Année universitaire 2008-2009

UNIVERSITÉ DE NANCY 1

Olivier GARET

Probabilités et Modélisation Stochastique

(Master 1ère année semestre 1)

Table des matières

Chapitre 1

Sommes de variables aléatoires indépendantes

1.1 Théorèmes de Lindeberg et Lyapounov

1.1.1 Théorème de Lindeberg

Théorème 1. *Soit* $(X_{n,k})_{n,k\geq 1}$ *des variables aléatoires centrées admettant un moment d'ordre 2,* $(N_n)_{n\geq 1}$ *un suite d'entiers tendant vers l'infini. On suppose que pour tout* n*, la suite* $(X_{n,k})_{1\leq k\leq N_n}$ *est formées par des variables indépendantes. On pose*

$$S_n = \sum_{k=1}^{N_n} X_{n,k}$$

et $s_n^2 = \operatorname{Var} S_n$*. On suppose que*

$$\forall \varepsilon > 0 \qquad \lim_{n\to+\infty} \sum_{k=1}^{N_n} \frac{1}{s_n^2} \mathbb{E} X_{n,k}^2 \mathbb{1}_{\{|X_{n,k}|\geq \varepsilon s_n\}} = 0 \quad \textit{(Condition de Lindeberg).}$$

alors S_n/s_n converge en loi vers $\mathcal{N}(0,1)$.

Démonstration. Notons $\varphi_{n,k}$ la fonction caractéristique de $\frac{X_{n,k}}{s_n}$: la fonction caractéristique de S_n/s_n est $\prod_{k=1}^{N_n} \varphi_{n,k}$. L'idée est bien sûr de montrer que pout tout t réel $\prod_{k=1}^{N_n} \varphi_{n,k}(t)$ converge vers $\exp(-\frac{t^2}{2})$ et d'appliquer le théorème de Lévy. Notons $\psi_{n,k}$ la fonction caractéristique d'une variable gaussienne qui a même espérance et même variance que $\frac{X_{n,k}}{s_n}$, soit

$\psi_{n,k}(t)=\exp(-\frac{t^2}{2}\frac{\operatorname{Var} X_{n,k}}{s_n^2})$: on a

$$\prod_{k=1}^{N_n}\varphi_{n,k}(t)-\exp(-\frac{t^2}{2})=\prod_{k=1}^{N_n}\varphi_{n,k}(t)-\prod_{k=1}^{N_n}\psi_{n,k}(t).$$

On va maintenant procéder comme dans la preuve du théorème central limite : la différence de deux produits de nombres complexes dont les modules ne dépassent pas un n'excède pas la somme des différences[1], d'où :

$$|\prod_{k=1}^{N_n}\varphi_{n,k}(t)-\exp(-\frac{t^2}{2})|\le\sum_{k=1}^{N_n}|\varphi_{n,k}(t)-\psi_{n,k}(t)|.$$

Soit α un réel positif[2] tel que

$$\forall x\in\mathbb{R}\quad |e^{ix}-(1+ix-x^2/2)|\le\alpha\min(x^2,|x|^3).$$

Appliquons cette inégalité à $x=\frac{tX_{n,k}}{s_n}$, puis prenons l'espérance : on obtient alors que

$$|\varphi_{n,k}(t)-(1-\frac{1}{2}t^2\frac{\operatorname{Var} X_{n,k}}{s_n^2})|\le\alpha\mathbb{E}Y,$$

où

$$Y=\min\left(t^2\left(\frac{X_{n,k}}{s_n}\right)^2,t^3\left(\frac{|X_{n,k}|}{s_n}\right)^3\right).$$

On peut écrire

$$\mathbb{E}Y=\mathbb{E}Y1\!\!1_{\{|X_{n,k}|/s_n\le\varepsilon\}}+\mathbb{E}Y1\!\!1_{\{|X_{n,k}|/s_n>\varepsilon\}}$$

