(u) = \min\{y \in \mathbb{R} : \ 1 - F^0(y) \leq u\}$ et

$$\forall x \in \mathbb{R}^n \quad \forall u \in \mathbb{R} \quad Q_n^*(x,u) = \min\{y \in \mathbb{R} : \ 1 - F^n_x(y) \leq u\},$$

puis on définit $(X_n)_{n\geq 0}$ par $X_0 = Q_0^*(U_0)$ et pour $n \geq 1$:

$$X_n = Q_n^*((X_0, \dots, X_{n-1}), U_n).$$

Montrons par récurrence que pour tout n, la loi de $(X_0, \dots, X_n)$ est $Q_{\{0,\dots,n\}}$. Pour $n = 0$, c'est une conséquence immédiate du théorème 38. Sinon, supposons que $(X_0, \dots, X_{n-1})$ a comme loi $Q_{\{0,\dots,n-1\}}$: comme le système est projectif $(X_0, \dots, X_{n-1})$ réalise la loi des n premières composantes de $Q_{\{0,\dots,n\}}$. Comme $(X_0, \dots, X_{n-1})$ est indépendant de U_n, le théorème de l'échantillonneur de Gibbs dit que $(X_0, \dots, X_{n-1}, \varphi_n((X_0, \dots, X_{n-1}), U_n)) = (X_0, \dots, X_{n-1}, X_n)$ suit la loi $Q_{\{0,\dots,n\}}$.

□

Au cours de la preuve, on a montré en particulier que l'espace $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1[, \mathcal{B}([0,1[), \lambda|_{[0,1[})$ est suffisamment gros pour y faire vivre tous les processus réels à temps discret que l'on peut imaginer. [1]

Corollaire 15. *Pour tout entier $n \geq 1$, soit $\mathbb{P}_n$ une mesure de probabilité sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Pour qu'il existe une suite $X = (X_n, n \geq 1)$ de variables aléatoires réelles simultanées telle que $\mathbb{P}_n$ soit la loi de $(X_1, X_2, ..., X_n)$ pour tout $n \geq 1$ il faut et il suffit que*

$$\forall B \in \mathcal{B}(\mathbb{R}^n), \mathbb{P}_{n+1}(B \times \mathbb{R}) = \mathbb{P}_n(B)(*).$$

1. L'existence de la loi de Π_n sachant $\Pi_{\{0,\dots,n-1\}} = x$ sous $Q_{\{0,\dots,n\}}$ repose sur le théorème 36 d'existence des lois conditionnelles. Énoncé dans le cadre des espaces polonais, ce théorème n'a été démontré dans ce cours que dans le cas $\Omega = [0,1]^n$. C'est suffisant pour faire vivre une suite de variables aléatoires à support dans $[0,1]$ dont les lois finidimensionnelles sont prescrites. Mais le cas d'une suite de variables aléatoires réelles quelconques s'en déduit : on commence par construire la suite $\frac{2}{\pi}\arctan X_0, \dots, \frac{2}{\pi}\arctan X_n, \dots$ sur $(\Omega, \mathcal{F}, \mathbb{P})$, puis $X_0, X_1, \dots, X_n, \dots$ en composant avec la fonction $x \mapsto \tan(\pi/2x)$.

Démonstration. En effet, s'il existe une telle suite X de variables aléatoires réelles simultanées, on a

$$\forall B \in \mathcal{B}(\mathbb{R}^n), \mathbb{P}_{n+1}(B\times\mathbb{R}) = \mathbb{P}((X_1, ..., X_{n+1}) \in B\times\mathbb{R}) = \mathbb{P}((X_1, ..., X_n) \in B) = \mathbb{P}_n(B).$$

Réciproquement, on suppose la condition $(*)$ satisfaite. Pour toute partie non vide S de $\mathbb{N}$ de cardinal p, notée $S = \{s_1, \dots, s_p\}$, on note $\mathbb{P}_S$ l'image de $\mathbb{P}_{\max(S)}$ par la projection $\Pi_S^{\{1,\dots,\max(S)\}}$. D'après le théorème d'existence de Kolmogorov, il suffit de vérifier que $(\mathbb{P}_S, S \in \mathcal{F}(T))$ est un système projectif de loi. Pour cela, on vérifie que pour tout S de $\mathcal{F}(T)$, et tout $n \geq \max(S)$, $\mathbb{P}_S$ est l'image de $\mathbb{P}_n$ par $\Pi_S^{\{1,\dots,n\}}$, ce qui se fait facilement par récurrence. □

8.2.1 Loi produit infini ; variables indépendantes

On a déjà remarqué que lorsque pour tout $i \in \mathbb{T}$, μ_i est une mesure de probabilité sur $\mathbb{R}$, la famille $(Q_S)_{S\in\mathcal{F}(T)}$ définie par $Q_S = \otimes_{i\in S}\mu_i$ est un système projectif de lois. D'après le théorème de Kolmogorov, il existe une loi sur $\mathbb{R}^{\mathbb{T}}$ dont la projection sur un ensemble fini d'indices S quelconques est $\otimes_{i\in S}\mu_i$: on notera désormais $\otimes_{i\in T}\mu_i$ cette loi.

Si on note $(X_i)_{i\in S}$ la famille des projections sur les différents indices, il est alors clair que sous $\otimes_{i\in T}\mu_i$, les variables X_i sont des variables aléatoires indépendantes, et que pour tout i, la loi de X_i est μ_i.

Ainsi, étant donnée une suite $(\mu_n, n \geq 1)$ de mesures de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ il existe toujours une suite de variables aléatoires réelles simultanées indépendantes $(X_n, n \geq 1)$ telle que $\forall n \geq 1, \mathbb{P}_{X_n} = \mu_n$.

8.2.2 Loi markovienne

Soit D un ensemble dénombrable, $P = (p_{i,j})_{(i,j)\in D^2}$ une matrice markovienne, c'est à dire que $p_{i,j} \geq 0$ pour tout couple (i, j) et $\sum_{k\in D} p_{i,k} = 1$ pour tout $i \in D$. Alors, pour toute loi μ sur D, on peut construire une unique loi $\mathbb{P}^\mu$ sur $D^{\mathbb{N}}$ telle que pour tout entier $n \geq 0$ et toute suite $x_0, \dots x_n$ d'éléments de D, on ait

$$\mathbb{P}^\mu(\Pi_0 = x_0, \dots, \Pi_n = x_n) = \mu(x_0) \prod_{i=0}^{n-1} p_{x_i,x_{i+1}}.$$

Démonstration. On définit par récurrence une suite de mesures $(\mathbb{P}_n)_{n\geq 0}$ avec $\mathbb{P}_0 = \mu$, puis

$$\mathbb{P}_{n+1}(\{(x_0, \dots, x_{n+1})\}) = \mathbb{P}_n(\{(x_0, \dots, x_n)\})p_{x_n,x_{n+1}}.$$

Soit $A \subset D^{n+1}$. On a

$$\begin{aligned}
\mathbb{P}_{n+1}(A \times D) &= \sum_{(x_0,\dots x_n)\in A} \sum_{x_{n+1}\in D} \mathbb{P}_{n+1}(\{(x_0,\dots,x_{n+1})\}) \\
&= \sum_{(x_0,\dots x_n)\in A} \sum_{x_{n+1}\in D} \mathbb{P}_{n}(\{(x_0,\dots,x_{n})\}) p_{x_n,x_{n+1}} \\
&= \sum_{(x_0,\dots x_n)\in A} \mathbb{P}_{n}(\{(x_0,\dots,x_{n})\}) 1 \\
&= \mathbb{P}_n(A)
\end{aligned}$$

L'identité qu'on vient de montrer permet de montrer par récurrence que $\mathbb{P}_n(D^{n+1}) = 1$. Elle exprime également que la loi de $(\Pi_0, \dots, \Pi_n)$ sous $\mathbb{P}_{n+1}$ est $\mathbb{P}_n$: on a donc un système projectif de lois, ce qui permet d'appliquer le (corollaire du) théorème d'existence de Kolmogorov. □

8.3 Processus réels stationnaires (temps discret)

On suppose ici que $\mathbb{T} = \mathbb{N}$ ou $\mathbb{T} = \mathbb{Z}$.

Définition: Un processus stochastique réel $(X_n, n \in \mathbb{T})$ est dit stationnaire si quels que soient les entiers $d \geq 1, n_1, \dots, n_d$ choisis dans $\mathbb{N}$, tels que $n_1 < \dots < n_d$, les vecteurs aléatoires réels d-dimensionnels $(X_{n_1}, \dots, X_{n_d})$ et $(X_{n_1+1}, \dots, X_{n_d+1})$ suivent la même loi.

Il en résulte évidemment que $(X_{n_1}, \dots, X_{n_d})$ et $(X_{n_1+h}, \dots, X_{n_d+h})$ suivent la même loi quel que soit l'entier $h \geq 0$.

Définition: Étant donné un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, on dit qu'une application $T : (\Omega, \mathcal{F}) \longrightarrow (\Omega, \mathcal{F})$ conserve la mesure $\mathbb{P}$ si $\forall A \in \mathcal{F}, \mathbb{P}(A) = \mathbb{P}(T^{-1}(A))$ ou si $\mathbb{P}$ est sa propre image par T.

Définition: On dit que $T : \Omega \longrightarrow \Omega$ est une bijection bimesurable si c'est une bijection $(\mathcal{F}, \mathcal{F})$-mesurable et si l'application réciproque T^{-1} est $(\mathcal{F}, \mathcal{F})$-mesurable .

Théorème 62. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et une application*

$$T : (\Omega, \mathcal{F}) \longrightarrow (\Omega, \mathcal{F}).$$

- *Si T conserve la mesure $\mathbb{P}$, pour tout entier $n \geq 1$, T^n conserve $\mathbb{P}$*
- *Si T est une bijection bimesurable qui conserve $\mathbb{P}$, T^{-1} conserve $\mathbb{P}$.*

Démonstration. — On raisonne par récurrence. Au rang initial, $\forall A \in \mathcal{F}, \mathbb{P}(A) = \mathbb{P}(T^{-1}(A))$ parce que T conserve $\mathbb{P}$. Si pour un entier $n \geq 1$, on a démontré que

$$\forall A \in \mathcal{F}, \mathbb{P}(A) = \mathbb{P}(T^{-n}(A)),$$

alors comme

$$\mathbb{P}(T^{-n}(A)) = \mathbb{P}(T^{-1}(T^{-n}(A))) = \mathbb{P}(T^{-(n+1)}(A)),$$

et il vient

$$\forall A \in \mathcal{F}, \mathbb{P}(A) = \mathbb{P}(T^{-(n+1)}(A)).$$

D'où la conclusion par récurrence : pour tout $n \geq 1$, T^n conserve $\mathbb{P}$.

— $\forall A \in \mathcal{F}, \mathbb{P}(T(A)) = \mathbb{P}(T^{-1}(T(A))) = \mathbb{P}(A)$.

□

Théorème 63. *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace de probabilité et* $\mathcal{C}$ *un* Π*-système d'événements engendrant* $\mathcal{F}$*. Alors une application* $T : (\Omega, \mathcal{F}) \longrightarrow (\Omega, \mathcal{F})$ *conserve la mesure* $\mathbb{P}$ *si et seulement si*

$$\forall A \in \mathcal{C}, \mathbb{P}(A) = \mathbb{P}(T^{-1}(A)).$$

Démonstration. Le sens direct est évident.
Réciproquement, on suppose que $\mathbb{P}$ et son image par T coïncident sur $\mathcal{C}$, donc sur $\mathcal{F}$. □

Définition: Pour $E = \mathbb{N}$ ou $E = \mathbb{Z}$, on notera θ l'application de $\mathbb{R}^E$ dans $\mathbb{R}^E$ appelée opérateur de translation définie par

$$\theta((x_n)_{n \in E}) = (y_n)_{n \in E},$$

où $\forall n \in E, y_n = x_{n+1}$.

Théorème 64. *Pour qu'un processus réel* $X = (X_n, n \in \mathbb{T})$ *soit stationnaire, il faut et il suffit que l'opérateur de translation* θ *conserve sa loi* $\mathbb{P}_X$*.*

Démonstration. Dire que θ préserve $\mathbb{P}_X$, c'est dire que X et $\theta \circ X$ ont même loi. Mais on sait que deux lois sur $\mathbb{R}^{\mathbb{T}}$ sont égales si et seulement si toutes les projections par les $\Pi_{s_1, \dots s_n}$ sont égales. Ainsi θ préserve $\mathbb{P}_X$ si et seulement si quels que soient $s_1, \dots s_n$, $\Pi_{s_1, \dots s_n} \circ X$ et $\Pi_{s_1, \dots s_n} \circ \theta \circ X$ ont même loi sous $\mathbb{P}$. Or la loi de $\Pi_{s_1, \dots s_n} \circ X$ sous $\mathbb{P}$ est $\mathbb{P}_{(X_{s_1}, \dots, X_{s_n})}$ et celle de $\Pi_{s_1, \dots s_n} \circ \theta \circ X$ sous $\mathbb{P}$ est $\mathbb{P}_{(X_{s_1+1}, \dots, X_{s_n+1})}$. Ainsi, par définition de la stationnarité, θ préserve $\mathbb{P}_X$ si et seulement si (X_n) est stationnaire. □

Les processus stationnaires fournissent ainsi un exemple fondamental de transformation conservant la mesure.

Théorème 65. *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace de probabilité et une application*

$$T : (\Omega, \mathcal{F}) \longrightarrow (\Omega, \mathcal{F})$$

qui conserve la mesure $\mathbb{P}$. *Alors*

— *quelle que soit la variable aléatoire* ξ *sur* $(\Omega, \mathcal{F}, \mathbb{P})$, $(\xi \circ T^n, n \geq 0)$ *est un processus stationnaire.*

— *si* T *est une bijection bimesurable de* $(\Omega, \mathcal{F})$, $(\xi \circ T^n, n \in \mathbb{Z})$ *est également un processus stationnaire.*

Démonstration. Prenons, suivant le cas $\mathbb{T} = \mathbb{N}$ ou $\mathbb{T} = \mathbb{Z}$ et prenons $n \geq 1$, puis $s_1 < \cdots < s_n$ dans $\mathbb{T}$. Considérons le vecteur aléatoire à valeurs dans $\mathbb{R}^n$:

$$V = (\xi \circ T^{s_1}, \ldots, \xi \circ T^{s_n}) : (\Omega, \mathcal{F}, \mathbb{P}) \quad \to \quad (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)).$$

La loi de $V \circ T$ sous $\mathbb{P}$ est la loi de V sous $\mathbb{P}_T$, mais $\mathbb{P}_T = \mathbb{P}$, donc V et $V \circ T$ ont même loi. Cependant, que $\mathbb{T} = \mathbb{N}$ ou $\mathbb{T} = \mathbb{Z}$, on a dans les deux cas :

$$V \circ T = (\xi \circ T^{s_1+1}, \ldots, \xi \circ T^{s_n+1}).$$

On vient de montrer que $(\xi \circ T^{s_1}, \ldots, \xi \circ T^{s_n})$ et $(\xi \circ T^{s_1+1}, \ldots, \xi \circ T^{s_n+1})$ ont même loi. Comme c'est vrai pour tout n et pour $s_1, \ldots, s_n$ quelconques, on vieut précisément de montrer que le processus $(\xi \circ T^n)_{n \in \mathbb{T}}$ est stationnaire. □

8.4 Processus gaussiens

8.4.1 Caractérisation

Définition: On dit qu'un processus $(X_t)_{t \in T}$ est gaussien si pour tout $S \in \mathcal{F}(T)$, le vecteur $(X_s)_{s \in S}$ est gaussien.

À tout processus gaussien, on peut associer son espérance $(\mathbb{E}X_t)_{t \in T}$ et sa fonction de covariance

$$C_X : (s, t) \mapsto \mathbb{E}(X_t - \mathbb{E}X_t)(X_s - \mathbb{E}X_s)).$$

Proposition 6. *Deux processus gaussiens ont même espérance et même fonction de covariance si et seulement si ils ont même loi.*

Démonstration. Notons (X_s) et (Y_s) les deux processus considérés.

— Le sens « même loi implique même espérance, même covariance »est "presque" évident. Arrétons nous y tout de même quelques instants. On a
$$\mathbb{E}X_s = \int_{\mathbb{R}^T} \omega_s \; d\mathbb{P}_X(\omega)$$
et
$$\mathbb{E}Y_s = \int_{\mathbb{R}^T} \omega_s \; d\mathbb{P}_Y(\omega).$$
Dire que (X_s) et (Y_s) ont même loi, c'est précisément dire que $\mathbb{P}_X = \mathbb{P}_Y$. Cela implique donc qu'ils ont les mêmes espérances. Les identités
$$\mathbb{E}X_sX_t = \int_{\mathbb{R}^T} \omega_s\omega_t \; d\mathbb{P}_X(\omega)$$
et
$$\mathbb{E}Y_sY_t = \int_{\mathbb{R}^T} \omega_s\omega_t \; d\mathbb{P}_Y(\omega)$$
permettent alors de compléter la preuve.

— Soit $F \subset T$, F fini. Les vecteurs $(X_s)_{s\in S}$ et $(Y_s)_{Y\in S}$ sont gaussiens. Par hypothèse, ils ont même espérance et même matrice de covariance. Des vecteurs gaussiens qui ont même espérance et même matrice de covariance ont même loi. Ainsi (X_s) et (Y_s) ont mêmes lois de dimension finie. Ils ont donc la même loi.

□

8.4.2 Condition d'existence

Théorème 66. *Soit $(m_t)_{t\in T}$ et $(c_{s,t})_{(s,t)\in T\times T}$ des réels. Il existe un processus gaussien de moyenne $(m_t)_{t\in T}$ et de covariance $(c_{s,t})_{(s,t)\in T\times T}$ si et seulement si*

— *Pour tous $s,t \in T$, on a $c_{s,t} = c_{t,s}$.*

— *Four tous S fini inclus dans T et tout $x \in \mathbb{R}^T$, on a*
$$\sum_{(s,t)\in S\times S} c_{s,t}(x_s - m_s)(x_t - m_t) \geq 0.$$

Démonstration. La nécessité des deux conditions provient du fait que $(c_{s,t})_{(s,t)\in S\times S}$ doit être la matrice de covariance du vecteur $(X_s)_{s\in S}$ Pour voir que ces conditions sont suffisantes, il suffit d'appliquer le théorème de Kolmogorov à la famille de mesures $\mathcal{N}(m_S, C_S)$ où $m_S = (m_t)_{t\in S}$ et $C_S = (c_{s,t})_{(s,t)\in S\times S}$ qui est compatible. □

8.4.3 Processus gaussiens stationnaires

Théorème 67. *Soit $(X_n)_{n\in\mathbb{Z}}$ un processus gaussien. $(X_n)_{n\in\mathbb{Z}}$ est stationnaire si et seulement si il existe une constante m et une fonction φ telle que*

- *Pour tout n $\mathbb{E}X_n = m$*
- *Pour tous n,p entiers on a $\mathbb{E}(X_n - m)(X_p - m) = \varphi(n-p)$.*

φ est appelée fonction d'autocovariance du processus.

Démonstration. Supposons que le processus est stationnaire et posons $m = \mathbb{E}X_0$ et $\varphi(n) = \mathbb{E}(X_n - m)(X_0 - m)$. Pour tout n X_0 et X_n ont même loi, donc $\mathbb{E}X_n = \mathbb{E}X_0 = m$. D'autre part, le couple (X_n, X_p) a même loi que le couple (X_{n-p}, X_0) : on a donc $\mathbb{E}(X_n - m)(X_p - m) = \mathbb{E}(X_{n-p})(X_0 - m)) = \varphi(n-p)$. Réciproquement, supposons que pour tout n $\mathbb{E}X_n = m$ et que pour tous n,p entiers on a $\mathbb{E}(X_n - m)(X_p - m) = \varphi(n-p)$. Il faut démontrer que le processus $(X_{n+1})_{n\in\mathbb{Z}}$ a même loi que le processus $(X_n)_{n\in\mathbb{Z}}$. Ces deux processus étant gaussiens, ils suffit de montrer qu'ils ont même espérance et même covariance. Or on a pour tout n :$\mathbb{E}X_{n+1} = m = \mathbb{E}X_n$ et pour tous n,p

$$\begin{aligned}
\mathbb{E}(X_{n+1} - \mathbb{E}X_{n+1})(X_{p+1} - \mathbb{E}X_{p+1}) &= \mathbb{E}(X_{n+1} - m)(X_{p+1} - m) \\
&= \varphi((n+1) - (p+1)) \\
&= \varphi(n-p) \\
&= \mathbb{E}(X_n - m)(X_p - m) \\
&= \mathbb{E}(X_n - \mathbb{E}X_n)(X_p - \mathbb{E}X_p),
\end{aligned}$$

ce qui achève la preuve. □

8.5 Exercices sur les processus

8.5.1 Exercices corrigés

Exercice 54. Soit $D \subset \mathbb{R}$ un ensemble dénombrable, $P = (p_{i,j})_{(i,j)\in D^2}$ une matrice markovienne. Pour μ mesure de probabilité sur D, on note $\mathbb{P}^\mu$ la loi markovienne associée. Si $i \in D$, on note simplement $\mathbb{P}^i$ pour $\mathbb{P}^{\delta_i}$.

1. Démontrer la propriété de Markov : pour toute mesure μ, pour tout entier n, pour tout $A \in \sigma(\Pi_0, \dots, \Pi_n)$, pour tout $B \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$, on a
$$\mathbb{P}^\mu(A, \Pi_n = i, \theta^{-n}(B)) = \mathbb{P}^\mu(A, \Pi_n = i)\mathbb{P}^i(B).$$
2. Montrer que $\mathbb{P}^\mu = \sum_{i\in D} \mu(i)\mathbb{P}^i$.
3. Montrer que si $\mu(j) = \sum_i \mu(i)p_{i,j}$ pour tout j, alors $\mathbb{P}^\mu$ est laissée invariante par θ.

lien vers l'indication lien vers la solution

Exercice 55. Soit φ une application mesurable de $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et $(X_n)_{n\geq 0}$ un processus stationnaire. Démontrer que le processus $(Y_n)_{n\geq 0}$ défini par $Y_n = \varphi(X_n, X_{n+1}, \dots) = \varphi(\theta^n \circ X)$ est stationnaire. lien vers l'indication lien vers la solution

Exercice 56. On dit d'une famille de variables $(X_n)_{n\geq 1}$ définies sur un espace $(\Omega, \mathcal{F}, \mathbb{P})$ quelconque qu'elles sont échangeables si pour tout n et pour tout $\sigma \in \mathcal{S}_n$, les vecteurs $(X_1, \dots, X_n)$ et $(X_{\sigma(1)}, \dots, X_{\sigma(n)})$ ont même loi (sous $\mathbb{P}$).

1. Montrer qu'une famille de variables échangeables est stationnaire.
2. Soit $(X_n)_{n\geq 1}$ une famille de variables échangeables de carré intégrables. Exprimer la variance de $X_1 + \cdots + X_n$ en fonction de $\operatorname{Var} X_1$ et $\operatorname{Covar}(X_1, X_2)$. En déduire que les X_i sont positivement corrélés.
3. Montrer qu'à partir d'un bruit blanc (une famille $(Y_i)_{i\geq 0}$ de variables indépendantes suivant la loi $\mathcal{N}(0, 1)$), on peut fabriquer (la loi de) n'importe quel processus gaussien de variables échangeables en posant $X_i = m + aY_0 + bY_i$.
4. Soit $(X_n)_{n\geq 1}$ un processus gaussien de variables échangeables. Montrer qu'il existe une variable aléatoire Z telle que, sachant Z, les variables aléatoires X_i, sont des variables aléatoires indépendantes. Ce résultat constitue un cas particulier d'un résultat plus général, le théorème de De Finetti–Hewitt–Savage, qui sera proposé un peu plus loin en exercice.

lien vers l'indication lien vers la solution

Exercice 57. *Le théorème des quatre couleurs stochastique, d'après Holroyd et Liggett*
Soit q un entier naturel non nul. On appelle mot ou coloriage sur l'alphabet $\{1,\dots,q\}$ une suite finie $x=(x_1,\dots,x_n)$ d'éléments de $\{1,\dots,q\}$. Le mot vide est l'unique mot de longueur 0. On dit qu'un mot $x=(x_1,\dots,x_n)$ est un mot propre si $x_i\neq x_{i+1}$ pour $1\le i<n$. On convient que le mot vide est un coloriage propre.

Le but de ce problème est de construire et d'étudier des processus stochastiques $(\Pi_n)_{n\ge 1}$ qui sont tels que

— Pour tout n, $(\Pi_1,\dots,\Pi_n)$ est un coloriage propre sur l'alphabet $\{1,\dots,q\}$
— $(\Pi_n)_{n\ge 1}$ est stationnaire

Si x et y sont deux mots, on note $x.y$ leur concaténation. Si $x=(x_1,\dots,x_n)$ est un mot et i un entier compris entre 1 et n, $\hat{x}_i$ désigne le mot x dont on a oté la i-ème lettre. Ainsi $(1,2,4).(7,5)=(1,2,4,7,5)$ et $\widehat{(1,2,7,8)}_3=(1,2,8)$.

On doit encore introduire la notion d'immeuble. Soit x un mot de taille n. Un immeuble propre de dernier étage x est une suite $(y_1,\dots y_n)$ de mots tels que

— Pour tout i entre 1 et n, y_i est un coloriage propre de taille i à q couleurs.
— Pour $1\le i<n$, y_i est obtenu en enlevant une lettre au mot y_{i+1}.
— $y_n=x$

Ainsi $((1),(1,2),(2,1,2))$ est un immeuble propre de dernier étage $(2,1,2)$. On note $B(x)$ l'ensemble des immeubles propres de dernier étage x.

1. Montrer que pour tout mot propre x de taille $n\ge 0$, on a
$$|B(x)|=\sum_{i=1}^n |B(\hat{x}_i)|.$$
2. Montrer par récurrence que pour tout $n\ge 0$ et tout mot x de longueur n, on a
$$\sum_{a\in\{1,\dots,q\}} |B(x.a)|=b_n(q)|B(x)|,\ \text{avec } b_n(q)=n(q-2)+q.$$
3. On note $S(q,n)$ le nombre d'immeubles propres de n étages. Montrer que $S(q,n+1)=b_n(q)S(q,n)$.
4. Montrer que la formule
$$\pi_n(\{x_1,\dots,x_n\})=\frac{|B((x_1,\dots,x_n))|}{S(q,n)}$$
définit une mesure de probabilité sur $\{1,\dots,q\}^n$.

5. À l'aide du théorème d'extension de Kolmogorov, démontrer qu'il existe une mesure de probabilité $\mathbb{P}_q$ sur $\{1,\dots,q\}^{\mathbb{N}_*}$ telle que pour tout entrier n et tout $x=(x_1,\dots,x_n)\in\{1,\dots,q\}^{\mathbb{N}}$, on ait
$$\mathbb{P}_q(\Pi_1=x_1,\dots\Pi_n=x_n)=\frac{|B((x_1,\dots,x_n))|}{S(q,n)}.$$
6. Montrer que $\mathbb{P}_q$ est réversible, c'est à dire que pour tout $n\geq 1$ et tout $x=(x_1,\dots,x_n)\in\{1,\dots,q\}^{\mathbb{N}}$, on a
$$\mathbb{P}_q(\Pi_1=x_1,\dots\Pi_n=x_n)=\mathbb{P}_q(\Pi_1=x_n,\dots\Pi_n=x_1).$$
7. En déduire que $\mathbb{P}_q$ est invariante par le décalage θ.
8. On s'intéresse maintenant au cas où $q=4$.
   (a) Montrer qu'il existe des constantes $(c_{n,p})_{n\geq 0,p\geq 0}$ telles que pour tout $n,p\geq 0$, pour tout $x\in\{1,\dots,q\}^n$ et tout $y\in\{1,\dots,q\}^p$, on ait
$$\sum_{a\in\{1,\dots,q\}}|B(x.a.y)|=c_{n,p}|B(x)|.|B(y)|.$$
   (b) En déduire que sous $\mathbb{P}_4$, $(\Pi_1,\dots,\Pi_n)$ est indépendant de $(\Pi_{n+2},\dots,\Pi_{n+p+1})$.

lien vers l'indication lien vers la solution

8.5.2 Exercices non corrigés

Exercice 58. *Coloriages propres : un théorème de Schramm, par la méthode de Fuxi Zhang*

Soit $\Omega=\{1,\dots,q\}^{\mathbb{N}}$, $\mathcal{F}=\mathcal{B}(\Omega)$, et $\mathbb{P}$ une probabilité sur Ω, invariante par le décalage θ et ne chargeant que des coloriages propres (voir l'exercice précédent). On suppose de plus que $\mathbb{P}$ est 1-dépendant, c'est à dire que pour tout entier naturel n, $\sigma(\Pi_i,i<n)$ est indépendante sous $\mathbb{P}$ de $\sigma(\Pi_i,i>n)$, où Π_i est l'opérateur de projection canonique : $\Pi_i(\omega)=\omega_i$). On suppose que la couleur c vérifie $p=\mathbb{P}(\omega_0=c)>0$. On note $\overline{\mathbb{P}}=\mathbb{P}(\cdot|\Pi_0=c)$ et $\overline{\mathbb{E}}$ l'intégrale sous $\overline{\mathbb{P}}$.

1. Posons, pour $s\in B(0,1)$: $S(s)=\sum_{i=0}^{+\infty}\mathbb{1}_{\{\Pi_i=c\}}s^i$. Montrer que
$$\overline{\mathbb{E}}(S(s))=\frac{1-s+ps}{1-s}.$$
2. Posons $T(\omega)=\inf\{n\geq 1;\omega_n=c\}$.
   Notons $\tilde{\theta}$ l'opérateur de Ω dans lui même défini par
$$\tilde{\theta}(x)=\begin{cases}\theta^{T(x)}(x) & \text{si } T(x)<+\infty\\ x & \text{sinon.}\end{cases}.$$

Montrer que pour tout $n \geq 1$ et $A \in \mathcal{F}$, on a

$$\overline{\mathbb{P}}(T = n, \tilde{\theta}^{-1}(A)) = \overline{\mathbb{P}}(T = n)\overline{\mathbb{P}}(A).$$

3. En déduire que les variables $(T \circ \tilde{\theta}^n)_{n \geq 0}$ sont indépendantes.
4. On note G_T la fonction génératrice de T sous $\overline{\mathbb{P}}$: pour $s \in B(0,1)$, $G_T(s) = \overline{\mathbb{E}}[s^T]$. En remarquant que $S(s) = \sum_{k=0}^{+\infty} s^{N_k}$, avec $N_0 = 0$ et $N_k = \sum_{i=0}^{k-1} T \circ \tilde{\theta}^i$, montrer que $\overline{\mathbb{E}}(S(s)) = \frac{1}{1-G_T(s)}$, puis que $G_T(s) = \frac{ps^2}{1-s+ps^2}$
5. On rappelle le théorème de Pringsheim–Hille : si une série entière à coefficients positifs a un rayon de convergence $R < +\infty$, alors la fonction somme n'admet de prolongement analytique sur aucun voisinage de R. À la lumière de ce résultat, montrer que $p \leq \frac{1}{4}$.
6. En déduire qu'il n'existe aucun champ stationnaire 1-dépendant de coloriages propres à 3 couleurs.

lien vers l'indication

Exercice 59. Soient $(X_n)_{n\geq 1}$, $(Y_n)_{n\geq 1}$ deux processus stationnaires indépendants. Montrer que pour toute application mesurable φ de $\mathbb{R} \times \mathbb{R}$ dans $\mathbb{R}$, le processus $\varphi(X_n, Y_n)$ est stationnaire. Donner un exemple de processus $(X_n)_{n\geq 1}$ et $(Y_n)_{n\geq 1}$ stationnaires tels que $X_n + Y_n$ ne soit pas stationnaire. lien vers l'indication

Exercice 60. Soit $(X_n)_{n\geq 1}$ un processus stationnaire. Montrer que le processus $(Y_n)_{n\geq 1}$ défini par $Y_n = X_n + 2X_{n+1}$ est stationnaire. lien vers l'indication

Exercice 61. Soit $(X_n)_{n\geq 0}$ une chaîne de Markov homogène dont l'espace d'état est fini. Montrer que $(X_n)_{n\geq 0}$ est stationnaire si et seulement si X_0 et X_1 ont même loi. lien vers l'indication

Exercice 62. Soit X_0 une variable aléatoire suivant la loi uniforme sur $\mathbb{Z}/7\mathbb{Z}$. Soit $(T_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes identiquement distribuées suivant la loi uniforme sur l'ensemble $\{-1, 1\}$. On définit par récurrence une suite $(T_n)_{n\geq 1}$ par $X_{n+1} = X_n + T_{n+1}$.On pose enfin $Z_n = \inf\{k \geq 0; X_{n+k} = 0\}$. Montrer que $(Z_n)_{n\geq 0}$ est un processus stationnaire. lien vers l'indication

Exercice 63. Montrer que l'application de $[0, 1]$ dans lui-même qui à x associe la partie fractionnaire de $2x$ laisse invariante la mesure de Lebesgue sur $[0, 1]$. lien vers l'indication

Exercice 64. Soit α un réel. Montrer que l'application de $[0,1]$ dans lui-même qui à x associe la partie fractionnaire de $x+\alpha$ laisse invariante la mesure de Lebesgue sur $[0,1]$. Comment interpréter ce résultat si l'on identifie $[0,1[$ au cercle unité par l'application $x \mapsto e^{2i\pi x}$? lien vers l'indication

Exercice 65. On appelle bruit blanc une suite $(Z_n)_{n\in\mathbb{Z}}$ de variables aléatoires indépendantes suivant la loi $\mathcal{N}(0,1)$. Soit $(Z_n)_{n\in\mathbb{Z}}$ un bruit blanc et $\beta = (\beta_0,\dots,\beta_q) \in \mathbb{R}^{q+1}$ avec $\beta_0 \neq 0$ et $\beta_q \neq 0$. On considère la moyenne mobile :

$$X_n = \sum_{k=0}^{q} \beta_k Z_{n-q}.$$

Démontrer que $(Z_n)_{n\in\mathbb{Z}}$ est un processus stationnaire dont on calculera la fonction d'autocovariance. lien vers l'indication

Chapitre 9

Chaînes de Markov

9.1 Définition et caractérisations

9.1.1 Définition

Soit S un ensemble fini ou dénombrable, ν une mesure de probabilité sur S et $P = (p_{i,j})_{(i,j)\in S\times S}$ une matrice à coefficients positifs. Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires définies sur un espace $(\Omega, \mathcal{F}, \mathbb{P})$. On dit que la suite $(X_n)_{n\geq 0}$ est une *chaîne de Markov* de loi initiale ν et de matrice de passage P si l'on a, pour tout entier $n \geq 1$ et toute suite $x_0, \dots x_n$ d'éléments de S :

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots X_n = x_n) = \nu(x_0) \prod_{i=0}^{n-1} p_{x_i, x_{i+1}}.$$

Exemple : une suite $(X_n)_{n\geq 0}$ de variables aléatoires indépendantes de même loi ν à valeurs dans S dénombrable est une chaîne de Markov. En effet, il suffit de poser pour $(i,j) \in S \times S$ $p_{i,j} = \nu(j)$.

9.1.2 Caractérisation par l'espérance conditionnelle

Théorème 68. *Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires à valeurs dans S. Les trois propriétés suivantes sont équivalentes :*

1. *$(X_n)_{n\geq 0}$ est une chaîne de Markov de matrice de passage P à valeurs dans S*
2. *Quels que soient $x_0, \dots, x_{n-1}$ dans S tels que $\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots X_{n-1} = x_{n-1}) > 0$, alors $\mathbb{P}(X_n = x_n | X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}) = p_{x_{n-1}, x_n}$.*

3.

$$\mathbb{P}(X_n = x_n | X_0, \dots, X_{n-1}) = p_{X_{n-1},x_n}. \tag{9.1}$$

Cela signifie que toute l'information que $X_0, \dots, X_{n-1}$ peuvent nous apporter sur X_n est comprise dans X_n. Remarque : (9.1) implique que $\mathbb{P}(X_n = x_n | X_{n-1}) = p_{X_{n-1},x_n}$

9.1.3 Dynamique markovienne

Qu'est ce concrètement, qu'une chaîne de Markov ? On va voir que c'est une suite de réalisations, au cours du temps, des états d'un système soumis à des transformations aléatoires, la suite des transformations est une suite de transformations indépendantes, de même loi. Évidemment, le résultat de la transformation dépend de la transformation choisie et de l'état du système avant la transformation.

Si $(\Omega, \mathcal{F})$ est un espace mesuré, on appelle "fonction aléatoire" toute application mesurable de $(\Omega, \mathcal{F})$ dans $(S^S, \mathcal{B}(S^S))$. Comme $\mathcal{B}(S^S)$ est engendrée par les projections sur les coordonnées, $f : \Omega \to S^S = \mathcal{F}(S, S)$ est une fonction aléatoire si et seulement si pour tout $i \in S$, l'application $\omega \mapsto f(\omega)(i)$ est une variable aléatoire. La tribu engendrée par une variable aléatoire f est la tribu engendrée par les variables $f(\cdot)(i)$, où i décrit S.

Si f est une fonction aléatoire et X une variable aléatoire, $f(X)$ est une variable aléatoire car

$$\{f(X) \in B\} = \cup_{i \in S} \{X = i\} \cap \{f(i) \in B\}.$$

Lemme 11. *Soit S un ensemble fini ou dénombrable, ν une loi sur S et χ une mesure sur $(S^S, \mathcal{B}(S^S))$.*

Soit $(f_n)_{n \geq 1}$ une suite de fonctions aléatoires indépendantes de loi χ et X_0 une variable aléatoire de loi μ indépendante de $(f_n)_{n \geq 1}$. On définit $(X_n)_{n \geq 1}$ par

$$\forall n \geq 0 \quad X_{n+1} = f_{n+1}(X_n)$$

Alors $(X_n)_{n \geq 0}$ est une chaîne de Markov de loi initiale ν et de matrice de transition M, où M est définie par

$$\forall (i, j) \in S \times S \quad m_{i,j} = \chi(\{f \in S^S ; f(i) = j\}).$$

Démonstration. Soit $A \subset S^{\{0,\dots,n\}}$.

$$\begin{aligned}
& \mathbb{P}(\{(X_0,\dots,X_n) \in A\} \cap \{X_n = i\} \cap \{X_{n+1} = j\}) \\
= \; & \mathbb{P}(\{(X_0,\dots,X_n) \in A\} \cap \{X_n = i\} \cap \{f_{n+1}(i) = j\}) \\
= \; & \mathbb{P}(\{(X_0,\dots,X_n) \in A\} \cap \{X_n = i\})\mathbb{P}(f_{n+1}(i) = j) \\
= \; & \mathbb{P}(\{(X_0,\dots,X_n) \in A\} \cap \{X_n = i\})\mathbb{P}(f_{n+1} \in S \times \dots \{j\} \times \dots S) \\
= \; & \mathbb{P}(\{(X_0,\dots,X_n) \in A\} \cap \{X_n = i\})\chi(S \times \dots \{j\} \times \dots S) \\
= \; & \mathbb{P}(\{(X_0,\dots,X_n) \in A\} \cap \{X_n = i\})m_{i,j}
\end{aligned}$$

□

Exemple : la marche de l'ivrogne (ou marche aléatoire sur $\mathbb{Z}$)
Un ivrogne sort du café passablement éméché. À chaque pas, il prend une décision (enfin, si tant est que cela lui soit possible...) : aller à gauche, ou aller à droite. Si on repère par X_n sa position dans la rue au temps n, on a $S = \mathbb{Z}$, $X_{n+1} = f_{n+1}(X_n)$, où f_n est une suite de translations indépendantes : $\mathbb{P}(f_n = (x \mapsto x+1)) = \mathbb{P}(f_n = (x \mapsto x-1)) = 1/2$.

Comme on va le voir, ce procédé permet de fabriquer toutes les chaînes de Markov.

9.2 Matrice stochastique

Définition: Soit S un ensemble dénombrable et $P = (p_{i,j})_{(i,j)\in S\times S}$ une matrice à coefficients positifs. On dit que P est une matrice stochastique si on a

$$\forall i \in S \quad \sum_{j\in S} p_{i,j} = 1.$$

9.2.1 Existence des chaînes de Markov

Théorème 69. *Soit S un ensemble dénombrable, $P = (p_{i,j})_{(i,j)\in S\times S}$ une matrice stochastique et ν une mesure de probabilité sur S. Alors, on peut construire une chaîne de Markov de loi initiale ν et de matrice de passage P.*

Démonstration. Définissons une mesure χ_P sur S^S par

$$\chi_P = \otimes_{i\in S}\mu_i,$$

où μ_i est la mesure sur S définie par $\mu_i(j) = p_{i,j}$. Alors χ_P vérifie $\chi_{\mathbb{P}}(S \times \dots \{j\} \times \dots S) = p_{i,j}$ et il suffit d'appliquer le lemme précédent. □

Lorsque la matrice P est fixée, on note souvent $\mathbb{P}^\nu$ une probabilité sous laquelle $(X_n)_{n\geq 0}$ est une chaîne de Markov de matrice de transition P telle que la loi de X_0 sous P^ν est ν. De même, on note $\mathbb{E}^\nu$ l'espérance correspondante. Dans le cas où la loi initiale est une masse de Dirac, on écrit simplement $\mathbb{P}^i$ (resp. $\mathbb{E}^i$) au lieu de $\mathbb{P}^{\delta_i}$ (resp. $\mathbb{E}^{\delta_i}$).

Remarque : on est souvent amené à réaliser une telle chaîne sur l'espace canonique $\Omega = S^{\mathbb{N}}$. Dans ce cas, les $(X_k)_{k\geq 0}$ sont les opérateurs de projection canonique : $X_k(\omega) = \omega_k$ et $\mathbb{P}^\nu$ est l'unique mesure sur Ω telle que pour tout entier $n \geq 1$ et toute suite $x_0, \dots x_n$ d'éléments de S :

$$\mathbb{P}^\nu(X_0 = x_0, X_1 = x_1, \dots X_n = x_n) = \nu(x_0) \prod_{i=0}^{n-1} p_{x_i,x_{i+1}}.$$

Corollaire 16. *Soit P une matrice markovienne sur S. Pour tout ν, on note P^ν la mesure markovienne sur $S^{\mathbb{N}}$ de loi initiale ν et de matrice de passage P, ainsi que $\mathbb{P}^i = \mathbb{P}^{\delta_i}$. Pour toute loi ν sur S, $\mathbb{P}^\nu$ admet la désintégration*

$$\mathbb{P}^\nu = \int \mathbb{P}^i \, d\nu \tag{9.2}$$

c'est à dire que pour tout borélien A de $S^{\mathbb{N}}$, on a

$$\mathbb{P}^\nu(A) = \int \mathbb{P}^i(A) \, d\nu \tag{9.3}$$

Démonstration. Il suffit de définir une mesure μ par

$$\mu(A) = \int \mathbb{P}^i(A) \, d\nu$$

et de vérifier que l'on a pour tout entier $n \geq 1$ et toute suite $x_0, \dots x_n$ d'éléments de S :

$$\mu(X_0 = x_0, X_1 = x_1, \dots X_n = x_n) = \nu(x_0) \prod_{i=0}^{n-1} p_{x_i,x_{i+1}}.$$

□

Remarques :

— On trouve parfois la notation $\mathbb{P}_i$ à la place de $\mathbb{P}^i$. Dans ce cas, il faut faire attention qu'il peut y avoir ambiguité sur le sens de la notation $\mathbb{P}_X$.

9.2.2 Point de vue fonctionnel (*)

Une matrice stochastique indexée par peut assez naturellement être vue comme un opérateur sur l'espace des fonctions bornées sur S. Rappelons que l'espace des fonctions bornées, que l'on note $\ell^\infty(S)$, est l'ensemble des fonctions f telles que $\|f\|_\infty = \sup\{|f(i)|; i \in S\} < +\infty$, et que $\|\cdot\|_\infty$ est une norme sur $\ell^\infty(S)$.

Maintenant, si $f \in \ell^\infty(S)$, la fonction Pf définie par

$$\forall i \in S \quad (Pf)(i) = \sum_{j \in S} p_{i,j} f(j)$$

est une fonction bornée. En effet, la majoration

$$\sum_{j \in S} |p_{i,j} f(j)| \leq \sum_{j \in S} p_{i,j} \|f\|_\infty = \|f\|_\infty$$

montre que la série converge absolument. De plus, comme pour tout i, $|Pf(i)| \leq \|f\|_\infty$, on a $\|Pf\|_\infty \leq \|f\|_\infty$: P est une contraction de l'espace des fonctions bornées.

Remarque : $(Pf)(i)$ est l'intégrale de la fonction f par rapport à la mesure de probabilité sur S qui affecte la j la probabilité $p_{i,j}$.

Théorème 70. *Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires à valeurs dans S. On a équivalence entre*

1. *$(X_n)_{n\geq 0}$ est une chaîne de Markov de matrice de passage P à valeurs dans S.*
2. *Pour toute fonction $f \in \ell^\infty(S)$ et pour tout entier $n \geq 0$*

$$\mathbb{E}[f(X_{n+1})|X_0, \dots, X_n] = (Pf)(X_n).$$

Démonstration. Pour le sens direct, il suffit de prendre $f = \delta_i$ et d'appliquer la caractérisation vue en début de chapitre. Regardons la réciproque. Si $f = \delta_i$, l'identité découle encore de la caractérisation vue en début de chapitre. Par linéarité, l'identité s'étend au cas des fonctions à support fini. Passons au cas dénombrable. Il existe une suite croissante d'ensembles finis $(S_p)_{p\geq 1}$ avec $S = \cup_{p\geq 1} S_p$. $f\mathbb{1}_{S_p}$ est une fonction bornée, donc

$$\mathbb{E}[(f\mathbb{1}_{S_p})(X_{n+1})|X_0, \dots, X_n] = (P(f\mathbb{1}_{S_p})(X_n).$$

Le théorème de convergence dominée pour l'espérance conditionnelle nous dit que

$$\lim_{p\to+\infty} \mathbb{E}[(f\mathbb{1}_{S_p})(X_{n+1})|X_0, \dots, X_n] = \mathbb{E}[f(X_{n+1})|X_0, \dots, X_n].$$

Pour conclure, il suffit de montrer que pour tout $i \in S$

$$\lim_{p \to +\infty} P(f \mathbb{1}_{S_p})(i) = P(f)(i).$$

Mais pour toute fonction bornée, le théorème de transfert donne

$$(Pg)(i) = \mathbb{E}^i g(X_1). \tag{9.4}$$

L'identité voulue découle alors immédiatement du théorème de convergence dominée. □

L'identité (9.4) est élémentaire, mais peut être utile. On déduit de ce théorème une autre remarque très simple, mais très puissante :

Corollaire 17. *Soit $(X_n)_{n\ge 0}$ est une chaîne de Markov de matrice de passage P à valeurs dans S, $f \in \ell^\infty(S)$. Si l'on pose $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$, alors la suite $(Y_n)_{n\ge 1}$ définie par*

$$\forall n \ge 0 \quad Y_{n+1} = f(X_{n+1}) - (Pf)(X_n)$$

est une suite de différences de martingales adaptée à la filtration $(\mathcal{F}_n)_{n\ge 0}$.

Démonstration. Y_{n+1} est $\mathcal{F}_{n+1}$-mesurable et

$$\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = \mathbb{E}[f(X_{n+1})|\mathcal{F}_n] - \mathbb{E}[(Pf)(X_n)|\mathcal{F}_n] = (Pf)(X_n) - (Pf)(X_n) = 0.$$

□

9.2.3 Puissances des matrices stochastiques

Théorème 71. *Soit (X_n) une chaîne de Markov de matrice de transition P et de loi initiale $\mathbb{P}_{X_0} = \nu$. Alors, la loi μ_n de la chaîne au temps n s'écrit $\mu_n = \nu P^n$, où on a écrit ν et μ_n comme des vecteurs lignes.*

Démonstration. Il suffit de montrer que $\mu_{n+1} = \mu_n P$, puis procéder par récurrence sur n. D'après le principe de partition, on a

$$\begin{aligned}
\mu_{n+1}(j) &= \mathbb{P}^\nu(X_{n+1} = j) \\
&= \sum_{i \in S} \mathbb{P}^\nu(X_n = i, X_{n+1} = j) \\
&= \sum_{i \in S} \mathbb{P}^\nu(X_n = i) p_{i,j} \\
&= \sum_{i \in S} \mu_n(i) p_{i,j} \\
&= (\mu_n M)(j)
\end{aligned}$$

□

En particulier, en prenant $\nu = \delta_i$, on a le corollaire important :

Corollaire 18. *Soit (X_n) une chaîne de Markov à valeur dans S, de matrice de transition P et de loi initiale δ_i, avec $i \in S$. Alors, pour tout $j \in S$, on a*

$$\mathbb{P}^i(X_n = j) = P^n(i,j).$$

9.2.4 Graphe associé à une matrice stochastique

Soit $P = (p_{i,j})_{(i,j)\in S\times S}$ une matrice stochastique. On peut associer à la matrice P (où aux chaînes de Markov correspondantes) un graphe orienté $G = (S, A)$ avec

$$A = \{(x,y) \in S \times S; p_{i,j} > 0\}.$$

Considérons une chaîne de Markov associée à la matrice stochastique P avec la condition initiale déterministe x_0, autrement dit $\nu = \delta_{x_0}$ et notons $\mathbb{P}^{x_0}$ la mesure de probabilité correspondante Alors, comme

$$\mathbb{P}^{x_0}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \prod_{i=0}^{n-1} p_{x_i,x_{i+1}},$$

il est clair que $\mathbb{P}^{x_0}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n)$ est non nul si et seulement si $(x_0, x_1, \dots, x_n)$ constitue un chemin dans le graphe G.

D'après le principe de partition, on a pour une chaîne de Markov avec une loi initiale δ_i

$$\mathbb{P}^i(X_n = x_n) = \sum_{(x_0,\dots x_{n-1})\in S^n} \mathbb{P}^i(X_0 = x_0, X_1 = x_1, \dots X_{n-1} = x_{n-1}, X_n = x_n). \tag{9.5}$$

En particulier, si l'on pose $p_{i,j}^{(n)} = \mathbb{P}^i(X_n = j)$, on a

$$p_{i,j}^{(n)} = \sum_{x\in S^{n-1}} \mathbb{P}^i(X_1 = x_1, X_2 = X_2, \dots X_{n-1} = x_{n-1}, X_n = j).$$

Donc $p_{i,j}^{(n)} > 0$, autrement dit il est possible d'aller en n étapes de l'état i à l'état j si et seulement si on peut trouver dans le graphe G un chemin de longueur n allant de i à j.

On en déduit que

$$\mathbb{P}^i(\exists n > 0;\ X_n = j) = \mathbb{P}^i(\cup_{n\geq 1}\{X_n = j\}),$$

qui représente la probabilité que, partant de i, on puisse arriver à j, est non nulle si et seulement si il existe dans le graphe G un chemin allant de i à j. Dans ce cas, on dit que j est *accessible* à partir de i et on écrit $i \to j$.

Si il y a à la fois un chemin de i vers j et un chemin de j vers i, on dit que les états i et j communiquent et on écrit $i \leftrightarrow j$.

Si tous les états communiquent, on dit que la chaîne de Markov est *irréductible.*

On appelle *période* d'un état x d'une chaîne de Markov et on note $d(x)$ le pgcd (plus grand commun diviseur) des longueurs des circuits du graphe G contenant x. Lorsque la période est 1, on dit que l'état x est *apériodique.*

Lemme 12. *Si deux états communiquent, alors ils ont même période.*

Démonstration. Soient i, j avec $i \leftrightarrow j$. Soit γ un chemin de i à j, γ' un chemin de j à i. Soit $\mathcal{C}$ un circuit quelconque (éventuellement vide) contenant j . $\gamma - \gamma'$ et $\gamma - \mathcal{C} - \gamma'$ sont deux circuits contenant i. Donc $d(i)$ divise leurs longueurs ainsi que la différence de leurs longueurs, soit la longueur de $\mathcal{C}$. Ainsi $d(i)$ divise les longueurs de tous les circuits contenant j, donc divise leur pgcd, soit $d(j)$. De la même manière, on montre que $d(j)$ divise $d(i)$, d'où $d(i) = d(j)$. $\square$

Définition: Si une chaîne irréductible a ses états de période 1, on dit qu'elle est apériodique.

Le lemme suivant et ses corollaires se révéleront très utiles par la suite

Lemme 13. *Soit x un état de période 1. Il existe un entier $N(x)$ tel que pour tout $n \geq N(x)$ le graphe associé à la chaîne de Markov possède un circuit de longueur n contenant x*

Soit A l'ensemble des valeurs de n telles que le graphe associé à la chaîne de Markov possède un circuit de longueur n contenant x. Il est clair que A est stable par addition (concaténation des circuits). Il existe $p \geq 1$ et $n_1, n_2, \dots, n_p$ tels que le pgcd de $n_1, n_2, \dots, n_p$ soit 1. D'après le lemme de Bezout, il existe des relatifs $a_1, \dots a_p$ tels que $1 = \sum_{k=1}^{p} a_k n_k$. Posons $P = \sum_{p:a_p>0} a_p n_p$ et $N = \sum_{p:a_p<0} (-a_p) n_p$. On a $P \in A, N \in A$ et $1 = P - N$. Soit $n \geq N(N-1)$. On peut écrire $n = bN + r$, avec $b \geq N - 1$ et $r \in \{0, 1, \dots N-1\}$. On a $n = bN + r = bN + r(P - N) = rP + (b - r)N \in A$ car $b - r \in \mathbb{N}$ et A est stable par addition.

Corollaire 19. *Si x est un état de période 1 et qu'il existe un chemin de longueur $d(x, y)$ allant de x à y, alors pour tout $n \geq N(x, y) = N(x) + d(x, y)$, il existe un chemin de longueur n allant de x à y. Ainsi, si P est la matrice associée, $P^n(x, y) > 0$.*

Démonstration. Il suffit de concaténer le chemin allant de x à x avec un chemin allant de x à y. □

Corollaire 20. *Si une chaîne de Markov est irréductible, apériodique, à valeurs dans un ensemble fini S, alors il existe un entier N tel que pour tout $n \geq N$ et tout couple (i, j), il existe un chemin de longueur n allant de i à j. Ainsi, si P est la matrice associée, P^n est à coefficients strictement positifs.*

Démonstration. Il suffit de prendre $N = \max(N(x), x \in S) + \operatorname{diam}(G)$. □

La définition suivante est très simple, mais sera abondamment utilisée dans les exercices.

Définition On appelle point absorbant d'une chaîne tout point x tel que $\mathbb{P}^x(X_1 = x) = 1$.

9.3 Propriété de Markov

9.3.1 Le théorème

Théorème 72. *Soit $(X_k)_{k\geq 0}$ une chaîne de Markov de matrice de passage P. Soit p un entier naturel. La suite $(X_{k+p})_{k\geq 0}$ est une chaîne de Markov de matrice de passage P et de loi initiale la loi de X_p. De plus, pour tout A $\mathcal{F}_p$-mesurable et tout $i \in S$, on a*

$$\mathbb{P}(A, X_p = i, X_{p+.} \in B) = \mathbb{P}(A, X_p = i)\mathbb{P}^i(X \in B).$$

De manière équivalente, on a $\mathbb{P}$ presque-sûrement :

$$\mathbb{P}(X_{p+.} \in B|\mathcal{F}_p) = f_B(X_p), \text{ avec } f_B(x) = \mathbb{P}^x(X \in B).$$

Démonstration. Comme être une chaîne de Markov est une propriété de la loi, on peut supposer que $(X_n)_{n\geq 0}$ est obtenue par le procédé décrit plus haut : $X_{n+1} = f_{n+1}(X_n)$ où $(f_n)_{n\geq 1}$ est une suite de variables aléatoires indépendantes de loi χ_M,$(f_n)_{n\geq 1}$ étant de plus supposée indépendante de X_0. Posons $Y_n = X_{n+p}$ Si l'on pose $g_n = f_{n+p}$, on a la récurrence $Y_{n+1} = g_{n+1}(Y_n)$. Mais la loi de $(g_n)_{\geq 1}$ est $\chi_M^{\otimes}\mathbb{N}^*$, ce qui montre bien que $(Y_n)_{\geq 0}$ est une chaîne de Markov de matrice de passage P et de loi initiale la loi de X_p. Maintenant, soit B un borélien de $S^{\mathbb{N}}$. On pose

$$G_i((h_n)_{n\geq 1}) = (i, h_1(i), h_2 \circ h_1(i), h_3 \circ h_2 \circ h_1(i), \dots)..$$

$$\mathbb{P}(A, X_p = i, X_{p+.} \in B) = \mathbb{P}(A, X_p = i, G_i(g) \in B)$$

$A \cap \{X_p = i\}$ est $\sigma(X_0, f_1, \dots, f_p)$-mesurable tandis que $\{G_i(g) \in B\}$ est $\sigma(f_k; k > p)$-mesurable, donc

$$\mathbb{P}(A, X_p = i, G_i(g) \in B) = \mathbb{P}(A, X_p = i)\mathbb{P}(G_i(g) \in B) = \mathbb{P}(A, X_p = i)\mathbb{P}^i(X \in B).$$

Pour la deuxième forme, il est clair que $f_B(X_p)$ est $\mathcal{F}_p$-mesurable : il suffit donc de vérifier donc que pour tout $A \in \mathcal{F}_p$, on a

$$\mathbb{E}\mathbb{1}_A \mathbb{1}_{\{X_{p+.} \in B\}} = \mathbb{E}\mathbb{1}_A f_B(X_p).$$

Or

$$\begin{aligned}
\mathbb{E}\mathbb{1}_A \mathbb{1}_{\{X_{p+.} \in B\}} &= \sum_i \mathbb{E}\mathbb{1}_A \mathbb{1}_{\{X_p = i\}} \mathbb{1}_{\{X_{p+.} \in B\}} \\
&= \sum_i \mathbb{P}(A, X_p = i, X_{p+.} \in B) \\
&= \sum_i \mathbb{P}(A, X_p = i)\mathbb{P}^i(X \in B) \\
&= \sum_i \mathbb{E}[\mathbb{1}_{A \cap \{X_p = i\}} \mathbb{P}^i(X \in B)] \\
&= \sum_i \mathbb{E}[\mathbb{1}_{A \cap \{X_p = i\}} f_B(X_p)] \\
&= \mathbb{E}[\mathbb{1}_A f_B(X_p)]
\end{aligned}$$

□

Remarque : on peut trouver dans la littérature l'écriture

$$\mathbb{P}(X_{p+.} \in B | \mathcal{F}_p) = \mathbb{P}^{X_p}(X \in B).$$

Je mets le lecteur en garde contre le fait que $\mathbb{P}^{X_p}$ ne signifie pas la même chose que $\mathbb{P}^{\mathbb{P}_{X_p}}$.

La propriété de Markov est souvent utilisée sous la forme simple suivant : si A est un borélien de $\mathbb{R}^n$, B un borélien de $\mathbb{R}^p$, alors

$$\begin{aligned}
& \mathbb{P}((X_0, \dots X_{n-1}) \in A, X_n = i, (X_{n+1}, \dots, X_{n+p}) \in B) \\
= \; & \mathbb{P}((X_0, \dots X_{n-1}) \in A, X_n = i)\mathbb{P}^i((X_1, \dots, X_p \in B)).
\end{aligned}$$

9.3.2 Analyse au premier pas

Corollaire 21. *Soit $(X_n)_{n \geq 0}$ une chaîne de Markov de matrice de passage $(p_{i,j})$, B un borélien de $\mathbb{R}^{\mathbb{N}}$. On note Θ l'opérateur de translation : $\Theta((x_n)_{n \geq 0}) = ((x_{n+1})_{n \geq 0})$.*

Alors

$$\begin{aligned}\mathbb{P}(\Theta(X) \in B) &= \mathbb{P}^{\mathbb{P}_{X_1}}(X \in B) \\ &= \sum_{j:\mathbb{P}(X_1=j)>0} \mathbb{P}(X_1 = j)\mathbb{P}^j(X \in B)\end{aligned}$$

En particulier, si B est invariant par l'opérateur de translation (c'est à dire que $\Theta^{-1}(B) = B$) , alors on a le système d'équations :

$$\mathbb{P}^i(X \in B) = \sum_{j:p_{i,j}>0} p_{i,j}\mathbb{P}^j(X \in B)$$

Démonstration. La première égalité traduit exactement la propriété de Markov : une chaîne de Markov observée à partir du temps 1 a la même loi qu'une chaîne de Markov de même dynamique commençant avec comme valeur initiale celle que prend la chaîne de Markov non décalée au temps 1. La deuxième égalité correspond à une décomposition suivant les valeurs que peut prendre X_1.

Passons au cas où B est invariant :

$$\begin{aligned}\mathbb{P}^i(X \in B) &= \mathbb{P}^i(X \in \Theta^{-1}(B)) \\ &= \mathbb{P}^i(\Theta(X) \in B) \\ &= \sum_{i:P^i(X_1=j)>0} \mathbb{P}^i(X_1 = j)\mathbb{P}^j(X \in B) \\ &= \sum_{i:p_{i,j}>0} p_{i,j}\mathbb{P}^j(X \in B)\end{aligned}$$

□

9.4 Exercices sur les chaînes de Markov

9.4.1 Exercices corrigés

Exercice 66. On lance un dé équilibré à 6 faces jusqu'à obtenir deux six consécutifs. Calculer l'espérance du nombre de lancers nécessaires. lien vers l'indication lien vers la solution

Exercice 67. Soit $(X_n)_{n\geq 0}$ une chaîne de Markov irréductible à valeurs dans un espace d'états fini ou dénombrable S. Soit $A \subset S$, avec A fini et $A \neq S$. On pose $\tau = \inf\{n \geq 0; X_n \notin A\}$. Montrer que $\tau < +\infty$ presque sûrement. lien vers l'indication lien vers la solution

Exercice 68. *L'image d'une chaîne de Markov n'est pas (toujours) une chaîne de Markov.*
On considère la chaîne de Markov (X_n) sur $E = \{0, 1, 2\}$ de matrice de transition $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$ et de loi initiale $\pi_0 = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Soit $f : E \to \{0, 1\}$ telle que $f(0) = f(1) = 0$, $f(2) = 1$. Pour $n \geq 0$, on pose $Y_n = f(X_n)$. Montrer que $(Y_n)_{n\geq 1}$ n'est pas une chaîne de Markov.

lien vers l'indication lien vers la solution

Exercice 69. *L'image d'une chaîne de Markov peut être une chaîne de Markov.*
Soit (X_n) une chaîne de Markov sur un ensemble dénombrable E de matrice de transition P. Soit ψ une application surjective de E dans un ensemble F telle que

$$\forall z \in F \quad \forall x, y \in E \quad \psi(x) = \psi(y) \Rightarrow \mathbb{P}^x(\psi(X_1) = z) = \mathbb{P}^y(\psi(X_1) = z).$$

Montrer que la suite (Y_n) définie par $Y_n = \psi(X_n)$ est une chaîne de Markov et déterminer sa matrice de transition. Montrer que si π est une probabilité stationnaire pour la chaîne (X_n) alors l'image de π par ψ est stationnaire pour (Y_n).

lien vers l'indication lien vers la solution

9.4.2 Exercices non corrigés

Exercice 70. On pose $Y_0 = 0$, puis, pour $n \geq 1$, Y_n est une suite de variables indépendantes suivant la loi de Bernoulli de paramètre p. On pose ensuite $X_0 = 0$, et pour tout $n \geq 1 : X_n = \max\{k : Y_n = Y_{n-1} = \dots Y_{n-k+1} = 1\} \wedge 0$.

1. Montrer que $(X_n)_{n\geq 0}$ est une chaîne de Markov.

2. Dans la suite, on note $\mathbb{P}^i$ la loi d'une chaîne de Markov avec la même dynamique partant de n, $\mathbb{E}^i$ l'espérance correspondante. On note $T^n = \inf\{n \geq 0; X_i = n\}$ Montrer que pour $i < n$, on a pour $i < n$
$$\mathbb{E}^i[T^n] = 1 + p\mathbb{E}^{i+1}[T^n] + (1-p)\mathbb{E}^0 T^n.$$
3. En déduire la valeur de $\mathbb{E}^0 T^n$.
4. Application : Une pièce de monnaie a pour probabilité p, de tomber sur face. On la lance indéfiniment. Calculer l'espérance du nombre de jets qu'il faudra jusqu'à ce qu'une chaîne de r résultats consécutifs de type face apparaisse.

lien vers l'indication

Exercice 71. *Chaîne à deux états.* Soit $\{X_n : n \geq 0\}$ une chaîne de Markov à valeurs dans $\{0,1\}$ et de probabilité de transition :
$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}, \, 0 \leq \alpha, \beta \leq 1.$$

1. Montrer que pour $(\alpha, \beta) \neq (0,0)$:
$$P^n = \frac{1}{\alpha+\beta}\begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{(1-\alpha-\beta)^n}{\alpha+\beta}\begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix}.$$
Que se passe-t-il lorsque $\alpha = 0$ ou $\beta = 0$ ou $\alpha = \beta = 0$? On supposera pour la suite de l'exercice que $(\alpha, \beta) \neq (0,0)$.
2. Vérifier que pour toute loi initiale μ, on a
$$\mathbb{P}^\mu(X_n = 0) = \frac{\beta}{\alpha+\beta} + (1-\alpha-\beta)^n\left(\mu(0) - \frac{\beta}{\alpha+\beta}\right).$$
3. Si $(\alpha, \beta) \neq (1,1)$, montrer que $(X_n)_{n\geq 0}$ converge en loi vers une loi ν que l'on déterminera. On supposera pour la suite de l'exercice que $(\alpha, \beta) \neq (1,1)$.
4. (*Mesure stationnaire*) Prouver que, pour tout $n \in \mathbb{N}$,
$$\mathbb{P}^\nu(X_n \in A) = \nu(A).$$

lien vers l'indication

Exercice 72. *Représentation canonique et simulation des chaînes de Markov.*

1. Soit $(Z_n)_{n\geq 1}$ une suite de vaiid à valeurs dans F, soit $g : E\times F \to E$ et soit X_0 une variable aléatoire à valeurs dans E indépendante de $(Z_n)_{n\geq 1}$. Montrer que la suite $(X_n)_{n\geq 0}$ définie par $X_{n+1} = g(X_n, Z_{n+1})$ est une chaîne de Markov homogène. Donner sa matrice de transition.
2. On suppose qu'on dispose d'un générateur de nombres aléatoire de loi uniforme sur $[0,1]$, noté 'rand'. Soit μ une mesure de probabilité sur $\mathbb{N}$. Donner un algorithme pour générer des nombres aléatoires suivant la loi μ.
3. Soit $P = (p_{i,j})$ une matrice de transition sur $\mathbb{N}$. On note $s_{i,k} = \sum_{j=0}^{k} p_{i,j}$. Soit $(Z_n)_{n\geq 1}$ une suite de vaiid de loi uniforme sur $[0,1]$ et X_0 une variable aléatoire à valeurs dans $\mathbb{N}$ indépendante de $(Z_n)_{n\geq 1}$. On construit la suite $(X_n)_{n\geq 0}$ par récurrence de la façon suivante :
$$\text{si } X_n(\omega) = i \text{ et } Z_{n+1}(\omega) \in]s_{i,j-1}, s_{i,j}] \text{ alors } X_{n+1} = j.$$
Montrer que la suite $(X_n)_{n\geq 0}$ ainsi définie est une chaîne de Markov homogène. Donner sa matrice de transition.
4. Application. Comment simuler une chaîne de Markov homogène de matrice de transition $P = (p_{i,j})$? Ecrire un algorithme explicite si
$$P = \begin{pmatrix} 0.25 & 0.5 & 0.25 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{pmatrix}.$$

lien vers l'indication

Exercice 73. *Temps d'atteinte d'un état absorbant.*

Soit (X_n) une chaîne de Markov sur un ensemble dénombrable E et $a \in E$ un état absorbant. On pose $T = \inf\{n \geq 0; X_n = a\}$. Montrer que $\mathbb{P}(X_n = a) = \mathbb{P}(T \leq n)$. lien vers l'indication

Exercice 74. *Temps d'entrée : une propriété d'invariance.*

Soit (X_n) une chaîne de Markov sur un ensemble dénombrable E de matrice de transition Q. Pour $f : E \to \mathbb{R}_+$, soit Qf la fonction définie par
$$Qf(x) = \sum_{y\in E} Q(x,y)f(y).$$
Pour $A \subset E$ on note $T_A = \inf\{n \geq 0; X_n \in A\}$ le temps d'entrée dans A. Montrer que la fonction f définie sur E par $f(x) = \mathbb{P}_x(T_A < +\infty)$ vérifie
$$f(x) = 1 \text{ pour } x \in A \text{ et } f(x) = (Qf)(x) \text{ pour } x \notin A.$$

lien vers l'indication

Exercice 75. *Chaîne de Markov arrêtée.*

Soit (X_n) une chaîne de Markov sur un ensemble dénombrable E de matrice de transition Q. Etant donné un ensemble $B \subset E$, on note

$$T_B = \inf\{n \geq 0; X_n \in B\}$$

le temps d'entrée dans B et on pose $Y_n = X_{n \wedge T_B}$. Montrer que $(Y_n)_{n \geq 0}$ est une chaîne de Markov sur E dont on précisera la matrice de transition. lien vers l'indication

Exercice 76. Soit $(X_n)_{n \geq 0}$ une chaîne de Markov à valeurs dans $\mathbb{N}$. On note A l'ensemble des points absorbant de la chaîne. Montrer que (X_n) ne peut converger que vers un élément de A.
Plus précisément : si il existe un événement B et une variable aléatoire Y telle que

$$\forall \omega \in B \quad \lim_{n \to +\infty} X_n(\omega) = Y(\omega),$$

alors $\mathbb{P}(B \cap \{Y \notin A\}) = 0$. lien vers l'indication

Exercice 77. *La ruine du joueur* Un joueur possédant une fortune de a unités joue à pile ou face jusqu'à ce qu'il ait fait sauter la banque ou qu'il soit ruiné. Les réserves de la banque sont de b unités. Chaque victoire rapporte une unité et chaque défaite en coûte une. On suppose que les lancers sont indépendants et que la probabilité de gain de la banque est $p = 1 - q$. On veut déterminer la probabilité p_g que la banque résiste.

On note $(X_n)_{n \geq 1}$ une suite de v.a.i.i.d. de loi $p\delta_1 + q\delta_{-1}$, puis $S_n = \sum_{k=1}^n X_k$ et $T = \inf\{n \geq 0; S_n = -b \text{ ou } S_n = a\}$. Si l'on pose $S'_n = S_{n \wedge T}$, il est aisé de constater que S'_n représente la suite des gains relatifs de la banque.

1. Montrer que S'_n est une chaîne de Markov homogène à espace d'états $E = \{-b, \dots, a\}$ dont on déterminera la loi initiale et la matrice de transition.
2. Considérons les chaînes de Markov ayant la même matrice de transition que $(S'_n)_{n \geq 0}$ Montrer que la suite $(u_n)_{-b \leq n \leq a}$ définie par

$$u_n = \mathbb{P}^n(\{\text{la banque résiste}\})$$

vérifie la récurrence linéaire

$$pu_{n+1} - u_n + qu_{n-1} = 0.$$

Que valent u_a et u_{-b} ?

3. Résoudre l'équation de récurrence et en déduire

$$p_g = \frac{(\frac{q}{p})^b - 1}{(\frac{q}{p})^{a+b} - 1}. \tag{9.6}$$

4. On note $v_n = \mathbb{E}^n[T]$. Montrer que si $-b < n < a$, on a

$$v_n = 1 + \frac{1}{2}(v_{n+1} + v_{n-1}).$$

5. Exprimer v_n en fonction de n.

lien vers l'indication

Exercice 78. *Le joueur inruinable*
Le problème est le même que le précédent, a ceci près que l'on suppose maintenant que le joueur est infiniment riche. On cherche toujours la probabilité que la banque résiste (ce qui ne signifie pas ici que le joueur est ruiné).

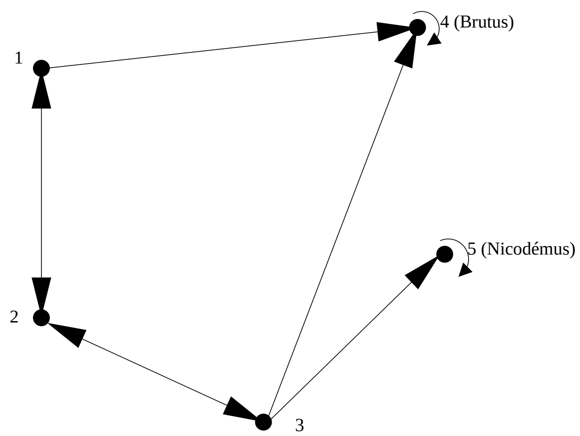
Intuitivement, il suffit de faire tendre a vers $+\infty$ dans la formule (9.6), le tout étant de le justifier…

On suggère de poser $T' = \inf\{n; S_n \leq -b\}$ et, pour tout $a > 0$, $U_a = \inf\{n; S_n \geq a\}$ et $G^a = \{U_a \leq T'\}$.

lien vers l'indication

Exercice 79. *Madame Brisby dans le labyrinthe*

Madame Brisby s'est perdue dans le labyrinthe que forment les galeries où vivent les rats de Nim. Quelle est la probabilité qu'elle rencontre le sage Nicodémus avant de croiser le belliqueux Rufus ?



lien vers l'indication

Exercice 80. Soit M la matrice d'une chaîne de Markov. Montrer que si $m_{i,i} > 0$, alors l'état i est apériodique. Qu'en déduire pour une chaîne irréductible ? lien vers l'indication

Exercice 81. Soit a et b des entiers supérieurs ou égaux à 2, $(D_n)_{n\geq 1}$ une suite de variables aléatoires i.i.d. à valeurs dans $\mathbb{Z}/a\mathbb{Z} \times \mathbb{Z}/b\mathbb{Z}$ vérifiant

$$P(D_1 = (0,1)) = P(D_1 = (1,0)) = \frac{1}{2}.$$

Soit $(D_n)_{n\geq 1}$ une suite de variables aléatoires et S_0 une variable aléatoire à valeurs dans $\mathbb{Z}/a\mathbb{Z} \times \mathbb{Z}/b\mathbb{Z}$ indépendante de $(D_n)_{n\geq 1}$ Pour $n \geq 1$, on pose

$$S_n = S_0 + \sum_{k=1}^{n} D_k.$$

Montrer que (S_n) est une chaîne de Markov . Est-elle irréductible, apériodique ?

lien vers l'indication

Exercice 82. *Propriété de Markov fonctionnelle*
Soit $(X_n)_{n\geq 1}$ une chaîne de Markov, F une application mesurable de $\mathbb{R}^{\mathbb{N}^*}$ dans $[0, +\infty[$. Montrer que pour tout entier $p \geq 1$, pour tout $A \in \mathcal{F}_p = \sigma(X_1, \dots, X_p)$, on a

$$\mathbb{E}[\mathbb{1}_A F((X_{n+p})_{n\geq 1})] = \mathbb{P}(A) g(X_p),$$

avec $g(x) = \mathbb{E}^x F((X_n)_{n\geq 1})$.

On rappelle que pour toute variable aléatoire Y positive et toute probabilité $\mathbb{P}$, on a $\mathbb{E}Y = \int_0^{+\infty} \mathbb{P}(Y > t)\ dt$. lien vers l'indication

Exercice 83. *Madame Brisby II*
On reprend la chaîne de Markov des aventures de madame Brisby. On note l'espace d'états $E = \{1, 2, 3, 4, 5\}$ et l'on pose $A = \{4, 5\}$ Soit $f : E \to \mathbb{C}$ telle que $\forall x \in A \quad f(x) = 0$.

On pose $F = \sum_{k=1}^{+\infty} f(X_k)$.

Montrer que $\mathbb{E}|F| \leq (\mathbb{E}T - 1)\|f\|_\infty$.

Montrer l'identité

$$(I - N) \begin{pmatrix} \mathbb{E}^1 F \\ \mathbb{E}^2 F \\ \mathbb{E}^3 F \end{pmatrix} = \begin{pmatrix} \mathbb{E}^1 f(X_1) \\ \mathbb{E}^2 f(X_1) \\ \mathbb{E}^3 f(X_1) \end{pmatrix},$$

où N est la matrice 3×3 telle que la matrice de la chaîne de Markov admette une écriture par blocs sous la forme

$$\begin{pmatrix} N & * \\ 0 & I_2 \end{pmatrix}$$

En déduire $\mathbb{E}^1 T, \mathbb{E}^2 T, \mathbb{E}^3 T$.
lien vers l'indication

Exercice 84. *Évolution d'un génotype avec fixation*
Nous travaillons ici sur une population de taille fixe formée de $2N$ gènes. Il y a deux types de gènes possibles : le type "a" et le type "A". Chacun des gènes au temps $n+1$ est engendré par deux des $2N$ gènes présents au temps N. Son type est celui d'un de ses deux parents (choisi au hasard).

On considère la variable aléatoire X_n égale au nombre d'individus de type "A" dans la population à l'étape n.

On admettra qu'on peut modéliser l'évolution par la récurrence suivante :

$$X_{n+1} = \sum_{k=1}^{2N} \mathbb{1}_{\{Y_{n+1,k} \le X_n\}},$$

où $(Y_{n,k})_{n \ge 1, k \in \{1,\dots,2N\}}$ est une suite de variables aléatoires indépendantes suivant la loi uniforme sur l'ensemble fini $\{1, \dots, 2N\}$. X_0 est indépendante des $(Y_{n,k})$.

1. Montrer que X_n est une chaîne de Markov à valeurs dans $E = \{0, \dots, 2N\}$.
2. Montrer que la loi de X_{n+1} sachant $X_n = k$ est une loi binomiale de paramètre $2N$ et $(k/2N)$. Identifier les éventuels points absorbants.
3. Montrer que $(X_n)_{n \ge 0}$ converge presque sûrement vers une variable aléatoire X_∞.
4. Déterminer la loi de X_∞ en fonction de la loi de X_0.

lien vers l'indication

Exercice 85. Soit ν, μ deux lois sur $\mathbb{N}$. ν est appelée loi de reproduction et μ est la loi de la taille de la population initiale.

On appelle chaîne de Galton-Waltson de loi initiale μ et de loi de reproduction ν la chaîne de Markov de loi initiale μ et de matrice de transition

$$p_{i,j} = \begin{cases} \nu^{*i}(j) \text{ si } i \neq 0 \\ \delta_0(j) \text{ si } i = 0 \end{cases}$$

Montrer que si $(X_n)_{n \ge 0}$ et $(Y_n)_{n \ge 0}$ sont deux chaînes de Markov indépendantes, $(X_n)_{n \ge 0}$ étant une chaîne de Galton-Waltson de loi initiale μ_1 et de loi de reproduction ν, et $(Y_n)_{n\geq 0}$ étant une chaîne de Galton-Waltson de loi initiale μ_2 et de loi de reproduction ν, alors $(X_n + Y_n)_{n\geq 0}$ est une chaîne de Galton-Waltson de loi initiale $\mu_1 * \mu_2$ et de loi de reproduction ν. lien vers l'indication

Chapitre 10

Récurrence et mesures invariantes

10.1 Temps d'arrêt et propriété de Markov forte

Avant d'énoncer la propriété de Markov forte, on va commencer par en donner une version simple dans le cas d'une famille de variables indépendantes.

Théorème 73. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé sur lequel vivent des variables aléatoires $(X_n)_{n\geq 1}$ indépendantes de loi μ. On suppose que $(\mathcal{F}_n)_{n\geq 1}$ est une filtration à laquelle sont adaptées $(X_n)_{n\geq 1}$ et un temps d'arrêt T qui est tel que $\mathbb{P}(T < +\infty) > 0$.*

On pose alors $\overline{\Omega} = \{T < +\infty\}$ et $\overline{\mathbb{P}} = \mathbb{P}(\cdot|\overline{\Omega})$ et on définit sur $(\overline{\Omega}, \mathcal{F}, \overline{\mathbb{P}})$: $Y_n(\omega) = X_{n+T(\omega)}(\omega)$. Alors, sur $(\overline{\Omega}, \mathcal{F}, \overline{\mathbb{P}})$, $(Y_n)_{n\geq 1}$ est une suite de variables aléatoires indépendantes de loi μ, et la tribu $\sigma((Y_i)_{i\geq 1})$ est indépendante de $\mathcal{F}_T$.

Démonstration. On commence par montrer que pour tout $n \geq 1$ et tout $A \in \mathcal{F}_T$, la loi de $(\mathbb{1}_A, Y_1, \dots, Y_n)$ sous $\overline{\mathbb{P}}$ est $\mathrm{Ber}(\overline{\mathbb{P}}(A)) \otimes \mu^{\otimes n}$. Soit $(t_0, \dots, t_n)$ des réels. On a

$$\begin{aligned}
\mathbb{1}_{\{T<+\infty\}} e^{it_0\mathbb{1}_A} \prod_{k=1}^{n} e^{it_k Y_k} &= \lim_{N\to+\infty} \mathbb{1}_{\{T\leq N\}} e^{it_0\mathbb{1}_A} \prod_{k=1}^{n} e^{it_k Y_k} \\
&= \lim_{N\to+\infty} \sum_{j=1}^{N} \mathbb{1}_{\{T=j\}} e^{it_0\mathbb{1}_A} \prod_{k=1}^{n} e^{it_k Y_k} \\
&= \lim_{N\to+\infty} \sum_{j=1}^{N} \mathbb{1}_{\{T=j\}} e^{it_0\mathbb{1}_A} \prod_{k=1}^{n} e^{it_k X_{j+k}}
\end{aligned}$$

Avec le théorème de convergence dominée, on en déduit que

$$\mathbb{E}\left(\mathbb{1}_{\{T<+\infty\}} e^{it_0 \mathbb{1}_A} \prod_{k=1}^n e^{it_k Y_k}\right) = \lim_{N\to+\infty} \sum_{j=1}^N \mathbb{E}\left(\mathbb{1}_{\{T=j\}} e^{it_0 \mathbb{1}_A} \prod_{k=1}^n e^{it_k X_{j+k}}\right)$$

On peut réécrire $\mathbb{1}_{\{T=j\}} e^{it_0 \mathbb{1}_A} = \mathbb{1}_{\{T=j\}\cap A} e^{it_0} + \mathbb{1}_{\{T=j\}\cap A^c}$. Comme T est un temps d'arrêt, $\{T=j\}\cap A$ et $\{T=j\}\cap A^c$ sont dans $\mathcal{F}_j$, ce qui montre que $\mathbb{1}_{\{T=j\}} e^{it_0 \mathbb{1}_A}$ est $\mathcal{F}_j$ mesurable. Par indépendance, on a donc

$$\begin{aligned}\sum_{j=1}^N \mathbb{E}\left(\mathbb{1}_{\{T=j\}} e^{it_0 \mathbb{1}_A} \prod_{k=1}^n e^{it_k X_{j+k}}\right) &= \sum_{j=1}^N \mathbb{E}\left(\mathbb{1}_{\{T=j\}} e^{it_0 \mathbb{1}_A}\right) \mathbb{E}\left(\prod_{k=1}^n e^{it_k X_{j+k}}\right) \\ &= \sum_{j=1}^N \mathbb{E}\left(\mathbb{1}_{\{T=j\}} e^{it_0 \mathbb{1}_A}\right)\left(\prod_{k=1}^n \varphi_\mu(t_k)\right) \\ &= \mathbb{E}\left(\mathbb{1}_{\{T\le N\}} e^{it_0 \mathbb{1}_A}\right)\left(\prod_{k=1}^n \varphi_\mu(t_k)\right),\end{aligned}$$

où on a posé $\varphi_\mu(t) = \mathbb{E}(e^{itX_1})$. En utilisant encore le théorème de convergence dominée, on obtient alors

$$\mathbb{E}\left(\mathbb{1}_{\{T<+\infty\}} e^{it_0 \mathbb{1}_A} \prod_{k=1}^n e^{it_k Y_k}\right) = \mathbb{E}\left(\mathbb{1}_{\{T<+\infty\}} e^{it_0 \mathbb{1}_A}\right)\left(\prod_{k=1}^n \varphi_\mu(t_k)\right)$$

En divisant par $\mathbb{P}(T<+\infty)$, on obtient enfin

$$\overline{\mathbb{E}}\left(e^{it_0 \mathbb{1}_A} \prod_{k=1}^n e^{it_k Y_k}\right) = \overline{\mathbb{E}}\left(e^{it_0 \mathbb{1}_A}\right)\left(\prod_{k=1}^n \varphi_\mu(t_k)\right),$$

ce qui nous permet, à l'aide de la fonction caractéristique, de reconnaître la loi de $(\mathbb{1}_A, Y_1, \dots, Y_n)$ sous $\overline{\mathbb{P}}$. Ainsi, pour tout $n \ge 1$ et tout $A \in \mathcal{F}_T$, la loi de $(\mathbb{1}_A, Y_1, \dots, Y_n)$ sous $\overline{\mathbb{P}}$ est $\mathrm{Ber}(\overline{\mathbb{P}}(A)) \otimes \mu^{\otimes n}$. Les variables $(Y_n)_{n\ge1}$ sont donc indépendantes de même loi μ et pour tout $A \in \mathcal{F}_T$, on sait que pour tout n, A est indépendant de $\sigma(Y_1, \dots, Y_n)$. On en déduit que A est indépendant de $\sigma((Y_k)_{k\ge1})$ qui est la tribu engendrée par l'algèbre $\cup_{n\ge1}\sigma(Y_1, \dots, Y_n)$. □

Exemple : : soit $(X_n)_{n\ge1}$ une suite de variables aléatoires suivant la loi de Bernoulli de paramètre $1/2$. On note $S_n = X_1 + \dots + X_n$, puis $T = \inf\{n \ge 3; \frac{S_n}{n} \le 1/3\}$. Les moments où $X_n = 1$ représentent les moments où on tire pile dans une série de lancers pile/face avec une pièce équilibrée. Sachant que $T < +\infty$, le théorème dit que les lancers après le temps T sont encore des lancers indépendants pile/face non biaisés, contrairement à une croyance populaire qui voudrait que les lancers « se rattrapent »en tirant davantage de piles pour respecter la loi des grands nombres.

Théorème 74. *Soit $(X_n)_{n\geq 0}$ une chaîne de Markov. On note $(\mathcal{F}_n)_{n\geq 0}$ sa filtration canonique. Soit T un temps d'arrêt adapté à la filtration et $A \in \mathcal{F}_T$ avec $\mathbb{P}(A\cap\{T < +\infty\}) > 0$. On pose alors $\overline{\Omega} = \{T < +\infty\}\cap A$ et $\overline{\mathbb{P}} = \mathbb{P}(\cdot|\overline{\Omega})$ et on définit sur $(\overline{\Omega}, \mathcal{F}, \overline{\mathbb{P}})$: $Y_n(\omega) = X_{n+T(\omega)}(\omega)$. Alors, sur $(\overline{\Omega}, \mathcal{F}, \overline{\mathbb{P}})$, $(Y_n)_{n\geq 1}$ est une chaîne de Markov suivant la même dynamique que $(X_n)_{n\geq 0}$ et de loi initiale $\overline{\mathbb{P}}_{X_T}$.*

En particulier, pour tout E borélien de $\mathbb{R}^{\mathbb{N}}$, on a

$$\overline{\mathbb{P}}(Y \in E) = \mathbb{P}^{\overline{\mathbb{P}}_{X_T}}(X \in E).$$

Démonstration. On peut supposer sans perte de généralité que $(X_n)_{n\geq 0}$ s'écrit sous la forme $X_{n+1} = F(X_n, U_{n+1})$, où $(U_n)_{n\geq 1}$ est une suite de variables aléatoires indépendantes suivant la loi uniforme sur $[0, 1]$. Dans ce cas la filtration canonique des (X_n) coincide avec $\mathcal{F}_n = \sigma(X_0, U_1, U_n)$. Sur $\{T < +\infty\}$, posons $U'_n = U_{T+n}$. Soit B un borélien de $\mathbb{R}^{\mathbb{N}_*}$. D'après le théorème précédent, on a

$$\mathbb{P}(A, U' \in B|T < +\infty) = \mathbb{P}(A|T < +\infty)\mathbb{P}(U \in B),$$

soit

$$\frac{\mathbb{P}(A, U' \in B, T < +\infty)}{\mathbb{P}(T < +\infty)} = \frac{\mathbb{P}(A, T < +\infty)}{\mathbb{P}(T < +\infty)}\mathbb{P}(U \in B),$$

ou encore

$$\mathbb{P}(U' \in B|(A \cap \{T < +\infty\})) = \mathbb{P}(U \in B).$$

Ainsi, sous $\mathbb{P}(\cdot|Y \in B, T < +\infty)$, (U'_n) a la même loi que (U_n). Maintenant, pour tout n, on a $X_{T+n+1} = F(X_{T+n}, U_{T+n+1})$, soit $Y_{n+1} = F(Y_n, U'_n)$: sous $\mathbb{P}(\cdot|Y \in B, T < +\infty)$, (Y_n) est donc une chaîne de Markov avec la même dynamique que (X_n). La loi initiale est la loi de Y_0 sous $\overline{\mathbb{P}}$, soit $\overline{\mathbb{P}}_{X_T}$. □

Comprendre le sens de la propriété de Markov forte demande un certain travail de réflexion, mais une fois cette réflexion faite, on constatera la redoutable efficacité de cette propriété, en particulier lorsque X_T est constante sur l'événement $\{T < +\infty\}$.

Corollaire 22. *Soit S un ensemble dénombrable, $(X_k)_{k\geq 0}$ une chaîne de Markov sur S de matrice de passage P et T un temps d'arrêt adapté à cette suite. Soit A un événement se produisant avant le temps T et $i \in S$.*

Conditionnellement à l'événement $\{T < +\infty\} \cap A' \cap \{X_T = i\}$, la suite $(X_{T+k})_{k\geq 0}$ est une chaîne de Markov de matrice de passage P partant de i.

Ainsi, pour tout borélien B de $S^{\mathbb{N}}$, on a

$$\mathbb{P}(T < +\infty, X_T = i, A', X_{T+.} \in B) = \mathbb{P}(T < +\infty, X_T = i, A')\mathbb{P}^i(X \in B).$$

Démonstration. Il suffit d'appliquer le théorème précédent avec $A = A' \cap \{X_T = i\}$ et remarquer que dans ce cas $\overline{\mathbb{P}}_{X_T} = \delta_i$. □

Corollaire 23. *Soit S un ensemble dénombrable, $(X_k)_{k\geq 0}$ une chaîne de Markov sur S de matrice de passage P et T un temps d'arrêt adapté à cette suite. On suppose que $\mathbb{P}(T < +\infty) > 0$.*

On définit la probabilité conditionnelle $\overline{\mathbb{P}}$ par $\overline{\mathbb{P}}(E) = \frac{\mathbb{P}(E, T<+\infty)}{\mathbb{P}(T<+\infty)}$.

Soit B un borélien de $\mathbb{R}^{\mathbb{N}}$. On a $\overline{\mathbb{P}}$ presque-sûrement :

$$\overline{\mathbb{P}}(X_{T+.} \in B | \mathcal{F}_T) = f_B(X_T), \text{ avec } f_B(x) = \mathbb{P}^x(X \in B).$$

Ces résultats sont souvent employés dans le cas où $\mathbb{P}(T < \infty) = 1$, puisqu'alors $\mathbb{P} = \overline{\mathbb{P}}$.

Démonstration. $f_B(X_T)$ est évidemment $\mathcal{F}_T$-mesurable, donc il s'agit de montrer que $\overline{\mathbb{E}}[\mathbb{1}_{\{X_{T+.}\in B\}}\mathbb{1}_A] = \overline{\mathbb{E}}[f_B(X_T)\mathbb{1}_A]$, soit en multipliant par $\mathbb{P}(T < +\infty)$, que

$$\mathbb{E}[\mathbb{1}_{\{X_{T+.}\in B\}}\mathbb{1}_{A\cap\{T<+\infty\}}] = \mathbb{E}[f_B(X_T)\mathbb{1}_{A\cap\{T<+\infty\}}],$$

soit encore $\mathbb{P}(X_{T+.} \in B | (A \cap \{T < +\infty\}) = \mathbb{E}[f_B(X_T) | (A \cap \{T < +\infty\})]$. Notons $\tilde{P} = \mathbb{P}(\cdot | (A \cap \{T < +\infty\})$. On a, avec les notations du théorème

$$\begin{aligned}
\mathbb{P}(X_{T+.} \in B | (A \cap \{T < +\infty\}) &= \tilde{\mathbb{P}}(Y \in B) \\
&= \mathbb{P}^{\tilde{P}_{X_T}}(X \in B) \\
&= \int \mathbb{P}^i(X \in B) \; d\tilde{\mathbb{P}}_{X_T}(i) \\
&= \int f_B(i) \; d\tilde{\mathbb{P}}_{X_T}(i) \\
&= \int f_B(X_T) \; d\tilde{\mathbb{P}} = \tilde{\mathbb{E}}(f_B(X_T)),
\end{aligned}$$

ce que l'on voulait démontrer. □

10.2 Classification des états

Définition: Soit $P = (p_{i,j})_{(i,j)\in S\times S}$ une matrice stochastique. Pour $i \in S$, on considère une chaîne de Markov $(X_n)_{n\geq 0}$ partant de l'état i et de matrice de passage P. On pose $T_i = \inf\{n \geq 1; X_n = i\}$. Si $\mathbb{P}^i(T_i < +\infty) = 1$, on dit que l'état i est *récurrent*. Inversement, si $\mathbb{P}^i(T_i < +\infty) < 1$, on dit que l'état i est *transient*.

Théorème 75. *Soit $P = (p_{i,j})_{(i,j)\in S\times S}$ une matrice stochastique. Pour $i \in S$, on considère une chaîne de Markov $(X_n)_{n\geq 0}$ partant de l'état i et de matrice de passage P. On pose $T_i = \inf\{n \geq 1; X_n = i\}$ et $N_i = \sum_{k=1}^{+\infty} \mathbb{1}_{\{i\}}(X_k)$. N_i représente le nombre de passage de la chaîne en i à partir de l'instant 1.*

— *Si i est transient, alors $1+N_i$ suit la loi géométrique de paramètre $1-\mathbb{P}^i(T_i<+\infty)$. En particulier N_i est presque sûrement fini et intégrable.*
— *Si i est récurrent, alors N_i est presque sûrement infinie sous $\mathbb{P}^i$. En particulier $\mathbb{E}^i[N_i]=+\infty$.*

Démonstration. Si $T_i<+\infty$ (ou de manière équivalente si $N_i>0$, on a

$$N_i = 1+\sum_{k=1}^{+\infty} \mathbb{1}_{\{i\}}(X_{T+k}).$$

Soit k un entier positif ou nul. On a

$$\mathbb{P}^i(N_i\geq k{+}1) = \mathbb{P}^i(N_i\geq k{+}1, T_i<+\infty) = \mathbb{P}^i(T_i<+\infty)\mathbb{P}^i(\sum_{k=1}^{+\infty}\mathbb{1}_{\{i\}}(X_{T+i}\geq k|T_i<+\infty).$$

Or T_i est un temps d'arrêt. Donc, d'après la propriété de Markov forte, sachant $T_i<+\infty$, $(X_{T+i})_{i\geq 0}$ a la loi d'une chaîne de Markov commençant en i et de matrice de transition P, c'est à dire la même loi que $(X_i)_{i\geq 0}$. On en déduit

$$\mathbb{P}^i(N_i\geq k+1) = \mathbb{P}^i(T_i<+\infty)\mathbb{P}^i(N_i\geq k).$$

Par récurrence, on en déduit

$$\mathbb{P}^i(N_i\geq k) = \mathbb{P}^i(T_i<+\infty)^k$$

D'après le théorème de continuité séquencielle décroissante, on a $P(N_i = +\infty) = \lim_{k\to+\infty}\mathbb{P}^i(N_i\geq k)$. Cette limite vaut donc 0 si i est transient, 1 si i est récurrent. Pour $k\geq 1$, on a

$$\begin{aligned}\mathbb{P}^i(1+N_i=k) &= \mathbb{P}^i(1+N_i\geq k) - P(N_i\geq k)\\ &= \mathbb{P}^i(T_i<+\infty)^{k-1} - \mathbb{P}^i(T_i<+\infty)^k\\ &= \mathbb{P}^i(T_i<+\infty)^{k-1}(1-\mathbb{P}^i(T_i<+\infty)),\end{aligned}$$

ce qui montre que $1+N_i$ suit bien une loi géométrique de paramètre $1-\mathbb{P}^i(T_i<+\infty)$. De plus

$$\mathbb{E}^iN_i = \sum_{k=1}^{+\infty}\mathbb{P}^i(N_i\geq k) = \sum_{k=1}^{+\infty} = \mathbb{P}^i(T_i<+\infty)^k = \frac{\mathbb{P}^i(T_i<+\infty)}{1-\mathbb{P}^i(T_i<+\infty)} < +\infty.$$

□

Corollaire 24. *Un état i est transient si et seulement si*

$$\sum_{k=1}^{+\infty}\mathbb{P}^i(X_k=i)<+\infty.$$

Démonstration. D'après le théorème précédent, i est transient si et seulement si N_i est intégrable sous $\mathbb{P}^i$. Or

$$\begin{aligned}\mathbb{E}^i N_i &= \mathbb{E}^i \sum_{k=1}^{+\infty} \mathbb{1}_{\{i\}}(X_k) \\ &= \sum_{k=1}^{+\infty} \mathbb{E}^i \mathbb{1}_{\{i\}}(X_k) \\ &= \sum_{k=1}^{+\infty} \mathbb{P}^i(X_k = i)\end{aligned}$$

(On utilise le théorème de Tonelli pour échanger la somme et l'espérance) Ceci achève la preuve. □

Corollaire 25. *Soient i et j deux états d'une chaîne de Markov de matrice de transition $P = (p_{i,j})_{(i,j)\in S\times S}$. On suppose que i et j communiquent. Alors i et j sont tous les deux transients ou tous les deux récurrents*

Démonstration. Soit n, p tels que $p_{i,j}^{(n)} > 0$ et $p_{j,i}^{(p)} > 0$. Pour tout $k \geq 0$, on a

$$p_{j,j}^{(n+p+k)} \geq p_{j,i}^{(p)} p_{i,i}^{(k)} p_{i,j}^{(n)}.$$

Ainsi, si la série de terme général $p_{i,i}^{(k)}$ diverge, la série de terme général $p_{j,j}^{(k)}$ aussi. Comme les rôles de i et j sont symétriques, les deux séries sont de même nature. Comme $p_{i,i}^{(k)} = \mathbb{P}^i(X_k = i)$, le résultat découle du corollaire précédent. □

Corollaire 26. *Considérons une chaîne de Markov irréductible de matrice de transition $P = (p_{i,j})_{(i,j)\in S\times S}$ et pour tous les $i \in S$, notons $\mathbb{P}^i$ les lois markoviennes correspondantes. Les propriétés suivantes sont équivalentes*

1. $\exists i, j \in S\ \mathbb{P}^j(N_i = +\infty) > 0.$
2. $\exists i \in S$, *i est récurrent*
3. $\forall i \in S$, *i est récurrent*
4. $\forall i, j \in S\ \mathbb{P}^j(N_i = +\infty) = 1.$

Démonstration.
— (1) $\Longrightarrow$ (2). Soit l tel que $\mathbb{P}^i(X_l = j) > 0$. On a $\mathbb{P}^i(N_i = +\infty) \geq \mathbb{P}^i(X_l = j, \sum_{k=1}^{+\infty} \mathbb{1}_{\{X_{k+l}=i\}}) \geq \mathbb{P}^i(X_l = j)\mathbb{P}^j(N_i = +\infty) > 0$, donc i est récurrent.
— (2) $\Longrightarrow$ (3). C'est une conséquence du corollaire précédent.