Lorsque $\frac{X_{n,k}}{s_n}$ est petit, on a intérêt à l'élever à une plus grande puissance, donc on écrit

$$\begin{aligned}\mathbb{E}Y &\le \mathbb{E}|t|^3\frac{|X_{n,k}|^3}{s_n^3}1\!\!1_{\{|\frac{X_{n,k}}{s_n}\le\varepsilon\}}+\mathbb{E}t^2X_{n,k}^2\frac{1}{s_n^2}1\!\!1_{\{|\frac{X_{n,k}}{s_n}\ge\varepsilon\}}\\ &\le \varepsilon|t|^3\frac{\operatorname{Var} X_{n,k}}{s_n^2}+t^2\frac{1}{s_n^2}\mathbb{E}X_{n,k}^21\!\!1_{\{|X_{n,k}|\ge\varepsilon s_n\}}.\end{aligned}$$

Ainsi

$$\sum_{k=1}^{N_n}|\varphi_{n,k}(t)-(1-\frac{1}{2}t^2\frac{\operatorname{Var} X_{n,k}}{s_n^2})|\le\alpha\varepsilon|t|^3+\alpha t^2\sum_{k=1}^{N_n}\frac{1}{s_n^2}\mathbb{E}X_{n,k}^21\!\!1_{\{|X_{n,k}|\ge\varepsilon s_n\}}$$

[1] Lorsqu'il n'y a que deux termes, la preuve résulte de l'identité $zu-z'u'=z(u-u')+u'(z-z')$, dans le cas général; on procède alors par récurrence.

[2] Vérifier l'existence...

.

Comme ε est arbitraire, en utilisant la condition de Lindeberg, on voit que

$$\forall t \in \mathbb{R} \lim_{n\to+\infty} \sum_{k=1}^{N_n} |\varphi_{n,k}(t) - (1 - \frac{t^2}{2}\frac{\operatorname{Var} X_{n,k}}{s_n^2})| = 0.$$

Il ne nous reste donc plus qu'à montrer que

$$\forall t \in \mathbb{R} \lim_{n\to+\infty} \sum_{k=1}^{N_n} |\psi_{n,k}(t) - (1 - \frac{t^2}{2}\frac{\operatorname{Var} X_{n,k}}{s_n^2})| = 0,$$

c'est à dire que

$$\forall t \in \mathbb{R} \lim_{n\to+\infty} \sum_{k=1}^{N_n} |\exp(-\frac{t^2}{2}\frac{\operatorname{Var} X_{n,k}}{s_n^2}) - (1 - \frac{t^2}{2}\frac{\operatorname{Var} X_{n,k}}{s_n^2})| = 0.$$

Mais $\forall z \in \mathbb{C} \quad |e^z - 1 - z| \le \frac{e^{|z|}}{2}|z|^2$ (il suffit de considérer le développement en série entière de l'exponentielle), d'où

$$\sum_{k=1}^{N_n} |\exp(-\frac{t^2}{2}\frac{\operatorname{Var} X_{n,k}}{s_n^2}) - (1 - \frac{t^2}{2}\frac{\operatorname{Var} X_{n,k}}{s_n^2})| \le \frac{t^4}{8}e^{t^2/2}\frac{1}{s_n^4}\sum_{k=1}^{N_n}(\operatorname{Var} X_{n,k})^2$$

Si l'on pose $M_n = \max(\operatorname{Var} X_{n,k}; 1 \le k \le N_n)$, on a

$$\frac{1}{s_n^4}\sum_{k=1}^{N_n}(\operatorname{Var} X_{n,k})^2 \le \frac{1}{s_n^4}\sum_{k=1}^{N_n} M_n \operatorname{Var} X_{n,k} = \frac{M_n}{s_n^2}.$$

Pour conclure la preuve, il suffit de montrer que $\frac{M_n}{s_n^2}$ tend vers 0.

Soit $\varepsilon \in]0, 1/2]$ et n un entier. Pour tout i entre 1 et N_n

$$\mathbb{E}\frac{X_{n,i}^2}{s_n^2} = \frac{1}{s_n^2}\mathbb{E}X_{n,i}^2 1\!\!1_{\{|X_{n,i}|<\varepsilon s_n\}} + \frac{1}{s_n^2}\mathbb{E}X_{n,i}^2 1\!\!1_{\{|X_{n,i}|>\varepsilon s_n\}},$$

d'où

$$\frac{M_n}{s_n^2} \le \varepsilon^2 + \sum_{k=1}^{N_n} \frac{1}{s_n^2} X_{n,k}^2 1\!\!1_{\{|X_{n,k}|\ge\varepsilon s_n\}}$$

D'après la condition de Lindeberg, pour n assez grand, le deuxième terme ne dépasse pas ε^2 et donc $\frac{M_n}{s_n^2} \le 2\varepsilon^2 \le \varepsilon$, ce qui montre bien que $\frac{M_n}{s_n^2}$ tend vers 0 $\square$

Comme corollaire, on récupère évidemment le Théorème Central Limite

Théorème 2. *Soit $X_1, \ldots, X_n$ une suite de variables aléatoires réelles i.i.d. admettant un moment d'ordre 2. On note alors m l'espérance et σ^2 la variance communes à ces variables. Alors*

$$\frac{(X_1 + \ldots X_n) - nm}{\sqrt{n}} \Longrightarrow \mathcal{N}(0, \sigma^2).$$

Démonstration. La preuve est laissée en exercice. □

1.1.2 Condition de Lyapounov

Soit $\delta > 0$. Si les $X_{n,k}$ ont un moment d'ordre $2 + \delta$ et que

$$\lim_{n \to +\infty} \sum_{k=1}^{N_n} \frac{1}{s_n^{2+\delta}} \mathbb{E}|X_{n,k}|^{2+\delta} = 0 \quad \textit{(Condition de Lyapounov)},$$

alors

$$\lim_{n \to +\infty} \sum_{k=1}^{N_n} \frac{1}{s_n^2} X_{n,k}^2 1\!\!1_{\{|X_{n,k}| \ge \varepsilon s_n\}} = 0 \quad \textit{(Condition de Lindeberg)}.$$

est vérifiée.

Démonstration. Il suffit de remarquer que

$$1\!\!1_{\{|X_n| \ge \varepsilon s_n\}} \le \left(\frac{|X_n|}{\varepsilon s_n} \right)^{\delta}.$$

□

Corollaire 1. *Soit $(X_{n,k})_{\ge 0}$ des variables aléatoires centrées indépendantes bornées par une constante M, $(N_n)_{n \ge 1}$ un suite d'entiers tendant vers l'infini. On pose*

$$S_n = \sum_{k=1}^{N_n} X_{n,k}$$

et $s_n^2 = \operatorname{Var} S_n$. Si s_n tend vers l'infini lorsque n tend vers l'infini, alors S_n / s_n converge en loi vers la loi $\mathcal{N}(0, 1)$

Démonstration. Soit $\delta > 0$ quelconque.

$$\sum_{k=1}^{N_n} \frac{1}{s_n^{2+\delta}} \mathbb{E} X_{n,k}^{2+\delta} \le \sum_{k=1}^{N_n} \frac{1}{s_n^{2+\delta}} \mathbb{E} X_{n,k}^2 M^{\delta} = \frac{M^{\delta}}{s_n^{\delta}},$$

qui tend bien vers 0 lorsque n tend vers l'infini, ainsi le critère de Lyapounov est vérifié, donc le critère de Lindeberg aussi et le théorème 1 s'applique. □

1.2 Sommes et séries de variables aléatoires indépendantes

1.2.1 Loi zéro-un

Théorème 3. *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes. L'événement "La série de terme général X_n converge" est de probabilité zéro ou un.*

Démonstration. Pour $k \geq 0$, notons A_k l'événement "La série de terme général X_{n+k} converge". Par construction, A_k est $\sigma(X_i, i \geq k)$-mesurable. Comme la nature d'une série ne dépend pas des premiers termes, on a pour tout k, $A_k = A_0$. Ainsi, pour tout $k \geq 0$, $A_0 \in \sigma(X_i, i \geq k)$, ce qui signifie que $A_0 \in \mathcal{Q} = \underset{k\geq 1}{\cap} \sigma(X_i, i \geq k)$.

Cette tribu est la tribu de queue associée à une famille de variables aléatoires indépendantes sous $\mathbb{P}$: d'après la loi $0-1$ de Kolmogorov, elle est triviale sous $\mathbb{P}$.