— (3) $\Longrightarrow$ (4). Considérons $\mathbb{P}^i(T_j < +\infty, \forall k > T_j \quad X_k \neq i)$. Comme i et j communiquent $\mathbb{P}^i(T_j < +\infty) > 0)$. D'après la propriété de Markov forte, on a

$$\begin{aligned} \mathbb{P}^i(T_j < +\infty, \forall k > T_j, X_k \neq i) &= \mathbb{P}^i(T_j < +\infty)\mathbb{P}^j(\forall k > 0 X_k \neq i) \\ &= \mathbb{P}^i(T_j < +\infty)\mathbb{P}^j(T_i = +\infty) \end{aligned}$$

Mais $\{T_j < +\infty, \forall k > T_j X_k \neq i\} \subset \{N_i < +\infty\}$ et, comme i est récurrent, $\mathbb{P}^i(N_i < +\infty) = 0$, donc $\mathbb{P}^i(T_j < +\infty)\mathbb{P}^j(T_i = +\infty) = 0$. Comme $\mathbb{P}^i(T_j < +\infty) > 0$, on a finalement $\mathbb{P}^j(T_i = +\infty) = 0$. Mais

$$N_i = 1 + \sum_{k=1}^{+\infty} \mathbb{1}_{\{i\}}(X_{k+T_i}),$$

Donc d'après la propriété de Markov forte

$$\mathbb{P}^j(N_i = +\infty) = \mathbb{P}^i(\sum_{k=1}^{+\infty} \mathbb{1}_{\{i\}}(X_k) = +\infty) = \mathbb{P}^i(N_i = +\infty) = 1.$$

— (4) $\Longrightarrow$ (1). Évident.

□

Définition: Si une chaîne de Markov vérifie une des 4 propriétés équivalentes ci-dessus, on dit que c'est une *chaîne récurrente*.

10.3 Mesures invariantes

Définition: On dit qu'une mesure μ est *invariante* sous l'action de la matrice de transition markovienne M si $\mu M = \mu$, c'est à dire.

$$\forall j \in S \quad \sum_{i \in S} \mu(i) m_{i,j} = \mu(j).$$

Si μ est *invariante* sous l'action de M, une récurrence immédiate donne $\forall n \geq 0 \quad \mu M^n = \mu$. Ainsi, si $(X_n)_{n \geq 0}$ est une chaîne de Markov de matrice de transition M et de mesure initiale $\mathbb{P}_{X_0} = \mu$, alors pour tout n, la loi de X_n est $\mathbb{P}_{X_n} = \mu$.

Définition: On dit qu'une mesure μ est réversible sous l'action de la matrice de transition markovienne M si

$$\forall i, j \in S \quad \mu(i) m_{i,j} = \mu(j) m_{j,i}.$$

Par extension, on dit d'une chaîne de Markov dont la loi au temps zéro est une mesure de probabilité réversible sous l'action de la matrice de transition de la chaîne qu'elle est une chaîne réversible.

Théorème 76. *Soit $(X_n)_{n\geq 0}$ une chaîne de Markov de loi initiale ν réversible sous l'action de M. Alors*

$$\forall n \geq 1 \quad (X_0, X_1, \dots X_n) \text{ et } (X_n, X_{n-1}, \dots, X_0) \text{ ont même loi sous } \mathbb{P}^\nu.$$

Démonstration. Il suffit de montrer par récurrence sur n que $\forall (x_0, \dots x_n) \in S^{n+1}$, on a

$$\mathbb{P}^\nu(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}^\nu(X_0 = x_n, X_1 = x_{n-1}, \dots, X_n = x_0).$$

Pour $n = 1$, il suffit de voir que

$$\mathbb{P}^\nu(X_0 = x_0, X_1 = x_1) = \nu(x_0) m_{x_0,x_1} = \nu(x_1) m_{x_1,x_0} = \mathbb{P}^\nu(X_0 = x_1, X_1 = x_0).$$

Ensuite

$$\begin{aligned}
\mathbb{P}^\nu(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) &= \mathbb{P}^\nu(X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}) m_{x_{n-1},x_n} \\
&= m_{x_{n-1},x_n} \mathbb{P}^\nu(X_0 = x_{n-1}, X_1 = x_{n-2}, \dots, X_{n-1} = x_0) \\
&= m_{x_{n-1},x_n} \nu(x_{n-1}) \prod_{i=1}^{n-1} m_{x_{n-i},x_{n-i-1}} \\
&= \nu(x_{n-1}) m_{x_{n-1},x_n} \prod_{i=1}^{n-1} m_{x_{n-i},x_{n-i-1}} \\
&= \nu(x_n) m_{x_n,x_{n-1}} \prod_{i=1}^{n-1} m_{x_{n-i},x_{n-i-1}} \\
&= \nu(x_n) \prod_{i=0}^{n-1} m_{x_{n-i},x_{n-i-1}} \\
&= \mathbb{P}^\nu(X_0 = x_n, X_1 = x_{n-1}, \dots, X_n = x_0).
\end{aligned}$$

□

Il est facile de voir que toute mesure réversible est invariante.

Théorème 77. *Si la matrice de transition M est irréductible et admet une probabilité μ invariante, alors les chaînes de Markov associées à M sont récurrentes. De plus, μ charge tous les points de l'espace d'états.*

Démonstration. Soit μ une probabilité invariante Pour tout $n \geq 0$, on a $\mu M^n = \mu$, soit

$$\forall j \in S \quad \forall n \geq 0 \quad \sum_{i \in S} \mu(i) m_{i,j}^{(n)} = \mu(j)$$

Si une chaîne de Markov irréductible n'est pas récurrente, les états sont tous transitoires et $\lim_{n\to+\infty}\mu(i)m_{i,j}^{(n)} = 0$ quels que soient i et j. D'après le théorème de convergence dominée, on a alors

$$\forall j \in S \quad 0 = \mu(j),$$

ce qui est impossible. Le premier point est donc démontré. Prenons maintenant $x \in E$ tel que $\mu(x) > 0$ et soit y un autre élément de E. Il existe un entier n et une suite $x = x_0, x_1, \dots x_n = y$ d'éléments de E avec pour tout i entre 0 et $n-1$, $m_{x_i,x_{i+1}} > 0$. Ainsi

$$\mu(y) = \mathbb{P}^\mu(X_n = y) \ge \mathbb{P}^\mu(X_0 = x_0, X_1 = x_1, \dots X_n = x_n) = \mu(x)\prod_{i=0}^{n-1} m_{x_i,x_{i+1}} > 0.$$

□

Théorème 78. *Toute chaîne de Markov sur un espace d'états S fini admet une probabilité invariante.*

Démonstration. L'ensemble $\mathcal{M}(S)$ des mesures de probabilité sur S s'identifie au compact $K = \{(x_1, \dots, x_n) \in \mathbb{R}_+^n; \sum_{k=1}^n x_k = 1\}$, avec $n = |S|$. $\mathcal{M}(S)$ est un convexe stable par $\mu \mapsto \mu M$. Ainsi, si μ est une mesure quelconque sur S, la suite $(\mu_n)_{n\ge0}$ définie par

$$\mu_n = \frac{1}{n}\sum_{k=0}^{n-1}\mu M^k$$

est à valeurs dans $\mathcal{M}(S)$. On a $\mu_n(I - M) = \frac{\mu(I-M^n)}{n}$. Comme la suite $(\mu I - M^n)_{n\ge0}$, est bornée, il s'ensuit que toute valeur d'adhérence de $(\mu_n)_{n\ge0}$ est laissée fixe par M. Comme $\mathcal{M}(S)$ est compacte, $(\mu_n)_{n\ge0}$ a au moins une valeur d'adhérence donc M au moins une mesure invariante. □

Corollaire 27. *Une chaîne de Markov irréductible dont l'espace d'états est fini est récurrente.*

Remarque-exercice : S'il est vrai qu'une chaîne de Markov avec un espace d'état fini admet toujours une mesure invariante ; en revanche elle n'admet pas toujours de probabilité réversible. Voici un exemple simple. Soit $N \ge 3$ un entier et $p \in]0,1]$. Considérons une suite $(X_n)_{n\ge1}$ de variables indépendantes à valeurs dans $\mathbb{Z}/N\mathbb{Z}$ avec $\mathbb{P}(X_n = 1) = 1 - \mathbb{P}(X_n = 0) = p$. On pose $S_n = X_1 + \dots + X_n$. On peut démontrer que $(S_n)_{n\ge1}$ est une chaîne de Markov, qui admet la probabilité uniforme comme probabilité invariante, mais la seule mesure réversible pour la dynamique est la mesure nulle.

10.4 Théorème de la probabilité stationnaire

Théorème 79. *Soit M la matrice de transition d'une chaîne de Markov irréductible apériodique admettant μ comme loi stationnaire. Alors pour toute loi ν sur S, la chaîne de Markov de matrice de transition M et de loi initiale ν converge vers μ.*

Démonstration. Soit X_0, X'_0 deux variables aléatoires indépendantes, X_0 suivant la loi μ, X'_0 la loi ν. On note également $Y_0 = X'_0$. Soit également $(f_n)_{n\geq 1}$ et $(f'_n)_{n\geq 1}$ deux suites de variables aléatoires i.i.i.d. de loi χ_M définie au lemme 1, ces deux suites étant indépendantes de X_0 et X'_0. On définit par récurrence les suites $(g_n)_{n\geq 1}, (X_n)_{n\geq 1}$ et $(X'_n)_{n\geq 1}, (Y_n)_{n\geq 1}$ par

$$\begin{cases} X_{n+1} &= f_{n+1}(X_n) \\ Y_{n+1} &= f'_{n+1}(Y_n) \\ g_{n+1} &= \begin{cases} f_{n+1} & \text{si } X_n = X'_n \\ f'_{n+1} & \text{sinon} \end{cases} \\ X'_{n+1} &= g_{n+1}(X'_n) \end{cases}$$

Il n'est pas difficile de voir qu'en tout point ω on a

$$(X_n(\omega) = X'_n(\omega)) \Longrightarrow (f_{n+1}(\omega) = g_{n+1}(\omega)) \Longrightarrow (X_{n+1}(\omega) = X'_{n+1}(\omega))$$

Ainsi, les processus X_n et X'_n évoluent de manière indépendante jusqu'au moment où ils se rencontrent. À partir de là, X'_n demeure scotché à X_n.

Lemme 14. *Soit $(X_n)_{n\geq 0}$ une chaîne de Markov de matrice de transition M et de loi initiale μ , $(Y_n)_{n\geq 0}$ une chaîne de Markov de matrice de transition N et de loi initiale ν. On suppose en outre que les suites $(X_n)_{n\geq 0}$ et $(Y_n)_{n\geq 0}$ sont indépendantes sous P. Alors la suite $(Z_n)_{n\geq 0}$ définie par $Z_n = (X_n, Y_n)$ est une chaîne de Markov de matrice de transition $M \otimes N$, où $M \otimes N$ est définie par*

$$\forall ((i,j),(k,l)) \in S^2 \times S^2 \quad (M \otimes N)(i,j),(k,l) = M(i,k)N(j,l).$$

Démonstration. Soient $(x_0, \dots, x_n) \in S^{n+1}$ et $(y_0, \dots, y_n) \in S^{n+1}$.

$$\begin{aligned}
\mathbb{P}(\forall i \in \{0,n\}(X_i, Y_i) = (x_i, y_i)) &= \mathbb{P}(\{\forall i \in \{0,n\}X_i = x_i\} \cap \{\forall i \in \{0,n\}Y_i = y_i\}) \\
&= \mathbb{P}(\forall i \in \{0,n\}X_i = x_i)\mathbb{P}(\forall i \in \{0,n\}Y_i = y_i) \\
&= \mu(\{x_0\}) \prod_{i=0}^{n-1} m_{x_i,x_{i+1}} \times \nu(\{y_0\}) \prod_{i=0}^{n-1} n_{y_i,y_{i+1}} \\
&= \mu(\{x_0\})\nu(\{y_0\}) \prod_{i=0}^{n-1} m_{x_i,x_{i+1}} n_{y_i,y_{i+1}} \\
&= (\mu \otimes \nu)(\{x_0, y_0\}) \prod_{i=0}^{n-1} (M \otimes N)((x_i, y_i), (x_{i+1}, y_{i+1}))
\end{aligned}$$

□

Lemme 15. *Soit U, V deux variables aléatoires de loi θ. On suppose que sous $\mathbb{P}$, U et V sont indépendantes de la tribu $\mathcal{A}$. Soit A un événement $\mathcal{A} -$ mesurable. On définit W par*

$$W(\omega) = \begin{cases} U(\omega) \ si \ \omega \in A \\ V(\omega) \ si \ \omega \notin A \end{cases}$$

Alors, sous $\mathbb{P}$, W suit la loi θ et W est indépendante de $\mathcal{A}$.

Démonstration. Soit A' un événement $\mathcal{A} - mesurable$ et B un borélien

$$\begin{aligned}
\mathbb{P}(A' \cap \{W \in B\}) &= \mathbb{P}(A \cap A' \cap \{W \in B\}) + \mathbb{P}(A^c \cap A' \cap \{W \in B\}) \\
&= \mathbb{P}(A \cap A' \cap \{U \in B\}) + \mathbb{P}(A^c \cap A' \cap \{V \in B\}) \\
&= \mathbb{P}(A \cap A')\mathbb{P}(U \in B) + \mathbb{P}(A^c \cap A')\mathbb{P}(V \in B) \\
&= \mathbb{P}(A \cap A')\theta(B) + \mathbb{P}(A^c \cap A')\theta(B) \\
&= (\mathbb{P}(A \cap A') + \mathbb{P}(A^c \cap A'))\theta(B) \\
&= \mathbb{P}(A')\theta(B)
\end{aligned}$$

En prenant $A' = \Omega$, on en déduit d'abord que $\mathbb{P}(W \in B) = \theta(B)$ pour tout borélien B. θ est donc la loi de W sous $\mathbb{P}$. En réinsérant dans la formule précédente, on a pour tout événement $\mathcal{A} - mesurable$ A' et pour tout borélien B :

$$\mathbb{P}(A' \cap \{W \in B\}) = \mathbb{P}(A')\mathbb{P}(W \in B),$$

ce qui veut dire que W est indépendante de $\mathcal{A}$. □

En appliquant le lemme précédent à $\mathcal{A} = \sigma(X_0, X'_0, f_1, \dots, f_n, f'_1, \dots, f'_n)$, $A = \{X_n = X'_n\}$, $U = f_{n+1}$, $V = f'_{n+1}$ et $W = g_{n+1}$ on voit que g_{n+1} suit la loi χ_M et que g_{n+1} est indépendante de $\sigma(X_0, X'_0, f_1, \dots, f_n, f'_1, \dots, f'_n)$. Comme $(g_1, \dots, g_n)$ est $\sigma(X_0, X'_0, f_1, \dots, f_n, f'_1, \dots, f'_n)$-mesurable, il s'ensuit que $(g_n)_{n\ge 1}$ est une suite de v.a.i.i.d de loi χ_M.

D'après le lemme 11, (X_n) est une chaîne de Markov de matrice de transition M et de loi initiale μ tandis que (X'_n) est une chaîne de Markov de matrice de transition M et de loi initiale ν.

On va maintenant montrer que $\tau = \inf\{n; X_n = X'_n\}$ est presque sûrement fini. Il est facile de voir que $\tau = \inf\{n; X_n = Y_n\}$. Ce qui est intéressant, c'est que $(X_n)_{n\ge 0}$ et $(Y_n)_{n\ge 0}$ sont indépendants.

Ainsi, d'après le lemme 14, (X_n, Y_n) est une chaîne de Markov de loi initiale $\nu \otimes \mu$ et de matrice de transition $M' = M \otimes M$. Soient $(x, y, z, t) \in S^4$. Comme M est la matrice d'une chaîne de Markov irréductible et apériodique, on peut, d'après le corollaire 19, trouver un entier $n_0 = \max(N(x, z), N(z, t))$ tel que $M^{n_0}(x, z)$ et $M^{n_0}(y, t)$ soient strictement positifs. Or $M'^{n_0} = (M \otimes M)^{n_0} = M^{n_0} \otimes M^{n_0}$: on a

$$M'^{n_0}((x, y), (z, t)) = M^{n_0}(x, z) M^{n_0}(y, t) > 0.$$

Ainsi $((Z_n)_{n\ge 0} = ((X_n, Y_n))_{n\ge 0}$ est une chaîne de Markov irréductible. Comme $M \otimes M$ admet $\mu \otimes \mu$ comme mesure invariante, la dynamique est donc récurrente : $(Z_n)_{n\ge 0}$ passe donc presque sûrement en tout point de $S \times S$. En particulier, elle passe presque sûrement sur sa diagonale, ce qui implique que $\mathbb{P}(\tau < +\infty) = 1$.

Soit f une fonction bornée de S dans $\mathbb{R}$. Pour $n \ge \tau$, on a $f(X_n) = f(X'_n)$. Donc $f(X_n) - f(X'_n)$ converge presque sûrement vers 0. D'après le théorème de convergence dominée, on en déduit que $\mathbb{E}(f(X_n) - f(X'_n))$ converge vers 0. Comme μ est invariante $\mathbb{E}(f(X_n) - f(X'_n)) = \int f \, d\mu - \mathbb{E} f(X'_n)$. Ainsi pour toute fonction f, $\mathbb{E} f(X'_n)$ converge vers $\int f \, d\mu$, ce qui veut dire que X'_n converge en loi vers μ.

□

Remarque-exercice : l'hypothèse d'apériodicité est importante. En effet, on peut construire deux chaînes de Markov indépendantes $(X_n)_{n\ge 0}$ et $(Y_n)_{n\ge 0}$ ayant la même matrice de transition irréductibles, telles que $(X_n, Y_n)_{n\ge 0}$ ne soit pas irréductible et que $(X_n, Y_n)_{n\ge 0}$ ne coupe jamais la diagonale. Donner deux exemples d'un tel phénomène, l'un avec S fini, l'autre avec S infini.

10.5 Théorème ergodique des chaînes de Markov

10.5.1 Convergence presque sûre des fréquences empiriques

Théorème 80. *Soit $(X_n)_{n\geq 0}$ une chaîne de Markov. Pour tout $x \in S$, on a*

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{\{x\}}(X_k) = \frac{\mathbb{1}_{\{T_x<+\infty\}}}{\mathbb{E}^x T_x} \quad p.s.$$

En particulier, si la chaîne est irréductible, on a

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{\{x\}}(X_k) = \frac{1}{\mathbb{E}^x T_x}.$$

Démonstration. Si $T_x = +\infty$, l'égalité est évidente, car $\frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{\{x\}}(X_k) = \frac{1}{n}\mathbb{1}_{\{X_0=x\}}$ pour tout n. Si $T_x < +\infty$, la suite de terme général $\frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{\{x\}}(X_k)$ a le même comportement asymptotique que la suite de terme général $\frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{\{x\}}(X_{T_x+k})$. Or, d'après la propriété de Markov forte, la loi de la suite de terme général $\frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{\{x\}}(X_{T_x+k})$ sous $\mathbb{P}$ est la même que la loi de la suite de terme général $\frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{\{x\}}(X_k)$ sous $\mathbb{P}^x$: on est ramené à étudier le cas où $\mathbb{P} = \mathbb{P}^x$.

Si x n'est pas récurrent, on a $\mathbb{E}^x[T_x] = +\infty$ et $\sum_{k=0}^{+\infty} \mathbb{1}_{\{x\}}(X_k)$ est fini $\mathbb{P}^x$-presque sûrement : cela donne l'identité voulue.

Passons donc au cas où x est récurrent. Pour tout $k \geq 1$, posons $T^k = \inf\{n > 0, \sum_{i=1}^{n} \mathbb{1}_{\{x\}}(X_i) \geq k\}$. Les T^k sont des temps d'arrêt adaptés à la filtration naturelle engendrée par $(X_n)_{n\geq 0}$. Comme x récurrent, les T^k sont presque sûrement finis. Il est aisé de constater que la suite $(T^k)_{k\geq 1}$ est croissante. Pour tout $i \in \{1, \dots k\}$, T^i est $\mathcal{F}_{T^i}$-mesurable (voir chapitre sur les martingales). Comme $T^i \leq T^k$, on a $\mathcal{F}_{T^i} \subset \mathcal{F}_{T^k}$. Finalement, $\sigma(T^1, \dots, T^k)$ est une sous-tribu de $\mathcal{F}_{T^k}$. Soit $k \geq 1$ et $A \in \sigma(T^1, \dots, T^k)$: il est clair que A se produit avant T^k. Ainsi, on va pouvoir utiliser la propriété de Markov forte :

$$\begin{aligned}
\mathbb{P}^x(A, T^{k+1} - T_k > n) &= \mathbb{P}^x(A, \cap_{j=T^k+1}^{T^k+n} \mathbb{1}_{\{x\}}(X_j) = 0) \\
&= \mathbb{P}^x(A)\mathbb{P}^x(\cap_{j=1}^{n} \mathbb{1}_{\{x\}}(X_j) = 0) \\
&= \mathbb{P}^x(A)\mathbb{P}^x(T^1 > n)
\end{aligned}$$

On en déduit que, sous la loi $\mathbb{P}^x$, les variables aléatoires $T^1, T^2 - T^1, T^3 - T^2, \dots$ forment une suite de variables aléatoires positives indépendantes ayant même loi que $T^1 = T_x$. D'après la loi forte des grands nombres on en déduit que $\frac{T^n}{n} = \frac{1}{n}\Big(T^1 + (T^2 - T^1) + (T^3 - T^2) + \dots (T^n - T^{n-1}\Big)$ converge presque sûrement vers $\mathbb{E}^x T^x$. Le résultat demeure si $\mathbb{E}^x[T^x] = +\infty$ (exercice classique de troncature) Posons $S_n = \sum_{k=0}^{n-1} \mathbb{1}_{\{x\}}(X_k)$. Un instant de réflexion montre que $T^{S_n} \le n < T^{S_n+1}$. On en déduit

$$\frac{T^{S_n}}{S_n} \le \frac{n}{S_n} < \frac{T^{S_n+1}}{S_n+1}\frac{S_n+1}{S_n}$$

Si x est récurrent $\lim_{n\to+\infty} S_n = +\infty$, donc $\lim_{n\to+\infty} \frac{n}{S_n} = \mathbb{E}^x T_x$, d'où le résultat. Si x est transient, l'inégalité $\frac{T^{S_n}}{S_n} \le \frac{n}{S_n}$ suffit a donner la convergence de S_n/n vers 0. □

Remarque. On peut observer que la limite presque sûre apparaissant dans ce théorème dépend assez peu de l'état initial de la chaîne. Dans le cas irréductible, $\frac{1}{\mathbb{E}^x[T_x]}$ représente la proportion asymptotique du temps passé par la chaîne dans l'état x.

Définition. Si x est un état récurrent, on convient de dire que x est *récurrent positif* si $\frac{1}{\mathbb{E}^x[T_x]} > 0$ (soit $\mathbb{E}^x[T_x] < +\infty$), et que x est *récurrent nul* si $\frac{1}{\mathbb{E}^x[T_x]} = 0$ (soit $\mathbb{E}^x[T_x] = +\infty$). [1]

10.5.2 Fréquences empiriques et probabilités invariantes

Théorème 81. *Soit $(X_n)_{n\ge 0}$ une chaîne de Markov irréductible admettant une probabilité invariante μ. Pour tout $x \in S$, on a*

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{\{x\}}(X_k) = \frac{1}{\mathbb{E}^x T_x} = \mu(x) > 0.$$

Démonstration. Une chaîne de Markov irréductible admettant une probabilité invariante est toujours récurrente (voir Théorème 77). Le théorème ergodique des chaînes de Markov s'applique donc, et on a pour toute loi initiale ν :

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{\{x\}}(X_k) = \frac{1}{\mathbb{E}^x T_x}\mathbb{P}^\nu\text{p.s.}$$

Comme $|\frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{\{x\}}(X_k)| \le 1$, le théorème de convergence dominée s'applique et on a

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{P}^\nu(X_k = x) = \frac{1}{\mathbb{E}^x T_x}.$$

1. On verra plus tard que pour une chaîne de Markov à espace d'états fini, les états récurrents sont tous récurrents positifs. Les notions de récurrence positive et de récurrence nulle que l'on vient d'introduire n'ont donc pas vraiment de pertinence dans le cas de l'étude des chaînes de Markov à espace d'état finis.

Si l'on prend pour ν la mesure invariante μ, on a pour tout $k \geq 0$ $\mathbb{P}^\nu(X_k = x) = \mu(x)$. On en déduit que $\frac{1}{\mathbb{E}^x T_x} = \mu(x)$, ce qui achève la preuve, puisque le fait qu'une mesure invariante d'une chaîne de Markov irréductible doive charger tous les points a déjà été démontré. □

Corollaire 28. *Une chaîne de Markov irréductible a au plus une probabilité invariante.*

Corollaire 29. *Une chaîne de Markov irréductible dont l'espace d'état est fini a exactement une mesure invariante. Ses états sont tous récurrents positifs.*

Démonstration. On sait déjà que tous les états d'une chaîne de Markov irréductible dont l'espace d'état est fini sont récurrents. On sait également qu'une chaîne de Markov sur un espace d'états fini admet au moins une musure invariante. D'après le précédent corollaire, cette mesure est unique. En réappliquant le théorème, on voit que les états sont tous récurrents positifs. □

Le théorème 81 admet une réciproque.

Théorème 82. *Soit $(X_n)_{n\geq 0}$ une chaîne de Markov irréductible et récurrente sur S. On suppose qu'il existe $x \in S$ tel que $\mathbb{E}^x[T_x] < +\infty$.*

Alors, la chaîne admet une unique probabilité invariante μ qui est donnée par $\forall y \in S \quad \mu(y) = \frac{1}{\mathbb{E}^y[T_y]} > 0$. Dans ce cas, on dit que la chaîne est récurrente positive.

Démonstration. Posons $m(y) = \frac{1}{\mathbb{E}^y[T_y]}$. m est une mesure positive sur S. Ce n'est pas la mesure nulle car $m(x) > 0$. Ainsi, si l'on montre que la mesure m est invariante sous la dynamique et que m est une mesure finie, la probabilité $\mu = m/m(S)$ sera une probabilité invariante sous la dynamique. D'après le théorème 81, on aura $\mu(y) = \frac{1}{\mathbb{E}^y[T_y]}$ pour tout y. De plus, d'après le théorème 77, $\mu(y) > 0$ pour tout y.

Montrons déjà que m est finie : soit S' une partie finie de S.

On a

$$\frac{1}{n}\sum_{k=0}^{n-1}\sum_{x\in S'} \mathbb{1}_x(X_k) \leq 1,$$

d'où en faisant tendre n vers l'infini, on a avec le théorème 80 :

$$\sum_{x\in S'} m(x) \leq 1,$$

ce qui montre que la somme des $m(x)$ est finie.

Il suffit alors de montrer que pour pour $x \in S$, on a

$$\sum_{y\in S} m(y)p_{y,x} = m(x).$$

Posons $Y_n = \frac{1}{n}\sum_{k=1}^n \mathbb{1}_{\{x\}}(X_k)$. On a vu que Y_n/n tend presque sûrement vers $m(x)$, donc par convergence dominée, $\mathbb{E}[Y_n]/n$ tend vers $m(x)$. On a aussi

$$\frac{Y_n}{n} = \sum_{y\in S}\frac{1}{n}\sum_{k=1}^n \mathbb{1}_{\{y\}}(X_{k-1})\mathbb{1}_{\{x\}}(X_k)$$

d'où

$$\begin{aligned}\frac{\mathbb{E}[Y_n]}{n} &= \sum_{y\in S}\frac{1}{n}\sum_{k=1}^n \mathbb{P}(X_{k-1}=y, X_{X_k}=x)\\ &= \sum_{y\in S}\frac{1}{n}\sum_{k=1}^n \mathbb{P}(X_{k-1}=y)p_{y,x}\end{aligned}$$

Soit S' une partie finie de S : on a

$$\frac{\mathbb{E}[Y_n]}{n} \ge \sum_{y\in S'}\frac{1}{n}\sum_{k=1}^n \mathbb{P}(X_{k-1}=y)p_{y,x},$$

et en faisant tendre n vers l'infini

$$m(x) \ge \sum_{y\in S'} m(y)p_{y,x},$$

d'où en passant au sup

$$m(x) \ge \sum_{y\in S} m(y)p_{y,x}.$$

Cependant

$$\sum_{x\in S}\sum_{y\in S} m(y)p_{y,x} = \sum_{y\in S}\sum_{x\in S} m(y)p_{y,x} = \sum_{y\in S} m(y)1$$

Ce qui entraîne donc que pour tout x, on a bien

$$m(x) = \sum_{y\in S'} m(y)p_{y,x},$$

ce qui achève la preuve. □

Bien sûr, cette réciproque est sans intérêt lorsque l'espace d'état est fini, puisque l'existence de la mesure invariante est assurée d'emblée.

On donne maintenant une preuve alternative, qui permet parfois de calculer la mesure invariante.

10.5.3 Calcul d'une mesure invariante à partir de la loi des trajectoires issues d'un point

Théorème 83. *Soit (X_n) une chaîne de Markov sur E et $x \in S$ tel que $\mathbb{E}^x[T_x] < +\infty$. Si l'on pose $N_y^x = \sum_{k=0}^{T_x-1} \mathbb{1}_{\{X_k=y\}}$, alors la mesure μ^x définie par*

$$\mu^x(y) = \frac{\mathbb{E}^x[N_y^x]}{\mathbb{E}^x[T_x]}.$$

est une mesure de probabilité invariante. En particulier $\mu^x(x) = \frac{1}{\mathbb{E}^x[T_x]}$

Démonstration. On pose, $S_0 = 0$, et pour $n \geq 1$: $S_n^y = \sum_{k=0}^{n-1} \mathbb{1}_{\{X_{k+1}=y\}} - p_{X_k,y}$. S_n^y est $\mathcal{F}_n$-mesurable et $S_n^y = S_{n-1}^y + \mathbb{1}_{\{X_n=y\}} - p_{X_{n-1},x}$, donc

$$\mathbb{E}[S_n^y|\mathcal{F}_{n-1}] = S_{n-1}^y + \mathbb{P}(X_n = x|\mathcal{F}_{n-1}) - p_{X_{n-1},x} = S_{n-1}^y,$$

Ainsi $(S_n^y)_{n\geq 1}$ est une martingale[2]. Comme T^x est un temps d'arrêt, le théorème d'arrêt dit que $(S_{n\wedge T_x}^y)_{n\geq 1}$ est aussi une martingale, en particulier $\mathbb{E}^x[S_{n\wedge T_x}^y] = 0$ pour tout n. $S_{n\wedge T_x}^y$ tend presque sûrement vers $S_{T_x}^y$. Comme $|S_{n\wedge T_x}^y| \leq T_x$, le théorème de convergence dominée nous donne $\mathbb{E}^x[S_{T_x}^y] = 0$. Cependant,

$$\begin{aligned} S_{T_x}^y &= \sum_{k=0}^{T_x-1} \mathbb{1}_{\{X_{k+1}=y\}} - p_{X_k,y} \\ &= \sum_{k=1}^{T_x} \mathbb{1}_{\{X_k=y\}} - \sum_{k=0}^{T_x-1} \sum_{z\in S} \mathbb{1}_{\{X_k=z\}} p_{z,y} \\ &= \sum_{k=1}^{T_x} \mathbb{1}_{\{X_k=y\}} - \sum_{z\in S} N_z^x p_{z,y} \\ &= -\mathbb{1}_{\{X_0=y\}} + \mathbb{1}_{\{x=y\}} + N_y^x - \sum_{z\in S} N_z^x p_{z,y} \end{aligned}$$

En prenant l'espérance sous $\mathbb{P}^x$, on obtient, pour tout $y \in S$:

$$0 = \mathbb{E}^y[N_y^x] - \sum_{z\in S} \mathbb{E}^x[N_z^x] p_{z,y},$$

2. Cette martingale ne sort pas de nulle part. Dans le corollaire 17, nous avions remarqué que pour une chaîne de Markov $(X_n)_{n\geq 0}$ de matrice de passage P et une fonction f, la suite $(Y_n)_{n\geq 1}$ définie par

$$\forall n \geq 0 \quad Y_{n+1} = f(X_{n+1}) - (Pf)(X_n)$$

est une suite de différences de martingales adaptée à la filtration canonique des X_i. Ici, on a simplement pris $f = \mathbb{1}_i$.

ce qui dit bien que $(\mathbb{E}^x[N_z^x])_{z\in E}$ est une mesure invariante. Pour avoir une mesure de probabilité, il suffit de diviser par

$$\sum_{z\in E}\mathbb{E}^x[N_z^x]=\mathbb{E}^x[\sum_{z\in E}N_z^x]=\mathbb{E}^x[T_x].$$

Enfin, comme $N_x^x=1$, $\mathbb{P}^x$-presque sûrement, on a $\mu^x(x)=\frac{1}{\mathbb{E}^x[T_x]}$. □

10.6 Retour à la classification des états (*)

Considérons une chaîne de Markov dont l'espace d'états est S On dit que x est un état essentiel si

$$\forall x\in S\quad (x\to y)\Longrightarrow (y\to x).$$

Il n'est pas difficile de voir qu'un état accessible depuis un état essentiel est lui-même un état essentiel.

Démonstration. Supposons en effet que x est essentiel et que $x\to y$. Soit z tel que que $y\to z$. On a $(x\to y)$ et $(y\to z)$ donc $(x\to z)$. Comme x est essentiel, $(z\to x)$, or $(x\to y)$, donc $(z\to y)$, ce qui montre que y est essentiel. □

Soit $(A_i)_{i\ I}$ la partition de l'ensemble des points essentiels induite par la relation d'équivalence "communique" ($\leftrightarrow$). Chaque ensemble A_i est appelé une *classe absorbante*. D'un point x appartenant à la classe absorbante A_i, on ne peut accéder qu'à des points de A_i.

Démonstration. En effet si $x\to y$, alors $y\to x$ car x est essentiel et y est essentiel : ainsi y est essentiel et $x\leftrightarrow y$, donc x et y sont dans la même classe d'équivalence : $y\in A_i$. □

Théorème 84. *Les points récurrents d'une chaîne de Markov sont terminaux.*

Démonstration. Soient i un état non terminal. Il existe j tel que $i\to j$ mais que j ne communique pas avec i. Soit n tel que $\mathbb{P}^i(X_n=j)>0$

$$\begin{aligned}\mathbb{P}^i(N_i<+\infty)\ &\geq\ \mathbb{P}^i(X_n=j,\forall k\geq n\quad X_k\neq i)\\ &=\mathbb{P}^i(X_n=j)\mathbb{P}^j(\forall k\geq 0\quad X_k\neq i)=\mathbb{P}^i(X_n=j).1>0,\end{aligned}$$

ce qui montre que i n'est pas récurrent. □

Théorème 85. *Une mesure de probabilité invariante d'une chaîne de Markov ne charge que des états récurrents positifs.*

Démonstration. Soit μ une mesure invariante et x chargé par μ. Posons $M_n = \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{\{x\}}(X_k)$. D'après le théorème 80, M_n converge $\mathbb{P}^\mu$-presque sûrement vers $\frac{\mathbb{1}_{\{T_x<+\infty\}}}{\mathbb{E}^x T_x}$, donc d'après le théorème de convergence dominée $\mathbb{E}^\mu[M_n]$ converge vers $\frac{\mathbb{P}^\mu(T_x<+\infty)}{\mathbb{E}^x T_x}$. Comme $\mathbb{E}^\mu[M_n] = \mu(x)$ pour tout n, on a $\mu(x) = \frac{\mathbb{P}^\mu(T_x<+\infty)}{\mathbb{E}^x T_x}$. Comme $\mu(x) > 0$, on a $\mathbb{E}^x[T_x] < +\infty$. □

Théorème 86. *Soit $p_{i,j}$ une matrice markovienne. Soient $(A_i)_{i\in I}$ les classes absorbantes associées à la chaîne, et $J \subset I$ l'ensemble des x tels que la chaîne de Markov de matrice $(p_{i,j})_{i,j\in A_x}$ soit récurrente positive. La chaîne $(p_{i,j})_{i,j\in E}$ possède au moins une probabilité invariante si et seulement si J est non-vide. Dans ce cas, les probabilités invariantes sont exactements les probabilités*

$$\sum_{x\in J} \alpha_x \tilde{\mu}^x,$$

où $\tilde{\mu}_x$ est le prolongement à S de l'unique mesure invariante de la la chaîne de markov de matrice $(p_{i,j})_{i,j\in A_x}$, et où les α_x sont des réels positifs de somme 1.

Démonstration. Soit μ une probabilité invariante, c'est à dire telle que

$$\forall i \in S \quad \mu(i) = \sum_{j\in S} \mu(j) p_{j,i}.$$

D'après les deux théorèmes précédents, μ ne charge que des états récurrents positifs, terminaux : J est donc non-vide. Soit $C = \cup_{x\in J} A_x$. On sait que $\mu(i)$ est nul pour $i \in E\backslash C$. Soit $x \in J$. On sait qu'on ne peut accéder de A_x qu'à des points de A_x : $\forall i \in A_x \forall j \in S\backslash A_x \quad p_{i,j} = 0$. Ainsi

$$\forall i \in A_x \sum_{j\in A_x} p_{i,j} = \sum_{j\in S} p_{i,j} - \sum_{j\in S\backslash A_x} p_{i,j} = 1 - 0 = 1.$$

La matrice $(p_{i,j})_{i,j\in A_x}$ est donc markovienne. On a également $\sum_{j\in S} \mu(j)p_{j,i} = \sum_{j\in A_x} \mu(j)p_{j,i}$. On a donc le système

$$\forall i \in A_x \quad \mu(i) = \sum_{j\in A_x} \mu(j) p_{j,i},$$

ce qui signifie que la restriction de μ à A_x est une mesure invariante finie pour $(p_{i,j})_{i,j\in A_x}$. D'après le théorème d'unicité pour les chaînes irréductibles, on a

$$\forall i \in A_x \quad \mu(i) = \mu(A_x)\mu^x(i).$$

On a ainsi la représentation voulue en posant $\alpha_x = \mu(A_x)$, en notant $\tilde{\mu}^x$ la mesure sur E qui coïncide avec μ^x sur A_x et est nulle sur $S\backslash A_x$.

On a montré que les probabilités invariantes étaient nécessairement de la forme annoncée. Montrons la réciproque. Montrons d'abord que l'extension de $\tilde{\mu}_x$ est invariante. On a d'abord que

$$\forall i \in A_x \quad \sum_{j\in S} \tilde{\mu}^j(j)p_{j,i} = \sum_{j\in A_x} \tilde{\mu}^x(j)p_{j,i} = \sum_{j\in A_x} \mu^x(j)p_{j,i} = \mu^x(j) = \tilde{\mu}^x(j).$$

Par ailleurs, pour $i \in S\backslash A_x$, observons les termes de la somme $\sum_{j\in S} \tilde{\mu}^x(j)p_{j,i}$: si $j \in A_x$, $p_{j,i}$ est nul, tandis que pour $j \in S\backslash A_x$, $\tilde{\mu}^j(j) = 0$: le produit $\tilde{\mu}^j(j)p_{j,i}$ est toujours nul, et on a

$$\sum_{j\in S} \tilde{\mu}^j(j)p_{j,i} = \sum_{j\in S} 0 = 0 = \tilde{\mu}^x(i).$$

La mesure de probabilités $\tilde{\mu}_x$ est donc bien invariante. Pour conclure, il est aisé de voir qu'un barycentre de probabilités invariantes est une probabilité invariante.

□

10.7 Algorithme de Propp et Wilson

Soit S un ensemble fini. On note $\mathcal{F}(S,S)$ l'ensemble des fonctions de S dans S. On remarquera que $\mathcal{F}(S,S)$ est fini.

Avant de décrire l'algorithme de Propp et Wilson, il convient de rappeler la représentation dynamique des chaînes de Markov. Une matrice de passage $(p_{i,j})_{(i,j)\in S\times S}$ étant donnée, on peut toujours construire une chaîne de Markov $(X_n)_{n\geq 0}$ à valeurs dans S de loi initiale donnée $\mathbb{P}_{X_0}$ quelconque, admettant $(p_{i,j})_{(i,j)\in S\times S}$ comme matrice de passage, et vérifiant la récurrence

$$X_{n+1} = f_{n+1}(X_n),$$

où $(f_n)_{n\geq 1}$ est une suite de fonctions aléatoires indépendantes, identiquement distribuées, indépendante de X_0. Comme $f_n \in \mathcal{F}(S,S)$, on peut toujours la représenter par un vecteur de taille $|S|$ dont les entrées sont à valeur dans S, mais c'est rarement comme cela qu'on procède.

Dans la pratique, la suite f_n est souvent générée de la manière suivante : on construit un ensemble X, une suite i.i.d. Z_n de variables aléatoires aisément simulables et une fonction déterministe $f : X \times S \to S$ telle que la suite f_n définie par $f_n(y) = f(Z_n, y)$ vérifie les conditions mentionnées ci-dessus.

On a donc pour tout $n \geq 1$, $X_n = f_n \circ f_{n-1} \circ \ldots f_1(X_0)$.

L'idée simple, mais diablement efficace, de Propp et Wilson, consiste à regarder la suite des les itérées prises dans l'autre sens : $(f_1 \circ f_2 \ldots f_n)_{n\geq 1}$.

Théorème 87. *Soit M la matrice d'une chaîne de Markov irréductible à valeurs dans S admettant μ comme mesure invariante et $(f_n)_{n\ge 1}$ une suite de fonctions aléatoires indépendantes suivant une loi θ sur $F = \mathcal{F}(S,S)$ telle que*

$$\forall (i,j) \in S \times S \quad \theta(f \in F, f(i) = j) = m_{i,j}.$$

On pose $g_0 = Id_S$, puis pour $n \ge 0$ $g_{n+1} = g_n \circ f_{n+1}$. Soit A une partie infinie de $\mathbb{N}$. On note

$$T = \inf\{n \in A, g_n \textit{ est une fonction constante}\}.$$

Si $\mathbb{P}(T < +\infty) = 1$, alors $g_T(x_0)$ suit la loi μ, où $x_0 \in S$ est quelconque.

Démonstration. Soit X_0 une variable aléatoire suivant la loi μ indépendante de la suite $(f_n)_{n\ge 1}$. Pour $n \ge T$, comme $g_n = g_T \circ (f_{T+1} \circ f_{T+2} \circ \cdots \circ f_n)$, on a $g_n(X_0) = g_T(x_0)$. On en déduit que pour $y \in S$ et $n \ge T$, on a $\mathbb{1}_{\{g_n(X_0)=y\}} = \mathbb{1}_{\{g_T(x_0)=y\}}$. On a donc presque sûrement $\lim_{n\to+\infty} \mathbb{1}_{\{g_n(X_0)=y\}} = \mathbb{1}_{\{g_T(x_0)=y\}}$. On appliquant le théorème de convergence dominée, on obtient

$$\lim_{n\to+\infty} P(g_n(X_0) = y) = P(g_T(x_0) = y).$$

Or $g_n(X_0) = f_1 \circ f_2 \cdots \circ f_n(X_0)$ a la même loi que $f_n \circ f_{n-1} \cdots \circ f_1(X_0)$, c'est à dire la loi que suit au temps n une chaîne de Markov de mesure initiale μ et de matrice de transition M, c'est à dire μ. (Attention : $(g_n(X_0))_{n\ge 0}$ n'est pas une chaîne de Markov !) On a donc pour tout $n \ge 0$ $P(g_n(X_0) = y) = \mu(y)$, d'où $\mathbb{P}(g_T(X_0) = y) = \mu(y)$. $\square$

On peut légitimement se demander pourquoi on ne prend pas tout simplement $A = \mathbb{N}$. Théoriquement, en effet, rien ne l'empêche. Cependant il faut voir que passer de n à $n+1$ n'augmente que très peu la probabilité que la fonction soit constante et tester si g_n est constante augmente de manière non négligeable le temps de calcul.

Explication : si $A = \{a_1, a_2, \dots,\}$ et que $\inf n \ge 0; n$ est constante $= n_0$. l'on a $a_{k-1} < t \le n_0 \le a_k$, le temps de calcul est proportionnel à $\sum_{i=1}^k a_i$. Ainsi, le choix $a_k = k$ (c'est à dire $A = \mathbb{N}$) conduit à un temps

$$\sum_{i=1}^{n_0} i = \frac{n_0(n_0+1)}{2} \sim \frac{n_0^2}{2},$$

tandis que le choix $a_k = 2^{k-1}$, conduit pour à $\sum_{k:2^{k-2}<n_0} 2^{k-1}$. Cette somme vaut $2n_0 - 1$ si n_0 est une puissance de 2, et au pire $4(n_0 - 1) - 1$.

Ainsi, on considère généralement que le choix $A = \{2^k; k \ge 0\}$ est un bon choix.

10.7.1 Loi 0-1 pour l'algorithme de Propp et Wilson

Théorème 88. *On reprend les hypothèses du théorème précédent*

- *Si il existe $n \geq 0$ tel que $\mathbb{P}(g_n$ est une fonction constante$) > 0$, alors $\mathbb{P}(T < +\infty) = 1$.*
- *Sinon $\mathbb{P}(T < +\infty) = 0$.*

En particulier, l'issue ne dépend pas du choix de A.

Démonstration. Il est clair que si

$$\forall n \geq 0 \quad \mathbb{P}(g_n) \text{ est une fonction constante}) = 0,$$

alors par réunion dénombrable on a $\mathbb{P}(T < +\infty) = 0$. Supposons donc qu'il existe n tel que $\mathbb{P}(g_n$ est une fonction constante$) > 0$. Comme $A \cap (n, +\infty)$ est une partie infinie de $\mathbb{N}$, il est possible d'en extraire un ensemble infini B tel que la différence entre deux éléments distincts quelconques de B dépasse n. Pour $k \in B$, notons

$$A_k = \{f_{k-n+1} \circ f_{k-n+2} \circ \cdots \circ f_{k-1} \circ f_k \text{ est une fonction constante}\}.$$

Les événements $(A_k)_{k\in B}$ sont des éléments indépendants, de même probabilité $\mathbb{P}(g_n$ est une fonction constante.$) > 0$. On en déduit que

$$\mathbb{P}(\cup_{k\in B} A_k) = 1.$$

Comme $g_k = f_1 \circ f_2 \cdots \circ f_k$, il est facile de voir que $A_k \subset \{g_k$ est une fonction constante$\}$, d'où

$$\begin{aligned} \mathbb{P}(T < +\infty) &= \mathbb{P}(\cup_{\{k\in A\}} g_k \text{ est une fonction constante}) \\ &\geq \mathbb{P}(\cup_{\{k\in B\}} g_k \text{ est une fonction constante}) = 1. \end{aligned}$$

□

10.7.2 Algorithme de Propp et Wilson pour des dynamiques monotones

On suppose ici que S est un ensemble fini muni d'un ordre partiel, possédant un plus grand élément et un plus petit élément. Un exemple classique de tel ensemble est $S = E^L$, où L est un ensemble fini et E une partie finie de $\mathbb{R}$. Dans le cas qui nous intéresse ici, on prendra simplement $S = E$.

Définition On dit qu'une dynamique associée à une matrice de Markov M est monotone si on peut construire un ensemble $F \subset \mathcal{F}(S, S)$ et une mesure θ sur $\mathcal{F}(S, S)$ tel que

— $\forall i, j \in S \quad \theta(f \in F, f(i) = j) = m_{i,j}$.
— F ne comprend que des fonctions croissantes.

Mise en oeuvre

Dans la pratique, la mesure θ est souvent construite comme l'image d'une mesure aisément simulable par l'application

$$x \mapsto (y \mapsto f(x, y)),$$

où f est une fonction de deux variables qui est croissante par rapport à chacune des variables.

En fait, la classe des dynamiques monotones recouvre un grand nombre de chaînes de Markov classiques. Par exemple, les marches aléatoires sur $\mathbb{Z}$ et les marches aléatoires sur $\mathbb{Z}$ avec barrières sont des dynamiques monotones.

Théorème 89. *Soit M la matrice d'une chaîne de Markov irréductible admettant μ comme mesure invariante. On suppose qu'on a construit un ensemble $F \subset \mathcal{F}(S, S)$ et une loi θ à support dans F telle que*
— $\forall i, j \in S \quad \theta(f \in F, f(i) = j) = m_{i,j}$.
— F ne contient que des fonctions croissantes.
(Ceci signifie que la dynamique est monotone).

Soit maintenant $(f_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes suivant la loi θ.

On pose $g_0 = Id_S$, puis pour $n \geq 0$ $g_{n+1} = g_n \circ f_{n+1}$.

Soit A une partie infinie de $\mathbb{N}$. On note

$$T = \inf\{n \in A, g_n \text{ est une fonction constante}\}.$$

Alors $\mathbb{P}(T < +\infty) = 1$ et $g_T(x_0)$ suit la loi μ, où $x_0 \in S$ est quelconque.

Démonstration. Notons min le plus petit élément de S et Max le plus grand élément de S. Commençons par une remarque simple : si une fonction croissante h de S dans S vérifie $h(\min) = \text{Max}$, alors elle est constante, car

$$\forall x \in S; \quad \text{Max} = h(\min) \leq h(x) \leq \text{Max},$$

ce qui montre que h est une fonction constante.

Maintenant, comme la chaîne de Markov est irréductible, il existe n tel que $\mathbb{P}(f_n \circ \dots f_1(\min) = \text{Max}) > 0$. Or, comme les f_n sont indépendantes identiquement distribuées, on a

$$\mathbb{P}(f_n \circ \dots f_1(\min) = \text{Max}) = \mathbb{P}(f_1 \circ \dots f_n(\min) = \text{Max}) > 0.$$

Finalement, on a

$$\mathbb{P}(f_1 \circ \dots f_n \text{ constante}) \geq \mathbb{P}(f_1 \circ \dots f_n(\min) = \text{Max}) > 0,$$

d'où $\mathbb{P}(T < +\infty)$ grâce au théorème 88.

$\square$

10.8 Exercices sur la récurrence et les mesures invariantes

10.8.1 Exercices corrigés

Exercice 86. On considère la chaîne de Markov dont la matrice de passage est donnée par

$$A = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \tag{10.1}$$

Montrer qu'il existe une unique mesure invariante, puis la donner. lien vers l'indication lien vers la solution

Exercice 87. On suppose que μ est la mesure invariante d'une chaîne de Markov sur S. Montrer que si les points i et j de S sont tels que $\mu(i) > 0$ et $i \to j$ pour cette chaîne de Markov, alors $\mu(j) > 0$. lien vers l'indication lien vers la solution

Exercice 88. *Retournement du temps et opérateurs associés*
Soit S un ensemble fini ou dénombrable, μ une mesure finie sur S.

1. Soit $P = (p_{i,j})_{(i,j)\in S^2}$, une matrice markovienne. On suppose que P laisse μ invariante. Pour $f \in \ell^2(\mu)$, on pose

$$\forall i \in S \quad (Pf)(i) = \sum_{j\in S} p_{i,j} f(j)$$

   dès que la série est absolument convergente. Montrer que c'est le cas lorsque $\mu(i) > 0$; montrer également que $Pf \in \ell^2(\mu)$.

2. Soit $P = (p_{i,j})_{(i,j)\in S^2}$ et $Q = (q_{i,j})_{(i,j)\in S^2}$ deux matrices markoviennes. On suppose que

$$\forall i, j \in S \quad \mu(i)p_{i,j} = \mu(j)q_{j,i}.$$

   Montrer que P et Q laissent μ invariante, puis que

$$\forall f, g \in \ell^2(\mu) \quad \int f(x)Pg(x)\ d\mu(x) = \int Qf(x)g(x)\ d\mu(x).$$

   On dit alors que P et Q sont conjugués dans $\ell^2(\mu)$.

3. Soit P une matrice markovienne laissant μ invariante. Montrer que P admet au moins une matrice markovienne conjuguée, et qu'il y a unicité si la chaîne associée est irréductible.

4. Soient P, Q deux matrices markoviennes conjuguées dans $\ell^2(\mu)$. On suppose que $(X_n)_{n\geq 0}$ et $(X'_n)_{n\geq 0}$ sont deux chaînes de Markov de loi initiale μ et de dynamiques respectives P et Q. Montrer que pour tout $n \geq 0$, $(X_0, X_1, \dots, X_n)$ et $(X'_n, X'_{n-1}, \dots, X'_0)$ ont même loi. Quel résultat du cours retrouve-t'on dans le cas où $P = Q$?
5. Soit $(X_n)_{n\geq 0}$ une chaîne de Markov stationnaire. Montrer que pour tout n, $(X_n, X_{n-1}, \dots, X_0)$ est (la restriction d')une chaîne de Markov.

lien vers l'indication lien vers la solution

Exercice 89. *Trace d'une chaîne sur un ensemble.*
Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov homogène d'espace d'états E dénombrable et de matrice de transition $P = (p_{i,j})_{i,j\in E}$. Soit A une partie de E. On observe cette chaîne de Markov seulement lors de ses passages par A, et on note Y_m la $m^{\text{ième}}$ observation. Plus formellement on note $T^0 = 0$ et, pour $m \geq 1$,

$$T^m = \inf\left\{n \geq 1 + T^{m-1} \;\middle|\; X_n \in A\right\}.$$

On suppose que $\forall x \in A$, $\mathbb{P}^x(T^1) < +\infty = +\infty$,

1. Soit $x \in A$. Montrer que pour tout $m \geq 1$, T^m est un temps d'arrêt $\mathbb{P}^x$-presque sûrement fini, que X_{T^m} est mesurable pour $\mathcal{F}_{T^m}$ et que pour $k \leq m$, T^k et X_{T^k} sont $\mathcal{F}_{T^m}$-mesurables.
2. Soit $x \in A$. On pose $Y_0 = X_0$ et pour $m \geq 1$, $Y_m = X_{T^m}$. Montrer que $(Y_n)_{n\in\mathbb{N}}$ est une chaîne de Markov homogène.
3. On suppose désormais que la chaîne est irréductible, que A est fini et que pour tout $x \in A$ $\mathbb{E}^x[T^1 < +\infty] < +\infty$. Montrer que la chaîne est récurrente positive.

lien vers l'indication lien vers la solution

Exercice 90. *Fonction de Lyapunov*

1. Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires à valeurs dans E. On suppose qu'il existe un réel $\varepsilon \geq 0$, une fonction $h : E \to \mathbb{R}_+$ et un ensemble M tel que pour tout $n \geq 1$, $f(X_n)$ est intégrable et

$$\mathbb{E}[f(X_n)|X_0, \dots, X_{n-1}] \leq f(X_{n-1}) - \varepsilon \text{ sur } \{X_{n-1} \notin M\}.$$

On pose alors $T = \inf\{n \geq 0; X_n \in M\}$.

(a) Montrer que pour tout $n \geq 1$, on a

$$\mathbb{E}\left(f(X_{n\wedge T}) - f(X_0)\right) \leq -\varepsilon \sum_{i=1}^{n} \mathbb{P}(T > i - 1).$$

(b) On suppose $\varepsilon > 0$. Montrer que $\mathbb{E}[T] \leq \frac{\mathbb{E}f(X_0)}{\varepsilon}$.

2. Soit $(X_n)_{n\geq 0}$ une chaîne de Markov à valeurs dans E, de matrice de passage $(p_{i,j})$. Dans toute cette question, on suppose qu'il existe un réel $\varepsilon \geq 0$, une fonction $f : E \to \mathbb{R}_+$ et un ensemble M tels que

$$\forall i \in E \quad (Pf)(i) = \sum_{j\in E} p_{i,j}f(j) < +\infty$$

et

$$\forall i \in E\backslash M \quad (Pf)(i) = \sum_{j\in E} p_{i,j}f(j) \leq f(i) - \varepsilon.$$

On pose $T = \inf\{n \geq 0; X_n \in M\}$.

(a) On suppose $\varepsilon > 0$. À l'aide du théorème 70, montrer que $\mathbb{E}[T] \leq \frac{\mathbb{E}f(X_0)}{\varepsilon}$.

(b) Ici $\varepsilon = 0$. On suppose que la chaîne est irréductible, que M est fini et que pour tout N, $\{x \in E; f(x) < N\}$ est fini. On note τ_N le temps d'entrée dans $\{x \in E; f(x) \geq N\}$. Montrer que pour pour tout N assez grand, on a $N\mathbb{P}(\tau_N < T) \leq \mathbb{E}(f(X_0))$. En déduire que $\mathbb{P}(T < +\infty) = 1$.

(c) *Théorème de Foster*
On suppose encore que la chaîne de Markov est irréductible et que M est fini. À l'aide de l'exercice précédent, montrer que la chaîne de Markov (X_n) est récurrente, et même récurrente positive si $\varepsilon > 0$.

lien vers l'indication lien vers la solution

10.8.2 Exercices non corrigés

Exercice 91. *Chaîne observée quand elle bouge ou est morte.*
Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov homogène sur l'espace d'états E, de matrice de transition P. Soit A l'ensemble des états absorbants. On définit la suite $(T_k)_{k\in\mathbb{N}}$ par récurrence comme suit : on pose $T_0 = 0$ et

$$T_{k+1} = T_k + \mathbb{1}_{\{X_{T_k}\notin A\}} \inf\{n \geq 0, X_{T_k+n} \neq X_{T_k}\},$$

avec la convention $0.\infty = 0$.

1. Montrer que les $(T_k)_{k\in\mathbb{N}}$ sont des temps d'arrêt pour $(X_n)_{n\in\mathbb{N}}$, finis presque sûrement.
2. On définit $Y_k = X_{T_k}$. Montrer que $(Y_k)_{k\in\mathbb{N}}$ est une chaîne de Markov homogène, donner son espace d'états et sa matrice de transition.

lien vers l'indication

Exercice 92. Soient $p \in]1/2,1[$ et $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes de loi commune $p\delta_1 + (1-p)\delta_{-1}$. On pose $S_0 = 0$, puis pour tout $n \geq 1$: $S_n = \sum_{k=1}^n X_k$. Soient a et b des entiers relatifs strictement positifs. On pose $V_\infty =]-\infty, -b] \cup [a, +\infty[$. Soit $T = \inf\{n \geq 0; n \in V_\infty\}$. Calculer

$$\mathbb{P}(\forall p \geq T; S_p \neq 0).$$

On pourra utiliser les résultats de l'exercice : "le joueur inruinable".

lien vers l'indication

Exercice 93. *Marche aléatoire sur $\mathbb{Z}/d\mathbb{Z}$.* On considère $X = \{X_n : n \geq 0\}$ la marche aléatoire sur $\mathbb{Z}/d\mathbb{Z}$ dont les pas sont indépendants de même loi $p\delta_1 + (1-p)\delta_{-1}$, avec $0 < p < 1$. La loi initiale de X_0 n'est *a priori* pas égale à δ_0.

1. Quelle est sa matrice de transition P ? La chaîne est-elle irréductible ? apériodique ?
2. On suppose que d est impair. Montrer que $(X_n)_{n\geq 0}$ converge en loi et préciser la limite.
3. On suppose que d est pair. Donner une condition nécessaire et suffisante sur $\mathbb{P}_{X_0}$ pour que $(X_n)_{n\geq 0}$ converge.

lien vers l'indication

Exercice 94. *Modèle d'Ehrenfest.* On cherche à modéliser la diffusion de N particules dans un système constitué de deux enceintes séparées par une paroi poreuse. A chaque instant, une particule prise au hasard (comprendre : avec équiprobabilité) change d'enceinte. On représente l'état du système à chaque instant par un vecteur $x = (x(1), \dots, x(N)) \in (\mathbb{Z}/2\mathbb{Z})^N$, où x_i représente le numéro de la boîte (0 ou 1) où est la particule i. Ainsi, si X_n est le vecteur des positions, on a la modélisation

$$X_{n+1} = X_n + \delta_{V_{n+1}},$$

où V_{n+1} est le numéro de la particule qui change de boîte. $(\delta_1, \dots, \delta_N)$ est la base canonique de $(\mathbb{Z}/2\mathbb{Z})^N$. $(V_n)_{n\geq 1}$ est une suite de variables aléatoires indépendantes suivant la loi uniforme sur l'ensemble fini $\{1, \dots, N\}$. $(V_n)_{n\geq 1}$ est indépendante de X_0.

1. Montrer que $(X_n)_{n\geq 0}$ est une chaîne de Markov.
2. On pose

$$Y_n = \sum_{i=1}^N \mathbb{1}_{\{X_n(i)=0\}}.$$

Y_n est donc le nombre de particules dans la boîte 0 au temps n. Montrer que $(Y_n)_{n\geq 0}$ est une chaîne de Markov de matrice de transition

$$p(x,x-1)=\frac{x}{N},\ p(x,x+1)=\frac{N-x}{N},\ 0\leq x\leq N,$$

les autres probabilités étant nulles.

3. On note U la loi uniforme sur $(\mathbb{Z}/2\mathbb{Z})^N$. Montrer que pour toute loi γ sur $(\mathbb{Z}/2\mathbb{Z})^N$, on a $U*\gamma=U$.
4. Montrer que si X_0 suit la loi U, alors $X_0(1),X_0(2),\dots X_0(N)$ sont indépendantes.
5. Montrer que U est une loi invariante pour la chaîne $(X_n)_{n\geq 0}$. En déduire que la loi binomiale $\mathcal{B}(N,1/2)$ est une loi invariante pour la chaîne $(Y_n)_{n\geq 0}$.
6. Retrouver ce dernier résultat par un calcul direct.

lien vers l'indication

Exercice 95. *Chaîne de Markov avec décision.*

Le $n^{\text{ème}}$ Lundi de l'année, une petite entreprise reçoit A_n propositions de travail de type A, et B_n propositions de travail de type B. Un travail de type A mobilise toute la capacité de travail de l'entreprise durant une semaine et lui rapporte 200 euros, alors qu'un travail de type B l'occupe deux semaines pour un rapport de 360 euros. Une semaine d'inactivité coûte 100 euros, un travail non traité pendant la semaine où il arrive est perdu. On suppose A_n,B_n indépendants, les couples $(A_n,B_n)_{n\geq 1}$ indépendants, et

$$\mathbb{P}(A_n=1)=1-\mathbb{P}(A_n=0)=0,5,\qquad \mathbb{P}(B_n=1)=1-\mathbb{P}(B_n=0)=0,6.$$

Modéliser la situation par une chaîne de Markov , avec si possible un nombre d'états minimal. Quelle est la meilleure stratégie, quand on reçoit simultanément une offre de chaque type : donner la préférence à celle de type A ou à celle de type B ? On pourra faire appel au Théorème ergodique pour départager les deux politiques.

lien vers l'indication

Exercice 96. *Un modèle de prédiction météo (!)* On suppose que le temps qu'il fera demain depend des deux jours précédents. On suppose que :

$$\begin{aligned}
&\mathbb{P}(\text{ il pleut demain} \mid \text{il a plu hier et aujourd'hui})=0,7\\
&\mathbb{P}(\text{ il pleut demain} \mid \text{il a plu aujourd'hui mais pas hier})=0,5\\
&\mathbb{P}(\text{ il pleut demain} \mid \text{il a plu hier mais pas aujourd'hui})=0,4\\
&\mathbb{P}(\text{ il pleut demain} \mid \text{il n'a pas plu ni hier, ni aujourd'hui})=0,2
\end{aligned}$$

Montrer qu'on peut modéliser ceci par une chaîne de Markov. Quelle est la probabilité, sachant qu'il a plu lundi et mardi qu'il pleuve jeudi ? Sur le long terme, quelle proportion de jours de pluie observe-t-on ?

lien vers l'indication

Exercice 97. *Chaîne de Markov réversible*

1. Soit P une matrice de transition sur un espace d'états E dénombrable. On suppose qu'il existe une probabilité π telle que
$$\pi_i p_{i,j} = \pi_j p_{j,i}.$$
Montrer que π est stationnaire pour la P.
2. Trouver rapidement la probabilité stationnaire de la marche aléatoire symétrique sur les sommets de l'hypercube de dimension d.
3. Marche aléatoire symétrique sur un échiquier (8×8). Calculer les temps de retours moyens des différents points de l'échiquier. (On trouvera 110 pour les coins, 220/3 pour les autres points du bord, 55 pour les autres points.)

lien vers l'indication

Exercice 98. *Modèle de Laplace–Bernoulli.*
N boules noires et N boules blanches sont placées dans deux urnes de sorte que chacune contienne N boules. Après chaque unité de temps on choisit au hasard une boule de chaque urne ; les deux boules ainsi choisies changent d'urne. On note Y_n le nombre de boules noires dans la première urne. Montrer que $(Y_n)_{n\geq 0}$ est une chaîne de Markov irréductible réversible et trouver sa mesure stationnaire.

lien vers l'indication

Exercice 99. *Loi de Pascal* Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires de Bernoulli de paramètre p. On pose, pour $n \geq 1$, $S_n = X_1 + \cdots + X_n$ et $\tau_n = \inf\{n \geq 1; S_n \geq n\}$. On appelle loi de Pascal (ou loi binomiale négative) de paramètre n et p.

1. Montrer que la suite $(\tau_i - \tau_{i-1})$ est une suite de variables aléatoires indépendantes dont on précisera la loi.
2. Donner une expression explicite de $\mathbb{P}(\tau_n = k)$, pour $k \geq n$.
3. Calculer la fonction génératrice de la loi de Pascal de paramètres n et p.

lien vers l'indication

Annexe A

Indications

A.1 Exercices sur les variables de Bernoulli

Indication 1 Noter que la connaissance de X donne celle des Ω_n.

A.2 Exercices sur l'équi-intégrabilité

Indication 2 On peut s'inspirer de la preuve de l'implication "bornée dans L^2 entraîne équi-intégrable".

Indication 3

1. Utiliser la loi forte des grands nombres.
2. Comme les X_i sont entre 0 et 1, on a $Z_n^2 \le n^2(X_1^2 + \dots X_n^2)^{-2}$.
3. Utiliser le théorème d'intégration des fonctions radiales.
4. On pourra montrer que $(Z_n)_{n\ge 1}$ est bornée dans L^2.

Indication 4 Deux méthodes possibles :

— Pour simplifier les calculs, on peut écrire $X_n = S_n + \lambda$, avec S_n centrée (mais pas symétrique !), de sorte que $X_n^3 = S_n^3 + \lambda^3 + 3\lambda S_n^2 + 3\lambda^2 S_n$.

— On peut calculer $\mathbb{E}(X_n(X_n-1)(X_n-2))$ (par un calcul direct ou par une interprétation combinatoire).

Indication 5 Introduire une somme de variables de Bernoulli indépendantes.

Indication 6 On pourra utiliser une caractérisation bien choisie de l'équi-intégrabilité.

Indication 7 On pourra utiliser une caractérisation bien choisie de l'équi-intégrabilité.

Indication 8 Partitionner suivant la valeur de N_θ.

Indication 9 Exprimer $\mathbb{E}|X_n|\mathbb{1}_{\{|X_n|>M\}}$ à l'aide de la fonction de queue.

A.3 Exercices sur l'espérance conditionnelle

Indication 10
1. Écrire $A = \cup_{n\geq 0} A \cap \{N = n\}$.
2. On peut utiliser la question précédente.

Indication 11 On pourra commencer par établir que $\mathbb{E}[T|U = k] = (k + 1)\frac{\sum_{n\geq k}\binom{n+1}{k+1}x^n}{\sum_{n\geq k}\binom{n}{k}x^n} - 1$, puis faire éventuellement une transformation d'Abel.

Indication 12
1. Calculer $\mathbb{Q}_{n,b}(X_2 = y_1, \dots X_{n+b} = y_{n+b-1})$ en distinguant suivant que $(y_1, \dots, y_{n+b-1}$ est dans $\Omega_{n-1,b}$ ou non.
2. Écrire T sous la forme $T = \mathbb{1}_{\{X_1=1\}} f(X_2, \dots, X_{n+b-1})$.

Indication 13 Revoir les propriétés de l'espérance conditionnelle.

Indication 14 On pourra écrire $f_i(x) = x_i \mathbb{1}_A(x_1 + \dots + x_n)$, remarquer que si $\theta_{i,j}$ est l'application qui échange la i-ème et la j-ème coordonnée de x, on a $f_i = f_1 \circ \theta_{1,i}$ et appliquer le théorème de transfert.

Indication 15 Pour la première méthode, on pourra poser $\varphi = \mathbb{E}[x^X y^Y]$, $g(x, y) = x\frac{\partial\varphi}{\partial x}(x, y)$, puis comparer les coefficients de degré s de $g(x, x)$ et de $\varphi(x, x)$.
Pour la deuxième méthode, quitte à changer d'espace, on pourra remarquer que X peut s'écrire comme une somme d'indicatrices.

Indication 16
1. $\Omega = \{X < h\} \cup \{X \geq h\}$.
2. Un max se réalise toujours en au moins un point.
3. L'indicatrice de l'intersection est le produit des indicatrices.
4. Écrire $Y_i(h)$ sous la forme $F(X_i, Z)$, avec Z indépendant de X_i.
5. Utiliser (c) et (d)
6. Prendre $h = \alpha\frac{\ln n}{n}$, avec α bien choisi.

Indication 17 On peut s'inspirer de la preuve du calcul de $\mathbb{E}[g(X, Y)|X]$ lorsque X et Y sont indépendantes.

Indication 18 On peut le voir comme un cas particulier de l'exercice précédent

Indication 19 On pourra remarquer que $X + Y - Z = 0$.

Indication 20
1. Calculer les mineurs.

2. On pourra écrire X sous la forme $X = \alpha Y + \beta Z + R$, avec R indépendant de $\sigma(Y, Z)$.
3. On pourra trouver une première expression de $f_{y,z}(x)$ à partir de l'exercice 7. On pourra ensuite remarquer que si

$$K \exp(-(ax^2 + bx + c))$$

est la densité d'une variable aléatoire, cette variable est nécessairement gaussienne. On pourra identifier les paramètres par un choix judicieux de φ.

Indication 21

Utiliser le théorème de convergence dominée conditionnel.

On peut s'inspirer de la preuve de l'inégalité de Markov.

On pourra remarquer que $\mathbb{E}[\mathbb{1}_{\{Y>M\}}|\mathcal{A}] \le Z/M$.

Il s'agit de montrer que $\lim_{n\to+\infty} \mathbb{P}(Y > M) = 0$.

Indication 22 1. pas d'indication

2. Elle est d'intégrale nulle (c'est même une martingale).
3. On pourra noter que $M_i = \sum_{k=1}^{i\wedge T}(S_k - 1) = (\sum_{k=1}^{i\wedge T} S_k) - (i \wedge T)$.
4. On utilisera la linéarité de l'espérance pour calculer le nombre moyen de points fixes d'une permutation.

A.4 Exercices sur les martingales

Indication 23 1. Remarquer que $x \le (x^2 + 1)/2$.

2. (a) On pourra remarquer que (X_n) prend ses valeurs dans $[0, 1]$
   (b) Appliquer le théorème de convergence dominée

Indication 24 1. (a) Procéder par récurrence sur n.

   (b) Remarquer que $1 = \sum_{k=1}^{d+nS} \mathbb{1}_{\{U_{n+1}=k\}}$.
   (c) Noter que U_{n+1} est indépendant de la tribu $\mathcal{F}_n$.
   (d) Revoir la définition d'une martingale.
   (e) Remarquer que la suite considérée prend ses valeurs dans $[0, 1]$.
   (f) On pourra calculer $\mathbb{P}(T_{n+1} = i|\mathcal{F}_n)$.

2. (a) Procéder par récurrence sur n.

(b) Expliciter le lien entre $\frac{V_n-d}{S}$ et les T_i et utiliser une partition de l'événement considéré.

(c) Appliquer le théorème de transfert et utiliser la formule du multinôme.

(d) Utiliser le théorème de Lévy. On sera amené à démontrer que si $y_1 \dots, y_m$ sont des réels positifs de somme 1, la suite $(\sum_{k=1}^m y_k e^{iu_k/n})^n$ converge vers $\exp(\sum_{k=1}^m iy_k u_k)$ lorsque n tend vers l'infini. Une preuve analytique ou une preuve probabiliste est possible.

(e) Noter que la convergence presque sûre entraîne la convergence en loi.

Indication 25 Soit A un événement $\mathcal{F}_\tau$ mesurable. On peut écrire $A = A_+ \cup A_-$ avec $A_+ = A \cap \{\tau_1 \le \tau_2\}$ et $A_- = A \cap \{\tau_1 > \tau_2\}$. Il suffit alors de montrer que A_+ est $\mathcal{F}_{\tau_2}$ mesurable, tandis que A_- est $\mathcal{F}_{\tau_1}$ mesurable.

Indication 26 Utiliser les propriétés de l'espérance conditionnelle

Indication 27
1. On rappelle que si $Y \ge 0$ et $\mathbb{E}Y = 0$, alors Y est nulle presque sûrement.
2. $x \mapsto (x-a)^+$ est une fonction convexe.
3. On pourra montrer que $(X_n - a)^-$ est presque sûrement nulle sur l'événement $\{X_p \ge a\}$.
4. Si deux nombres sont distincts, il y a un rationnel entre les deux.

Indication 28
1. Classique.
2. Utiliser l'inégalité de Cauchy-Schwarz.
3. On pourra montrer que $Y_n^t = o(Y_n^{t/2})$.
4. On pourra commencer par montrer que $\varphi(t) > \mathbb{E}(X_1)t$, en commençant par traiter le cas où X_1 prend des valeurs positives.

Indication 29 On pourra utiliser une caractérisation adaptée de l'équi-intégrabilité.

Indication 30 On rappelle que $\sin x \le x$ pour tout $x \ge 0$.

Indication 31
1. $-g$ est convexe.
2. On pourra utiliser le théorème d'arrêt.
3. Utiliser la question précédente.

4. L'inégalité $\Psi(\theta s+(1-\theta)t) \geq \theta f(s)+(1-\theta)f(t)$ découle de la concavité de Ψ. Pour l'inégalité inverse, on peut considérer la fonction $\tilde{\Psi}$ qui coïncide avec Ψ à extérieur de $]s,t[$ et qui est affine sur $[s,t]$.
5. On pourra montrer successivement

$$\mathbb{E}f(X_T) = \mathbb{E}\Psi(X_T) = \Psi(\mathbb{E}X_T) = \Psi(i_0).$$

Indication 32 On pourra remarquer que les Y_n sont non corrélées et que la suite des sommes partielles forme une martingale.

A.5 Exercices sur les complements

Indication 33 Il suffit (pourquoi ?) de montrer que l'image réciproque d'un fermé est dans $\mathcal{F}$.

Indication 34 Fixer $y_0 \in Y$ et poser $H(x) = Z(x, y_0)$.

Indication 35
1. Pour déterminer la loi conditionnelle, il s'agit de déterminer $\mathbb{P}(X = k|X + Y = n)$.
2. L'espérance conditionnelle peut se calculer par intégration de la loi conditionnelle. Ici, la loi conditionnelle est une loi dont les moments sont bien connus.

Indication 36 Utiliser le théorème fondamental de la mesurabilité : $\sigma(f^{-1}(\mathcal{C})) = f^{-1}(\sigma(\mathcal{C}))$.

Indication 37 Utiliser le théorème de transfert.

Indication 38 On pourra s'inspirer de la preuve du théorème de Radon-Nicodým et montrer que pour f $\mathcal{F}_n$ mesurable positive bornée, on a

$$C^{-1}\mathbb{E}_{\mathbb{Q}}[f] \leq \mathbb{E}_{\mathbb{P}}[f] \leq C\mathbb{E}_{\mathbb{Q}}[f].$$

A.6 Exercices sur les inégalités

Indication 39
1. Une permutation est une bijection, donc peut être utile à un changement d'indice.
2. Il suffit montrer que $h(X_1, \dots, X_n)$ est à valeurs dans $\mathfrak{S}_n$ et qu'elle charge également tous les points.
3. Appliquer le principe de Maurey grâce à la représentation construite.

4. Appliquer la question précédente avec $n = b + r$.

Indication 40 1. On trouvera $X_k^n = \sum_{k=1}^n \mathbb{1}_{\{N_{k,p}=1\}}$. Noter que l'espérance est linéaire et la loi de $N_{k,p}$ connue.

2. Noter que $v_n \geq \mathbb{E}[X_n^n]$ et que $1 - x \leq \exp(-x)$ pour tout $x \geq 0$.
3. On pourra traiter séparément les événements $\{X_*^n - n/e \geq n\varepsilon\}$ et $\{X_*^n - n/e \leq -n\varepsilon\}$.
4. Appliquer le principe de Maurey. Comment est modifié X^n si on change un seul tirage ?
5. Une fois que tout a été tiré deux fois, X^n est connu.
6. Remarquer que X_n/n est bornée.
7. Modéliser le problème.

Indication 41 On pourra utiliser la méthode de l'exercice 39 en considérant sur $\mathfrak{S}_{2n}$ la fonction $g(\sigma) = \max\{|\sum_{j=1}^k (-1)^{\sigma(j)}|; 1 \leq k \leq n\}$.

Indication 42 On pourra écrire S_A et $S_{\{1\dots,n\}\backslash A}$ comme des fonctions de $X_1, \dots, X_n$.

Indication 43 Notons $D = \mathcal{B}_2(\{1, \dots, n\})$ et posons pour $x \in \{0,1\}^D$

$$\varphi(x) = \inf\left\{N \geq 1; \exists c \in \{1, \dots, N\}^n; \sum_{\{i,j\}\in C} x_{\{i,j\}}\mathbb{1}_{\{c(i)=c(j)\}} = 0\right\}.$$

On a $\chi = \varphi((X_e)_{e\in D})$ où les $(X_e)_{e\in D}$ sont des variables de Bernoulli indépendantes de paramètre p.

A.7 Exercices sur les statistiques exhaustives

Indication 44 1. Exprimer $\mathbb{E}_\theta\varphi(Z)$ comme une intégrale par rapport à m et utiliser le théorème de factorisation de Neyman pour la dominante privilégiée.

2. On pourra montrer qu'il existe une constante $c(\varphi)$ telle que $\mathbb{E}_\theta(\varphi(Z)|S) = c(\varphi)$.

Indication 45 On pourra par exemple considérer $\sum_{i=1}^n X_i^2$ et $\sum_{i=1}^n X_i$.

Indication 46 1. On pourra utiliser le théorème de Neyman–Fisher.

2. On trouvera que la loi de S_n sous $\mathbb{P}_\theta$ est $\mathcal{P}(n\theta)$.

3. Une limite uniforme sur tout compact de fonctions holomorphes est holomorphe.
4. Noter que $\mathbb{E}_\theta(f(S_n)) = e^{-n\theta}F(\theta)$ et appliquer le principe des zéros isolés.

Indication 47 1. On pourra appliquer le théorème de Neyman-Fisher.
2. On montrera que $\mathbb{E}_\theta(\varphi(M_n)) = \frac{1}{\theta^n}\int_0^\theta \varphi(t)nt^{n-1}\, dt$.
3. Calculer $\mathbb{E}[X_1 + \cdots + X_n]$.
4. Utiliser la procédure classique d'amélioration des estimateurs sans biais.

Indication 48 1. On notera que le modèle est exponentiel.
2. $\mathbb{E}_\theta[\mathbb{1}_{\{X_i \le t\}}] = \ldots$
3. Utiliser la procédure classique d'amélioration des estimateurs sans biais.

Indication 49 1. $\overline{X}_n$ est une statistique exhaustive complète.
2. S_n^2 est libre.

A.8 Exercices sur l'information de Fisher

Indication 50 1. Un changement de variable peut également être utile.
2. Sous de bonnes hypothèses W_α est centré.
3. Remarquer que le modèle est exponentiel.
4. Si $f(X)$ est un estimateur sans biais de θ, on pourra s'intéresser à la transformée de Laplace de mesures de probabilités construites à l'aide de $f \circ \exp$.

Indication 51 1. Appliquer le théorème de transfert.
2. Remarquer que le modèle est exponentiel.
3. On pourra commencer par déterminer la loi de S_d.
4. Utiliser le lemme de Lehman-Scheffé.
5. $\mathrm{Var}_\lambda \frac{S_d(S_d-1)}{d^2} = \frac{1}{d^4}\mathrm{Var}_\lambda S_d(S_d-1)$. La loi de S_d étant connu, on est ramené à un calcul de série. On conseille d'exprimer $X^2(X-1)^2$ dans une base appropriée de polynômes.

Indication 52

1. On trouve $W_\theta(X) = (-1 + \frac{X}{\theta}) = \frac{1}{\theta}(X - \mathbb{E}_\theta(X))$. On est amené à calculer (ou à se souvenir) de la variance d'une loi de Poisson.
2. Un théorème du cours donne le résultat sans calcul.

Indication 53

1. Exprimer Y_n^2 comme une fonction de deux variables aléatoires indépendantes.
2. Appliquer le théorème de convergence dominé à $\mathcal{I}(\theta + U\varepsilon_n)$.
3. Noter que Y_n^2 est positive et converge presque sûrement.
4. Utiliser l'hypothèse de bornitude locale pour $\theta \mapsto \mathbb{E}_\theta[h^2(X)]$.
5. Utiliser un critère approprié d'équi-intégrabilité.
6. Z_n converge presque sûrement et ...
7. Appliquer le théorème de transfert.
8. Utiliser la caractérisation de la dérivée par les suites.

A.9 Exercices sur les processus

Indication 54

Tout événement A s'écrit comme réunion dénombrable d'événements de la forme $\{\Pi_0 = x_0, \dots, \Pi_n = x_n\}$.

Des probabilités sur $\mathbb{R}^{\mathbb{N}}$ sont caractérisées par leurs lois de dimensions finies.

Utiliser la question 1 avec $n = 1$.

Indication 55 On peut considérer l' application

$$\begin{array}{rcl} \psi : \mathbb{R}^{\mathbb{N}} & \to & \mathbb{R}^{\mathbb{N}} \\ ((x_n)_{n\geq 0}) & \mapsto & (\varphi(\theta^n \circ x)) \end{array}$$

Indication 56

1. Considérer le cycle de $\mathcal{S}_{n+1} : \sigma = (1 \quad \dots n \quad n+1)$.
2. On trouvera que $n \operatorname{Var} X_1 + n(n-1) \operatorname{Covar}(X_1, X_2) \geq 0$ pour tout n.
3. Noter qu'un processus gaussien est caractérisé par sa fonction de corrélation, et que la valeur de celle-ci est maximale en zéro. On a intérêt à considérer la permutation qui échange 2 et n de manière à comparer $\operatorname{Covar}(X_1, X_2)$ et $\operatorname{Covar}(X_1, X_n)$.
4. Utiliser le concept de loi d'un processus et reconstruire Y_0 à partir des X_i.

Indication 57 1. Considérer une partition de $B(x)$.

2. Si x n'est pas un mot propre, l'identité est évidente. Sinon, on peut remarquer que

$$\sum_{a\in\{1,\dots,q\}} B(x.a) = \sum_{a\in\{1,\dots,q\}\setminus\{x_n\}} B(x.a)$$

et maintenant le $x.a$ apparaissant dans la somme est un mot propre.

3. Sommer en x l'identité obtenue à la question précédente.
4. Il suffit de vérifier que $\pi_n(\{1,\dots,q\}^n) = 1$.
5. On pourra calculer $\pi_{n+1}(\{(x_1,\dots,x_n)\} \times \{1,\dots,q\})$.
6. Si $x = (x_1,\dots,x_n)$ et $\tilde{x} = (x_n,\dots,x_1)$, il y a une bijection simple entre $B(x)$ et $B(\tilde{x})$.
7. On peut noter que

$$\begin{aligned}\mathbb{P}_q(\Pi_1 = x_1,\dots,\Pi_n = x_n) &= \mathbb{P}_q(\Pi_1 = x_n,\dots,\Pi_n = x_1)\\ &= \sum_{a=1}^{q} \mathbb{P}_q(\Pi_1 = x_n,\dots,\Pi_n = x_1, \Pi_{n+1} = a)\end{aligned}$$

8. (a) Procéder par récurrence sur $n+p$ en s'inspirant de la preuve de la question 2.
   (b) On pourra sommer l'identité précédente sur tous les $(x,y) \in \{1,\dots,q\}^{n+p}$.

Indication 58 1. Dans le calcul de la série, on traitera séparément les cas $i=0$, $i=1$, $i\geq 2$.

2. On peut noter que $\{T = n, \tilde{\theta}^{-1}(A)\} = \{\Pi_1 \neq c, \Pi_2 \neq c, \Pi_{n-1} \neq c, \Pi_n = c, \tilde{\theta}^{-n}(A)\}$, mais que $\overline{\mathbb{P}}$-presque sûrement, les autres conditions entraînent la $n-1$-ième.
3. Il suffit de montrer que pour tout n, les variables $(T\circ\tilde{\theta}^k)_{0\leq\leq n}$, ce qui peut se faire par récurrence sur n.
4. On pourra montrer que $(G_T)^k$ est la fonction génératrice de N_k.
5. Si $p > 1/4$, on pourra remarquer que $s \mapsto \frac{ps^2}{1-s+ps^2}$ est un prolongement analytique de G_T sur un voisinage de la droite réelle.
6. Noter que $\frac{3}{4} < 1$.

Indication 59 On peut considérer les applications

$$\begin{array}{rcl}\psi_n : \mathbb{R}^n \times \mathbb{R}^n & \to & \mathbb{R}^n\\ ((x_1,\dots,x_n),(y_1,\dots,y_n)) & \mapsto & (x_1+y_1,\dots,x_n+y_n)\end{array}$$

Pour le contre-exemple, on peut prendre pour (X_n) un bruit blanc et poser $Y_n = (-1)^n X_n$.

Indication 60 On peut considérer les applications

$$\begin{array}{rcl}\psi_n : \mathbb{R}^{n+1} & \to & \mathbb{R}^n \\ (x_1, \dots, x_n) & \mapsto & (x_1 + 2x_2, x_2 + 2x_3, \dots, x_{n-1} + 2x_n)\end{array}$$

Indication 61 Si $(X_n)_{n\geq 0}$ est une chaîne de Markov, $(X_{n+1})_{n\geq 0}$ est aussi une chaîne de Markov, ayant la même matrice de passage.

Indication 62 On pourra commencer par montrer que $(X_n)_{n\geq 0}$ est stationnaire. En particulier, il faut montrer que X_0 et X_1 ont même loi.

Indication 63 On rappelle que deux mesures sur $\mathbb{R}$ sont égales si elles coïncident sur les ensembles $]-\infty, a]$, où a décrit $\mathbb{R}$.

Indication 64 Même remarque que pour l'exercice précédent. Par ailleurs, on pourra considérer la mesure image de la loi uniforme sur $[0, 1]$ par l'application $x \mapsto e^{2i\pi x}$.

Indication 65 Remarquer que $\mathbb{E}[Z_i Z_j] = \delta_{i,j}$.

A.10 Exercices sur les chaînes de Markov

Indication 66 Le système peut se représenter par une chaîne de Markov à 3 états, chaque état étant induit par situation six/pas six de deux instants de temps consécutifs. On pourra par exemple utiliser la technique de l'analyse au premier pas

Indication 67 Commencer par déterminer n et $\alpha > 0$ tels que que pour tout $x \in S$, $\mathbb{P}^x(\tau \leq n) \geq \alpha$.

Indication 68 Si $(Y_n)_{n\geq 0}$ est une chaîne de Markov avec $\mathbb{P}^a(Y_1 = b) > 0$ et $\mathbb{P}^b(Y_1 = c) > 0$, alors $\mathbb{P}^a(Y_1 = b, Y_2 = c) > 0$.

Indication 69 Commencer par trouver en candidat pour la matrice.

Indication 70
1. Écrire $X_{n+1} = F(X_n, Y_{n+1})$.
2. Appliquer la méthode de l'analyse au premier pas.
3. Si l'on pose $u_i = E^0[T^n] - E^i[T^n]$, nous cherchons $u_n = E^0[T^n] - E^n[T^n] = E^0[T^n]$. Comme on a la récurrence $u_i = pu_{i+1} - 1$, il vient facilement $u_n = \frac{p^{-n}-1}{1-p}$.
4. C'est une application immédiate de la question précédente.

Indication 71 On peut remarquer que si M est une matrice 2×2 dont λ est une valeur propre, alors les matrices $(M - \lambda I)$ et $(M - \lambda I)^2$ sont liées. On rappelle que si $(X_n)_{n \geq 0}$ est une chaîne de Markov dont la loi initiale est donnée par le vecteur ligne x, alors, la loi de X_n est donnée par le vecteur ligne xP^n, où P est la matrice de passage.

Indication 72

1. Revoir le cours.
2. Découper l'intervalle $[0, 1]$ en morceaux de différentes longueurs.
3. Utiliser le (a).
4. Recoller les morceaux.

Indication 73 On pourra remarquer que $\{X_n = a\} \subset \{T \leq n\}$ et que

$$\{T \leq n\} \backslash \{X_n = a\} \subset \{\exists k < n; X_k = a \text{ et } X_{k+1} \neq a\}.$$

Indication 74 Si on pose $B = \{u \in \mathbb{R}^{\mathbb{N}}; \exists n \geq 0; u_n \in A\}$, on peut remarquer que pour $x \notin A$, on a $\mathbb{P}^x(X \in A) = \mathbb{P}^x((X_{n+1})_{n \geq 0} \in A)$ et utiliser la propriété de Markov.

Indication 75 Il y a plusieurs méthodes possibles. Le plus simple est sans doute de se ramener au cas où $(X_n)_{n \geq 0}$ est donnée par la représentation canonique $X_{n+1} = f_{n+1}(X_n)$ ou $X_{n+1} = g(X_n, Z_{n+1})$.

Indication 76 Si $E_x = \{X_n \to x\}$, alors $E_x = \cup_{N \geq 1} \cap_{n \geq N} \{X_n = x\}$, donc pour montrer que $\mathbb{P}(E_x) = 0$, il suffit de montrer que pour tout n, $\mathbb{P}(\cap_{n \geq N}\{X_n = x\}) = 0$.

Indication 77

1. On peut établir que $S'_{n+1} = S'_n + \mathbb{1}_{\{S'_n \notin \{-b, a\}\}} X_{n+1}$.
2. Comme dans l'exercice 4, on utilisera la propriété de Markov.
3. On rappelle que les solutions d'une récurrence linéaire forment un espace vectoriel.
4. Comme dans l'exercice 4, on utilisera la propriété de Markov.
5. On peut remarquer que $n^2 = 1 + \frac{(n+1)^2 + (n-1)^2}{2}$.

Indication 78 On peut montrer que $\{T' = +\infty\} = \cap_{a \geq 1} G^a$.

Indication 79 Raisonner comme dans l'exercice 4 et résoudre le système linéaire.

Indication 80 Relire les définitions.

Indication 81 On pourra remarquer que si $\mathbb{P}(\sum_{k=1}^n D_k = (0,0)) > 0$, alors il existe i, j positifs ou nuls avec $i + j = n$, $a|i$ et $b|j$.

Indication 82 À t fixé, on pourra remarquer que $\mathbb{P}(\mathbb{1}_A F((X_n)_{n\geq 1} > t) = \mathbb{P}(A, F((X_n)_{n\geq 1} > t)$ et appliquer la propriété de Markov.

Indication 83 Pour $x \in \{1, 2, 3\}$, on a $\mathbb{E}^x F(X) = f(x) + \mathbb{E}^x F((X_{n+1})_{n\geq 1})$. On peut alors appliquer le résultat de l'exercice précédent.

Indication 84

1. La suite $(Y_{n,1}, Y_{n,2}, \dots, Y_{n,2N})_{n\geq 1}$ est une suite de vecteurs aléatoires indépendants de même loi que l'on peut utiliser pour obtenir une représentation canonique.
2. Se ramener à l'étude d'une somme de variables de Bernoulli.
3. Il y a plusieurs méthodes. On peut par exemple calculer $[E_{X_{n+1}}|\sigma(X_1, \dots, X_n)]$, ou utiliser des techniques générales sur les chaînes de Markov dont l'espace d'état est fini et qui possèdent des points absorbants .
4. On pourra remarquer que la suite $(\mathbb{E}X_n)_{n\geq 1}$ est constante.

Indication 85 On peut commencer par montrer que $(X_n, Y_n)_{n\geq 0}$ est une chaîne de Markov, puis utiliser l'exercice précédent.

A.11 Exercices sur la récurrence et les mesures invariantes

Indication 86 On pourra éventuellement utiliser le théorème 83.

Indication 87 Considérer $\mathbb{P}^\mu(X_n = j)$, avec n bien choisi.

Indication 88

1. $Pf(j)$ peut s'interpréter comme une intégrale.
2. Le premier point ne pose pas de difficulté particulière. Pour le second point, on pourra remarquer que les fonctions $(\delta_i)_{i\in S}$ engendrent un sous-espace dense de $\ell^2(\mu)$.
3. On commencera par chercher des conditions nécessaires.
4. Le point délicat est de montrer que $\mathbb{E}[f(X_k)|X_{k+1}, \dots, X_n]$ ne dépend que de X_{k+1}. À cet effet, on pourra utiliser la représentation canonique dynamique des chaînes de Markov.
5. Combiner les questions précédentes.

Indication 89

1. Penser à utiliser la propriété de Markov forte. On peut remarquer que $T^m = \inf\{n \geq 1; \sum_{k=1}^n \mathbb{1}_{X_k\in A} \geq m\}$.

2. On pourra commencer par calculer $\mathbb{P}^x(X_{T^{m+1}} = y|\mathcal{F}_{T^m})$.
3. Remarquer que $T_x = \sum_{k=0}^{+\infty}(T_A^{k+1} - T_A^k)\mathbb{1}_{\{S>k\}}$, où $S = \inf\{n \geq 1; Y_n = x\}$.

Indication 90 1. (a) On pourra remarquer que $f(X_{n\wedge T}) - f(X_0) = \sum_{i=1}^n \mathbb{1}_{\{i-1<T\}}\left(f(X_i) - f(X_{i-1})\right)$.

   (b) Ne pas oublier que f est positive.