Ainsi, la probabilité $\mathbb{P}(A_0)$ ne peut valoir que 0 ou 1. □

1.2.2 Une inégalité maximale

Théorème 4. *Soit $X_1, \dots, X_n$ des variables aléatoires indépendantes. On note $(S_k)_{k\in\{0,\dots,n\}}$ la suite des sommes partielles : $S_0 = 0$, $S_k = X_1 + \dots + X_k$. Pour tout $\alpha > 0$, on a*

$$\mathbb{P}(\max_{1\leq k\leq n} |S_k| \geq 4\alpha) \leq 4 \max_{1\leq k\leq n} P(|S_k| \geq \alpha).$$

Démonstration. Notons $T = \inf\{k \in \{1, \dots, n\}; |S_k| \geq 4\alpha\}$.

$$\begin{aligned}
\{T < +\infty\} &= \{T < +\infty\} \cap \{|S_n| \geq 2\alpha\} \cup \{T < +\infty\} \cap \{|S_n| < 2\alpha\} \\
&\subset \{|S_n| \geq 2\alpha\} \cup \{T < +\infty\} \cap \{|S_n| < 2\alpha\} \\
&= \{|S_n| \geq 2\alpha\} \cup \bigcup_{k=1}^{n-1} \{T = k\} \cap \{|S_n| < 2\alpha\} \\
&\subset \{|S_n| \geq 2\alpha\} \cup \bigcup_{k=1}^{n-1} \{T = k\} \cap \{|S_n - S_k| \geq 2\alpha\}
\end{aligned}$$

D'où

$$\mathbb{P}(T < +\infty) \leq \mathbb{P}(|S_n| \geq 2\alpha) + \sum_{k=1}^{n-1} \mathbb{P}(\{T = k\} \cap \{|S_n - S_k| \geq 2\alpha\}).$$

Mais $T = k$ et $S_n - S_k$ sont indépendants, donc

$$\begin{aligned}
\mathbb{P}(T < +\infty) &\leq \mathbb{P}(|S_n| \geq 2\alpha) + \sum_{k=1}^{n-1} \mathbb{P}(\{T = k\})\mathbb{P}(|S_n - S_k| \geq 2\alpha) \\
&\leq \beta + \sum_{k=1}^{n-1} \mathbb{P}(\{T = k\})\beta \\
&\leq 2\beta,
\end{aligned}$$

où l'on a posé $\beta = \max_{0 \leq k \leq n} P(|S_n - S_k| \geq 2\alpha)$. La probabilité d'une réunion est plus grande que le max des probabilités, donc

$$P(T < +\infty) = \mathbb{P}(\max_{1 \leq k \leq n} |S_k| \geq 4\alpha) \leq 2 \max_{0 \leq k \leq n} P(|S_n - S_k| \geq 2\alpha)$$

Posons maintenant pour tout $k \in \{0, \ldots, n\}$: $S'_k = S_n - S_{n-k}$. En d'autres termes, on a $S'_k = X'_1 + X'_2 + \ldots X'_k$, avec $X'_k = X_{n-k}$. On a bien sûr $S'_0 = 0$ et $S'_n = S_n$. On répète le raisonnement précédent, cette fois avec $T' = \inf\{k \in \{1, \ldots, n\}; |S'_k| \geq 2\alpha\}$.

$$\begin{aligned}
\{T' < +\infty\} &= \{T' < +\infty\} \cap \{|S'_n| \geq \alpha\} \cup \{T' < +\infty\} \cap \{|S'_n| < \alpha\} \\
&\subset \{|S'_n| \geq \alpha\} \cup \{T' < +\infty\} \cap \{|S'_n| < \alpha\} \\
&= \{|S'_n| \geq \alpha\} \cup \bigcup_{k=1}^{n-1} \{T' = k\} \cap \{|S'_n| < \alpha\} \\
&\subset \{|S'_n| \geq \alpha\} \cup \bigcup_{k=1}^{n-1} \{T' = k\} \cap \{|S'_n - S'_k| < \alpha\}
\end{aligned}$$

D'où

$$\mathbb{P}(T' < +\infty) \leq \mathbb{P}(|S'_n| \geq \alpha) + \sum_{k=1}^{n-1} \mathbb{P}(\{T' = k\} \cap \{|S'_n - S'_k| < \alpha\}).$$