2. (a) Si $\mathbb{E}[f(X_0)] = +\infty$, il n'y a rien à montrer. Sinon, commencer par montrer que pour tout i, $f(X_i)\mathbb{1}_{\{i-1<T\}}$ est intégrable et $\mathbb{E}[f(X_i)\mathbb{1}_{\{i-1<T\}}|\mathcal{F}_{i-1}] = (Pf)(X_{i-1})\mathbb{1}_{\{i-1<T\}}$.

   (b) Appliquer le 1.(a) à $M' = M \cup \{x \in E; f(x) \geq n\}$.

   (c) Faire partir la chaîne d'un point $x \in M$.

Indication 91 1. Il y a plusieurs manières de montrer que T_k est un temps d'arrêt. On peut procéder par récurrence ou remarquer que

$$T_k = \inf\{n \geq 0; n \in A \text{ ou } \sum_{i=0}^{n-1} \mathbb{1}_{\{}X_i \neq X_{i+1} \geq k\}.$$

Pour montrer que les $(T_k)_{k\in\mathbb{N}}$ sont finis presque sûrement, on peut remarquer que $T_{k+1} = T_k + F(X_{T_K+.})$, où $F(x) = \inf\{n \geq 0, x_n \neq x_0\}$ et montrer par récurrence sur k que pour toute mesure initiale μ, T_k est presque sûrement fini et $(X_{T_k+.})$ est une chaîne de Markov de loi initiale $\mathbb{P}^{\mu}_{X_{T_k}}$.

2. On pourra calculer $\mathbb{P}(X_{T_{k+1}} = j|\mathcal{F}_{T_k})$. Pour ce faire, on peut remarquer qu'il existe $\Psi : \mathbb{R}^{\mathbb{N}} \to S$, avec $X_{T_{k+1}} = \Psi(X_{T+.})$.

Indication 92 On remarque que si $S_T = -b$, la suite repassera nécessairement par zéro. Si $S_T = a$, on est ramené à un problème de lutte contre un joueur inruinable.

Indication 93 1. On discutera suivant la parité de d.

2. On pourra exhiber une mesure invariante.
3. Si X_0 est pair, la suite $(X_{2n})_{n\geq 0}$ ne prend que des valeurs “paires”.

Indication 94 1. Utiliser la représentation canonique.

2. On remarque que $Y_{n+1} = Y_n + (-1)^{\mathbb{1}_{\{X_n(V_{n+1})=1\}}}$.
3. Commencer par étudier le cas où γ est une masse de Dirac.
4. Calculer la probabilité de chaque N-uplet.

5. Utiliser les questions précédentes.
6. On pourra montrer que la chaîne est réversible.

Indication 95 Si l'on note D_n l'indicatrice de l'événement “la société est disponible le lundi de la semaine n”, alors on a
— pour la stratégie A : $D_{n+1} = 1 - D_n(1 - A_n)B_n$.
— pour la stratégie B : $D_{n+1} = 1 - D_nB_n$.
Quant au gain des travaux commencés à la semaine n, il est
— pour la stratégie A : $G_n = D_n(200A_n + 360B_n - 360A_nB_n)$.
— pour la stratégie B : $G_n = D_n(200A_n + 360B_n - 200A_nB_n)$.
On peut remarquer que $(D_n)_{n\geq 0}$ et $(A_n, B_n, D_n)_{n\geq 0}$ sont des chaînes de Markov.

Indication 96 Utiliser le théorème ergodique des chaînes de Markov

Indication 97 1. C'est fait dans le cours !

2. De manière plus générale, on peut remarquer qu'une marche aléatoire symétrique sur un groupe fini admet toujours une probabilité réversible très simple.
3. Si μ est une mesure réversible et que x et y sont deux voisins, on a $\frac{\mu(x)}{d(x)} = \frac{\mu(y)}{d(y)}$, où $d(x)$ est le nombre de voisins de x.

Indication 98 On peut représenter l'état du système au temps n par un vecteur $(x(1), x(2), \dots, x(2N))$, où $x(i)$ est le numéro de l'urne (0 ou 1) où est la boule i. On peut supposer que les boules numérotées de 1 à N sont noires. Pour passer du temps n au temps $n+1$, on choisit un “1” au temps n que l'on remplace par un “0” et un “0” au temps n que l'on remplace par un “1”. X_n est une chaîne de Markov irréductible à valeurs dans $\{x \in \{0,1\}^{2N}; \sum_{i=1}^{2N} x_i = N\}$. Si deux états communiquent, la probabilité de passer de l'un à l'autre est $1/N^2$.

Indication 99 1. Utiliser la propriété de Markov forte.

2. On pourra noter que $\{\tau_n = k\} = \{S_{k-1} = n-1, X_k = 1\}$.
3. Utiliser la première question.

Annexe B

Solutions des exercices corrigés

B.1 Solutions des exercices corrigés sur les variables de Bernoulli

Solution 1 Si $p = 1$, on a presque sûrement $X = 1$, donc $\nu_1 = \delta_1$ qui n'a pas de densité par rapport à la mesure de Lebesgue. Supposons donc $p \in [0, 1[)$. Posons $A_n(x) = r(\lfloor 2^n x \rfloor, 2)$, où $r(n, 2)$ est le résultat de la division euclidienne de n par 2, puis $E_p = \{x \in [0, 1[; \lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^n A_n(x) = p\}$. On a

$$\begin{aligned}\nu_p(E_p) = \mathbb{P}(X \in E_p) &\geq \mathbb{P}(X \in E_p, \lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^n \omega_k(x) = p) \\ &\geq \mathbb{P}(\forall k \geq 1 \quad A_k(X) = \omega_k; \lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^n \omega_k(x) = p) \\ &= \mathbb{P}(\lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^n \omega_k(x) = p) = 1\end{aligned}$$

où la dernière égalité provient de la loi forte des grands nombres. L'avant-dernière égalité provient de l'identité

$$\{\forall k \geq 1 \quad A_k(X) = \omega_k; \lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^n \omega_k(x) = p)\} = \{\lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^n \omega_k(x) = p\},$$

qui mérite sans doute quelques mots : soit $(\omega_n)_{n\geq 1}$ une suite à valeurs dans $\{0, 1\}$ telle que $\frac{1}{n}\sum_{k=1}^n \omega_k(x) \to p$.

Pour $n \geq 1$, on peut écrire

$$2^n X = \sum_{k=0}^{n-1} 2^{n-k}\omega_k + \omega_n + \sum_{k=n+1}^{+\infty} \frac{\omega_k}{2^{k-n}}$$

On a pour tout $k \geq n$, $0 \leq \omega_k \leq 1$, d'où $0 \leq \sum_{k=n+1}^{+\infty} \frac{\omega_k}{2^{k-n}} \leq 1$. Cependant, il existe au moins un $k \geq n+1$ tel que $\omega_k < 1$, sinon on aurait $\lim_{n\to+\infty} \frac{1}{n} \sum_{k=1}^{n} \omega_k(x) = 1$. Ainsi, on a $\sum_{k=n+1}^{+\infty} \frac{\omega_k}{2^{k-n}} < 1$, d'où comme $\sum_{k=0}^{n-1} 2^{n-k}\omega_k + \omega_n$ est entier :

$$\lfloor 2^n X \rfloor = \sum_{k=0}^{n-1} 2^{n-k}\omega_k + \omega_n.$$

Mais $\sum_{k=0}^{n-1} 2^{n-k}\omega_k$ est pair et $\omega_n \in \{0,1\}$, donc $\omega_n = r(\lfloor 2^n X \rfloor, 2) = A_n(X)$.

Pour $q \neq p$, $\nu_q(E_p) = \nu_q(E_p \cap E_q) + \nu_q(E_p \cap E_q^c) \leq \nu_q(\varnothing) + \nu_q(E_q^c) = 0$, soit $\nu_q(E_p) = 0$. E_p est de probabilité 1 sous ν_p, 0 sous ν_q si $q \neq p$: ces mesures de probabilité sont étrangères deux à deux . En particulier pour $p \neq 1/2$, p est étrangère à $\nu_{1/2}$ qui est la restriction à $[0,1]$ de la mesure de Lebesgue. Finalement, ν_p a une densité par rapport à la mesure de Lebesgue si et seulement si $p = 1/2$.

B.2 Solution des exercices sur l'équi-intégrabilité

Solution 2 On a

$$X_n \mathbb{1}_{\{X_n \geq M\}} \leq \frac{X_n \log(1+X_n)}{\log(1+M)}$$

Donc

$$\sup_{n\geq 1} \mathbb{E}[X_n \mathbb{1}_{\{X_n \geq M\}}] \leq \frac{\sup_{n\geq 1} \mathbb{E}[X_n \log(1+X_n)]}{\log(1+M)},$$

qui tend vers 0 quand M tend vers l'infini.

Solution 3

1. D'après la loi forte des grands nombres $\frac{X_1+\dots X_n}{n}$ converge presque sûrement vers $\mathbb{E}[X_1] = \int_0^1 x dx = \frac{1}{2}$ et $\frac{X_1^2+\dots X_n^2}{n}$ converge presque sûrement vers $\mathbb{E}[X_1^2] = \int_0^1 x^2 \, dx = \frac{1}{3}$. En faisant le quotient, on a le résultat.
2. Comme les X_i sont entre 0 et 1, on a $Z_n^2 \leq n^2(X_1^2 + \dots X_n^2)^{-2}$.
   — Si $N_n \geq n/3$, alors $n^2(X_1^2 + \dots X_n^2)^{-2} \leq 144$,
   — sinon, on a $n^2(X_1^2 + \dots X_n^2)^{-2} \leq n^2(X_1^2 + \dots X_N^2)^{-2}\mathbb{1}_{\{N_n < n/3\}}$.
3. $\mathbb{E}Q_N^2 = \int_{[0,1]^N} \frac{1}{(x_1^2+\dots+x_N^2)^4} dx_1 \dots\dots dx_N \leq \frac{1}{2^N} \int_{B(0,\sqrt{N})} \frac{1}{\|x\|_2^8} dx_1 \dots dx_N$, où la dernière boule est une boule pour la norme euclidienne de $\mathbb{R}^N$. D'après le théorème d'intégration des fonctions radiales , on a $\int_{B(0,\sqrt{N})} \frac{1}{\|x\|_2^8} dx_1 \dots dx_N = V_N \int_0^{\sqrt{N}} t^{-8} N t^{N-1} dt$, où V_N est le volume de la boule unité de dimension N. Ainsi pour $N = 9$, on a bien l'intégrabilité.

4. En prenant $N = 9$ dans l'inégalité de 2. et en appliquant l'inégalité de Cauchy-Schwarz, on a pour $n \geq 9$
$\mathbb{E}(Z_n^2) \leq 144 + n^2\|Q_9\|_2\sqrt{\mathbb{P}(N_n < n/3)}$. Ainsi pour voir que $\mathbb{E}(Z_n^2)$ est bornée, il suffit de remarquer que $\mathbb{P}(N_n < n/3)$ décroît exponentiellement vite. On peut utiliser par exemple l'inégalité de Hoeffding :

$$\begin{aligned}\mathbb{P}(N_n < n/3) &\leq \mathbb{P}(|N_n - \mathbb{E}N_n| \geq n/6)\\ &\leq 2\exp(-2\frac{(n/6)^2}{n}) = 2\exp(-\frac{n}{18}),\end{aligned}$$

mais n'importe quelle inégalité de type "grandes déviations" convient. Ici, on peut facilement les retrouver "à la main" : soit $\theta > 0$. On a

$$\begin{aligned}\mathbb{P}(N_n \leq n/3) &\leq \mathbb{P}(e^{-\theta N_n} \geq e^{-\theta n/3})\\ &\leq e^{\theta n/3}\mathbb{E}(e^{-\theta N_n}) = e^{\theta n/3}(\frac{1+e^{-\theta}}{2})^n = C_\theta^n,\end{aligned}$$

avec $C_\theta = \frac{1}{2}(e^{\theta/3} + e^{-2\theta/3})$. Je choisis $\theta > 0$ tel que $e^{\theta/3} = 1,5$: j'ai alors $C_\theta = \frac{1}{2}(\frac{3}{2} + \frac{4}{9}) < 1$. La suite Z_n est bornée dans L^2, donc équi-intégrable. La convergence presque sûre ajoutée à l'équi-intégrabilité entraîne la convergence dans L^1.

Solution 4 Comme $(X_n)_{n\geq\lambda}$ est à valeurs positives [1], pour montrer que $(X_n)_{n\geq\lambda}$ est bornée dans L^3, il suffit de majorer $\mathbb{E}X_n^3$ indépendamment de n.

— Méthode 1.
Si $Y_1, \dots, Y_n$ sont des variables de Bernoulli de paramètre λ/n et que l'on pose $Z_k = Y_k - \lambda/n$, puis $S_k = Z_1 + \dots + Z_k$, on a

$$\mathbb{E}(X_n^3) = \mathbb{E}(S_n+\lambda)^3 = \mathbb{E}(S_n^3)+\lambda^3+3\lambda\mathbb{E}(S_n^2)+3\lambda^2\mathbb{E}(S_n) = \mathbb{E}(S_n^3)+\lambda^3+3\lambda\mathbb{E}(S_n^2).$$

On a $\mathbb{E}S_n^2 = \operatorname{Var} S_n = \operatorname{Var} X_n = n\lambda/n(1-\lambda/n) \leq \lambda$.
Comme $S_{k+1}^3 = S_k^3 + Z_{k+1}^3 + 3S_k^2 Y_{k+1} + 3Z_{k+1}S_k^2$, on a par récurrence sur $k \leq n$: $\mathbb{E}S_k^3 = \sum_{i=1}^k \mathbb{E}(Z_i^3)$, d'où en particulier

$$\begin{aligned}\mathbb{E}S_n^3 &= n\mathbb{E}Z_1^3 = n(\lambda/n(1-\lambda/n)^3 + (1-\lambda/n)(0-\lambda/n)^2)\\ &= n\lambda/n(1-\lambda/n)((1-\lambda/n)^2 + \lambda/n) \leq \lambda\end{aligned}$$

Ainsi $\mathbb{E}X_n^3 \leq \lambda + \lambda^3 + 3\lambda^2$, donc (X_n) est bornée dans L^3.

1. L'oubli de cette vérification est une source classique d'erreurs. Ainsi, si X_n suit la loi uniforme sur $[-n, n]$, $\mathbb{E}X_n^3$ est identiquement nulle, mais (X_n) n'est pas bornée dans L^3.

— Méthode 2.
Si $B_1, \dots, B_n$ sont des nombres valant 0 ou 1, le nombre de triplets (i,j,k) avec $i<j<k$ et $B_i = B_j = B_k = 1$, vaut $\sum_{1\le i<j<k\le n} B_iB_jB_k$. Mais c'est aussi le nombre de parties de cardinal 3 de l'ensemble des i tels que $B_i = 1$. Comme il y a exactement $S_n = B_1 + \dots + B_n$ tels i, on a

$$\sum_{1\le i<j<k\le n} B_iB_jB_k = \frac{S_n(S_n-1)(S_n-2)}{6}.$$

C'est un raisonnement purement déterministe. Maintenant, si $B_1, \dots, B_n$ sont indépendants suivant la loi de Bernoulli de paramètre p, S_n suit alors la loi binomiale $\mathcal{B}(n,p)$; par ailleurs, par linéarité de l'espérance, on a

$$\binom{n}{3}p^3 = \frac{1}{6}\mathbb{E}[S_n(S_n-1)(S_n-2)].$$

Ainsi $\mathbb{E}(X_n(X_n-1)(X_n-2)) = (\lambda/n)^3 n(n-1)(n-2) \le \lambda^3$. Pour tout $x \ge 3$, on a

$$x^3 = x(x-1)(x-2)\frac{x}{x-1}\frac{x}{x-2} \le \frac{9}{2}x(x-1)(x-2).$$

On en déduit que pour tout $k \in \mathbb{N}$, $x^3 \le 8 + \frac{9}{2}x(x-1)(x-2)$. Ainsi $\mathbb{E}(X_n^3) \le 8 + \frac{9}{2}\mathbb{E}(X_n(X_n-1)(X_n-2)) \le 8 + \frac{9}{2}\lambda^3$. (X_n) est donc bornée dans L^3.

Soit X_λ^* une variable suivant la loi de Poisson de paramètre λ. X_n est bornée dans L^3, donc équiintégrable, et converge en loi vers X_λ, donc $\mathbb{E}X_n$ converge vers $\mathbb{E}(X_\lambda^*)$. Comme $\mathbb{E}X_n = \lambda$, $\mathbb{E}(X_\lambda^*) = \lambda$. X_n est bornée dans L^3, donc X_n^2 est bornée dans $L^{3/2}$, donc $(X_n^2)_{n\ge1}$ est équi-intégrable. $(X_n^2)_{n\ge1}$ converge en loi vers X_λ, donc $\mathbb{E}X_n^2$ converge vers $\mathbb{E}((X_\lambda^*)^2)$. Comme la variance est l'espérance du carré moins le carré de l'espérance, la variance de X_λ^* est la limite de $\operatorname{Var} X_n = n\lambda/n(1-\lambda/n)$, soit λ.

Solution 5 Soient $(X_n)_{n\ge1}$ une suite de variables de Bernoulli de paramètre 1/2. On pose $S_n = X_1 + \dots X_n$. Si l'on pose $\varphi(x) = x_+$, on a

$$\begin{aligned}
\mathbb{E}\varphi(S_{2n}-n) &= \sum_{k=-n}^{n} \mathbb{P}(S_{2n}-n=k)\varphi(k) \\
&= \sum_{k=-n}^{n} \mathbb{P}(S_{2n}-n=k).0 + \sum_{k=1}^{n} \mathbb{P}(S_{2n}-n=k).k \\
&= \sum_{k=1}^{n} 2^{-2n}\binom{2n}{n+k}k \\
&= 2^{-2n}u_n
\end{aligned}$$

On a $\mathbb{E}(S_{2n}) = 2n \times \frac{1}{2} = n$ et $\operatorname{Var} S_{2n} = 2n \operatorname{Var} X_1 = n/2$. Ainsi

$$u_n = 2^{2n}\mathbb{E}\varphi(S_{2n} - n) = 2^{2n}\sqrt{n/2}\mathbb{E}\varphi\left(\frac{S_{2n} - \mathbb{E}(2n)}{\sqrt{\operatorname{Var} S_n}}\right).$$

Soit Z suivant la loi normale $\mathcal{N}(0,1)$, d'après le théorème central limite, la suite $\frac{S_{2n}-\mathbb{E}(2n)}{\operatorname{Var} S_n}$ tend en loi vers Z.
Comme φ est continue, la suite $Y_n = \varphi(\frac{S_{2n}-\mathbb{E}(2n)}{\sqrt{\operatorname{Var} S_n}})$ converge en loi vers $\varphi(Z)$.
On a $Y_n^2 \le \frac{(S_{2n}-\mathbb{E}(2n))^2}{\operatorname{Var} S_n}$, d'où $\mathbb{E}(Y_n^2) \le \frac{\mathbb{E}(S_{2n}-\mathbb{E}(2n))^2}{\operatorname{Var} S_n} = 1$. (Y_n) est bornée dans L^2 donc équi-intégrable. Comme elle converge en loi vers $\varphi(Z)$, $\mathbb{E}(Y_n)$ converge vers $\mathbb{E}\varphi(Z)$. On a

$$\begin{aligned}
\mathbb{E}\varphi(Z) &= \frac{1}{\sqrt{2\pi}}\int_0^{+\infty} xe^{-x^2/2}\,dx \\
&= \lim_{M\to+\infty} \frac{1}{\sqrt{2\pi}}\int_0^{M} xe^{-x^2/2}\,dx \\
&= \lim_{M\to+\infty} \frac{1}{\sqrt{2\pi}}\left(1 - e^{-M^2/2}\right) = \frac{1}{\sqrt{2\pi}}
\end{aligned}$$

Finalement

$$u_n \sim 2^{2n}\sqrt{n/2}\mathbb{E}\varphi(Z) = 2^{2n-1}\sqrt{n/\pi}.$$

B.3 Solutions des exercices sur l'espérance conditionnelle

Solution 10 1. La condition est évidemment nécessaire d'après la définition de l'espérance conditionnelle car $\{N = n\}$ est $\sigma(N)$-mesurable
Si A est $\sigma(N)$-mesurable, A s'écrit $A = \{N \in B\}$, où B est un borélien de $\mathbb{R}$. notons que comme N est à valeurs dans $\mathbb{N}$, on a presque sûrement $\mathbb{1}_A = \mathbb{1}_{\{N\in B\cap\mathbb{N}\}}$. On a

$$\begin{aligned}
X\mathbb{1}_A &= X\mathbb{1}_{\{N\in B\cap\mathbb{N}\}} \\
&= \sum_{n=0}^{+\infty} \mathbb{1}_{n\in B} X\mathbb{1}_{\{N=n\}} \qquad = \lim_{N\to+\infty}\sum_{n=0}^{N} \mathbb{1}_{n\in B} X\mathbb{1}_{\{N=n\}}
\end{aligned}$$

Comme $|\sum_{n=0}^{N} \mathbb{1}_{n\in B} X\mathbb{1}_{\{N=n\}}| \le X$, avec le théorème de convergence dominée, on a $\mathbb{E}[X\mathbb{1}_A] = \lim_{N\to+\infty}\sum_{n=0}^{+\infty} \mathbb{1}_{n\in B}\mathbb{E}[X\mathbb{1}_{\{N=n\}}]$. Les mêmes arguments donnent $\mathbb{E}[Y\mathbb{1}_A] = \lim_{N\to+\infty}\sum_{n=0}^{+\infty} \mathbb{1}_{n\in B}\mathbb{E}[Y\mathbb{1}_{\{N=n\}}]$. Comme $\mathbb{E}[X\mathbb{1}_{\{N=n\}}] = \mathbb{E}[Y\mathbb{1}_{\{N=n\}}]$, on a finalement $\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_A]$, ce qui donne le résultat voulu puisque Y est $\sigma(N)$-mesurable.

2. $\mathbb{E}(U\mathbb{1}_{\{N=n\}}) = \mathbb{E}((X_1+\cdots+X_n)\mathbb{1}_{N=n}) = \mathbb{E}(X_1+\cdots+X_n)\mathbb{E}(\mathbb{1}_{N=n})$ par indépendance. On a donc $\mathbb{E}(U\mathbb{1}_{\{N=n\}}) = n\mathbb{E}(X_1)\mathbb{E}(\mathbb{1}_{N=n}) = \mathbb{E}(X_1)\mathbb{E}(n\mathbb{1}_{N=n}) = \mathbb{E}(X_1)\mathbb{E}(N\mathbb{1}_{N=n}) = \mathbb{E}[\mathbb{E}(X_1)N\mathbb{1}_{N=n}]$, où la dernière égalité est encore obtenue par indépendance. Avec la première question, on en déduit $\mathbb{E}(U|N) = \mathbb{E}(X_1)N = qN$.

Solution 11 On a vu que $\mathbb{E}[U|T] = \mathbb{E}[X_1]T = qT$; on en déduit $\mathbb{E}[U] = \mathbb{E}(\mathbb{E}[U|T]) = \mathbb{E}(qT) = q\mathbb{E}[T] = q(\frac{1}{p} - 1)$, car $1 + T$ suit un loi géométrique de paramètre p.

$$\begin{aligned}
\mathbb{E}[T|U=k] &= \frac{\mathbb{E}T\mathbb{1}_{\{U=k\}}}{\mathbb{P}(U=k)} = \frac{\sum_{n\ge k} n\mathbb{P}(U=k, N=n)}{\sum_{n\ge k} \mathbb{P}(U=k, N=n)} \\
&= \frac{\sum_{n\ge k} np(1-p)^n\binom{n}{k}q^k(1-q)^{n-k}}{\sum_{n\ge k} p(1-p)^n\binom{n}{k}q^k(1-q)^{n-k}} \\
&= \frac{\sum_{n\ge k} nx^n\binom{n}{k}}{\sum_{n\ge k} x^n\binom{n}{k}} \text{ avec } x = (1-p)(1-q) \\
&= \frac{\sum_{n\ge k}(n+1)x^n\binom{n}{k}}{\sum_{n\ge k} x^n\binom{n}{k}} - 1 = \frac{\sum_{n\ge k}(k+1)\binom{n+1}{k+1}x^n}{\sum_{n\ge k}\binom{n}{k}x^n} - 1 \\
&= (k+1)\frac{\sum_{n\ge k}\binom{n+1}{k+1}x^n}{\sum_{n\ge k}\binom{n}{k}x^n} - 1
\end{aligned}$$

On fait une transformation d'Abel : on pose $r_n = \sum_{k\ge n} x^n$, ainsi que $c_n = \binom{n+1}{k+1}$ pour $n \ge k$ et $c_n = 0$ pour $n < k$:

Alors

$$\begin{aligned}
\sum_{n\ge k}\binom{n+1}{k+1}x^n &= \sum_{n\ge k} c_n(r_n - r_{n+1}) = \sum_{n\ge k} c_n r_n - \sum_{n\ge k} c_n r_{n+1} \\
&= \sum_{n\ge k} c_n r_n - \sum_{n\ge k+1} c_{n-1}r_n = \sum_{n\ge k} c_n r_n - \sum_{n\ge k} c_{n-1}r_n \\
&= \sum_{n\ge k}(c_n - c_{n-1})r_n
\end{aligned}$$

Pour $n = k$ on a $c_n - c_{n-1} = c_k - c_{k-1} = c_k = 1 = \binom{n}{n}$. Pour $n > k$, on a $c_n - c_{n-1} = \binom{n+1}{k+1} - \binom{n}{k+1} = \binom{n}{k}$. Par ailleurs $r_n = \frac{x^n}{1-x}$ Finalement

$$\begin{aligned}\mathbb{E}[T|U=k] &= (k+1)\frac{\sum_{n\geq k}\binom{n}{k}\frac{x^n}{1-x}}{\sum_{n\geq k}\binom{n}{k}x^n} - 1\\ &= \frac{k+1}{1-x} - 1 = \frac{k}{1-x} + \frac{x}{1-x}\end{aligned}$$

Finalement $\mathbb{E}[T|U] = \frac{1}{1-x}U + \frac{x}{1-x} = \frac{x+\mathbb{E}U}{1-x}$.

Pour mettre à l'épreuve nos calculs, on va vérifier que $\mathbb{E}(\mathbb{E}[T|U]) = \mathbb{E}(T)$. On va vérifier que $\frac{\mathbb{E}U+x}{1-x} = \mathbb{E}[T]$ ou, de manière équivalente $x = \frac{\mathbb{E}T-\mathbb{E}U}{\mathbb{E}T+1}$:

$$\begin{aligned}\frac{\mathbb{E}T-\mathbb{E}U}{\mathbb{E}T+1} &= \frac{(\frac{1}{p}-1)(1-q)}{(\frac{1}{p}-1)+1}\\ &= \frac{(\frac{1}{p}-1)(1-q)}{\frac{1}{p}}\\ &= (1-p)(1-q)\end{aligned}$$

Solution 12 1. Il est clair que $\mathbb{Q}_{n,b}$ est une mesure de probabilités (c'est une application probabilité conditionnelle).

Par construction $\mathbb{Q}_{n,b}$ est à support dans $\{0,1\}^{n+b-1}$. Soit $(y_1,\dots,y_{n+b-1}) \in \{0,1\}^{n+b-1}$. On pose $A = \{X_2 = y_1, \dots X_{n+b} = y_{n+b-1}\}$.

Par définition,

$$\begin{aligned}\mathbb{Q}_{n,b}(A) &= \frac{\mu_{n,b}(\{1,y_1,\dots,y_{n+b-1}\})}{\mu_{n,b}(X_1=1)}\\ &= \frac{\mathbb{1}_{1+y_1+\dots+y_{n+b-1}=n\}}}{\mu_{n,b}(X_1=1)|\Omega_{n,b}|}\\ &= \frac{\mathbb{1}_{y_1+\dots+y_{n+b-1}=n-1\}}}{\mu_{n,b}(X_1=1)|\Omega_{n,b}|}\end{aligned}$$

On a donc

$$\mathbb{Q}_{n,b}(X_2 = y_1, \dots X_{n+b} = y_{n+b-1}) = \frac{\mathbb{1}_{y_1+\dots+y_{n+b-1}=n-1\}}}{\mu_{n,b}(X_1=1)|\Omega_{n,b}|} \text{ tandis que}$$

$$\mu_{n-1,b}(\{y_1,\dots,y_{n+b-1}\}) = \frac{\mathbb{1}_{y_1+\dots+y_{n+b-1}=n-1\}}}{|\Omega_{n-1,b}|}.$$

Ainsi, la loi de $(X_2, \dots X_{n+b})$ sous $\mathbb{Q}_{n,b}$ coïncide, à un facteur près, avec $\mu_{n-1,b}$. Mais deux mesures de probabilités qui coïncident à un facteur près sont égales.

Par ailleurs, on obtient que $\mu_{n,b}(X_1 = 1) = \frac{|\Omega_{n-1,b}|}{|\Omega_{n,b}|}$. L'application $x \mapsto x^{-1}(\{1\})$ établit une bijection entre $\Omega_{n,b}$ et $\mathcal{B}_n(\{1,\dots,n+b\})$, donc $|\Omega_{n,b}| = |\mathcal{B}_n(\{1,\dots,n+b\})| = \frac{(n+b)!}{n!b!}$, soit après calculs,

$$\mu_{n,b}(X_1 = 1) = \frac{n}{n+b},$$

qui sera utile dans la suite.

2. Notons

$$\begin{aligned}\Psi_n : \{0,1\}^n &\to \{0,\dots,n-1\} \cup \{+\infty\} \\ (x_i)_{1\le i\le n} &\mapsto \inf\{i \in \{0,\dots,n-1\}; x_{i+1} = 0\}.\end{aligned}$$

On a

$$\Psi_{n+b}(X_1,\dots,X_{n+b}) = \mathbb{1}_{\{X_1=1\}}(1 + \Psi_{n+b-1}(X_2,\dots,X_{n+b})),$$

d'où

$$\mathbb{E}_{\mu_{n,b}}\Psi_{n+b}(X_1,\dots,X_{n+b}) = \mu_{n,b}(X_1 = 1)\mathbb{E}_{\mu_{n,b}}(1+\Psi_{n+b-1}(X_2,\dots,X_{n+b})|X_1 = 1),$$

Cependant, posant $f = 1 + \Psi_{n+b-1}$ et découpant suivant les valeurs prises par $(X_2,\dots,X_{n+b})$, on voit que

$$\begin{aligned}&\mathbb{E}_{\mu_{n,b}}(f(X_2,\dots,X_{n+b})|X_1 = 1) \\ &= \frac{1}{\mu_{n,b}(X_1 = 1)} \sum_{y\in\Omega_{n-1,b}} \mathbb{E}_{\mu_{n,b}}(\mathbb{1}_{\{(X_2,\dots,X_{n+b})=y\}}\mathbb{1}_{\{X_1=1\}}f(y) \\ &= \sum_{y\in\Omega_{n-1,b}} \mathbb{P}_{\mu_{n,b}}((X_2,\dots,X_{n+b}) = y|X_1 = 1)f(y) \\ &= \sum_{y\in\Omega_{n-1,b}} \mathbb{Q}_{n,b}((X_2,\dots,X_{n+b}) = y)f(y) \\ &= \sum_{y\in\Omega_{n-1,b}} \mu_{n-1,b}(\{y\})f(y) \\ &= \mathbb{E}_{\mu_{n-1,b}}(1 + \Psi_{n+b-1}(X_1,\dots,X_{n+b-1}))\end{aligned}$$

Ce qui, posant $u_n = \mathbb{E}_{\mu_{n,b}}$, et b étant fixé, nous donne la jolie récurrence

$$u_n = \frac{n}{n+b}(1 + u_{n-1}),$$

Si $n = 0$, $\psi_{n+b} = \psi_b$ est constante égale à 0, d'où $u_0 = 0$, et on peut alors montrer par une récurrence simple que $u_n = \frac{n}{b+1}$ pour tout $n \geq 0$.

B.4 Solutions des exercices sur les martingales

Solution 23 1. Posons $\mathcal{F}_n = \sigma(U_1, \dots, U_n)$. Par récurrence, on voit facilement que X_n est mesurable par rapport à $\mathcal{F}_n$. Ainsi $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \mathbb{E}[U_{n+1}|\mathcal{F}_n] + X_n^2\mathbb{E}[(1-U_{n+1})|\mathcal{F}_n]$. D'un autre côté, U_{n+1} est indépendant de $\mathcal{F}_n$. On en déduit que

$$\mathbb{E}[X_{n+1}|\mathcal{F}_\backslash] = \mathbb{E}[U_{n+1}] + X_n^2\mathbb{E}[1-U_{n+1}] = \frac{1+X_n^2}{2} \geq X_n.$$

La suite $(X_n)_{n\geq 0}$ est bien une sous-martingale.

2. (a) X_{n+1} est une combinaison convexe de X_n^2 et 1. Ainsi, si X_n est dans $[0,1]$, X_n^2 sera encore dans $[0,1]$, et par combinaison convexe, X_{n+1} sera encore dans $[0,1]$. Ainsi, pour $a \in [0,1]$, la suite (X_n) sera à valeurs dans $[0,1]$, donc bornée dans L^1. Comme $(X_n)_{n\geq 0}$ est une sous-martingale, le théorème de Doob assure que (X_n) converge presque sûrement vers une variable X^*.

(b) $\mathbb{E}[X_{n+1}] = \mathbb{E}[\mathbb{E}[X_{n+1}|\mathcal{F}_n]] = \mathbb{E}\frac{1+X_n^2}{2}$. Par convergence dominée (par 1), le membre de gauche converge vers $\mathbb{E}[X^*]$ tandis que le membre de droite converge vers $\mathbb{E}\frac{1+(X^*)^2}{2}$. On a donc

$$\mathbb{E}X^* = \mathbb{E}\frac{1+(X^*)^2}{2}.$$

Mais comme $X^* \leq \frac{1+(X^*)^2}{2}$, on obtient que $X^* = \frac{1+(X^*)^2}{2}$ presque sûrement, d'où $X^* = 1$ presque sûrement.

Solution 24 1. (a) On procède par récurrence sur n. Pour $n = 1$ et i entre 1 et S, on a $B_{d+i} = T_1 = B_{U_1}$. Or, on a $1 \leq U_1 \leq d$, or $B_1, \dots, B_d$ sont déterministes, donc $B_{d+i} = B_{U_1}$ est $\sigma(U_1) = \mathcal{F}_1$-mesurable. Supposons le résultat acquis jusqu'au rang $n-1$. Comme $B_{d+i+(n-1)S} = T_n$, il s'agit de montrer que T_n est $\mathcal{F}_n$-mesurable. Or $T_n = \sum_{k=1}^{d+(n-1)S} \mathbb{1}_{\{U_n=k\}} B_k$. Les variables $B_1, \dots, B_d$ sont déterministes. Si on fait varier i de 1 à S et k de 1 à $n-1$, $d+(k-1)S+i$ prend toutes les valeurs entre $d+1$ et $d+(n-2)S+S = d+(n-1)S$ Comme la suite des tribus $(\mathcal{F}_k)_{k\geq 1}$ est croissante et que des variables constantes sont mesurables par rapport à n'importe quelle tribu, les variables $B_1, \dots, B_{d+(n-1)S}$ sont $\mathcal{F}_{n-1}$-mesurables, donc en particulier $\mathcal{F}_n$-mesurables. Comme pour tout k entre 1 et $d+(n-1)S$, la variable $\mathbb{1}_{\{U_n=k\}}$ est $\mathcal{F}_n$-mesurable, il s'ensuit que T_n est $\mathcal{F}_n$-mesurable, ce qui montre la propriété d'hérédité recherchée.

(b) Comme $1 = \sum_{k=1}^{d+nS} \mathbb{1}_{\{U_{n+1}=k\}}$, on a

$$\begin{aligned}
V_{n+1} &= V_n + Se_{T_{n+1}} = V_n + Se_{B_{U_{n+1}}} \\
&= V_n + S(\sum_{k=1}^{d+nS} \mathbb{1}_{\{U_{n+1}=k\}})e_{B_{U_{n+1}}} \\
&= V_n + S(\sum_{k=1}^{d+nS} \mathbb{1}_{\{U_{n+1}=k\}}e_{B_{U_{n+1}}}) \\
&= V_n + S(\sum_{k=1}^{d+nS} \mathbb{1}_{\{U_{n+1}=k\}}e_{B_k})
\end{aligned}$$

En faisant le produit scalaire par e_i, on obtient

$$\begin{aligned}
V_{n+1}^i &= V_n^i + S(\sum_{k=1}^{d+nS} \mathbb{1}_{\{U_{n+1}=k\}}\langle e_{B_k}, e_i\rangle) \\
&= V_n^i + S(\sum_{k=1}^{d+nS} \mathbb{1}_{\{U_{n+1}=k\}}\mathbb{1}_{\{B_k=i\}}
\end{aligned}$$

(c) Comme on l'a déjà noté, pour $k \leq d + nS$, B_k est $\mathcal{F}_n$-mesurable. On en déduit que V_n, et donc V_n^i, est $\mathcal{F}_n$-mesurable. On a donc avec la question précédente

$$\begin{aligned}
\mathbb{E}(V_{n+1}^i|\mathcal{F}_n) &= V_n^i + S\sum_{k=1}^{d+nS} \mathbb{E}(\mathbb{1}_{\{U_{n+1}=k\}}\mathbb{1}_{\{B_k=i\}}|\mathcal{F}_n) \\
&= V_n^i + S\sum_{k=1}^{d+nS} \mathbb{E}(\mathbb{1}_{\{U_{n+1}=k\}}\mathbb{1}_{\{B_k=i\}}|\mathcal{F}_n) \\
&= V_n^i + S\sum_{k=1}^{d+nS} \mathbb{1}_{\{B_k=i\}}\mathbb{E}(\mathbb{1}_{\{U_{n+1}=k\}}|\mathcal{F}_n) \\
&= V_n^i + S\sum_{k=1}^{d+nS} \mathbb{1}_{\{B_k=i\}}\mathbb{E}(\mathbb{1}_{\{U_{n+1}=k\}})
\end{aligned}$$

où la dernière inégalité vient du fait que U_{n+1} est indépendant de la tribu $\mathcal{F}_n = \sigma(U_1, \dots, U_n)$. On a donc

$$\begin{aligned}
\mathbb{E}(V_{n+1}^i|\mathcal{F}_n) &= V_n^i + \frac{S}{d+Sn}\sum_{k=1}^{d+nS} \mathbb{1}_{\{B_k=i\}} \\
&= V_n^i + \frac{S}{d+Sn}V_n^i = \frac{d+S(n+1)}{d+Sn}V_n^i
\end{aligned}$$

(d) En divisant par $d+S(n+1)$, on obtient

$$\mathbb{E}\left(\frac{V_{n+1}^i}{d+S(n+1)}|\mathcal{F}_n\right) = \frac{V_n^i}{d+Sn};$$

comme $\frac{V_n^i}{d+Sn}$ est $\mathcal{F}_n$ mesurable et que $(\mathcal{F}_n)_{n\geq 0}$ est une filtration, cela montre que $(\frac{V_n^i}{d+Sn})_{n\geq 0}$ est une $\mathcal{F}_n$-martingale.

(e) Pour tout i, la suite $(\frac{V_n^i}{d+Sn})_{n\geq 1}$ est une martingale bornée (comme $0 \leq V_n^i \leq d+Sn$, la martingale prend ses valeurs dans $[0,1]$), elle converge donc presque sûrement. Comme ses composantes convergent, la suite de vecteurs aléatoires $(\frac{V_n}{d+Sn})_{n\geq 1}$ converge presque sûrement.

(f) Soit $n \geq 0$. On a $\mathbb{1}_{\{T_{n+1}=i\}} = \frac{V_{n+1}^i - V_n^i}{S}$, d'où

$$\mathbb{E}(\mathbb{1}_{\{T_{n+1}=i\}}|\mathcal{F}_n) = \mathbb{E}(\frac{V_{n+1}^i - V_n^i}{S}|\mathcal{F}_n) = \frac{V_n^i}{d+Sn},$$

d'après 1c. En réintégrant, on a $\mathbb{P}(T_{n+1} = i) = \mathbb{E}(\frac{V_n^i}{d+Sn})$. Or, $(\frac{V_n^i}{d+Sn})_{n\geq 0}$ est une martingale, donc $\mathbb{E}(\frac{V_n^i}{d+Sn}) = \mathbb{E}(\frac{V_0^i}{d}) = \frac{d_i}{d}$. Ainsi, pour tout $n \geq 1$, $\mathbb{P}(T_n = i) = \frac{d_i}{d}$.

2. (a) Les événement $\{U_n = k\}$, où k varie de 1 à $d+(n-1)S$ forment une partition de l'espace. On a donc

$$\begin{aligned}\mathbb{1}_{\{T_1=t_1,\dots,T_n=t_n\}} &= \sum_{k=1}^{d+(n-1)S} \mathbb{1}_{\{T_1=t_1,\dots,T_n=t_n\}}\mathbb{1}_{\{U_n=k\}} \\ &= \sum_{k=1}^{d+(n-1)S} \mathbb{1}_{\{T_1=t_1,\dots,T_{n-1}=t_{n-1},B_k=t_n\}}\mathbb{1}_{\{U_n=k\}}\end{aligned}$$

Donc, comme $\{T_1 = t_1,\dots,T_{n-1} = t_{n-1}, B_k = t_n\} \in \mathcal{F}_{n-1}$ et que U_n est indépendant de $\mathcal{F}_n$, on a

$$\begin{aligned}\mathbb{E}(\mathbb{1}_{\{T_1=t_1,\dots,T_n=t_n\}}|\mathcal{F}_{n-1} &= \sum_{k=1}^{d+(n-1)S} \mathbb{1}_{\{T_1=t_1,\dots,T_{n-1}=t_{n-1},B_k=t_n\}}\mathbb{P}(U_n = k) \\ &= \sum_{k=1}^{d+(n-1)S} \mathbb{1}_{\{T_1=t_1,\dots,T_{n-1}=t_{n-1}\}}\mathbb{1}_{\{B_k=t_n\}}\mathbb{P}(U_n = k) \\ &= \mathbb{1}_{\{T_1=t_1,\dots,T_{n-1}=t_{n-1}\}}V_{n-1}^{t_n}\frac{1}{d+(n-1)S} \\ &= \mathbb{1}_{\{T_1=t_1,\dots,T_{n-1}=t_{n-1}\}}\frac{d_{t_n}+Sa_{t_n}^{n-1}}{d+(n-1)S}\end{aligned}$$

En réintégrant, on obtient

$$\mathbb{P}(T_1 = t_1, \dots, T_n = t_n) = \mathbb{P}(T_1 = t_1, \dots, T_{n-1} = t_{n-1})\frac{d_{t_n} + Sa_{t_n}^{n-1}}{d + (n-1)S}.$$

Montrons la formule demandée par récurrence.
Comme $\mathbb{P}(T_1 = t_1) = \frac{d_{t_1}}{d}$, la formule est vraie pour $n = 1$. Si elle est vraie au rang $n-1$, on a alors

$$\begin{aligned}
&\mathbb{P}(T_1 = t_1, \dots, T_n = t_n)\\
&= \frac{\prod_{i=1}^m d_i(d_i+S)\dots(d_i + (a_i^{n-1}-1)S)}{d(d+S)\dots(d+(n-2)S)} \times \frac{d_{t_n} + Sa_{t_n}^{n-1}}{d+(n-1)S}\\
&= \frac{\prod_{i=1}^m d_i(d_i+S)\dots(d_i + (a_i^{n-1}-1)S)}{d(d+S)\dots(d+(n-1)S)}(d_{t_n} + Sa_{t_n}^{n-1})
\end{aligned}$$

Pour $i \in \{1, \dots, m\}\backslash\{t_n\}$, on a $a_i^{n-1} = a_i^n$, donc

$$d_i(d_i+S)\dots(d_i+(a_i^{n-1}-1)S) = d_i(d_i+S)\dots(d_i+(a_i^n-1)S).$$

Pour l'unique i tel que $i = t_n$, $a_i^{n-1} = a_i^n - 1$, donc

$$\begin{aligned}
&d_i(d_i+S)\dots(d_i+(a_i^{n-1}-1)S) \times (d_{t_n} + Sa_{t_n}^{n-1})\\
&= d_i(d_i+S)\dots(d_i+(a_i^n-2)S) \times (d_i + S(a_i^n - 1)),
\end{aligned}$$

ce qui donne

$$\mathbb{P}(T_1 = t_1, \dots T_n = t_n) = \frac{\prod_{i=1}^m d_i(d_i+S)\dots(d_i+(a_i^n-1)S)}{d(d+S)\dots(d+(n-1)S)}.$$

En divisant le numérateur et le dénominateur par S^n, on obtient

$$\mathbb{P}(T_1 = t_1, \dots T_n = t_n) = \frac{\prod_{i=1}^m (d_i/S)(d_i/S+1)\dots(d_i/S+(a_i^n-1))}{(d/S)(d/S+1)\dots(d/S+(n-1))}$$

En écrivant $x = \frac{\Gamma(x+1)}{\Gamma(x)}$ et en simplifiant les produits apparaissant, on obtient $(d_i/S)(d_i/S+1)\dots(d_i/S+(a_i^n-1)) = \frac{\Gamma(d_i/S+a_i^n)}{\Gamma(d_i/S)}$ et $(d/S)(d/S+1)\dots(d/S+(n-1)) = \frac{\Gamma(d/S+n)}{\Gamma(d/S)}$, d'où

$$\begin{aligned}
\mathbb{P}(T_1 = t_1, \dots T_n = t_n) &= \frac{\Gamma(d/S)}{\Gamma(d/S+n)}\prod_{i=1}^m \frac{\Gamma(d_i/S+a_i^n)}{\Gamma(d_i/S)}\\
&= \frac{(\prod_{i=1}^m \Gamma(d_i/S+a_i^n))/\Gamma(d/S+n)}{(\prod_{i=1}^m \Gamma(d_i/S))/\Gamma(d/S)}\\
&= \frac{\tilde{B}(d/S+a)}{\tilde{B}(d/S)},
\end{aligned}$$

ce qui donne le résultat voulu compte tenu des résultats admis sur les lois de Dirichlet.

(b) $V_n - d = \sum_{j=d+1}^{d+nS} e_{B_j} = \sum_{k=1}^{n} \sum_{i=1}^{S} e_{B_{d+(k-1)S}} = \sum_{k=1}^{n} S e_{T_k}$, donc $\frac{V_n-d}{S} = \sum_{k=1}^{n} e_{T_k}$, et

$$\begin{aligned}\mathbb{P}(\frac{V_n - d}{S} = (a_1, \dots, a_m)) &= \mathbb{P}(\sum_{k=1}^{n} e_{T_k} = (a_1, \dots, a_m))\\ &= \sum_{t \in A(a)} \mathbb{P}(T_1 = t_1, \dots, T_n = t_n),\end{aligned}$$

où $A(a)$ est l'ensemble des applications de $\{1, \dots, n\}$ dans $\{1, \dots, m\}$ prenant a_i fois la valeur i. Or, d'après la question précédente $\mathbb{P}(T_1 = t_1, \dots, T_n = t_n) = \mathbb{E}\left(\prod_{i=1}^{m} Y_i^{a_i}\right)$ pour tout $t \in A(a)$, On en déduit

$$\mathbb{P}(\frac{V_n - d}{S} = (a_1, \dots, a_m)) = |A(a)| \mathbb{E}\left(\prod_{i=1}^{m} Y_i^{a_i}\right) = \binom{n}{a_1, \dots, a_m} \mathbb{E}[\prod_{i=1}^{m} Y_i^{a_i}].$$

(c) Posons $W_n = \frac{V_n-d}{S}$. D'après le théorème de transfert

$$\begin{aligned}\mathbb{E}(\exp(i\langle W_n, u\rangle)) &= \sum_{a \in \{1,\dots,m\}^n} \mathbb{P}(W_n = a) \exp(i\langle a, u\rangle)\\ &= \sum_{a \in \{1,\dots,m\}^n} \binom{n}{a_1, \dots, a_m} \mathbb{E}[\prod_{i=1}^{m} Y_i^{a_i}] \prod_{k=1}^{m} e^{i a_k u_k}\\ &= \sum_{a \in \{1,\dots,m\}^n} \binom{n}{a_1, \dots, a_m} \mathbb{E}[\prod_{k=1}^{m} Y_k^{a_k}] \prod_{k=1}^{m} (e^{i u_k})^{a_k}\\ &= \sum_{a \in \{1,\dots,m\}^n} \binom{n}{a_1, \dots, a_m} \mathbb{E}[\prod_{k=1}^{m} (Y_k e^{i u_k})^{a_k}]\\ &= \mathbb{E}[\sum_{a \in \{1,\dots,m\}^n} \binom{n}{a_1, \dots, a_m} \prod_{k=1}^{m} (Y_k e^{i u_k})^{a_k}]\\ &= \mathbb{E}\left(\sum_{j=1}^{m} Y_j e^{i u_j}\right)^n,\end{aligned}$$

où la dernière égalité vient de la formule du multinôme.

(d) On commence par montrer que $\frac{V_n-d}{Sn} \Longrightarrow Y$. D'après le théorème de Lévy, il suffit de montrer que pour tout $u \in \mathbb{R}^m$,

$$\mathbb{E} \exp(i\langle \frac{V_n - d}{Sn}, u\rangle) = \varphi_n(u/n) = \mathbb{E}\left(\sum_{j=1}^{m} Y_j e^{i u_j/n}\right)^n$$

converge vers $\mathbb{E}\exp(i\langle Y,u\rangle)$. Or, pour tout ω, le vecteur $(Y_1(\omega),\dots,Y_m(\omega))$ est un vecteur de nombres positifs de somme 1 : on a donc

$$|\sum_{j=1}^m Y_j e^{iu_j/n}| \le \sum_{j=1}^m |Y_j e^{iu_j/n}| = \sum_{j=1}^m Y_j = 1,$$

d'où

$$\left|\left(\sum_{j=1}^m Y_j e^{iu_j/n}\right)^n\right| \le 1.$$

Comme 1 est intégrable, il suffit de montrer que si $y_1\dots,y_m$ sont des réels positifs de somme 1, la suite $(\sum_{k=1}^m y_k e^{iu_k/n})^n$ converge vers $\exp(\sum_{k=1}^m iy_k u_k)$ lorsque n tend vers l'infini. Le théorème de convergence dominée permettra alors de conclure. Si je pose $z_n = \sum_{k=1}^m y_k e^{iu_k/n}$ et $z'_n = \exp(\sum_{k=1}^m iy_k u_k/n)$, on a $|z_n|\le 1$ et $|z'_n|\le 1$, donc $|z_n^n - z'^n_n| \le n|z_n - z'_n|$. Cependant, pour tout k, on a $e^{iu_k/n} = 1 + i\frac{u_k}{n} + O(\frac{1}{n^2})$, donc

$$\begin{aligned}\sum_{k=1}^m y_k e^{iu_k/n} &= \sum_{k=1}^m y_k(1+i\frac{u_k}{n}) + O(\frac{1}{n^2})\\ &= 1 + \frac{i}{n}(\sum_{k=1}^m y_k u_k) + O(\frac{1}{n^2}) \qquad = z'_n + O(\frac{1}{n^2})\end{aligned}$$

Ainsi $|z_n^n - z'^n_n| \le n|z_n - z'_n| = O(1/n)$, ce qui donne la convergence voulue. Ceci achève la preuve analytique.

Alternativement, on peut considérer une suite (R_n) de variables aléatoires indépendantes de même loi $\sum_{k=1}^n y_k\delta_{u_k}$. D'après la loi faible des grands nombres, $(R_1+\dots+R_n)/n$ converge en probabilité vers $\mathbb{E}(R_1) = \sum_{k=1}^n y_k u_k$. Elle converge donc en loi vers $\sum_{k=1}^n y_k u_k$, donc la fonction caractéristique de $(R_1+\dots+R_n)/n$ converge vers la fonction caractéristique de la variable constante $\sum_{k=1}^n y_k u_k$. En particulier, il y a convergence au point 1, et il est aisé de de constater que $\varphi_{(R_1+\dots+R_n)/n}(1) = (\sum_{k=1}^m y_k e^{iu_k/n})^n$, ce qui donne le résultat voulu.

On sait donc que $\frac{V_n - d}{Sn} \Longrightarrow Y$. Comme $d/(Sn)$ tend vers 0, le lemme de Slutsky entraîne que $\frac{V_n}{Sn} \Longrightarrow Y$. Comme $\frac{Sn}{Sn+d}$ tend vers 1, le lemme de Slutsky entraîne encore que $\frac{V_n}{Sn+d} \Longrightarrow Y$

(e) On sait $\frac{V_n}{Sn+d} \to W$ presque sûrement, et la convergence presque sûre entraîne la convergence en loi, donc $\frac{V_n}{Sn+d} \Longrightarrow W$. D'après la question précédente, W a la même loi que Y, à savoir la loi de Dirichlet de paramètre d/S.

B.5 Solutions des exercices sur les compléments

Solution 33 Comme les fermés engendrent la tribu borélienne $\mathcal{B}$, il suffit de montrer que l'image réciproque d'un fermé est dans $\mathcal{F}$. Soit F un fermé. À x fixé, $x \mapsto d(x,F)$ est 1-lipschtzienne, donc continue : on a donc $d(X,F) = \lim_{n\to+\infty} d(X_n,F)$. Comme F est fermé, on a donc

$$\{X \in F\} = \{d(X,F) = 0\} = \{d(X_n,F) \to 0\} = \cap_{p\ge 1} \cup_{n\ge 1} \cap_{k\ge n} \{d(X_k,F) \le 1/p\}.$$

Comme $x \mapsto d(x,F)$ est continue, elle est borélienne ; par composition, l'application $\omega \mapsto d(X_n(\omega),F)$ est borélienne, et donc $\{d(X_n,F) \le 1/p\} \in \mathcal{F}$. Comme on effectue ensuite des opérations d'union et d'intersection dénombrable, on reste dans $\mathcal{F}$ et $\{X \in F\} = X^{-1}(F) \in \mathcal{F}$.

Solution 34 Fixons $y_0 \in Y$ et posons $H(x) = Z(x,y_0)$. L'application $x \mapsto (x,y_0)$ est $(\mathbb{R},\mathcal{B}(\mathbb{R}) - (\mathbb{R}^2,\mathcal{B}(\mathbb{R}^2)$ mesurable et l'application $(x,y) \mapsto Z(x,y)$ est $(\mathbb{R}^2,\mathcal{B}(\mathbb{R}^2) - (\Omega,\mathcal{F})$-mesurable, donc par composition, H est $(\mathbb{R},\mathcal{B}(\mathbb{R}) - (\Omega,\mathcal{F})$ mesurable. Reste à voir que $Z = H(X)$. Soit $(x,y)^2 \in \mathbb{R}^2$. On pose $u = Z(x,y)$. Comme $\{u\} \in \mathcal{F}$, $Z^{-1}(\{u\}) \in \sigma(X)$ car Z est $\sigma(X)$-mesurable. Or $\sigma(X) = \{B \times \mathbb{R}; B \in \mathcal{B}(R)\}$, donc il existe B tel que $Z^{-1}(\{u\}) = B \times \mathbb{R}$. Bien sûr $(x,y) \in Z^{-1}(\{u\})$, donc $x \in B$, ce qui entraîne que $(x,y_0) \in= B \times \mathbb{R} = Z^{-1}(\{u\})$, d'où $Z(x,y_0) = u = Z(x,y)$, soit $H(X(x,y)) = Z(x,y)$.

Solution 35

1. Soit $n \in \mathbb{N}$. Pour k entre 0 et n, on a

$$\begin{aligned}
\mathbb{P}(X = k|X+Y = n) &= \frac{\mathbb{P}(X = k, Y = n-k)}{\mathbb{P}(X+Y = n)} \\
&= \frac{\mathbb{P}(X = k)}{\mathbb{P}\left(\cup_{i=0}^n \{X = i, Y = n-i\}\right)} = \frac{\mathbb{P}(X = k)}{\sum_{i=0}^n \mathbb{P}(X = i, Y = n-i)} \\
&= \frac{e^{-\lambda}e^{-\mu}\frac{\lambda^k}{k!}\frac{\mu^{n-k}}{(n-k)!}}{\sum_{i=0}^n e^{-\lambda}e^{-\mu}\frac{\lambda^i}{i!}\frac{\mu^{n-i}}{(n-i)!}} = \frac{\binom{n}{k}\lambda^k\mu^{n-k}}{\sum_{i=0}^n \binom{n}{i}\lambda^i\mu^{n-i}} = \frac{\binom{n}{k}\lambda^k\mu^{n-k}}{(\lambda+\mu)^n} \\
&= \binom{n}{i}\left(\frac{\lambda}{\lambda+\mu}\right)^i \left(\frac{\mu}{\lambda+\mu}\right)^{n-i}
\end{aligned}$$

   Comme $X \le X+Y$, les autres coefficients sont nuls : la loi de X sachant $X+Y = n$ est la loi binomiale $\mathcal{B}(n, \frac{\lambda}{\lambda+\mu})$.

2. L'espérance du carré, c'est le carré de l'espérance plus la variance. Pour une loi binomiale, ces quantités sont bien connues. On a donc

$$\mathbb{E}[X^2|X+Y = n] = \left(\frac{n\lambda}{\lambda+\mu}\right)^2 + n\frac{\lambda\mu}{(\lambda+\mu)^2},$$

$$\text{d'où } \mathbb{E}[X^2|X+Y] = \left(\frac{\lambda(X+Y)}{\lambda+\mu}\right)^2 + (X+Y)\frac{\lambda\mu}{(\lambda+\mu)^2}.$$

B.6 Solutions des exercices corrigés sur les inégalités

Solution 39

1.
$$\begin{aligned} h(\sigma.x)(i) &= |\{j \in \{1,\dots,n\}, x_{\sigma(j)} \le x_{\sigma(i)}\}| \\ &= |\{j' \in \{1,\dots,n\}, x_{j'} \le x_{\sigma(i)}\}| \\ &= [h(x)](\sigma(i)) \end{aligned}$$

2. On a $\{X \notin \delta_n\} \subset \cup_{1\le i<j\le n}\{X_i = x_j\}$. Mais $X_i - X_j$ est la somme de deux variables à densité : elle est à densité, donc sans atome : on a donc $\mathbb{P}(X_i - X_j) = 0$ d'où $\mathbb{P}(X \notin \delta_n) = 0$. Ainsi $\sigma_n = h(X)$ est à valeurs dans $\mathfrak{S}_n$. Comme $\sigma.X$ a même loi que X, σ_n a même loi que $h(\sigma.X) = h(X)\circ\sigma$. Ainsi, pour toute permutation σ,
$$\mathbb{P}(\sigma_n = \sigma) = \mathbb{P}(h(X)\circ\sigma = \sigma) = \mathbb{P}(h(X) = \mathrm{Id}),$$
ce qui montre bien que toutes les permutations ont la même masse.
3. Quitte à changer d'espace de probabilité, on peut supposer que σ_n s'écrit $\sigma_n = h(X_1,\dots,X_n)$, où $X_1,\dots,X_n$ sont des variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. On veut alors appliquer le principe de Maurey à $f = g\circ h$. Soient $(x_1,\dots,x_n)$ et $(x'_1,\dots,x'_n)$ tels que $x_j = x'_j$ pour $j \ne i_0$. Posons $\sigma = h(x)$ et $\sigma' = h(x')$. On a $x_{\sigma^{-1}(1)} < \dots x_{\sigma^{-1}}$ et $x'_{\sigma'^{-1}(1)} < \dots x'_{\sigma'^{-1}}$. Si $x'_{i_0} \in]x_{\sigma^{-1}(\sigma(i_0)-1)}, x_{\sigma^{-1}(\sigma(i_0)+1)}[$, on a simplement $\sigma = \sigma'$. Sinon, si $x'_{i_0} < x_{\sigma^{-1}(\sigma(i_0)-1)}$, σ'^{-1} est de la forme
$$\begin{pmatrix} 1 & \dots & i & i+1 & i+2 & \dots & i+1 & i+k+1 & \dots & n \\ \sigma^{-1}(1) & \dots & \sigma^{-1}(i) & \sigma^{-1}(i+k) & \sigma^{-1}(i+1) & \dots & \sigma^{-1}(i+k-1) & \sigma^{-1}(i+k+1) & \dots & \sigma^{-1}(n) \end{pmatrix},$$
soit
$$\sigma'^{-1} = \sigma^{-1}\circ\begin{pmatrix} 1 & \dots & i & i+1 & i+2 & \dots & i+1 & i+k+1 & \dots & n \\ 1 & \dots & i & i+k & i+1 & \dots & i+k-1 & i+k+1 & \dots & n \end{pmatrix},$$
c'est à dire $\sigma' = \gamma_{i+1,i+k}\circ\sigma$. On en déduit que $|f(x) - f(x')| = |g(\sigma) - g(\sigma')| \le M$. Le cas où $x'_{i_0} > x_{\sigma^{-1}(\sigma(i_0)+1)}$ se traite de manière analogue. Il suffit alors d'appliquer l'hypothèse sur g et le principe de Maurey pour avoir le résultat voulu.

4. On numérote les boules de 1 à $n = b+r$, les boules rouges étant numérotées de 1 à r. Ordonner de manière aléatoire uniforme les n boules et ne garder que les k premières est la même chose qu'en prendre k sans remise. On pose donc $g(\sigma) = |\sigma(1,\dots,k) \cap \{1,\dots,r\}$. Ainsi $g(\sigma)$ a la loi de X. On a $g(\sigma) = \sum_{i=1}^k \mathbb{1}_{\{\sigma(k)\in\{1,\dots,r\}}$, d'où
$$\mathbb{E}[X] = \mathbb{E}[g(\sigma)] = \sum_{i=1}^k \mathbb{P}(\sigma(k) \in \{1,\dots,r\}) = \sum_{i=1}^k \frac{r}{b+r} = \frac{kr}{b+r}.$$
g vérifie l'hypothèse de la question précédente avec $M = 1$, d'où le résultat.

Solution 40 1. On a immédiatement $X_k^n = \sum_{k=1}^n \mathbb{1}_{\{N_{k,p}=1\}}$. Comme $N_{k,p}$ est une somme de k variables de Bernoulli indépendantes de paramètre $1/n$, $N_{k,p} \sim \mathcal{B}(k, 1/n)$ et
$$\mathbb{E}[\mathbb{1}_{\{N_{k,p}=1\}}] = \mathbb{P}(N_{k,p} = 1) = \binom{k}{1}\frac{1}{n}(1-\frac{1}{n})^{k-1} = \frac{k}{n}(1-\frac{1}{n})^{k-1},$$
et par linéarité, on a $\mathbb{E}[X_k^n] = n\frac{k}{n}(1-\frac{1}{n})^{k-1}$.

2. On a
$$\begin{aligned}(1-\frac{1}{n})^n = n^{-1}(1-\frac{1}{n})\mathbb{E}[X_n^n] \le n^{-1}(1-\frac{1}{n})v_n \\ = \max_{0\le k\le \lfloor n^{3/2}\rfloor} \frac{k}{n}\left(1-\frac{1}{n}\right)^k \\ \le \max_{0\le k\le \lfloor n^{3/2}\rfloor} \frac{k}{n}\exp(-1/n)^k \\ \le \sup_{x\in\mathbb{R}_+} x\exp(-x)\end{aligned}$$
Comme $\log(1-1/n)^n = n\log(1-1/n) \sim n(-1/n) = 1$, on a classiquement $\lim_{n\to+\infty}(1-1/n)^n = e^{-1}$. D'autre part, une petite étude de fonction montre que $\sup_{x\in\mathbb{R}_+} x\exp(-x) = \exp(-1)$, donc par comparaison $\lim_{n\to+\infty} n^{-1}(1-\frac{1}{n})v_n = 1/e$, d'où $\lim_{n\to+\infty} n^{-1}v_n = 1/e$, soit $v_n \sim \frac{n}{e}$.

3. Soit N tel que $(1-1/n)^{n-1} \ge \frac{1}{e} - \varepsilon/2$ et $\frac{v_n}{n} \le \frac{1}{e} + \varepsilon/2$ pour $n \ge N$. On a alors pour $n \ge N$:
$$\begin{aligned}\mathbb{P}\left(X_n^* \le n(\frac{1}{e}-\varepsilon)\right) &\le \mathbb{P}\left(X_n^n \le n(\frac{1}{e}-\varepsilon)\right) \\ &\le \mathbb{P}\left(X_n^n \le n((1-\frac{1}{n})^{n-1}-\varepsilon/2)\right) \\ &= \mathbb{P}\left(X_n^n - \mathbb{E}[X_n^n] \le \varepsilon/2\right)\end{aligned}$$

D'autre part

$$\begin{aligned}\mathbb{P}\left(X_n^* \geq n(\frac{1}{e}+\varepsilon)\right) &= \mathbb{P}\left(\cup_{k=1}^n \{X_n^k \geq n(\frac{1}{e}+\varepsilon)\}\right)\\ &\leq \sum_{k=1}^{\lfloor n^{3/2}\rfloor} \mathbb{P}\left(X_n^k \geq n(\frac{1}{e}+\varepsilon)\right)\\ &\leq \sum_{k=1}^{\lfloor n^{3/2}\rfloor} \mathbb{P}\left(X_n^k \geq (v_n + n\varepsilon/2)\right)\\ &\leq \sum_{k=1}^{\lfloor n^{3/2}\rfloor} \mathbb{P}\left(X_n^k \geq \mathbb{E}[X_n^k] + n\varepsilon/2)\right)\end{aligned}$$

En additionnant les deux majorations, on obtient l'inégalité voulue.

4. Posons, pour $y \in \{1, \dots, n\}^k$:

$$\varphi_k(y) = \sum_{u=1}^n \mathbb{1}_{\{1=\sum_{j=1}^k \mathbb{1}_{\{y_j=u\}}\}}.$$

On a ainsi $X_k^n = \varphi((Y_i)_{1\leq i \leq k})$. Si y et y' diffèrent au temps $t \leq k$ et seulement à ce moment là, mettons que la boule au temps t est mise en i dans la configuration y, alors qu'elle est mise au temps t est mise en i' dans la configuration y'. Au temps k, seuls les effectifs des effectifs des boîtes $u = i$ et $u = i'$ diffèreront, donc la différence entre $\varphi_k(y)$ et $\varphi_k(y')$ ne peut excéder deux unités. On peut donc appliquer le principe de Maurey avec $k_i = 2$. Avec le principe de Maurey, on a

$$\mathbb{P}(|X_n^k - \mathbb{E}[X_n^k]| \geq \varepsilon/2n) \leq 2\exp(-\frac{(\varepsilon/2n)^2}{8k}),$$

de sorte que pour $k \leq n^{3/2}$, on a

$$\mathbb{P}(|X_n^k - \mathbb{E}[X_n^k]| \geq \varepsilon/2n) \leq 2\exp(-\frac{\varepsilon^2}{32}\sqrt{n}).$$

Ainsi, finalement, on a

$$\mathbb{P}(|X_/^n n * - e^{-1}| > \varepsilon) \leq 2n^{3/2}\exp(-\frac{\varepsilon^2}{32}\sqrt{n}),$$

ce qui entraîne que X_*^n/n tend en probabilité vers e^{-1}.

5. On a

$$\begin{aligned}\{N_{\lfloor n^{3/2}\rfloor,u} \geq 2\} &\subset \{\forall k \geq \lfloor n^{3/2}\rfloor; N_{k,u} \geq 2\}\\ &\subset \{\forall k \geq \lfloor n^{3/2}\rfloor; \mathbb{1}_{\{N_{k,u}=1\}} = 0\},\end{aligned}$$

donc

$$\begin{aligned}\cap_{u=1}^{n} N_{\lfloor n^{3/2}\rfloor} \geq 2\} &\subset \{\forall k \geq \lfloor n^{3/2}\rfloor; X_k^n = 0\}\\ &\leq \cup\{X^n = X_*^n\}\end{aligned}$$

En passant au complémentaire, on a l'inclusion

$$\{X^n \neq X_*^n\} \subset \cup_{p=1}^{n}\{N_{\lfloor n^{3/2}\rfloor,p} < 2\},$$

6. On a alors

$$\begin{aligned}\mathbb{P}(X^n \neq X_*^n) &\leq n\mathbb{P}(N_{\lfloor n^{3/2}\rfloor,1} < 2)\\ &= n\mathcal{B}(\lfloor n^{3/2}\rfloor, 1/n)(\{0;1\})\\ &= n((1-1/n)^{\lfloor n^{3/2}\rfloor} + \lfloor n^{3/2}\rfloor(1/n)(1-1/n)^{\lfloor n^{3/2}\rfloor-1})\\ &\leq n(\lfloor n^{3/2}\rfloor + 1)(1-1/n)^{\lfloor n^{3/2}\rfloor}\\ &\leq Kn^{5/2}(1-1/n)^{n^{3/2}} \leq Kn^{5/2}\exp(-\sqrt{n})\end{aligned}$$

Comme pour tout $\varepsilon > 0$,

$$\mathbb{P}(|X^n/n - 1/e| > \varepsilon) \leq \mathbb{P}(|X_*^n/n - 1/e| > \varepsilon) + \mathbb{P}(X^n \neq X_*^n),$$

et que les deux termes de la somme de droite ont une limite nulle quand n tend vers l'infini, $|X^n/n$ tend en probabilité vers $1/e$.

7. On a $0 \leq X^n/n \leq 1$. $\mathbb{P}_{X^n/n}$ tend en loi vers $\delta_{1/e}$ et est équi-intégrable car bornée dans L_∞, donc la suite des moments converge vers $1/e$
8. Pour se ramener au problème précédent, il suffit de considérer que chaque numéro tiré est le numéro d'une pilule. Si on tire un numéro d'une pilule pas complètement mangée, on en mange la moitié, sinon l'essai est en échec et on retire un numéro jusqu'à atteindre un numéro correspondant à une pilule pas complètement mangée ou qu'il n'y ait plus de pilule. On a donc le résultat voulu avec $K = 1/e$.

B.7 Solutions des exercices sur les statistiques exhaustives

Solution 44 1. D'après le théorème de factorisation de Neyman, $\mathbb{P}_\theta$ a une densité f_θ par rapport à m. Pour φ mesurable bornée, on a alors $\mathbb{E}_\theta(\varphi(Z)) = \int \varphi(Z)f_\theta(T)\, dm$. Comme Z et S sont indépendantes sous m

$$\mathbb{E}_\theta(\varphi(Z)) = \mathbb{E}_m[\varphi(Z)f_\theta(S)] = \mathbb{E}_m[\varphi(Z)]\mathbb{E}_m[f_\theta(S)]$$

En prenant $\varphi = 1$, on trouve $\mathbb{E}_m[f_\theta(S)] = 1$, d'où

$$\mathbb{E}_\theta(\varphi(Z)) = \mathbb{E}_m[\varphi(Z)]$$

La loi de Z sous $\mathbb{P}_\theta$ est donc sa loi sous m, qui ne dépend pas de θ : Z est libre.

2. Soit φ mesurable bornée. Comme $\mathbb{Z}$ est libre, il existe une constante $c(\varphi)$ telle que $\mathbb{E}_\theta(\varphi(Z)) = c(\varphi)$ pour tout $\theta \in \Theta$. Comme S est exhaustive, $\mathbb{E}_\theta(\varphi(Z)|S)$ est une fonction $\psi(S)$ avec ψ indépendante de θ. L'application $\psi(S) - c(\varphi)$ est $\sigma(S)$-mesurable et d'espérance nulle sous tout $\mathbb{P}_\theta$. Comme S est complète, $\psi(S) - c(\varphi)$ est nulle. On a donc $\mathbb{E}_\theta(\varphi(Z)|S) = c(\varphi)$, ce qui signifie que Z et S sont indépendantes sous $\mathbb{P}_\theta$. En effet pour tout g,

$$\begin{aligned}
\mathbb{E}_\theta[g(S)\varphi(Z)] &= \mathbb{E}_\theta[\mathbb{E}_\theta[g(S)\varphi(Z)|S]] \\
&= \mathbb{E}_\theta[g(S)\mathbb{E}_\theta[\varphi(Z)|S]] \\
&= \mathbb{E}_\theta[g(S)c(\varphi)] \\
&= c(\varphi)\mathbb{E}_\theta[g(S)] = \mathbb{E}_\theta(\varphi(Z))\mathbb{E}_\theta[g(S)]
\end{aligned}$$

B.8 Solution des exercices sur l'information de Fisher

Solution 50 Les lois considérées ont toutes le même support $[1, +\infty[$. Un calcul simple donne $W_\alpha = \frac{1}{\alpha} - \log x$. On a donc

$$\begin{aligned}
\mathcal{I} &= \int_{[1,+\infty[} \frac{\alpha}{x^{\alpha+1}} \left(\frac{1}{\alpha} - \log x\right)^2 \, d\lambda(x) \\
&= \int_{[0,+\infty[} \alpha e^{-\alpha u} (\frac{1}{\alpha} - u)^2 \, du
\end{aligned}$$

On reconnait la variance de la loi exponentielle de paramètre α : c'est $\frac{1}{\alpha^2}$.

$\sqrt{\mathcal{I}(\alpha)} = \frac{1}{\alpha}$ est C^1 sur $]0, +\infty[$, donc le modèle vérifie les bonnes conditions : on a donc $\mathbb{E}_\alpha[W_\alpha] = 0$. Comme $W_\alpha = \frac{1}{\alpha} - \log X$, on constate bien que T est un estimateur sans biais de $1/\alpha$.

$\mathrm{Var}_\alpha T = \mathbb{E}_\alpha(E - \mathbb{E}_\alpha[T])^2 = \mathbb{E}_\alpha(W_\alpha^2) = \mathcal{I}(\alpha)$. On note que $\mathrm{Var}_\alpha T = \alpha^{-2} = \frac{g'(\alpha)^2}{\mathcal{I}(\alpha)}$: on a égalité dans l'inégalité de Cramer-Rao. La densité $f_\alpha(x) = \frac{1}{x^{\alpha+1}}$ s'écrit aussi $\frac{\alpha}{x} \exp(-\alpha \log X)$: on a un modèle exponentiel, et on retrouve le résultat du cours.

Si f est un estimateur sans biais de α, on a pour tout α

$$\int_{[1,+\infty[} \frac{1}{x^{\alpha+1}} f(x)\ d\lambda(x) = \alpha,$$

soit $\int_{[0,+\infty[} \alpha e^{-\alpha u} f(e^u) d\lambda(u) = \alpha$, ou encore

$$\int_{[0,+\infty[} \alpha e^{-\alpha u} f(e^u) d\lambda(u) = 1.$$

Posons $g(u) = f(e^u)$. Fixons $\alpha_0 > 0$. Pour tout $\alpha \in]\alpha_0, +\infty[$, on a $\frac{\partial}{\partial\alpha} e^{-\alpha u} g(u) = -ue^{-\alpha u} g(u)$, d'où

$$|\frac{\partial}{\partial\alpha} e^{-\alpha u}| \le ue^{-\alpha_0 u} g(u) \le \frac{2}{\alpha_0} e^{-\alpha_0/2u} g(u)$$

Comme cette fonction est intégrable, on obtient sur $]\alpha_0, +\infty[$:

$$0\frac{\partial}{\partial\alpha} 1 = \int_{[0,+\infty[} \alpha u e^{-\alpha u} g(u) d\lambda(u).$$

Comme $]\alpha_0$ peut être pris aussi petit que l'on veut, l'identité est encore vraie sur $]0, +\infty[$. Posons $A = \int_] 0, +\infty[x f_+(x) e^{-x}\ dx$. On a $A = \int_] 0, +\infty[x f_-(x) e^{-x}\ dx$, donc $\mathbb{P}_1 = \frac{1}{A} x f_+ e^{-x} \lambda$ et $\mathbb{P}_2 = \frac{1}{A} x f_- e^{-x}$ sont deux mesures de probabilités sur $\mathbb{R}_+$. Elles ont la même transformée de Laplace, donc $f_+ = f_-$ et $f = 0$. Contradiction.

Solution 51 1. C'est un calcul classique :

$$\begin{aligned}
\mathbb{E}X_1(X_1 - 1) &= \sum_{k=0}^{+\infty} k(k-1)\mathbb{P}(X = k) \\
&= \sum_{k=2}^{+\infty} k(k-1)\mathbb{P}(X = k) = \sum_{k=2}^{+\infty} k(k-1)e^{-\lambda}\frac{\lambda^k}{k!} \\
&= e^{-\lambda}\lambda^2 \sum_{k=2}^{+\infty} \frac{\lambda^{k-2}}{(k-2)!} = e^{-\lambda}\lambda^2 e^{\lambda} = \lambda^2
\end{aligned}$$

Comme les X_i ont même loi, le résultat s'ensuit par linéarité.

2. La densité de $\mathcal{P}(\lambda)$ par rapport à la mesure de comptage est $n \mapsto e^{-\lambda}\frac{\lambda^n}{n!} = e^{-\lambda}\frac{1}{n!}\exp((\log\lambda)n)$: la famille $(\mathcal{P}(\lambda))_{\lambda>0}$ est une famille exponentielle et X est la statistique naturelle associée. Ainsi $X_1 + \dots + X_d$ est la statistique naturelle associée au modèle exponentiel $(\mathcal{P}(\lambda)^{\otimes d})_{\lambda>0}$. Comme $\log(]0, +\infty[) = \mathbb{R}$ est d'intérieur non-vide, la statistique est complète.

3. La loi de S_d sous $\mathbb{P}_\lambda$ est la loi de X_1 sous $\mathbb{P}_{d\lambda}$: on a donc $\mathbb{E}_\lambda S_d(S_d-1) = \mathbb{E}_{d\lambda} X_1(X_1-1) = (d\lambda)^2$, d'où on déduit que $\frac{S_d(S_d-1)}{d^2}$ est un estimateur sans biais de λ^2. Mais $\frac{S_d(S_d-1)}{d^2}$ est $\sigma(S_d)$ mesurable et S_d est une statistique exhaustive complète, donc d'après le théorème de Lehman-Scheffé, $\frac{S_d(S_d-1)}{d^2}$ est le meilleur estimateur quadratique de λ^2.
4. E_d étant un estimateur sans biais de λ^2 et S_d une statistique exhaustive complète, l'amélioration de E_d : $\mathbb{E}_\lambda[E_d|S_d]$ coïncide avec le meilleur estimateur quadratique de λ^2. On a $\mathbb{E}_\lambda[E_d|S_d] = \frac{S_d(S_d-1)}{d^2}$, d'où on déduit aisément la formule voulue.
5. L'information de Fisher du modèle a été calculée précédemment : c'est $\frac{d}{\lambda}$. Avec la fonction $g(\lambda) = \lambda^2$, la borne de Cramer-Rao est

$$\frac{(g'(\lambda))^2}{\mathcal{I}(\lambda)} = \frac{(2\lambda)^2}{d/\lambda} = \frac{4\lambda^3}{d}.$$

D'autre part

$$\begin{aligned}\operatorname{Var}_\lambda \frac{S_d(S_d-1)}{d^2} &= \frac{1}{d^4}\operatorname{Var}_\lambda S_d(S_d-1)\\ &= \frac{1}{d^4}\operatorname{Var}_{n\lambda} X_1(X_1-1).\end{aligned}$$

On a déjà vu que $\mathbb{E}_\lambda X_1(X_1-1) = \lambda^2$. On a aussi les identités $\mathbb{E}_\lambda X_1(X_1-1)(X_1-2) = \lambda^3$ et $\mathbb{E}_\lambda X_1(X_1-1)(X_1-2)(X_1-3) = \lambda^4$. On décompose alors $X^2(X-1)^2$ sur cette base :

$$\begin{aligned}X^2(X-1)^2 &= X(X-2+2)(X-1)(X-3+2)\\ &= X(X-1)(X-2)(X-3) + 2X(X-1)(X-3)\\ &\quad + 2X(X-1)(X-2) + 4X(X-1)\\ &= X(X-1)(X-2)(X-3)\\ &\quad + 2X(X-1)(X-2) - 2X(X-1)\\ &\quad + 2X(X-1)(X-2) + 4X(X-1)\\ &= X(X-1)(X-2)(X-3)\\ &\quad + 4X(X-1)(X-2) + 2X(X-1)\end{aligned}$$

D'où $\mathbb{E}_\lambda X^2(X-1)^2 = \lambda^4 + 4\lambda^3 + 2\lambda^2$ et $\operatorname{Var}_\lambda X^2(X-1)^2 = 4\lambda^3 + 2\lambda^2$. On a finalement

$$\operatorname{Var}_\lambda \frac{S_d(S_d-1)}{d^2} = \frac{1}{d^4}\operatorname{Var}_{n\lambda} X_1(X_1-1) = \frac{4(d\lambda)^3 + 2(d\lambda)^2}{d^4} = 4\frac{\lambda^3}{d} + 2\frac{\lambda^2}{d^4},$$

qui dépasse la borne de Cramer-Rao : l'estimateur n'est pas efficace.

B.9 Solutions des exercices sur les processus

Solution 54 1. La loi du vecteur $(\Pi_0, \dots, \Pi_n)$ étant discrète, tout ensemble $\sigma(\Pi_0, \dots, \Pi_n)$-mesurable peut s'écrire (à un négligeable près) comme réunion dénombrable disjointe d'événements de la forme

$$A = \{\Pi_0 = x_0, \dots, \Pi_n = x_n\}.$$

Ainsi, avec le principe de partition, il suffit de démontrer l'égalité lorsque A s'écrit $A = \{\Pi_0 = x_0, \dots, \Pi_n = x_n\}$, avec $(x_0 \dots, x_n) \in D^{n+1}$. Si $i \neq x_n$, les deux membres de l'égalité voulue sont nuls : il n'y a rien à démontrer. On doit donc montrer que pour tout $B \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$, on a

$$\mathbb{P}^{\mu}(\Pi_0 = x_0, \dots, \Pi_n = x_n, \theta^{-n}(B)) = \mathbb{P}^{\mu}(\Pi_0 = x_0, \dots, \Pi_n = x_n)\mathbb{P}^{x_n}(B).$$

Bien sûr, si $\mathbb{P}^{\mu}(\Pi_0 = x_0, \dots, \Pi_n = x_n) = 0$, il n'y a encore rien à démontrer puisque les deux membres sont nuls. Soit $\mathcal{C}$ l'ensemble des événements qui s'écrivent

$$B = \{\Pi_0 = y_0, \dots, \Pi_k = y_k\}$$

pour un certain k : il est aisé de constater que les mesures de probabilité $B \mapsto \mathbb{P}^{\mu}(\theta^{-n}(B)|\Pi_0 = x_0, \dots, \Pi_n = x_n)$ et $B \mapsto \mathbb{P}^{x_n}(B)$ coïncident sur $\mathcal{C}$. Comme $\mathcal{C}$ est un π-système qui engendre la tribu $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$, ces deux probabilités sont égales, ce qui donne le résultat voulu.

2. Ici encore, il est aisé de constater que $\mathbb{P}^{\mu}$ et $\sum_{i \in D} \mu(i)\mathbb{P}^i$ sont deux mesures de probabilité qui coïncident sur $\mathcal{C}$.

3. On a

$$\begin{aligned}
\mathbb{P}^{\mu}(\theta^{-1}(B)) &= \sum_{i \in D} \mathbb{P}^{\mu}(\Pi_1 = i, \theta^{-1}(B)) \\
&= \sum_{i \in D} \mathbb{P}^{\mu}(\Pi_1 = i)\mathbb{P}^i(B)) \\
&= \sum_{i \in D} \mu(i)\mathbb{P}^i(B)) \\
&= \mathbb{P}^{\mu}(B))
\end{aligned}$$

Solution 55 Considérons l' application

$$\begin{array}{rcl}
\psi : \mathbb{R}^{\mathbb{N}} & \to & \mathbb{R}^{\mathbb{N}} \\
((x_n)_{n \geq 0}) & \mapsto & (\varphi(\theta^n \circ x))
\end{array}$$

Comme θ est mesurable de $(\mathbb{R}^{\mathbb{N}},\mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ dans lui même et φ $(\mathbb{R}^{\mathbb{N}},\mathcal{B}(\mathbb{R}^{\mathbb{N}}))-(\mathbb{R},\mathcal{B}(R))$ mesurable, pour tout n $\varphi(\theta^n\circ X)$ est bien $(\mathbb{R}^{\mathbb{N}},\mathcal{B}(\mathbb{R}^{\mathbb{N}}))-(\mathbb{R},\mathcal{B}(R))$ mesurable, et donc ψ est bien mesurable de $(\mathbb{R}^{\mathbb{N}},\mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ dans lui même. Ainsi $Y=\psi(X)$. On peut noter que $\psi\circ\theta=\theta\circ\psi$. Ainsi

$$\begin{aligned}
\mathbb{P}_Y(\theta^{-1}(A)) &= \mathbb{P}(Y^{-1}(\theta^{-1}(A)))\\
&= \mathbb{P}((\theta\circ Y)^{-1}(A))\\
&= \mathbb{P}((\theta\circ\psi\circ X)^{-1}(A))\\
&= \mathbb{P}((\psi\circ\theta\circ X)^{-1}(A))\\
&= \mathbb{P}(X^{-1}(\theta^{-1}(\psi^{-1}(A))))\\
&= \mathbb{P}_X(\theta^{-1}(\psi^{-1}(A))\\
&= \mathbb{P}_X(\psi^{-1}(A)) = \mathbb{P}(X^{-1}(\psi^{-1}(A))\\
&= \mathbb{P}((\psi\circ X)^{-1}(A))\\
&= \mathbb{P}(Y^{-1}(A)) = \mathbb{P}_Y(A)
\end{aligned}$$

Solution 56

1. Il suffit de montrer que pour tout n, $(X_1,\dots,X_n)$ et $(X_2,\dots,X_{n+1})$ ont même loi. Or si on considère la permutation cyclique de $\mathcal{S}_{n+1}$: $\sigma=(1\quad 2\dots n+1)$, l'hypothèse d'échangeabilité entraîne que $(X_1,\dots,X_n,X_{n+1})$ a même loi que $(X_{\sigma(1)},\dots,X_{\sigma(n)},X_{\sigma(n+1)})=(X_2,\dots,X_{n+1},X_1)$. Par projection sur les n premières composantes, $(X_1,\dots,X_n)$ et $(X_2,\dots,X_{n+1})$ ont même loi.
   Remarque : alternativement, on peut remarquer que échangeable $\Longrightarrow$ réversible $\Longrightarrow$ stationnaire.
2. Par biléarité et symétrie

   $$VarS_n=\sum_{k=1}^n \operatorname{Var} X_k+2\sum_{1\le i<j\le n}\operatorname{Covar}(X_i,X_j).$$

   Par stationnarité, $\operatorname{Var} X_k=\operatorname{Var} X_1$ pour tout k. Pour $1\le i<j\le n$, considérons la permutation de $\mathcal{S}_n$: $\sigma=(1\quad i)\circ(2\quad j)$. Par projection, les vecteurs (X_1,X_2) et $(X_{\sigma(1)},X_{\sigma(2)})=(X_i,X_j)$ ont même loi, donc $\operatorname{Covar}(X_1,X_2)=\operatorname{Covar}(X_i,X_j)$, d'où

   $$VarS_n=n\operatorname{Var} X_1+n(n-1)\operatorname{Covar}(X_1,X_2)$$

   En divisant par n^2, on a

   $$0\le(\operatorname{Var} S_n)/n^2=(\operatorname{Var} X_1)+(n-1)/n\operatorname{Covar}(X_1,X_2).$$

   En faisant tendre n vers l'infini, on obtient $0\le\operatorname{Covar}(X_1,X_2)=\operatorname{Covar}(X_i,X_j)$: les X_i sont positivement corrélés.

3. Soit $(X_n)_{n\ge 1}$ un processus gaussien échangeable. D'après ce qui précède, il existe m tel que $\mathbb{E}(X_i) = m$ pour tout i (stationnarité), $\sigma^2 \ge 0$ tel que $\operatorname{Var} X_i = \sigma^2$ pour tout i, et c tel que $\operatorname{Covar}(X_i, X_j) = c$ pour $i \ne j$. Évidemment $c = \operatorname{Covar}(X_1, X_2) \le \sqrt{\operatorname{Var} X_1 \operatorname{Var} X_2} = \sigma^2$. Réciproquement, supposons que nous sont donnés des réels $m, \sigma^2 \ge 0$ et c avec $c \le \sigma^2$. Soient a, b réels, $(Y_i)_{i\ge 0}$ un bruit blanc. Si l'on pose $X_i = m + aY_0 + bY_i$, $(X_i)_{i\ge 1}$ est un processus gaussien. On a $\mathbb{E}(X_i) = m$ et $\operatorname{Var} X_i = a^2 \operatorname{Var} Y_0 + b^2 \operatorname{Var} Y_i = a^2 + b^2$, tandis que pour $i \ne j$

$$\begin{aligned}\operatorname{Covar}(X_i, X_j) &= \operatorname{Covar}(aY_0 + bY_i, aY_0 + bY_j)\\ &= a^2 \operatorname{Var} Y_0 + ab(\operatorname{Covar}(X_0, X_j) + \operatorname{Covar}(X_0, X_j)) + b^2 \operatorname{Covar}(Y_i, Y_j)\\ &= a^2 \operatorname{Var} Y_0 = a^2\end{aligned}$$

Ainsi, si on prend $a = \sqrt{c}$ et $b = \sqrt{\sigma^2 - c}$, le processus $(X_i)_{i\ge 1}$ aura bien la fonction d'espérance et la famille des covariances espérées. Comme la fonction d'espérance et la famille des covariances caractérisent les processus gaussiens, on a construit le processus voulu.

4. Traitons d'abord le cas où le processus est construit, comme précédemment, à partir d'un bruit blanc : $X_i = m + aY_0 + bY_i$. Dans ce cas, la loi forte des grands nombres entraîne que $\frac{1}{n}(X_1 + \dots X_n)$ converge presque sûrement vers $Z = aY_0$. Si $a = 0$, il n'y a rien à démontrer car les X_i sont alors indépendants. Soit $\varphi_1, \dots, \varphi_n$ des fonctions bornées. Comme Z est indépendant de $Y_1, \dots, Y_n$

$$\begin{aligned}\mathbb{E}(\varphi_1(X_1)\varphi_2(X_2)\dots\varphi_n(X_n)|Z) &= \mathbb{E}(\varphi_1(Z + bY_1)\varphi_2(Z + bY_2)\dots\varphi_n(Z + bY_n)|Z)\\ &= \psi(Z),\end{aligned}$$

avec

$$\begin{aligned}\psi(z) &= \mathbb{E}(\varphi_1(z + bY_1)\varphi_2(z + bY_2)\dots\varphi_n(z + bY_n))\\ &= \prod_{i=1}^n \mathbb{E}\varphi_i(z + bY_1)\end{aligned}$$

De manière similaire, on montre que $\mathbb{E}(\varphi_i(X_i)|Z) = \psi_i(Z)$, avec $\psi_i(z) = \mathbb{E}(\varphi_i(z + bY_1))$. Ainsi

$$\mathbb{E}(\varphi_1(X_1)\varphi_2(X_2)\dots\varphi_n(X_n)|Z) = \prod_{i=1}^n \mathbb{E}(\varphi_i(X_i)|Z),$$

ce qui donne l'indépendance voulue.

Passons au cas général. On peut toujours définir $Z = \overline{\lim}\frac{1}{n}(X_1 + \cdots + X_n)$. On veut montrer que sachant Z les X_i sont indépendants, autrement dit, pour tout n, la loi conditionnelle de $(X_i, \dots, X_n)$ sachant Z est une loi produit. Or, la loi conditionnelle de $(X_i, \dots, X_n)$ sachant Z est complètement déterminée par la loi de $(X_i, \dots, X_n, Z)$, qui elle-même est complètement déterminée par la loi du processus $(X_n)_{n\ge 1}$, puisque c'est la loi image de $\mathbb{P}_X$ par $\omega \mapsto (\omega_1, \dots, \omega_n, \overline{\lim}(\omega_1 + \dots \omega_k)/k)$. Comme on peut toujours construire un processus X'_n sous la forme $X'_i = m + aY_0 + bY_i$ avec $\mathbb{P}_X = \mathbb{P}_{X'}$, le résultat s'ensuit.

Solution 57 1. Pour tout $y = (y_1, \dots, y_n) \in B(x)$, il existe i tel que $y_{n-1} = \hat{x}_i$. Cet entier i est unique, car s'il existait deux tels entiers i et j, on aurait $x_i = x_{i+1}$, ce qui est impossible car x est propre. Notons le $i(y)$. On peut alors poser $B_i(x) = \{y \in B(x); i(y) = i\}$. On a évidemment

$$|B(x)| = \sum_{i=1}^{n} |B_i(x)|.$$

Or l'application $y \mapsto (y_1, \dots, y_{n-1})$ réalise une bijection de $B_i(x)$ dans $B(\hat{x}_i)$, d'où la formule voulue.

2. On va procéder par récurrence sur n. Si $n = 0$, chaque terme de la somme de gauche est égale 1 et le second membre vaut $b_0(q)B(\varnothing) = q$: l'identité est vérifiée. Supposons la propriété acquise au rang n et montrons là au rang $n+1$ Si x n'est pas un mot propre, l'identité est évidente car les deux termes sont nuls. Sinon, on peut remarquer que

$$\begin{aligned}
\textstyle\sum_{a\in\{1,\dots,q\}}|B(x.a)| &= \textstyle\sum_{a\in\{1,\dots,q\}\backslash\{x_n\}}|B(x.a)| \\
&= \textstyle\sum_{a\in\{1,\dots,q\}\backslash\{x_n\}} \left(\left(\displaystyle\sum_{i=1}^{n} |B(\hat{x}_i.a)| \right) + B(x) \right) \\
&= -\sum_{i=1}^{n} |B(\hat{x}_i.x_n)| + \textstyle\sum_{a\in\{1,\dots,q\}} \left(\displaystyle\sum_{i=1}^{n} |B(\hat{x}_i.a)| \right) + (q-1)B(x) \\
&= -|B(x)| + \sum_{i=1}^{n} b_{n-1}(q)|B(\hat{x}_i)| + (q-1)|B(x)| \\
&= (b_{n-1}(q) + q - 2)|B(x)| = b_n(q)|B(x)|,
\end{aligned}$$

ce qui montre l'hérédité.

3. Tous les mots à $n+1$ lettres s'écrivent $x.a$, où x est un mot de n lettre et $a \in \{1, \dots, q\}$. Ainsi, les $B(x.a)$ forment une partition de

l'ensemble des immeubles propres de hauteur $n+1$: on a

$$\begin{aligned} S(n+1,q) &= \sum\nolimits_{x\in\{1,\dots,q\}^n}\sum\nolimits_{a\in\{1,\dots,q\}}|B(x.a)| \\ &= \sum\nolimits_{x\in\{1,\dots,q\}^n} b_n(q)|B(x)| \\ &= b_n(q)S(n,q) \end{aligned}$$

4. π_n est une mesure positive. On a

$$\begin{aligned} \pi_n(\{1,\dots,q\}^n) &= \sum\nolimits_{x\in\{1,\dots,q\}^n}\pi_n(x) \\ &= \sum\nolimits_{x\in\{1,\dots,q\}^n} S(n,q)^{-1}|B(x)| = S(n,q)^{-1}S(n,q) = 1, \end{aligned}$$

ce qui donne le résultat

5. On a

$$\begin{aligned} \pi_{n+1}(\{(x_1,\dots,x_n)\}\times\{1,\dots,q\}) &= \sum\nolimits_{a\in\{1,\dots,q\}}\pi_{n+1}(\{(x_1,\dots,x_n,a\}) \\ &= \sum\nolimits_{a\in\{1,\dots,q\}}\frac{|B(x.a)|}{S(n+1,q)} \\ &= \frac{1}{b_n(q)S(n,q)}\sum\nolimits_{a\in\{1,\dots,q\}}|B(x.a)| \\ &= \frac{1}{b_n(q)S(n,q)}b_n(q)|B(x)| \\ &= \pi_n(\{(x_1,\dots,x_n)\}) \end{aligned}$$

Il suffit alors d'appliquer le théorème d'extension de Kolmogorov.

6. Si on note $x=(x_1,\dots,x_n)$ et $\tilde{x}=(x_n,\dots,x_1)$, l'application $y=(y_1,\dots,y_n)\mapsto(\tilde{y_1},\dots,\tilde{y_n})$ réalise une bijection entre $B(x)$ et $B(\tilde{x})$, ce qui entraîne immédiatement le résultat voulu.

7. Soient $n\ge 1$ et $x\in\{1,\dots,q\}^n$. En appliquant deux fois la réversibilité puis le principe de partition, on a

$$\begin{aligned} \mathbb{P}_q(\Pi_1=x_1,\dots,\Pi_n=x_n) &= \mathbb{P}_q(\Pi_1=x_n,\dots,\Pi_n=x_1) \\ &= \sum_{a=1}^{q}\mathbb{P}_q(\Pi_1=x_n,\dots,\Pi_n=x_1,\Pi_{n+1}=a) \\ &= \sum_{a=1}^{q}\mathbb{P}_q(\Pi_1=a,\pi_2=x_1,\dots,\Pi_n=x_1,\Pi_{n+1}=x_n) \\ &= \mathbb{P}_q(\pi_2=x_1,\dots,\Pi_n=x_1,\Pi_{n+1}=x_n), \end{aligned}$$

ce qui donne l'invariance par le décalage.

8. (a) On va construire $c_{n,p}$ par récurrence sur $n+p$. Pour $n+p=0$, il est facile de voir que la formule est vérifiée avec $c_{0,0}=q=4$. Ensuite, on a

$$\begin{aligned}\sum_{a=1}^{q}|B(x.a.y)| &= \textstyle\sum_{a\neq x_n, a\neq y_1}|B(x.a.y)| \\ &= \textstyle\sum_{a\neq x_n, a\neq y_1}\left(\displaystyle\sum_{i=1}^{n}|B(\hat{x}_i.a.y)| + B(x.y) + \sum_{i=1}^{p}|B(x.a.\hat{y}_i)|\right) \\ &= \textstyle\sum_{a\neq x_n}\displaystyle\sum_{i=1}^{n}|B(\hat{x}_i.a.y)| + \textstyle\sum_{a\neq y_1}\displaystyle\sum_{i=1}^{p}|B(x.a.\hat{y}_i)| \\ &\quad + (q-2+1_{\{x_n=y_1\}})B(x.y)\end{aligned}$$

On a alors

$$\begin{aligned}\textstyle\sum_{a\neq x_n}\displaystyle\sum_{i=1}^{n}|B(\hat{x}_i.a.y)| &= -\sum_{i=1}^{n}|B(\hat{x}_i.x_n.y)| + \textstyle\sum_{a\in\{1,\ldots,q\}}\displaystyle\sum_{i=1}^{n}|B(\hat{x}_i.a.y)| \\ &= -B(x.y) + \sum_{i=1}^{n}c_{n-1,p}|B(\hat{x}_i)|.|B(y)| \\ &= -B(x.y) + c_{n-1,p}|B(x)|.|B(y)|\end{aligned}$$

En effet, dans la première somme, seul le terme pour $i=n$ est non-nul. La deuxième somme est identifiée par l'hypothèse de récurrence.
De même

$$\textstyle\sum_{a\neq y_1}\displaystyle\sum_{i=1}^{p}|B(x.a.\hat{y}_i)| = -B(x.y) + c_{n,p-1}|B(x)|.|B(y)|,$$

d'où

$$\begin{aligned}\sum_{a=1}^{q}|B(x.a.y)| &= (c_{n-1,p}+c_{n,p-1})|B(x)|.|B(y)| + (q-4+1_{\{x_n=y_1\}})|B(x.y)| \\ &= (c_{n-1,p}+c_{n,p-1})|B(x)|.|B(y)| + 1_{\{x_n=y_1\}}|B(x.y)| \\ &= (c_{n-1,p}+c_{n,p-1})|B(x)|.|B(y)|,\end{aligned}$$

ce qui donne le résultat voulu en posant $c_{n,p}=c_{n-1,p}+c_{n,p-1}$.

(b) En sommant l'identité précédente sur tous les $(x,y)\in\{1,\ldots,q\}^{n+p}$, on a $S(n+p+1,4)=c_{n,p}S(n,4)S(p,4)$. En divisant l'identité de la question précédente par les deux membres de la nouvelle identité, on a

$$\sum_{a=1}^{q}\frac{|B(x.a.y)|}{S(n+p+1,4)} = \frac{|B(x)|}{S(n,4)}.\frac{|B(y)|}{S(p,4)},$$

, soit

$$\sum_{a=1}^{q} \mathbb{P}_q(\pi_1 = x_1, \pi_n = x_n, \pi_{n+1} = a, \pi_{n+2} = y_1, \dots, \pi_{n+p+1} = y_p)$$
$$= \mathbb{P}_q(\Pi_1 = x_1, \dots, \Pi_n = x_n)\mathbb{P}_q(\Pi_1 = y_1, \dots, \Pi_n = y_p),$$

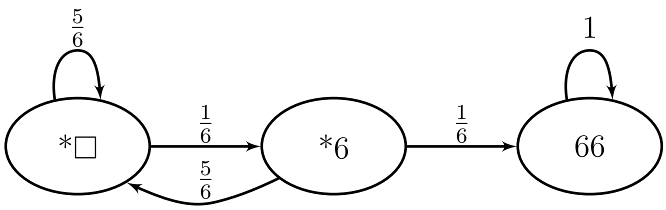
soit enfin en utilisant une partition et la stationnarité :

$$\mathbb{P}_q(\pi_1 = x_1, \pi_n = x_n, \pi_{n+2} = y_1, \dots, \pi_{n+p+1} = y_p)$$
$$= \mathbb{P}_q(\Pi_1 = x_1, \dots, \Pi_n = x_n)\mathbb{P}_q(\Pi_{n+2} = y_1, \dots, \Pi_{n+p+1} = y_p).$$

B.10 Solutions des exercices sur les chaînes de Markov

Solution 66 On choisit comme espace d'états l'ensemble $E = \{*\square, *6, 66\}$. (Il faut penser que $*$ représente un chiffre quelconque, $\square$ un chiffre différent de 6.)

On a le graphe de transition probabiliste suivant :



Je note $x = (x_n)_{n \geq 0}$ une suite possible d'éléments de E, et

$$T(x) = \inf\{n \geq 0; x_n =' 66'\}.$$

Il est facile de voir que si $x_0 =' 66'$, alors $T(x) = 0$, sinon

$$T(x) = 1 + T(\theta x),$$

où θx est la suite obtenue par décalage :

$$\forall n \geq 0 \quad (\theta x)_n = x_{n+1}.$$

Jusqu'ici il n'y a pas de probabilités. Notons maintenant $X = (X_i)_{i \geq 0}$ la suite des états obtenus, et $\mathbb{P}^i$ (resp. $\mathbb{E}^i$) la probabilité (l'espérance) sachant que le système part de l'état i ($X_0 = i$) et posons $x_i = \mathbb{E}^i[T(X)]$. Evidemment $x_{66} = 0$ et pour $i \neq 66$, on a

$$\begin{aligned}x_i &= \mathbb{E}^i[1+T(\theta X)]\\ &= 1+\mathbb{E}^i[T(\theta X)]\\ &= 1+\mathbb{E}^i[\sum_{j\in E}\mathbb{1}_{X_1=j}T(\theta X)]\\ &= 1+\sum_{j\in E}\mathbb{E}^i[\mathbb{1}_{X_1=j}T(\theta X)]\end{aligned}$$

Maintenant la propriété de Markov dit que pour toute fonction F de la trajectoire, on a

$$\mathbb{E}^i[\mathbb{1}_{X_1=j}F(\theta X)] = \mathbb{P}^i(X_1=j)\mathbb{E}^j[F(X)].$$

J'obtiens alors le système

$$\begin{aligned}x_{66} &= 0\\ x_{*6} &= 1+\frac{5}{6}x_{*\square}+\frac{1}{6}x_{66}\\ x_{*\square} &= 1+\frac{5}{6}x_{*\square}+\frac{1}{6}x_{*6}\end{aligned}$$

On résoud et on trouve $x_{*\square}=42$ et $x_{*6}=36$.

Le problème initial pouvant être représenté avec un départ en $*\square$, le nombre moyen de lancers nécessaires est donc $x_{*\square}=42$.

Solution 67 Comme la chaîne est irréductible et que le complémentaire de A est non-vide, pour tout $x\in A$, il existe n_x tel que $\mathbb{P}_x(X_{n_x}\notin A)>0$.

Ainsi $\alpha=\min\{\mathbb{P}_x(X_{n_x}\notin A); x\in A\}>0$ et, comme A est fini, $n=\max\{n_x; x\in A\}<+\infty$. Pour tout $x\in A$, on a $\mathbb{P}^x(T\le n)\ge \mathbb{P}^x(X_{n_x}\notin A)\ge \alpha$. Mais si $x\notin A$, on a $\mathbb{P}^x(T\le n)\ge\mathbb{P}^x(T=0)=1\ge\alpha$, donc finalement

$$\forall x\in S\quad \mathbb{P}^x(T\le n)\ge\alpha.$$

Posons $u_k=\mathbb{P}^x(\forall i\le nk, X_i\in A)$. Avec la propriété de Markov, on a

$$\begin{aligned}\mathbb{P}^x(\forall i\le n(k+1), X_i\in A|X_n) &= \mathbb{P}^x(\forall i\le nk, X_i\in A)\mathbb{P}^{X_n}(\forall i\le n, X_i\in A)\\ &\le \mathbb{P}^x(\forall i\le nk, X_i\in A)(1-\alpha),\end{aligned}$$

d'où en réintégrant $u_{k+1}\le(1-\alpha)u_k$, et, par récurrence $u_k\le(1-\alpha)^k$. Ainsi $\mathbb{P}^x(\tau=+\infty)\le\mathbb{P}^x(\tau>kn)\le(1-\alpha)^k$, donc en faisant tendre k vers l'infini, $\mathbb{P}^x(\tau=+\infty)=0$. Mais en fait, on a un peu plus,

$$\mathbb{P}^x(\tau>k)\le\mathbb{P}^x(\tau>n\lfloor k/n\rfloor)\le(1-\alpha)^{n\lfloor k/n\rfloor}\le(1-\alpha)^{k-n},$$

donc la variable aléatoire τ a une queue sous-exponentielle.

Solution 68 De trois choses l'une :

— Si $X_0 = 0$, alors $(X_0, X_1, X_2) = (0, 2, 0)$ et $(Y_0, Y_1, Y_2) = (0, 1, 0)$

— Si $X_0 = 1$, alors $(X_0, X_1, X_2) = (1, 1, 1)$ et $(Y_0, Y_1, Y_2) = (0, 0, 0)$

— Si $X_0 = 2$, alors $(X_0, X_1, X_2) = (2, 0, 0)$ et $(Y_0, Y_1, Y_2) = (1, 0, 1)$

En particulier $\mathbb{P}(Y_0 = 1, Y_1 = 0) \geq \mathbb{P}(X_0 = 2) > 0$, donc $\mathbb{P}(Y_1 = 0|Y_0 = 1) > 0$. De même $\mathbb{P}(Y_1 = 0, Y_2 = 0) \geq \mathbb{P}(X_0 = 1) > 0$, donc $\mathbb{P}(Y_2 = 0|Y_1 = 0) > 0$. Si $(Y_n)_{n\geq 0}$ était une chaîne de Markov, on aurait $\mathbb{P}(Y_0 = 1, Y_1 = 0, Y_2 = 0) = \mathbb{P}(Y_0 = 1, Y_1 = 0)\mathbb{P}(Y_2 = 0|Y_1 = 0) > 0$. Or $\mathbb{P}(Y_0 = 1, Y_1 = 0, Y_2 = 0) = 0$, donc $(Y_n)_{n\geq 0}$ n'est pas une chaîne de Markov.

Solution 69 Soient z et z' dans F. Comme ψ est surjective, il existe $x_0 \in E$ $\psi(x_0) = z$. On pose alors $q_{z,z'} = \mathbb{P}^{x_0}(\psi(X_1) = z')$. D'après l'hypothèse particulière faite sur f, on a $q_{z,z'} = \mathbb{P}^{x}(\psi(X_1) = z')$ pour tout $x \in E$ tel que $\psi(x) = z$. D'après la propriété de Markov, on a pour tout n,

$$\begin{aligned}
\mathbb{P}(Y_n = z'|X_0, \dots, X_{n-1}) &= \mathbb{P}(X_n \in \psi_{-1}(z')|X_0, \dots, X_{n-1}) \\
&= \mathbb{P}^{X_{n-1}}(X_1 \in \psi^{-1}(z')) \\
&= \mathbb{P}^{X_{n-1}}(\psi(X_1) = z') \\
&= q_{\psi(X_{n-1}),z'} = q_{Y_{n-1},z'}
\end{aligned}$$

Comme $\sigma(Y_0, \dots Y_{n-1})$ est une sous-tribu de $\sigma(X_0, \dots X_{n-1})$, on obtient que $\mathbb{P}(Y_n = z'|Y_0, \dots, Y_{n-1}) = q_{Y_{n-1},z'}$, ce qui montre que $(Y_n)_{n\geq 0}$ est une chaîne de Markov de matrice de passage q.

Si la chaîne (X_n) est invariante, soit (X_n) une chaîne partant sous la loi invariante. On a $\mathbb{P}(Y_1 \in A) = \mathbb{P}(X_1 \in \psi^{-1}(A)) = \mathbb{P}(X_0 \in \psi^{-1}(A)) = \mathbb{P}(Y_0 \in A)$, donc (Y_n) est invariante, et la loi de Y_0 est la loi image de P_{X_0} par ψ.

Solution 81 Quitte à échanger a et b, avec Bezout il existe u et v positifs avec $au - bv = 1$.

$$\frac{nu}{b} - \frac{nv}{a} = n\frac{au - bv}{ab} = \frac{n}{ab} \geq 1,$$

donc il existe p entier avec $\frac{nu}{b} \geq p \geq \frac{nv}{a}$. On a alors l'écriture

$$a(nu - bp) + b(ap - nv) = n.$$

B.11 Solutions des exercices sur la récurrence et les mesures invariantes

Solution 86 On peut aller en un coup de 3 vers 2, de 2 vers 1, de 1 vers 3 : la chaîne est donc irréductible. Par ailleurs on peut aller en un coup de 1 vers

1, donc la chaîne est apériodique : il y a donc une unique mesure invariante, qui est donnée par $\mu(y) = \frac{\mathbb{E}^1[N_y^1]}{\mathbb{E}^1[T_1]}$. Sous $\mathbb{P}^1$, on a $\mathbb{N}_1 = 1$, $N_2 = \mathbb{1}_{\{X_1 \neq 1\}}$, $N_3 = \mathbb{1}_{\{X_1 = 3}$ et $T_1 = X_1$, ce qui nous donne $\mathbb{E}^1[N^1] = 1$, $\mathbb{E}^1[N^2] = 2/3$, $\mathbb{E}^1[N^3] = 1/3$, $[\mathbb{E}[T_1] = 2$ puis $\mu(1) = \frac{1}{2}$, $\mu(2) = \frac{1}{3}$, $\mu(3) = \frac{1}{6}$.

Solution 87 Soit $P = (p_{i,j})$ la matrice de la chaîne. Comme $i \to j$, il existe n entier naturel tel que $p_{i,j}^{(n)} > 0$. Comme la mesure est invariante, on a $\mu(j) = \mathbb{P}^\mu(X_n = j) \geq \mathbb{P}^\mu(X_0 = i, X_n = j) = \mu(i)p_{i,j}^{(n)} > 0$.

Solution 88

1. La quantité $\sum_{j\in S} p_{i,j}|f(j)|$ est toujours bien définie, éventuellement infinie. C'est l'intégrale de la fonction $j \mapsto |f(j)|$ par rapport à la mesure de probabilité qui affecte la valeur $p_{i,j}$ au point j. On a donc
$$\left(\sum_{j\in S} p_{i,j}|f(j)|\right)^2 \leq \sum_{j\in S} p_{i,j}|f(j)|^2,$$
d'où
$$\mu(i)\left(\sum_{j\in S} p_{i,j}|f(j)|\right)^2 \leq \sum_{j\in S} p_{i,j}\mu(i)|f(j)|^2,$$
et en sommant
$$\sum_{i\in S}\mu(i)\left(\sum_{j\in S} p_{i,j}|f(j)|\right)^2 \leq \sum_{i\in S}\sum_{j\in S} p_{i,j}\mu(i)|f(j)|^2,$$
D'après le théorème de Tonelli des séries,
$$\sum_{i\in S}\sum_{j\in S} p_{i,j}\mu(j)|f(j)|^2 = \sum_{j\in S}\left(\sum_{i\in S} p_{i,j}\mu(i)\right)|f(j)|^2 = \sum_{j\in S}\mu(j)|f(j)|^2 = \|f\|_{2,\mu}^2 < +\infty.$$
En particulier, pour tout $i \in S$, on a $\mu(i)\left(\sum_{j\in S} p_{i,j}|f(j)|\right)^2 < +\infty$. Si $\mu(i) > 0$, cela entraîne que la série de terme général $(p_{i,j}f(j))_j$ converge absolument, donc converge, ce qui montre que $(Pf)(i)$ est bien défini. On a bien sûr $|(Pf)(i)|^2 \leq \left(\sum_{j\in S} p_{i,j}|f(j)|\right)^2$, d'où en combinant avec les inégalités ci-dessus : $\|P(f)\|_{2,\mu}^2 \leq |f\|_{2,\mu}^2$.
2. Pour tout $j \in S$, on a
$$\sum_{i\in S}\mu(i)p_{i,j} = \sum_{i\in S}\mu(j)p_{j,i}\mu(j)\sum_{i\in S}\mu(j)\mu(j).1 = \mu(j),$$
donc P laisse μ invariante. De même, pour tout $j \in S$, on a
$$\sum_{i\in S}\mu(i)q_{i,j} = \sum_{i\in S}\mu(j)p_{j,i}\mu(j)\sum_{i\in S}\mu(j)\mu(j).1 = \mu(j)$$

et Q laisse μ invariante. D'après la question précédente, Pf et Pg sont des éléments de $\ell^2(\mu)$. Comme le produit de deux éléments de $\ell^2(\mu)$ est dans $\ell^1(\mu)$, les intégrales considérées sont bien définies.

$$(f,g) \mapsto \int f(x)Pg(x)\ d\mu(x)$$

est une forme bilinéaire sur $\ell^2(\mu)$. C'est aussi une forme continue car, d'après Cauchy-Schwarz,

$$|\int f(x)Pg(x)\ d\mu(x)| \le \|f\|_{2,\mu}\|Pg\|_{2,\mu} \le \|f\|_{2,\mu}\|g\|_{2,\mu}.$$

Il en est de même pour $(f,g) \mapsto \int g(x)Qf(x)\ d\mu(x)$. Ainsi, l'ensemble des $(f,g) \in \ell^2(\mu) \times \ell^2(\mu)$ tels que

$$\int f(x)Pg(x)\ d\mu(x) = \int Qf(x)g(x)$$

est un fermé de $\ell^2(\mu)\times\ell^2(\mu)$ muni de la topologie produit. Si on trouve une partie D de $\ell^2(\mu)$ telle que

$$\forall f,g \in D \quad \int f(x)Pg(x)\ d\mu(x) = \int Qf(x)g(x),$$

on aura gagné car l'identité se prolongera alors à $\overline{D \times D} = \overline{D} \times \overline{D} = \ell^2(\mu) \times \ell^2(\mu)$.
On a $P\delta_j(x) = \sum_{i\in S} p_{x,i}\delta_j(i) = p_{x,j}$, ce qui nous donne

$$\int \delta_i(x)P\delta_j(x)\ d\mu(x) = \sum_{x\in S} \delta_i(x)p_{x,j}\mu(x) = p_{i,j}\mu(i).$$

De même $Q\delta_i(x) = q_{x,i}$ et

$$\int \delta_j(x)Q\delta_i(x)\ d\mu(x) = \sum_{x\in S} \delta_j(x)q_{x,i}\mu(x) = q_{j,i}\mu(j).$$

Or par hypothèse $\mu(i)p_{i,j} = \mu(j)q_{j,i}$, donc l'équation est vérifiée pour f et g de la forme $(\delta_i)_{i\in S}$. Par bi-linéarité, elle est encore vérifiée si f et g sont dans l'ensemble D des suites à support fini. Or cet ensemble est dense dans $\ell^2(\mu)$, ce qui donne le résultat voulu.

3. Soit $j \in S$.
   — Si $\mu(j) > 0$, on doit nécessairement avoir $q_{j,i} = \frac{\mu(i)}{\mu(j)}p_{i,j}$
   — si $\mu(j) = 0$, comme μ est invariante, on a $\mu(j) = \sum_k \mu(k)p_{k,j} \ge \mu(i)p_{i,j}$, donc $\mu(i)p_{i,j} = 0$ pour tout i, ce qui fait que pour n'importe quel choix de $q_{j,i}$, on aura $\mu(i)p_{i,j} = 0 = \mu(j)q_{j,i}$.

Ce dernier cas ne peut se produire si la chaîne est irréductible, car alors μ charge tous les points : il y a donc unicité de l'éventuelle solution. On fait ici le choix de poser $q_{j,i} = \delta_{i,j}$ si $\mu(j) = 0$. Il reste à voir qu'un candidat ainsi construit nous donne bien une matrice markovienne. Bien sûr, les coefficients sont tous positifs. Posons $S(j) = \sum_{i\in S} q_{j,i}$. Si $\mu(j) = 0$, $\mu(j) = 1$. Sinon,

$$S(j) = \sum_{i\in S} \frac{\mu(i)}{\mu(j)} p_{i,j} = \frac{1}{\mu(j)} \sum_{i\in S} \mu(i) p_{i,j}.$$

Mais comme μ est invariante, $\sum_{i\in S} \mu(i)p_{i,j} = \mu(j)$, d'où $S(j) = 1$: $Q = (q_{i,j})$ est bien une matrice markovienne.

4. Quelles que soient les fonctions mesurables f et g, on a

$$\begin{aligned}\mathbb{E}[f(X_k)g(X_{k+1})] &= \mathbb{E}[f(X_k)Pg(X_k)] \\ &= \int_0^1 f(x)Pg(x)\ d\lambda(x) \\ &= \int_0^1 Qf(x)g(x)\ d\lambda(x) \\ &= \mathbb{E}[Qf(X_{k+1})g(X_{k+1})]\end{aligned}$$

Ce qui entraîne $\mathbb{E}[f(X_k)|X_{k+1}] = Qf(X_{k+1})$. Soit $n > gek$. D'après le lemme de Doob, il existe une fonction F telle que $\mathbb{E}[f(X_k)|X_{k+1}, \dots, X_n] = F(X_{k+1}, \dots, X_n)$. Montrons que F ne dépend que de sa première coordonnée. Quitte à changer d'espace de probabilité, on peut supposer que (X_n) est construite par une dynamique aléatoire $X_{n+1} = f_{n+1}(X_n)$, où X_0 suit la loi μ et $(f_n)_{n\geq 1}$ une suite de fonctions aléatoires indépendantes, indépendantes de X_0. Comme $(f_{k+2}, \dots, f_n)$ est indépendant de $(f(X_k), X_{k+1})$, on a $\mathbb{E}[f(X_k)|X_{k+1}] = \mathbb{E}[f(X_k)|X_{k+1}, f_{k+2}, \dots, f_n]$. Mais $\sigma(X_{k+1}, f_{k+2}, \dots, f_n) = \sigma(X_{k+1}, X_{k+2}, \dots, X_n)$, donc on obtient

$$\mathbb{E}[f(X_k)|X_{k+1}, \dots X_n] = \mathbb{E}[f(X_k)|X_{k+1}] = Qf(X_{k+1}).$$

Si pour $0 \leq k \leq n$, on pose $Z'_k = X_{n-k}$, on a donc

$$\mathbb{E}[f(Z'_{k+1})|Z'_k, \dots Z'_0] = Qf(Z'_k),$$

ce qui montre que $(Z'_k)_{0\leq k\leq n})$ est une chaîne de Markov qui a le même opérateur de transition que $(X'_n)_{n\geq 0}$. Comme la loi initiale est la même, on a l'égalité en loi entre $(Z'_0, \dots Z'_n)$ et $(X'_0, X'_1, \dots, X'_n)$, soit donc entre (X_n, X_{n-1}, X_0) et $(X'_0, X'_1, \dots, X'_n)$.

Si $P = Q$, le système est alors réversible : on retrouve le fait que pour une mesure initiale réversible, (X_n, X_{n-1}, X_0) et $(X'_0, X'_1, \dots, X'_n)$ ont même loi.

5. D'après la question 3., on peut toujours construire une matrice markovienne conjuguée à P. La question 4. donne alors le résultat voulu.

Solution 89 1. Notons $\mathcal{F}_m = \sigma(X_0, \dots, X_m)$. T^m prend a priori ses valeurs dans $\mathbb{N}^* \cup \{+\infty\}$. Pour tout entier $k \geq 1$, on a

$$\{T^m = k\} = \{\sum_{i=0}^{k-1} \mathbb{1}_A(X_k) = m-1, X_k \in A\} \in \mathcal{F}_k,$$

Donc T^m est un temps d'arrêt adapté à la filtration $(\mathcal{F}_m)_{m\geq 0}$. Montrons par récurrence que T^m est $\mathbb{P}^x$-presque sûrement fini. C'est vrai pour $m = 0$ et $m = 1$. Supposons $T^{m-1} < +\infty$ $\mathbb{P}^x$-presque sûrement. On a

$$\{T_m < +\infty\} = \{T_{m-1} < +\infty, \exists; n \geq 1; X_{T_{m-1}+n} \in A\},$$

donc avec la propriété de Markov forte, on a $\mathbb{P}^x$ presque sûrement

$$\begin{aligned}\mathbb{P}_x(T_m < +\infty | \mathcal{F}_{T_m}) &= \mathbb{P}^x(T_{m-1} < +\infty)\mathbb{P}^{X_{T_{m-1}}}(\exists; n \geq 1; X_{T_{m-1}+n} \in A) \\ &= 1.\mathbb{P}^{X_{T_{m-1}}}(T^1 < +\infty) = 1,\end{aligned}$$

et en réintégrant $\mathbb{P}^x(T_m < +\infty) = \mathbb{E}^x[\mathbb{P}_x(T_m < +\infty | \mathcal{F}_{T_m})] = 1$, ce qui montre par récurrence la propriété voulue. Enfin, pour tout borélien B, on a

$$\{X_{T^m} \in B, T^m \leq n\} = \cup_{i=1}^n \{X_i \in B\} \cap \{T^m = i\}.$$

Mais pour tout i entre 1 et n, $\{X_i \in B\} \cap \{T^m = i\} \in \mathcal{F}_i \subset \mathcal{F}_n$, donc $\{X_{T^m} \in B, T^m \leq n\} \in \mathcal{F}_n$, ce qui montre que $\{X_{T^m} \in B\} \in \mathcal{F}_{T^m}$. Comme c'est vrai pour B borélien quelconque, X_{T^m} est $\mathcal{F}_{T^m}$-mesurable. On a vu en cours que T^m était $\mathcal{F}_{T^m}$-mesurable. Maintenant, si $k \leq m$, comme $T^k \leq T^m$, on a l'inclusion $\mathcal{F}_{T^k} \subset \mathcal{F}_{T^k}\mathcal{F}_{T^m}$, ce qui entraîne que pour $k \leq m$, T^k et X_{T^k} sont $\mathcal{F}_{T^m}$-mesurables.

2. En utilisant la propriété de Markov forte, on a

$$\begin{aligned}\mathbb{P}^x(X_{T^{m+1}} = y | \mathcal{F}_{T^m}) &= \sum_{n=1}^{+\infty} \mathbb{P}^x(\inf\{i \geq 1; X_{T^m+i} \in A\} = n, X_{T^m+n} = y | \mathcal{F}_{T^m}) \\ &= \sum_{n=1}^{+\infty} \mathbb{P}^{X_{T^m}}(\inf\{i \geq 1; X_i \in A\} = n, X_n = y) \\ &= \mathbb{P}^{X_{T^m}}(X_{T^1} = y)\end{aligned}$$

Ainsi, si on pose $q_{x,y} = \mathbb{P}^x(X_{T^1} = y)$, on a $\mathbb{P}^x(X_{T^{m+1}} = y|\mathcal{F}_{T^m}) = q_{X_{T^m},y}$. D'après la question précédente, $\sigma(X_{T^0}, \dots X_{T^m})$ est une sous-tribu de $\mathcal{F}_{T^m}$, donc

$$\begin{aligned}\mathbb{P}^x(X_{T^{m+1}} = y|\sigma(X_{T^0}, \dots X_{T^m})) &= \mathbb{E}^x(\mathbb{P}^x(X_{T^{m+1}} = y|\mathcal{F}_{T^m})|\sigma(X_{T^0}, \dots X_{T^m})) \\ &= \mathbb{E}^x(q_{X_{T^m},y}|\sigma(X_{T^0}, \dots X_{T^m})) \\ &= q_{X_{T^m},y}\end{aligned}$$

3. Comme les $(T_A^m)_{m\geq 1}$ sont tous les moments où (X_n) passe dans A, le premier moment (s'il existe) où X_n vaut x est un T_k, d'où

$$T_x = \sum_{k=0}^{+\infty}(T_A^{k+1} - T_A^k)\mathbb{1}_{\{S>k\}},$$

où $S = \inf\{n \geq 1; Y_n = x\}$. Comme les termes sont positifs, on a

$$\mathbb{E}^x(T_x) = \sum_{k=0}^{+\infty}\mathbb{E}^x[(T_A^{k+1} - T_A^k)\mathbb{1}_{\{S>k\}}],$$

$$\{S > k\} = \cap_{i=1}^k\{X_{T_i} \neq x\},$$

donc $\{S > k\}$ est $\mathcal{F}_{T^k}$ mesurable, comme intersection d'événements $\mathcal{F}_{T^k}$ mesurables. On a donc

$$\mathbb{E}^x[(T_A^{k+1} - T_A^k)\mathbb{1}_{\{S>k\}}|\mathcal{F}_{T_k}] = \mathbb{1}_{\{S>k\}}\mathbb{E}^x[(T_A^{k+1} - T_A^k)|\mathcal{F}_{T_k}].$$

Mais $T_A^{k+1} - T_A^k = \inf\{n \geq 1; X_{T_A^k+n} \in A\}$, donc avec la propriété de Markov forte, on a

$$\mathbb{E}^x[(T_A^{k+1} - T_A^k)|\mathcal{F}_{T_k}] = \mathbb{E}^{X_{T_A^k}}\inf\{n \geq 1, X_n \in A\} = \mathbb{E}^{X_{T_A^k}}T_A^1.$$

Ainsi, si on pose $\alpha = \max\{E^x(T^1); x \in A\}$, on a $\mathbb{E}^x[(T_A^{k+1} - T_A^k)\mathbb{1}_{\{S>k\}}|\mathcal{F}_{T_k}] \leq \alpha\mathbb{1}_{\{S>k\}}$, et en réintégrant

$$\mathbb{E}^x[(T_A^{k+1} - T_A^k)\mathbb{1}_{\{S>k\}}] \leq \alpha\mathbb{P}^x(S > k),$$

d'où en faisant la somme $\mathbb{E}^x(T^x) \leq \sum_{k=0}^{+\infty}\alpha\mathbb{P}^x(S > k) = \alpha\mathbb{E}^x(S)$. Mais S est le temps de retour en x pour une chaîne de Markov irréductible sur un espace d'état fini : il est donc intégrable, car une chaîne de Markov irréductible sur un espace d'états fini est toujours récurrente positive. On a donc $E^x(T_x) < +\infty$, ce qui montre que la chaîne (X_n) elle-même est récurrente.

Solution 90 1. (a) On a

$$\begin{aligned} f(X_{n\wedge T}) - f(X_0) &= \sum_{i=1}^{n\wedge T} (f(X_i) - f(X_{i-1})) \\ &= \sum_{i=1}^{n} \mathbb{1}_{\{i\le T\}} (f(X_i) - f(X_{i-1})) \\ &= \sum_{i=1}^{n} \mathbb{1}_{\{i-1< T\}} (f(X_i) - f(X_{i-1})) \end{aligned}$$

En prenant l'espérance, on a

$$\begin{aligned} \mathbb{E}(f(X_{n\wedge T}) - f(X_0)) &= \sum_{i=1}^{n} \mathbb{E}\left(\mathbb{1}_{\{i-1< T\}} (f(X_i) - f(X_{i-1}))\right) \\ &= \sum_{i=1}^{n} \mathbb{E}\left(\mathbb{E}\left(\mathbb{1}_{\{i-1< T\}} f(X_i) - f(X_{i-1}) | \mathcal{F}_{i-1}\right)\right), \end{aligned}$$

où l'on a posé $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$. Mais

$$\begin{aligned} \mathbb{E}\left(\mathbb{1}_{\{i-1< T\}}(f(X_i) - f(X_{i-1}) | \mathcal{F}_{i-1}\right) &= \mathbb{1}_{\{i-1< T\}} \mathbb{E}\left((f(X_i) - f(X_{i-1}) | \mathcal{F}_{i-1}\right) \\ &\le -\varepsilon \mathbb{1}_{\{i-1< T\}}, \end{aligned}$$

ce qui en réintégrant donne l'inégalité voulue.

(b) On a pour tout $n \ge 1$,

$$\sum_{i=1}^{n} \mathbb{P}(T > i-1) = \sum_{i=1}^{+\infty} \mathbb{P}(T \wedge (n-1) > i-1) = \mathbb{E}[T \wedge (n-1)],$$

d'où $\varepsilon \mathbb{E}(T \wedge (n-1)) \le \mathbb{E} f(X_0) - \mathbb{E} f(X_{T\wedge n}) \le \mathbb{E} f(X_0)$, soit

$$\mathbb{E}(T \wedge (n-1)) \le \frac{\mathbb{E} f(X_0)}{\varepsilon}.$$

En faisant tendre n vers l'infini, on obtient par convergence monotone $\mathbb{E}(T) \le \frac{\mathbb{E} f(X_0)}{\varepsilon}$.

2. (a) On suppose $\mathbb{E}[f(X_0)] < +\infty$. On montre l'intégrabilité de $f_M(X_i)\mathbb{1}_{\{i-1<T\}}$ par récurrence sur i. Pour $i = 0$, c'est évident Posons $f_M = f \wedge M$. Avec le théorème 70, on a

$$\begin{aligned} \mathbb{E}[f_M(X_i)\mathbb{1}_{\{i-1<T\}} | \mathcal{F}_{i-1}] &= \mathbb{1}_{\{i-1<T\}} \mathbb{E}[f_M(X_i) | \mathcal{F}_{i-1}] &\text{(B.1)} \\ &= \mathbb{1}_{\{i-1<T\}} (P f_M)(X_{i-1}) &\text{(B.2)} \end{aligned}$$

On en déduit

$$\begin{aligned}\mathbb{E}[f_M(X_i)\mathbb{1}_{\{i-1<T\}}|\mathcal{F}_{i-1}] &\le \mathbb{1}_{\{i-1<T\}}(Pf_M)(X_{i-1})\\ &\le f(X_{i-1})\mathbb{1}_{\{i-1<T\}} \le f(X_{i-1})\mathbb{1}_{\{i-2<T\}},\end{aligned}$$

d'où en intégrant $\mathbb{E}([f_M(X_i)\mathbb{1}_{\{i-1<T\}}) \le \mathbb{E}[f(X_{i-1})\mathbb{1}_{\{i-2<T\}}]$. En passant au sup en M, on obtient

$$\mathbb{E}([f(X_i)\mathbb{1}_{\{i-1<T\}}) \le f(X_{i-1})\mathbb{1}_{\{i-2<T\}} < +\infty.$$

Avec le théorème de convergence dominée conditionnel,(B.1) entraîne $\mathbb{E}[f(X_i)\mathbb{1}_{\{i-1<T\}}|\mathcal{F}_{i-1}] = (Pf)(X_{i-1})\mathbb{1}_{\{i-1<T\}}$. Par linéarité, on a alors

$$\begin{aligned}\mathbb{E}\left(\mathbb{1}_{\{i-1<T\}}(f(X_i) - f(X_{i-1})|\mathcal{F}_{i-1}\right) &= \mathbb{1}_{\{i-1<T\}}((Pf)(X_{i-1}) - f(X_{i-1}))\\ &\le -\varepsilon\mathbb{1}_{\{i-1<T\}},\end{aligned}$$

et la preuve de 1) et 2) se poursuit sans modification notable, puisqu'on a vérifié les intégrabilités plus subtiles.

(b) Soit $N > \max(f(x); x \in M)$. Si T'_N est le temps d'entrée dans $M' = M \cup \{x \in E : f(x) \ge N\}$, le 1.(a) nous donne pour tout n : $E(f(X_{n\wedge T'_N})) \le \mathbb{E}f(X_0)$, d'où

$$N\mathbb{P}(\tau_N < T, T'_N \le n) = \mathbb{E}(f(X_{n\wedge T'})\mathbb{1}_{\{\tau_N<T,T'_N\le n\}})) \le \mathbb{E}f(X_0)$$

T' est le temps d'entrée d'une chaîne irréductible dans le complémentaire d'un ensemble fini : il est presque sûrement fini. En faisant tendre n vers l'infini, on obtient $N\mathbb{P}(\tau_N < T) \le \mathbb{E}f(X_0)$. Maintenant

$$P(T = +\infty) = \mathbb{P}(T = +\infty, T'_N < +\infty) \le \mathbb{P}(\tau_N < T) \le (\mathbb{E}f(X_0))/N.$$

En faisant tendre N vers l'infini, on obtient $\mathbb{P}(T = +\infty) = 0$.

(c) D'après la première question $\mathbb{P}^x(T_M < +\infty)$ pour tout $x \in M$, donc pour $x \in M$, on peut faire partir une chaîne (X_n) de x et construire avec l'exercice précédent la suite (Y_n) des sites de M successivement visités. (Y_n) est une chaîne de Markov irréductible d'un espace d'état fini : elle est donc récurrente et passe une infinité de fois en x. Par suite, la chaîne (X_n) elle-même passe une infinité de fois en x et donc l'état x est récurrent. Si $\varepsilon > 0$, on a pour tout $x \in M$, $\mathbb{E}^xT_M < +\infty$, et donc avec l'exercice précédent x est récurrent positif. Comme la chaîne est irréductible la récurrence (ou la récurrence positive) d'un état entraîne celle de la chaîne elle-même.

Annexe C

Problèmes

C.1 Problème 1 : nombres de Stirling

Partie I Soit X_0 une variable aléatoire à valeurs dans $\mathbb{N}$, $(U_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. On suppose de plus que X_0 est indépendante de $(U_n)_{n\geq 1}$. On définit une suite de variables aléatoires $(X_n)_{n\geq 0}$ par la donnée de X_0 et la récurrence $X_{n+1} = \mathrm{Ent}(U_n X_n)$.

1. Montrer que l'on obtient ainsi une chaîne de Markov.
   Dans la suite, on notera $\mathbb{P}^N$ la loi sur l'espace canonique de la chaîne de Markov associée à cette dynamique qui vérifie $\mathbb{P}^N(X_0 = N) = 1$.
2. On note $T = \inf\{n \geq 0; X_n = 0\}$. Montrer que $\mathbb{P}^N(T < +\infty) = 1$.
3. On note φ_n la fonction génératrice de T sous la loi $\mathbb{P}^n$, c'est à dire $\varphi_n(x) = \mathbb{E}^n[x^T]$. À l'aide de la propriété de Markov, établir la formule de récurrence
$$\varphi_n(x) = \frac{x}{n}\sum_{i=0}^{n-1}\varphi_i(x).$$
4. Montrer que
$$\forall n \geq 0 \quad \forall x \in [0,1] \quad \varphi_n(x) = P_n(x),$$
   où la suite de polynômes $(P_n)_{n\geq 0}$ est définie par $P_0 = 1$ et $P_n(X) = \frac{X(X+1)\dots(X+n-1)}{n!}$ pour $n \geq 1$.
5. On définit les *nombres de Stirling de première espèce* $\begin{bmatrix} n \\ k \end{bmatrix}$ par l'identité
$$X(X-1)\dots(X-n+1) = \sum_{k=1}^{n}\begin{bmatrix} n \\ k \end{bmatrix} X^k.$$

Montrer que

$$\mathbb{P}^n(T = k) = \frac{(-1)^{n+k}}{n!} \begin{bmatrix} n \\ k \end{bmatrix}.$$

6. Montrer que pour tout $n \geq 1$, on a

$$\mathbb{E}^n[T] = \sum_{k=1}^{n} \frac{1}{k}.$$

Partie II On a N cartes avec des chiffres numérotés de 1 à N. On tire au hasard une première carte et on la garde. Ensuite, on tire au hasard les cartes, et si le chiffre est inférieur à la première carte, on la garde, sinon, on la jette. Et ainsi de suite, on jette toute carte de valeur supérieure à la plus grande des cartes tirées précédemment, jusqu'à épuisement du paquet. On s'intéresse au nombre final F de cartes gardées.

1. Montrer comment on peut modéliser le problème à l'aide de la loi uniforme sur $\mathfrak{S}(N)$, avec

$$F(\sigma) = |\{i \in \{1, \ldots, N\}; j < i \Longrightarrow \sigma(i) < \sigma(j)\}|.$$

2. On définit une suite $(Z_n)_{n\geq 0}$ par récurrence comme suit :
On pose $Z_0 = N$ et pour $k \geq 1$,

$$Z_k = \min(\sigma(1), \sigma(2), \ldots, \sigma(k)) - 1,$$

avec la convention $\sigma(i) = +\infty$ pour $i > N$.
Montrer que $\forall n \in \mathbb{N}^* \quad \forall j \in \{0, \ldots, N\}$

$$\mathbb{P}(Z_{n+1} = j | \sigma(1), \ldots, \sigma(n)) = \frac{1}{N-n} \mathbb{1}_{\{j<Z_n\}} + \frac{N-n-Z_n}{N-n} \mathbb{1}_{\{j=Z_n\}}.$$

En déduire que $(Z_n)_{n\geq 0}$ est une chaîne de Markov.

3. On définit la suite $(T_k)_{k\in\mathbb{N}}$ par récurrence comme suit : on pose $T_0 = 0$ et

$$T_{k+1} = T_k + \mathbb{1}_{\{Z_{T_k}>0\}} \min\{n \geq 0, Z_{T_k+n} \neq Z_{T_k}\},$$

avec la convention $0.\infty = 0$. Montrer que la suite $(Z_{T_k})_{k\geq 0}$ a la même loi que la suite $(X_n)_{n\geq 0}$ étudiée en I sous la loi $\mathbb{P}^N$.

4. Montrer que $F = \inf\{k \geq 1;\ Z_{T_k} = 0\}$.
5. Montrer que $\{F = k\}$ est en bijection avec les éléments de $\mathfrak{S}(N)$ possédant exactement k cycles.
6. Démontrer alors l'interprétation combinatoriale des nombres de Stirling de première espèce : $(-1)^{n+k} \begin{bmatrix} n \\ k \end{bmatrix}$ est le nombre de permutations de n objets ayant exactement k cycles.

C.2 Problème 2 : théorème d'Erdös, Feller et Pollard

Processus de renouvellement

Soit $(U_n)_{n\geq 1}$ une suite de variables aléatoires indépendante suivant la loi μ. On suppose que $\text{pgcd}(n \geq 1; \mu(n-1) > 0) = 1$. Soit ν une loi quelconque sur $\mathbb{N}$. On définit une suite $(X_n)_{n\geq 0}$ par la donnée de X_0 suivant la loi ν et indépendante de $(U_n)_{n\geq 1}$, puis, pour $n \geq 0$

$$X_{n+1} = \begin{cases} U_{n+1} & \text{si } X_n = 0 \\ X_n - 1 & \text{sinon} \end{cases}$$

1. Montrer que $(X_n)_{n\geq 0}$ est une chaîne de Markov de loi initiale ν. Écrire sa matrice.
2. Montrer que la chaîne est irréductible et apériodique.
3. La chaîne est-elle transiente, récurrente ?
4. Montrer que
$$\forall k \in \mathbb{N} \quad \mathbb{P}^\nu(X_1 = k) = \nu(k+1) + \nu(0)\mu(k).$$
5. Montrer que si $\nu(0) = 0$, alors la chaîne n'est pas stationnaire.
6. Pour $k \geq 0$, on note
$$T_k = \inf\{n \geq 1; X_n = k\}.$$
Montrer que sous $\mathbb{P}^0$, T_0 et $1 + X_1$ ont même loi.
7. En utilisant les questions précédentes, montrer que si la chaîne admet une mesure invariante, alors μ admet un moment d'ordre 1 et on a la relation $\nu(0) = \frac{1}{1+\mathbb{E}[U_1]}$.
8. Réciproquement, montrer que si μ admet un moment d'ordre 1, alors la chaîne admet une mesure invariante.
Indication : on pourra considérer la mesure ν définie par
$$\nu(k) = \frac{\mathbb{P}(U_1 \geq k)}{1 + \mathbb{E}[U_1]}.$$
9. Montrer que si μ admet un moment d'ordre 1, la suite $\frac{1}{n}\sum_{k=0}^{n-1} X_k$ converge presque sûrement vers une limite que l'on déterminera.
10. On pose $p'_k = \mathbb{P}^0(T_0 = k)$ et $q'_k = \mathbb{P}^0(X_k = 0)$. Établir la récurrence
$$q'_0 = 1 \text{ et } \forall n \geq 1 \quad q'_n = \sum_{k=1}^{n} p'_k q'_{n-k}.$$

11. Démontrer le théorème de renouvellement, appelé également, en l'honneur de ses auteurs, théorème d'Erdös, Feller et Pollard :
   Soient $X_1, \dots, X_n \dots$ des variables aléatoires indépendantes identiquement distribuées avec $\mathbb{P}(X_1 = k) = p_k$. On suppose que $0 < \mathbb{E}[X_1] < +\infty$ et que le pgcd des éléments du support de X_1 est 1. On note

$$S_n = X_1 + X_2 + \cdots + X_n \text{ et } N_k = \sum_{n=1}^{+\infty} \mathbb{1}_{\{S_n = k\}}.$$

   Alors, si on pose

$$\forall z \in B_F(0,1) \quad P(z) = \sum_{n=0}^{+\infty} p_n z^n$$
$$\text{et } Q(z) = \frac{1}{1 - P(z)} = \sum_{n=0}^{+\infty} q_n z^n,$$

   on a $\mathbb{E}[N_k] = q_k$ et $\lim_{k\to+\infty} q_k = \frac{1}{\mathbb{E}[X_1]}$.

C.3 Problème 3 : théorème de De Finetti–Hewitt–Savage

Soit $\Omega = \mathbb{R}^{\mathbb{N}^*}$ l'espace canonique. Pour $n \in \mathbb{N}^*$, on note X_k la projection de Ω sur la k-ième coordonnée : $X_k(\omega) = \omega_k$. On note $\mathcal{S}_n$ l'ensemble des permutations de $\mathbb{N}^*$ qui laissent invariants point par point les éléments de $\{n+1, \dots, \}$. Pour $\sigma \in \mathcal{S}_n$ et $\omega \in \Omega$, on note $T_\sigma(\omega) = (\omega_{\sigma(k)})_{k\geq 1}$ et

$$\mathcal{I}_n = \{A \in \mathcal{B}(\Omega); \forall \sigma \in \mathcal{S}_n \quad T_\sigma(A) = A\}.$$

On note enfin

$$\mathcal{I} = \cap_{n\geq 1} \mathcal{I}_n \text{ et } \mathcal{S} = \cup_{n\geq 1} \mathcal{S}_n.$$

On dit qu'une mesure $\mathbb{P}$ sur $(\Omega, \mathcal{B}(\Omega))$ est échangeable si

$$\forall \sigma \in \mathcal{S} \quad \mathbb{P}_{T_\sigma} = \mathbb{P}. \tag{C.1}$$

On dit d'une famille de variables $(X_n)_{n\geq 1}$ définies sur un espace $(\Omega, \mathcal{F}, \mathbb{P})$ quelconque qu'elle est échangeable si pour tout n et pour tout $\sigma \in \mathcal{S}_n$, les vecteurs $(X_1, \dots, X_n)$ et $(X_{\sigma(1)}, \dots, X_{\sigma(n)})$ ont même loi (sous $\mathbb{P}$).

1. Vérifier que la famille de variables $(X_k)_{k\geq 1}$ est échangeable si et seulement si sa loi est échangeable. Dès lors, pour étudier les propriétés des familles de variables échangeables, il suffit d'étudier les propriétés des lois échangeables sur l'espace canonique.

2. On se donne donc une mesure de probabilité échangeable sur l'espace canonique $\Omega = \mathbb{R}^{\mathbb{N}}$. On se propose de montrer que, conditionnellement à la tribu $\mathcal{I}$ des événements invariants par les permutations à support fini, les variables X_i sont indépendantes, c'est à dire que pour tout p, si $f_1, \dots, f_p$ sont des fonctions mesurables bornées de $\mathbb{R}$ dans $\mathbb{R}$, on a

$$\mathbb{E}(\prod_{i=1}^{p} f_i(X_i)|\mathcal{I}) = \prod_{i=1}^{p} \mathbb{E}(f_i(X_i)|\mathcal{I}).$$

(a) Montrer que $(\mathcal{I}_n)_{n\geq 1}$ est une suite décroissante de tribus.

(b) Soit $n \geq 1$, F une fonction $(\Omega, \mathcal{B}(\Omega)) - (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable telle que $F \circ T_\sigma = F$ pour tout $\sigma \in \mathcal{S}_n$. Montrer que F est $\mathcal{I}_n$-mesurable.

(c) Soit n un entier naturel non nul et f une fonction bornée de $\mathbb{R}$ dans $\mathbb{R}$. Montrer que pour tout $A \in \mathcal{I}_n$ et tout k entre 1 et n, $\mathbb{E}[\mathbb{1}_A f(X_1)] = \mathbb{E}[\mathbb{1}_A f(X_k)]$. En déduire que pour tout i entre 1 et n,

$$\mathbb{E}[f(X_i)|\mathcal{I}_n] = \frac{1}{n}\sum_{k=1}^{n} f(X_k).$$

(d) À l'aide d'un théorème de martingales, montrer que $\frac{1}{n}\sum_{k=1}^{n} f(X_k)$ converge presque sûrement vers $\mathbb{E}[f(X_1)|\mathcal{I}]$, qui coïncide avec $\mathbb{E}[f(X_i)|\mathcal{I}]$ pour tout entier naturel non nul i.

(e) Soient $f_1, \dots, f_p$ des fonctions mesurables bornées de $\mathbb{R}$ dans $\mathbb{R}$. Pour $n \geq p$, Notons $I_{n,p}$ l'ensemble des injections de $\{1, \dots, p\}$ dans $\{1, \dots, n\}$. Montrer que

$$\mathbb{E}\left(\prod_{i=1}^{p} f_i(X_i)|\mathcal{I}_n\right) = \frac{1}{|I_{n,p}|}\sum_{g\in I_{n,p}}\prod_{i=1}^{p} f_i(X_{g(i)})$$

et étudier le comportement asymptotique lorsque n tend vers l'infini.

(f) Comparer $\frac{1}{|I_{n,p}|}\sum_{g\in I_{n,p}}\prod_{i=1}^{p} f_i(X_{g(i)})$ et $\frac{1}{n^p}\sum_{g\in\{1,\dots,n\}^p}\prod_{i=1}^{p} f_i(X_{g(i)})$. En déduire que

$$\lim_{n\to+\infty}\frac{1}{|I_{n,p}|}\sum_{g\in I_{n,p}}\prod_{i=1}^{p} f_i(X_{g(i)}) = \prod_{i=1}^{p}\mathbb{E}(f_i(X_1)|\mathcal{I}).$$

(g) Montrer enfin que

$$\mathbb{E}(\prod_{i=1}^{p} f_i(X_i)|\mathcal{I}) = \prod_{i=1}^{p} \mathbb{E}(f_i(X_i)|\mathcal{I}).$$

(h) En suivant les notations introduites dans le théorème 36, pour $\omega \in \Omega$, on note $\mathbb{P}_\omega^{\mathcal{I}}$ la valeur au point ω d'une loi conditionnelle de $\mathbb{P}$ sachant la tribu $\mathcal{I}$. On note alors $M_\omega = ((\mathbb{P}_\omega^{\mathcal{I}})_{X_1})^{\otimes \mathbb{N}^*}$. Montrer que pour tout $A \in \mathcal{B}(\Omega)$, $A \mapsto M_\omega(A)$ est mesurable et que

$$\mathbb{P}(A) = \int M_\omega(A) \; d\mathbb{P}(\omega),$$

Ainsi, la mesure $\mathbb{P}$ apparaît comme un mélange de mesures produits.

3. *Retour sur l'urne de Pólya*

On suppose maintenant que $\mathbb{P}$ est la loi du processus des tirages (T_i) dans le modèle d'urne de Pólya vu à l'exercice 24.

(a) Vérifier que $\mathbb{P}$ est échangeable.

(b) Montrer que sachant $\mathcal{I}$, les variables aléatoires $(T_i)_{i \geq 1}$ sont des variables aléatoires indépendantes dont la loi est donnée par le vecteur de probabilité W.

Annexe D

Solutions des problèmes

D.1 Solution du problème 1

Partie I

1. La récurrence s'écrit $X_{n+1} = F(X_n, U_n)$, avec $(U_n)_{n\geq 0}$ iid et indépendant de X_0, avec $F(x, y) = \mathrm{Ent}(x, y)$
2. Si $X_n > 0$, on a presque sûrement $U_n X_n < X_n$, or X_n est entier, donc $X_{n+1} = \mathrm{Ent}(U_n X_n) < X_n \leq X_n - 1$, d'où $T \leq X_0 = N$ et on a bien $\mathbb{P}^N(T < +\infty) = 1$.
3. Pour tout n, on a $\mathbb{P}^n(0 \leq X_1 < N) = 1$, donc

$$\begin{aligned}
\mathbb{E}^n[x^T] &= \sum_{i=0}^{n-1} \mathbb{E}^n[\mathbb{1}_{\{X_1=i\}} x^T] \\
&= \sum_{i=0}^{n-1} \mathbb{E}^n[\mathbb{1}_{\{X_1=i\}} x^{1+T\circ\theta}] \\
&= \sum_{i=0}^{n-1} \mathbb{E}^n[(\mathbb{1}_{\{X_1=i\}} x) x^T \circ \theta] \\
&= \sum_{i=0}^{n-1} \mathbb{E}^n[(\mathbb{1}_{\{X_1=i\}} x] \mathbb{E}^i[x^T] \\
&= \sum_{i=0}^{n-1} x \mathbb{P}^n(X_1 = i) \varphi_i(x) \\
&= \frac{x}{n} \sum_{i=0}^{n-1} \varphi_i(x)
\end{aligned}$$

4. Preuve par récurrence.

5. $P_n(x) = \varphi_n(x) = \mathbb{E}^n[x^T] = \sum_{k=0}^n \mathbb{P}^n(T=k)x^k$, donc $\mathbb{P}^n(T=k)$ est le coefficient en x^k du polynôme $P_n(x)$. Mais comme
$$X(X-1)\dots(X-n+1) = \sum_{k=1}^n \begin{bmatrix} n \\ k \end{bmatrix} X^k,$$
par substitution
$$(-X)(-X-1)\dots(-X-n+1) = \sum_{k=1}^n \begin{bmatrix} n \\ k \end{bmatrix} (-1)^k X^k,$$
et en multipliant par $(-1)^n$:
$$X(X+1)\dots(X+n-1) = \sum_{k=1}^n \begin{bmatrix} n \\ k \end{bmatrix} (-1)^{n+k} X^k,$$
ce qui nous permet d'identifier le coefficient voulu et donne bien
$$\mathbb{P}^n(T=k) = \frac{(-1)^{n+k}}{n!} \begin{bmatrix} n \\ k \end{bmatrix}.$$
6. D'après les propriétés de la fonction génératrice, on a $\mathbb{E}^n[T] = \varphi_n'(1) = P_n'(1)$. Or
$$\frac{P_n'}{P_n} = \sum_{i=0}^{n-1} \frac{1}{X+i}, \text{ donc } \frac{P_n'(1)}{P_n(1)} = \sum_{i=0}^{n-1} \frac{1}{1+i} = \sum_{k=1}^n \frac{1}{k},$$
Comme $P_n(1) = 1$, on a donc finalement
$$\mathbb{E}^n[T] = \sum_{k=1}^n \frac{1}{k}.$$

Partie II

1. $F(\sigma)$ est le nombre de “records” descendants de la permutation σ.
2. On a
$$\begin{aligned} & \mathbb{P}(\sigma(n+1) = a_{n+1} | \sigma(1) = a_1, \dots, \sigma(n) = a_n) \\ = & \frac{|\{\sigma \in S_N : \sigma(1) = a_1, \dots, \sigma(n+1) = a_{n+1}\}|/N!}{|\{\sigma \in S_N : \sigma(1) = a_1, \dots, \sigma(n) = a_n\}|/N!} \\ = & \frac{(N-(n+1))!}{(N-n)!} \mathbb{1}_{\{a_{n+1} \notin \{a_1, \dots, a_n\}\}} \\ = & \frac{1}{N-n} \mathbb{1}_{\{a_{n+1} \notin \{a_1, \dots, a_n\}\}} \end{aligned}$$

Notons $m = \min(a_1, \dots, a_n)$. Si $j = m - 1$,

$$\begin{aligned} & \mathbb{P}(Z_{n+1} = j | \sigma(1) = a_1, \dots, \sigma(n) = a_n) \\ = & \mathbb{P}(X_{n+1} > m | \sigma(1) = a_1, \dots, \sigma(n) = a_n) \\ = & \sum_{a > m, a \notin \{a_1, \dots, a_n\}} \mathbb{P}(X_{n+1} = a | \sigma(1) = a_1, \dots, \sigma(n) = a_n) \\ = & \frac{N - m - (n-1)}{N - n} = \frac{N - n - j}{N - n} \end{aligned}$$

et si $j < m - 1$, on a

$$\begin{aligned} & \mathbb{P}(Z_{n+1} = j | \sigma(1) = a_1, \dots, \sigma(n) = a_n) \\ = & \mathbb{P}(X_{n+1} = j + 1 | \sigma(1) = a_1, \dots, \sigma(n) = a_n) \\ = & \frac{1}{N - n}, \end{aligned}$$

ce qui donne l'identité

$$\mathbb{P}(Z_{n+1} = j | \sigma(1), \dots, \sigma(n)) = \frac{1}{N-n} \mathbb{1}_{\{j < Z_n\}} + \frac{N - n - Z_n}{N - n} \mathbb{1}_{\{j = Z_n\}}.$$

La tribu $\mathcal{F}_n$ engendrée par $Z_1, \dots Z_n$ est une sous-tribu de la tribu engendrée par $\sigma(1), \dots \sigma(n)$, donc

$$\begin{aligned} \mathbb{P}[Z_{n+1} = j | \mathcal{F}_n] &= \mathbb{E}[\mathbb{P}(Z_{n+1} = j | \sigma(1), \dots, \sigma(n)) | \mathcal{F}_n] \\ &= \mathbb{E}[\frac{1}{N-n} \mathbb{1}_{\{j < Z_n\}} + \frac{N - n - Z_n}{N - n} \mathbb{1}_{\{j = Z_n\}} | \mathcal{F}_n] \\ &= \frac{1}{N-n} \mathbb{1}_{\{j < Z_n\}} + \frac{N - n - Z_n}{N - n} \mathbb{1}_{\{j = Z_n\}} \\ &= p_{Z_n, j}, \end{aligned}$$

avec $p_{i,j} = \frac{1}{N-n} \mathbb{1}_{\{j<i\}} + \frac{N-n-i}{N-n} \mathbb{1}_{\{j=i\}}$, ce qui montre que $(Z_n)_{n \geq 0}$ est une chaîne de Markov.

3. 0 est l'unique point absorbant de la chaîne $(Z_n)_{n \geq 0}$. $Y_k = Z_{T_k}$ est la suite $(Z_n)_{n \geq 0}$ observée quand elle bouge ou est morte (voir exercice 91). Ainsi, $(Y_k)_{k \geq 0}$ est une chaîne de Markov de matrice $(q_{i,j})$ donnée par

$$q_{i,j} = \begin{cases} \delta_{0,j} & \text{si } i = 0 \\ 0, & \text{si } i = j > 0 \\ \frac{p_{i,j}}{1 - p_{i,i}} & \text{si } i > 0, i \neq j \end{cases},$$

ce qui nous donne ici

$$q_{i,j} = \begin{cases} \delta_{0,j} & \text{si } i = 0 \\ 0, & \text{si } j \geq i > 0 \\ \frac{1}{i} & \text{si } i > 0, j < i \end{cases} .$$

4. La chaîne (Z_n) bouge losqu'elle descend : le nombre de mouvements est donc le nombre de records descendants.

5. Soient $r_1, r_2, \dots r_k$ les k records de σ : on peut associer à σ la permutation $\gamma = (1 = r_1 \dots r_2 - 1)(r_2 \dots r_3 - 1) \dots (r_k \dots N)$.

   On retrouve les r_k, et donc σ, à partir de γ : $r_k = \gamma(1)$, $r_{k-1} = \gamma(r_k)$, $r_{k-2} = \gamma(r_{k-1}), \dots$. Ainsi $\sigma \mapsto \gamma$ est injective.

   Voyons la surjectivité. Je pose $r_k = 1$, $r_{k-1} = \min\{1, \dots N\} \backslash O_\gamma(r_k)$, $r_{k-2} = \min\{1, \dots N\} \backslash (O_\gamma(r_{k-1}) \cup O_\gamma(r_k))$,...À γ on associe la permutation

$$\begin{pmatrix} 12 \dots\dots\dots\dots & \dots\dots\dots\dots & \dots & \dots\dots\dots N \\ r_1\sigma(r_1)\sigma^2(r_1)\dots & r_2\sigma(r_2)\sigma^2(r_2)\dots & \dots & r_k\sigma(r_k)\dots \end{pmatrix}$$

   C'est la réciproque.

6. D'après la question II.3., comme $(X_n)_{n\geq 0}$ a la même loi que $(Y_n))_{n\geq 0}$, F a la même loi que T. D'où avec I.5 :

$$\mathbb{P}(F = k) = \mathbb{P}^N(T = k) = \frac{(-1)^{N+k}}{N!} \begin{bmatrix} N \\ k \end{bmatrix}.$$

   Donc $(-1)^{N+k} \begin{bmatrix} N \\ k \end{bmatrix}$ est le nombre de permutations avec k records, donc d'après II.5 le nombre de permutations possédant exactement k cycles.

D.2 Solution du problème 2

1. Si l'on pose $f_n(x) = x - 1 + (1 - x + U_n)\delta_0(x)$, (f_n) est une suite de fonctions indépendantes identiquement distribuées, indépendantes de X_0 et on a $X_{n+1} = f_{n+1}(X_n)$. $(X_n)_{n\geq 0}$ est donc une chaîne de Markov de loi initiale $\mathbb{P}_{X_0} = \nu$. La loi de X_{n+1} sachant $X_n = x$ est δ_{x-1} si

$x > 0$, μ sinon, ce qui donne la matrice de passage :

$$\begin{pmatrix} \mu(0) & \mu(1) & \mu(2) & \dots & \dots & \dots \\ 1 & 0 & \dots & \dots & \dots & \dots \\ 0 & 1 & \ddots & & & \\ \vdots & \ddots & \ddots & \ddots & & \\ \vdots & & \ddots & \ddots & \ddots & \end{pmatrix}$$

2. Soit $n \in \mathbb{N}^*$. $0 \to x$ car $\mathbb{P}^0(X_1 = n) = \mu(n) > 0$. D'autre part, $n \to 0$ car $\mathbb{P}^n(X_1 = n-1, X_2 = n-2, \dots, X_n = 0) = 1$. La chaîne est donc irréductible. Comme la chaîne est irréductible, elle admet la période de 0 comme période, mais la période de 0 est 1 car $\mathbb{P}^0(X_1 = 0) = \mu(0) > 0$.
3. Notons $A = \{\exists k \geq 1; X_k = 0\}$. On a vu que pour tout $n > 0$, $\mathbb{P}^n(X_1 = n-1, X_2 = n-2, \dots, X_n = 0) = 1$ On en déduit que pour tout $n > 0$ $\mathbb{P}^n(A) = 1$. Maintenant

$$\begin{aligned} \mathbb{P}^0(A) &= \sum_{j=0}^{+\infty} \mathbb{P}^0(X_1 = j; \exists k \geq 1; X_k = 0) \\ &= \mathbb{P}^0(X_1 = 0) + \sum_{j=1}^{+\infty} \mathbb{P}^0(X_1 = j; \exists k \geq 2; X_k = 0) \\ &= \mathbb{P}^0(X_1 = 0) + \sum_{j=1}^{+\infty} \mathbb{P}^0(X_1 = j)\mathbb{P}^j(A) \end{aligned}$$

Mais on a vu que pour tout $j > 0$, $\mathbb{P}^j(A) = 1$, donc finalement

$$\mathbb{P}^0(A) = \sum_{j=0}^{+\infty} \mathbb{P}^0(X_1 = j) = 1,$$

ce qui montre que 0 est récurrent. Comme la chaîne est irréductible, la chaîne est donc récurrente. Elle n'est donc évidemment pas transiente.
4. Pour tout $j \notin \{0, k+1\}$,

$$\mathbb{P}^\nu(X_0 = j, X_1 = k) = \nu(j)\mathbb{P}^j(X_1 = k) = \nu_j \delta_{j-1}(k) = 0.$$

On a donc

$$\begin{aligned} \mathbb{P}^\nu(X_1 = k) &= \mathbb{P}^\nu(X_0 = 0, X_1 = k) + \mathbb{P}^\nu(X_0 = k+1, X_1 = k) \\ &= \nu(0)\mathbb{P}^0(X_1 = K) + \nu(k+1)\mathbb{P}^{k+1}(X_1 = k) \\ &= \nu(0)\mu(k) + \nu(k+1) \end{aligned}$$

5. La chaîne est stationnaire si et seulement si $P^{\nu}_{X_1} = \nu$. D'après la question précédente, ν est stationnaire si et seulement si
$$\forall k \in \mathbb{N} \quad \nu(k) = \nu(0)\mu(k) + \nu(k+1).$$
Si la chaîne était stationnaire avec $\nu(0) = 0$, on aurait
$$\forall k \in \mathbb{N} \quad \nu(k) = \nu(k+1),$$
ce qui donne par récurrence $\forall k \in \mathbb{N} \quad \nu(k) = 0$: contradiction.
6. En fait sous $\mathbb{P}^0$, on a $T_0 = 1 + X_1$. En effet, si $X_1 = 0$, on a $T_0 = 1 = 1 + X_1$. D'autre part si $X_1 = k$, avec $k > 0$, on a : $X_2 = k-1, X_3 = k-2, \dots, K_k = k-(k-1) = 1, X_{k+1} = k-k = 0$, d'où $T_0 = k+1 = X_0 + 1$.
7. Comme la chaîne de Markov est irréductible, si elle admet une probabilité invariante ν, on a nécessairement $\nu(0) = \frac{1}{\mathbb{E}^0 T_0}$. Mais on a vu que si la chaîne admettait une probabilité invariante ν, on aurait nécessairement $\nu(0) > 0$. Cela impose que $\mathbb{E}^0 T_0 < +\infty$. Mais comme on l'a vu, $1 + X_1$ et T_0 ont même loi sous P^0. Donc $(\mathbb{E}^0 T_0 < +\infty) \iff (\mathbb{E}^0 1 + X_1) < +\infty) \iff \mathbb{E}^0 X_1 < +\infty$, ce qui signifie que μ admet un moment d'ordre 1. D'autre part, on a bien
$$\nu(0) = \frac{1}{\mathbb{E}^0 T_0} = \frac{1}{1 + \mathbb{E}X_1}.$$
8. Considérons la mesure ν définie par
$$\nu(k) = \frac{\mathbb{P}(X_1 \geq k)}{1 + \mathbb{E}X_1}.$$
Montrons que ν est une probabilité : elle est évidemment positive et
$$\begin{aligned}\sum_{k \geq 0} \nu(k) &= \frac{1}{1 + \mathbb{E}X_1} \sum_{k=0}^{+\infty} \mathbb{P}(X_1 \geq k) \\ &= \frac{1}{1 + \mathbb{E}X_1} \sum_{k=0}^{+\infty} \mathbb{P}(1 + X_1 > k) \\ &= \frac{1}{1 + \mathbb{E}X_1} \mathbb{E}(1 + X_1) \\ &= 1\end{aligned}$$
Il reste à vérifier que
$$\forall k \in \mathbb{N} \quad \nu(k) = \nu(0)\mu(k) + \nu(k+1),$$

c'est à dire que

$$\forall k \in \mathbb{N} \quad \nu(k) = \frac{1}{\mathbb{E}(1+X_1)}\mu(k) + \nu(k+1),$$

en multipliant par $\mathbb{E}(1+X_1)$, on se ramène à vérifier la condition équivalente :

$$\forall k \in \mathbb{N} \quad \mathbb{P}(X_1 \geq k) = \mu(k) + \mathbb{P}(X_1 \geq k+1),$$

soit

$$\forall k \in \mathbb{N} \quad \mathbb{P}(X_1 \geq k) = \mathbb{P}(X_1 = k) + \mathbb{P}(X_1 \geq k+1),$$

ce qui est évident car X_1 est à valeurs entières.

9. Si μ a un moment d'ordre 1, on a vu que la chaîne de Markov est irréductible, de loi invariante ν, et on peut appliquer le théorème ergodique des chaînes de Markov : $\frac{X_1+\dots+X_n}{n}$ converge presque sûrement (quelque soit la loi initiale) vers $\int_{\mathbb{R}_+} x \, d\nu(x)$.

10. Si $X_n = 0$, on a $T_0 \leq n$. Ainsi les événements $(\{T_0 = k\} \cap \{X_n = 0\})_{1 \leq k \leq n}$ forment une partition de $\{X_n = 0\}$. On a donc

$$\begin{aligned}
\mathbb{P}^0(X_n = j) &= \sum_{k=1}^{n} \mathbb{P}^0(T_0 = k, X_n = 0) \\
&= \sum_{k=1}^{n} \mathbb{P}^0(T_0 = k, X_n = j) \\
&= \sum_{k=1}^{n} \mathbb{P}^0(T_0 = k)\mathbb{P}^j(X_{n-k} = 0) (\text{propriété de Markov forte})
\end{aligned}$$

On applique ici la propriété de Markov forte avec le temps d'arrêt T_0.

11. Préliminaire : si $f(z) = \sum_{n=0}^{+\infty} a_n z^n$ est une série entière de rayon de convergence $\geq R$, alors pour tout $r \in]0, R[$ et tout entier naturel k, on a

$$a_k = r^{-k}\frac{1}{2\pi}\int_0^{2\pi} f(re^{i\theta})e^{-ik\theta} \, d\theta.$$

Cela peut être vu comme une version simple de la formule de Cauchy, mais on peut aussi le voir simplement à partir de l'identité

$$f(re^{i\theta})e^{-ik\theta} = \sum_{n=0}^{+\infty} a_n r^n e^{i(n-k)\theta}.$$

En effet

$$|\sum_{n=0}^{N} r^n e^{i(n-k)\theta}| \leq \sum_{n=0}^{+\infty} |a_n| r^n$$

et on peut appliquer le théorème de convergence dominée, puisque la fonction constante $\sum_{n=0}^{+\infty} |a_n| r^n$ est bien sûr intégrable sur $[0, 2\pi]$.
La fonction génératrice de X_1 est P, donc par indépendance, la fonction génératrice de S_n est $P(z)^n$, soit

$$P(z)^n = \sum_{k=0}^{+\infty} \mathbb{P}(S_n = k) z^k,$$

avec la remarque précédente, pour $r \in]0, 1[$ quelconque, on a

$$\mathbb{P}(S_n = k) = r^{-k} \frac{1}{2\pi} \int_0^{2\pi} P(re^{i\theta})^n e^{-ik\theta} \, d\theta.$$

Avec l'inégalité triangulaire, on a

$$|P(re^{i\theta})| \leq \sum_{n=0}^{+\infty} p_n r^n \leq \sum_{n=0}^{+\infty} p_n 1^n = 1,$$

car $p_n r^n \leq p_n 1^n$ pour tout n, mais il existe au moins un $n \geq 1$ tel que $p_n \neq 0$ et donc $p_n r^n < p_n 1^n$: finalement $\sum_{n=0}^{+\infty} p_n r^n = P(r) < 1$.
On peut donc écrire :

$$\sum_{n=0}^{N-1} \mathbb{P}(S_n = k) = r^{-k} \frac{1}{2\pi} \int_0^{2\pi} \frac{1 - P(re^{i\theta})^N}{1 - P(re^{i\theta})} e^{-ik\theta} \, d\theta.$$

Il n'y a pas de problème d'intégrabilité : $|1 - P(re^{i\theta})| \geq 1 - |P(re^{i\theta})| \geq 1 - P(r) > 0$. En faisant tendre N vers l'infini (par exemple convergence dominée, ou à la main), on obtient

$$\begin{aligned}
\sum_{n=0}^{+\infty} \mathbb{P}(S_n = k) &= r^{-k} \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{1 - P(re^{i\theta})} e^{-ik\theta} \, d\theta \\
&= r^{-k} \frac{1}{2\pi} \int_0^{2\pi} Q(re^{i\theta}) e^{-ik\theta} \, d\theta \\
&= q_k.
\end{aligned}$$

Avec Tonelli, on a alors

$$\mathbb{E}[N_k] = \sum_{n=1}^{+\infty} \mathbb{E}[\mathbb{1}_{\{S_n = k\}}] = \sum_{n=1}^{+\infty} \mathbb{P}(S_n = k) = q_k.$$

Reste à étudier le comportement asymptotique. On s'intéresse d'abord au cas où $p_0 = 0$.

Posons pour $n \in \mathbb{N}$, $\mu(n) = p_{n+1}$. On a $\text{pgcd}(n \geq 1; \mu(n-1) > 0) = \text{pgcd}(n \geq 1; p_n > 0) = 1$, donc on peut appliquer les questions précédentes. On avait remarqué que la suite (p'_k) vérifie $p'_k = \mu(k-1)$, donc on a en fait $p'_k = p_k$ pour tout k. Montrons que $q'_n = q_n$ pour tout n : on a l'identité $(1-P(z))Q(z) = 1$, soit $Q(z) = 1 + P(z)Q(z)$. En identifiant les coefficients (produit de Cauchy), on a

$$q_n = \delta_0(n) + \sum_{k=0}^{n} p_k q_{n-k}.$$

$q_0 = Q(0) = \frac{1}{1-P(0)} = \frac{1}{1-p_0} = \frac{1}{1-0} = 1$. Pour $n \geq 1$, on a

$$q_n = \sum_{k=1}^{n} p_k q_{n-k}$$

Comme

$$q'_n = \sum_{k=1}^{n} p'_k q'_{n-k} = \sum_{k=1}^{n} p_k q'_{n-k},$$

il est alors facile de montrer par récurrence que $q'_n = q_n$ pour tout n. Or, d'après le théorème de convergence des chaînes de Markov irréductibles,

$$\lim_{n \to +\infty} q'_n = \nu(0) = \frac{1}{1+\mathbb{E}[U_1]} = \frac{1}{\mathbb{E}[(1+U_1)]} = \frac{1}{\mathbb{E}[X_1]},$$

ce qui démontre le théorème de Erdös, Feller et Pollard dans le cas où $p_0 = 0$.

Passons au cas général. On pose $\tilde{p}_0 = 0$ et $\tilde{p}_k = \frac{p_k}{1-p_0}$ pour $k \geq 1$. Les $(\tilde{p}_k)$ définissent bien une probabilité sur $\mathbb{N}$, qui vérifie les hypothèses de la partie précédente : si on pose

$$\forall z \in B_F(0,1) \quad \tilde{P}(z) = \sum_{n=0}^{+\infty} \tilde{p}_n z^n \text{ et } \forall z \in B(0,1) \quad \tilde{Q}(z) = \frac{1}{1-\tilde{P}(z)} = \sum_{n=0}^{+\infty} \tilde{q}_n z^n,$$

on a $\lim_{k \to +\infty} \tilde{q}_k = \frac{1}{\mathbb{E}[\tilde{X}_1]}$, avec $\mathbb{P}(\tilde{X}_1 = k) = \tilde{p}_k$. On a facilement $\mathbb{E}[\tilde{X}_1] = \frac{\mathbb{E}[X_1]}{1-p_0}$, d'où $\lim_{k \to +\infty} \tilde{q}_k = \frac{1-p_0}{\mathbb{E}[X_1]}$ Cependant $P(z) = p_0 + (1+p_0)\tilde{P}(z)$, d'où $Q(z) = \frac{1}{1-p_0}\tilde{Q}(z)$, ce qui entraîne $q_n = \frac{1}{1-p_0}\tilde{q}_n$. Finalement, on a bien $\lim_{k \to +\infty} q_k = \frac{1}{\mathbb{E}[X_1]}$.

D.3 Solution du problème 3

1. Notons Π_n la projection canonique de Ω sur $\mathbb{R}^n$.

   Supposons que $\mathbb{P}_X$ est échangeable. Soit $n \geq 1$ et $\sigma \in \mathcal{S}_n$ quelconque. On a $(X_{\sigma(1)}, \dots, X_{\sigma(n)}) = \Pi_n \circ T_\sigma$, donc $\mathbb{P}_{(X_{\sigma(1)},\dots,X_{\sigma(n)})}$ est la loi image de $\mathbb{P}_{T_\sigma}$ par Π_n. Comme $\mathbb{P}$ est échangeable, c'est la loi image de $\mathbb{P}$ par Π_n, c'est à dire $\mathbb{P}_{(X_1,\dots,X_n)}$. Comme n et $\sigma \in \mathcal{S}_n$ sont quelconques, on a bien montré que la famille $(X_n)_{n\geq 1}$ est échangeable.

   Réciproquement, supposons que la famille $(X_n)_{n\geq 1}$ est échangeable. On sait bien que pour identifier les probabilités $\mathbb{P}_{T_\sigma}$ et $\mathbb{P}$, il suffit de les identifier sur les événements cylindriques, $(X_1, \dots, X_n) \in A$, où n décrit $\mathbb{N}^*$ et A l'ensemble des boréliens de $\mathbb{R}^n$. Or $\mathbb{P}((X_1, \dots, X_n) \in A) = \mathbb{P}_{(X_1,\dots,X_n)}(A)$ et

$$\begin{aligned}
\mathbb{P}_{T_\sigma}((X_1, \dots, X_n) \in A) &= \mathbb{P}((X_1, \dots, X_n) \circ T_\sigma \in A) \\
&= \mathbb{P}((X_{\sigma(1)}, \dots, X_{\sigma(n)}) \in A) \\
&= \mathbb{P}_{(X_{\sigma(1)},\dots,X_{\sigma(n)})}(A),
\end{aligned}$$

   et $\mathbb{P}_{(X_1,\dots,X_n)}$ et $\mathbb{P}_{(X_{\sigma(1)},\dots,X_{\sigma(n)})}$ coïncident, d'après l'hypothèse d'échangeabilité de la famille $(X_n)_{n\geq 1}$. Cela montre le résultat voulu.

2. (a) C'est une conséquence de la définition de $\mathcal{I}_n$ et de l'inclusion $\mathcal{S}_n \subset \mathcal{S}_{n+1}$.

   (b) Soit $B \in \mathcal{B}(\mathbb{R})$; pour tout $\sigma \in \mathcal{S}_n$, on a $\sigma^{-1} \in \mathcal{S}_n$, avec $T_\sigma^{-1} = T_{\sigma^{-1}}$. On a alors

$$F^{-1}(B) = (F \circ T_{\sigma^{-1}})^{-1}(B) = T_{\sigma^{-1}}^{-1}(F^{-1}(B)) = T_\sigma(F^{-1}(B)).$$

   Comme σ peut être pris quelconque dans $\mathcal{S}_n$, cela montre que $F^{-1}(B) \in \mathcal{I}_n$. Comme B est un borélien quelconque, F est $\mathcal{I}_n$-mesurable.

   (c) Soit σ la transposition qui échange les points 1 et k. Comme $\mathbb{P}$ est invariante par T_σ,

$$\mathbb{E}(1_A f(X_1)) = \mathbb{E}(1_A f(X_1) \circ T_\sigma) = \mathbb{E}(\mathbb{1}_A \circ T_\sigma . f(X_1) \circ T_\sigma).$$

   Or $f(X_1) \circ T_\sigma = f \circ X_1 \circ T_\sigma = f \circ X_k = f(X_k)$ et $\mathbb{1}_A \circ T_\sigma = \mathbb{1}_{T_\sigma^{-1}(A)} = \mathbb{1}_A$ car $A \in \mathcal{I}_n$. Finalement, $\mathbb{E}(1_A f(X_1)) = \mathbb{E}(\mathbb{1}_A f(X_k))$. Soit i entre 1 et n : on a $\mathbb{E}(1_A f(X_i)) = \mathbb{E}(1_A f(X_1)) = \mathbb{E}(\mathbb{1}_A f(X_k))$. En sommant pour k variant de 1 à n et en divisant par n, on a

$$\mathbb{E}(\mathbb{1}_A f(X_i)) = \mathbb{E}(\mathbb{1}_A \frac{1}{n}\sum_{k=1}^n f(X_k)).$$

Ainsi, si l'on pose $F = \frac{1}{n}\sum_{k=1}^n f(X_k)$,on a $\mathbb{E}(\mathbb{1}_A f(X_i)) = \mathbb{E}(\mathbb{1}_A F)$ pour tout $A \in \mathcal{I}_n$. Si on montre que F est $\mathcal{I}_n$-mesurable, on saura alors que F est un représentant de l'espérance conditionnelle de $f(X_i)$ sachant $\mathcal{I}_n$.

Pour $\sigma \in \mathcal{S}_n$ et k quelconque, on a

$$X_k \circ T_\sigma = X_{\sigma^{-1}(k)} \tag{D.1}$$

Notons cette formule, qui resservira. Maintenant,

$$F \circ T_\sigma = \frac{1}{n}\sum_{k=1}^n f(X_k)\circ T_\sigma = \frac{1}{n}\sum_{k=1}^n f(X_{\sigma^{-1}(k)}) = \frac{1}{n}\sum_{k=1}^n f(X_k) = F.$$

En effet, un élément de $\mathcal{S}_n$ induit en restriction une permutation de $\{1,\dots,n\}$. Le résultat découle alors de (2b).

(d) Soit i un entier naturel fixé. Comme la suite $(\mathcal{I}_n)_{n\geq 1}$, la suite $\mathbb{E}(f(X_i)|\mathcal{I}_n)$, le théorème de convergence des martingales inverses entraîne que la suite $\mathbb{E}(f(X_i)|\mathcal{I}_n)$ converge presque sûrement vers $\mathbb{E}(f(X_i)|\mathcal{I})$. Or pour $n \geq i$, $\mathbb{E}(f(X_i)|\mathcal{I}_n) = \frac{1}{n}\sum_{k=1}^n f(X_k)$, donc la suite $(\frac{1}{n}\sum_{k=1}^n f(X_k))_{n\geq 1}$ converge presque sûrement vers $\mathbb{E}(f(X_i)|\mathcal{I})$. Par unicité de la limite presque sûre, on a pour tout i $\mathbb{E}(f(X_i)|\mathcal{I}) = \mathbb{E}(f(X_1)|\mathcal{I})$ presque sûrement.

(e) Soit $A \in \mathcal{I}_n$ et $g \in I_{n,p}$. Soit σ une permutation de $\{1,\dots,n\}$, que l'on étend en un élément de $\mathcal{S}_n$. On a

$$(\prod_{i=1}^p f_i(X_{g(i)}))\circ T_\sigma = \prod_{i=1}^p f_i(X_{\sigma^{-1}(g(i))})).$$

On peut construire une permutation σ de $\{1,\dots,n\}$ qui envoie 1 sur $g(1)$, 2 sur $g(2)$,...p sur $g(p)$. Dans ce cas, on a l'identité

$$(\prod_{i=1}^p f_i(X_{g(i)}))\circ T_\sigma = \prod_{i=1}^p f_i(X_i)).$$

Comme $\mathbb{1}_A \circ T_\sigma = \mathbb{1}_A$, en multipliant on a

$$(\mathbb{1}_A \prod_{i=1}^p f_i(X_{g(i)}))\circ T_\sigma = \mathbb{1}_A \prod_{i=1}^p f_i(X_i)),$$

soit, en utilisant le théorème de transfert et l'invariance de $\mathbb{P}$ par T_σ :

$$\mathbb{E}(\mathbb{1}_A \prod_{i=1}^{p} f_i(X_{g(i)})) = \mathbb{E}(\mathbb{1}_A \prod_{i=1}^{p} f_i(X_i)).$$

En sommant pour g variant dans $I_{n,p}$ et en divisant par $|I_{n,p}|$, on obtient

$$\mathbb{E}(\mathbb{1}_A G) = \mathbb{E}(\mathbb{1}_A \prod_{i=1}^{p} f_i(X_i)),$$

où

$$G = \frac{1}{|I_{n,p}|} \sum_{g\in I_{n,p}} \prod_{i=1}^{p} f_i(X_{g(i)}),$$

Comme A est quelconque dans $\mathcal{I}_n$, comme précédémment, pour montrer que G est un représentant de l'espérance conditionnelle de $\prod_{i=1}^{p} f_i(X_{g(i)})$ sachant $\mathcal{I}_n$, il suffit de montrer que G est $\mathcal{I}_n$-mesurable. D'après (2b), il suffit de montrer que $G = G \circ T_\sigma$ pour tout $\sigma \in \mathcal{S}_n$. Or pour $\sigma \in \mathcal{S}_n$, on a

$$F \circ T_\sigma = \frac{1}{|I_{n,p}|} \sum_{g\in I_{n,p}} \prod_{i=1}^{p} f_i(X_{\sigma^{-1}(g(i))}) = \frac{1}{|I_{n,p}|} \sum_{g\in I_{n,p}} \prod_{i=1}^{p} f_i(X_{(\sigma^{-1}\circ g)(i)}),$$

Or $g \mapsto \sigma^{-1} g \circ g$ réalise une bijection de $I_{n,p}$ dans lui-même, donc

$$\frac{1}{|I_{n,p}|} \sum_{g\in I_{n,p}} \prod_{i=1}^{p} f_i(X_{(\sigma^{-1}\circ g)(i)}) = \frac{1}{|I_{n,p}|} \sum_{g\in I_{n,p}} \prod_{i=1}^{p} f_i(X_i) = F,$$

ce qui donne donc l'identité voulue. D'après le théorème des martingales inverses, la suite $(\mathbb{E}(\prod_{i=1}^{p} f_i(X_i)|\mathcal{I}_n))_{n\geq 1}$ converge presque sûrement vers $\mathbb{E}(\prod_{i=1}^{p} f_i(X_i)|\mathcal{I})$ lorsque n tend vers l'infini.

(f) Soit $M = \max(\|f_1\|_\infty, \dots, \|f_p\|_\infty)$. On pose $A_n = \sum_{g\in I_{n,p}} \prod_{i=1}^{p} f_i(X_{g(i)})$ et $B_n = \sum_{g\in\{1,\dots,n\}^p} \prod_{i=1}^{p} f_i(X_{g(i)})$.

$$B_n - A_n = \sum_{g\in\{1,\dots,n\}^p\setminus I_{n,p}} \prod_{i=1}^{p} f_i(X_{g(i)}),$$

donc

$$|B_n - A_n| \leq \sum_{g\in\{1,\dots,n\}^p\setminus I_{n,p}} M^p = (n^p - |I_{n,p}|)M^p = o(n^p),$$

car $|I_{n,p}| = n(n-1)\dots(n-p+1) \sim n^p$. Ainsi $\frac{A_n}{n^p} = \frac{B_n}{n^p} + o(1)$, et comme $\frac{n^p}{|I_{n,p}|} = 1+o(1)$, on a $\frac{A_n}{|I_{n,p}|} = \frac{B_n}{n^p}+o(1)$ Comme $\frac{A_n}{|I_{n,p}|}$ converge presque sûrement vers $\mathbb{E}\left(\prod_{i=1}^p f_i(X_i)|\mathcal{I}\right)$, il en est de même pour $\frac{B_n}{n^p}$. Or, on peut réécrire

$$\frac{B_n}{n^p} = \sum\nolimits_{g\in\{1,\dots,n\}^p} \prod_{i=1}^p (f_i(X_{g(i)})/n) = \prod\nolimits_{i=1}^p \sum_{j=1}^n (f_i(X_j)/n).$$

On a utilisé ici la formule de développement d'un produit de sommes en somme de produits.

Pour i quelconque entre 1 et p, on sait d'après (2d) que $\sum_{j=1}^n (f_i(X_j)/n) = \frac{1}{n}\sum_{j=1}^n f_i(X_j)$ converge presque sûrement vers $\mathbb{E}(f_i(X_1)|\mathcal{I})$. Finalement, $\frac{B_n}{n^p}$ converge vers $\prod_{i=1}^p \mathbb{E}(f_i(X_1)|\mathcal{I})$, ce qui donne le résultat voulu.

(g) On a vu en (2d) que pour f quelconque bornée $\mathbb{E}(f(X_i)|\mathcal{I}) = \mathbb{E}(f(X_1)|\mathcal{I})$. C'est vrai en particulier pour $f = f_i$, donc la limite de $\frac{B_n}{n^p}$ se réécrit $\prod_{i=1}^p \mathbb{E}(f_i(X_1)|\mathcal{I})$. Comme c'est aussi la limite de $\frac{A_n}{|I_{n,p}|}$ qui est $\mathbb{E}\left(\prod_{i=1}^p f_i(X_i)|\mathcal{I}\right)$, les deux expressions coïncident presque sûrement.

(h) Soit $\mathcal{E}$ l'ensemble des $A \in \mathcal{B}(\Omega)$ tels que $A \mapsto M_\omega(A)$ est mesurable, avec $\mathbb{P}(A) = \int M_\omega(A)\ d\mathbb{P}(\omega)$.

— $\varnothing \in \mathcal{E}$, car $\omega \mapsto M_\omega(\varnothing)$ est identiquement nulle, donc mesurable et on a évidemment $0 = \int 0\ d\mathbb{P}$.

— Si $A \in \mathcal{E}$, $M_\omega(A^c) = 1 - M_\omega(A)$, donc $\omega \mapsto M_\omega(A^c)$ est mesurable, et $\int M_\omega(A^c)\ d\mathbb{P} = \int(1-M_\omega)\ d\mathbb{P}(\omega) = \int 1\ d\mathbb{P} - \int M_\omega(A)\ d\mathbb{P}(\omega) = 1-\mathbb{P}(A) = \mathbb{P}(A^c)$, donc $A^c \in \mathcal{E}$.

— Soient $(A_n)_{n\geq 1}$, une suite d'éléments de $\mathcal{E}$ deux à deux disjoints. Si on pose $A = \cup_{n\geq 1} A_i$, on a pour tout ω a $M_\omega(A) = \lim_{n\to+\infty} \sum_{k=1}^n M_\omega(A_k)$, et donc $\omega \mapsto M_\omega(A)$ est mesurable comme limite d'applications mesurables. Par convergence monotone, il vient

$$\int M_\omega(A)\ d\mathbb{P} = \lim_{n\to+\infty} \int \sum_{k=1}^n M_\omega(A_k)\ d\mathbb{P}(\omega).$$

Or

$$\int \sum_{k=1}^n M_\omega(A_k)\ d\mathbb{P}(\omega) = \sum_{k=1}^n \int M_\omega(A_k)\ d\mathbb{P}(\omega) = \sum_{k=1}^n \mathbb{P}(A_k),$$

d'où

$$\int M_\omega(A)\ d\mathbb{P} = \lim_{n\to+\infty} \sum_{k=1}^n \mathbb{P}(A_k) = \sum_{k=1}^{+\infty} \mathbb{P}(A_k).$$

Ainsi $A \in \mathcal{E}$.

On a donc montré que $\mathcal{E}$ est une tribu. Montrons que $\mathcal{E}$ contient les cylindres $C = \times_{i=1}^p B_i \otimes \mathbb{R}^{\mathbb{N}^*\setminus\{1,\dots,p\}}$, où les B_i sont des boréliens de $\mathbb{R}$. Posons $f_i = \mathbb{1}_{B_i}$. On a

$$\begin{aligned}
\mathbb{P}(C) &= \mathbb{E}(\prod_{i=1}^p f_i(X_i)) = \mathbb{E}(\mathbb{E}(\prod_{i=1}^p f_i(X_i)|\mathcal{I})) \\
&= \int \mathbb{E}(\prod_{i=1}^p f_i(X_i)|\mathcal{I})\ d\mathbb{P} \\
&= \int \prod_{i=1}^p \int \mathbb{E}(f_i(X_i)|\mathcal{I})\ d\mathbb{P}, \text{ avec (2g).} \\
&= \int \prod_{i=1}^p \int \mathbb{E}(f_i(X_1)|\mathcal{I})\ d\mathbb{P}, \text{ avec (2d).} \\
&= \int \prod_{i=1}^p \int \mathbb{E}(\mathbb{1}_{X_1^{-1}(B_i)}|\mathcal{I})\ d\mathbb{P},
\end{aligned}$$

Cependant, par définition de C et de la mesure produit infini, on a pour tout ω, $M_\Omega(C) = \prod_{i=1}^p (\mathbb{P}_\omega^{\mathcal{I}})_{X_1}(B_i) = \prod_{i=1}^p (\mathbb{P}_\omega^{\mathcal{I}})(X_1^{-1}(B_i))$ Mais on sait que $\mathbb{P}_\omega^{\mathcal{I}}(X_1^{-1}(B_i))$ est un représentant de $\mathbb{E}(\mathbb{1}_{X_1^{-1}(B_i)}|\mathcal{I})$, donc

$$M_\Omega(C) = \prod_{i=1}^p (\mathbb{P}_\omega^{\mathcal{I}})(X_1^{-1}(B_i))$$

est un représentant de $\prod_{i=1}^p \int \mathbb{E}(\mathbb{1}_{X_1^{-1}(B_i)}|\mathcal{I})$. Comme $\omega \mapsto \mathbb{P}_\omega^{\mathcal{I}})(X_1^{-1}(B_i))$ est mesurable pour tout i, $\omega \mapsto M_\omega(E)$ est bien mesurable comme produit d'applications mesurables, et il découle de ce qui précède que

$$\mathbb{P}(C) = \int M_\Omega(C)\ d\mathbb{P}(\omega),$$

ce qui montre bien que $C \in \mathcal{E}$. $\mathcal{E}$ est une tribu qui contient les événements cylindriques : c'est $\mathcal{B}(\Omega)$ tout entier.

3. (a) D'après la quesion 1), il suffit de montrer que la suite $(T_i)_{i\geq 1}$ est échangeable. Soit $n \geq 1$ et σ une permutation de $\{1,\dots,n\}$. On a $\mathbb{P}(X_{\sigma(1)} = t_1, \dots, X_{\sigma(n)} = t_1) = \mathbb{P}(X_1 = t'_1, \dots, X_n = t'_n)$, avec $t'_i = t_{\sigma^{-1}(i)}$. Le nombre a_i d'apparition de i dans la suite $t_1, \dots, t_n$ coïncide avec le nombre a'_i d'apparition de i dans la suite $t'_1, \dots, t'_n$, donc d'après la formule (3.2) ou la formule (3.3), les probabilités $\mathbb{P}(X_1 = t_1, \dots, X_n = t_n)$ et $\mathbb{P}(X_1 = t'_1, \dots, X_n = t'_n)$ coïncident, ce qui donne le résultat voulu.

(b) Fixons $t_1, \dots t_p$. Avec (2d), $\mathbb{P}(X_i = t_i|\mathcal{I})$ est la limite presque sûre de
$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\{X_k = t_i\}}.$$
Or
$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\{X_k = t_i\}} = \frac{(V_n^{t_i} - V_0^{t_i})/S}{n},$$
qui converge presque sûrement vers W_{t_i} d'après la question 1.(e) de l'exercice 24 sur l'urne de Pólya. Ainsi $\mathbb{P}(X_i = t_i|\mathcal{I}) = W_{t_i}$. En particulier, W_{t_i} est $\mathcal{I}$-mesurable. Maintenant, avec (2g),
$$\mathbb{P}(T_1 = t_1, \dots, T_p = t_p|\mathcal{I}) = \prod_{i=1}^{p} \mathbb{P}(X_i = t_i|\mathcal{I}) = \prod_{i=1}^{p} W_{t_i},$$
ce qui montre bien que, conditionnellement à $\mathcal{I}$, les T_i sont des variables aléatoires indépendantes dont la loi discrète donnée par le vecteur W. Comme W est $\mathcal{I}$-mesurable, $\sigma(W) \subset \mathcal{I}$, et en reconditionnant, on a la formule
$$\mathbb{P}(T_1 = t_1, \dots, T_p = t_p|W) = \prod_{i=1}^{p} W_{t_i},$$
qui avait été annoncée en commentaire de l'exercice sur l'urne de Pólya.

Bibliographie

[1] M. Briane and G. Pagès. *Théorie de l'intégration.* Ellipses, Vuibert, 2009.

[2] O. Garet and A. Kurtzmann. *De l'intégration aux probabilités.* Ellipses, Paris, 2011.

[3] Francis Hirsch and Gilles Lacombe. *Éléments d'analyse fonctionnelle.* Enseignement des Mathématiques. Masson, Paris, 1997. Cours et exercices.

[4] W. Rudin. *Analyse réelle et complexe.* Masson, Paris, 1980.

[5] Daniel W. Stroock. *Probability theory.* Cambridge University Press, Cambridge, second edition, 2011. An analytic view.

Index