Mais $T' = k$ et $S'_n - S'_k$ sont indépendants, donc

$$\begin{aligned}
\mathbb{P}(T' < +\infty) &\leq \mathbb{P}(|S'_n| \geq \alpha) + \sum_{k=1}^{n-1} \mathbb{P}(\{T' = k\})\mathbb{P}(|S'_n - S'_k| \geq \alpha) \\
&= \mathbb{P}(|S_n| \geq \alpha) + \sum_{k=1}^{n-1} \mathbb{P}(\{T' = k\})\mathbb{P}(|S_{n-k}| \geq \alpha) \\
&\leq \beta' + \sum_{k=1}^{n-1} \mathbb{P}(\{T' = k\})\beta \\
&\leq 2\beta',
\end{aligned}$$

où l'on a posé $\beta' = \max_{0 \le k \le n} P(|S_k| \ge \alpha)$. Finalement,

$$\mathbb{P}(\max_{1 \le k \le n} |S_k| \ge 4\alpha) = P(T < +\infty) \le 4\beta' = 4 \max_{1 \le k \le n} P(|S_k| \ge \alpha).$$

□

1.2.3 Lien entre les modes de convergence

Théorème 5. *Soit $X_1, \dots, X_n$ une suite de variables aléatoires indépendantes. On note $(S_k)_{k \ge 0}$ la suite des sommes partielles : $S_0 = 0$, $S_k = X_1 + \cdots + X_k$. S_n converge presque sûrement si et seulement S_n converge en probabilité.*

Démonstration. Bien sûr, comme la convergence presque sûre implique la convergence en probabilité, il suffit de démontrer l'autre implication. Supposons donc que S_n converge en probabilité vers une variable notée S. Soit $\varepsilon > 0$. On pose $Y_n = \sup\{|S_i - S_j|; i, j \ge n\}$ et $Y = \lim_{n \to +\infty} Y_n$. Soit $\varepsilon > 0$ et n un entier. On a

$$\mathbb{P}(Y > \varepsilon) \le \mathbb{P}(Y_n > \varepsilon)$$

On a

$$\begin{aligned} \mathbb{P}(Y_n > \varepsilon) &= \mathbb{P}(\exists i, j \ge n; \quad |S_i - S_j| > \varepsilon) \\ &\le \mathbb{P}(\exists i \ge n; \quad |S_n - S_i| > \varepsilon/2) \end{aligned}$$

D'après le théorème de continuité séquentielle croissante

$$\mathbb{P}(\exists i \ge n;\ |S_n - S_i| > \varepsilon/2) = \lim_{p \to +\infty} \mathbb{P}(\max_{1 \le k \le p} |S_{n+k} - S_n| > \varepsilon/2)$$

Mais d'après le théorème 4

$$\mathbb{P}(\max_{1 \le k \le p} |S_{n+k} - S_n| > \varepsilon/2) \le 4 \max_{1 \le k \le p} \mathbb{P}(|S_{n+k} - S_n| > \varepsilon/8)$$

Par ailleurs

$$\mathbb{P}(|S_{n+k} - S_n| > \varepsilon/8) \le \mathbb{P}(|S_n - S| > \varepsilon/16) + \mathbb{P}(|S_{n+k} - S| > \varepsilon/16) \le 2 \max_{r \ge n} \mathbb{P}(|S_r - S| > \varepsilon/16).$$

Ainsi

$$\mathbb{P}(Y_n > \varepsilon) \le 8 \max_{r \ge n} \mathbb{P}(|S_r - S| > \varepsilon/16),$$

D'où

$$\forall n \ge 1 \quad \mathbb{P}(Y > \varepsilon) \le 8 \max_{r \ge n} \mathbb{P}(|S_r - S| > \varepsilon/16).$$

En faisant tendre n vers $+\infty$, on obtient que $P(Y > \varepsilon) = 0$. Maintenant $\mathbb{P}(Y > 0) = \mathbb{P}(\cup_{n=1}^{+\infty} \{Y > 1/n\}) = 0$.

□

1.2.4 Critères de convergence

Convergence L^2

Théorème 6. *Soit $X_1, \dots, X_n$ une suite de variables aléatoires indépendantes, centrées, admettant un moment d'ordre 2. On note $(S_k)_{k\geq 0}$ la suite des sommes partielles : $S_0 = 0$, $S_k = X_1 + \dots + X_k$.*
Si

$$\sum_{k=1}^{+\infty} \operatorname{Var} X_k < +\infty,$$

alors S_n converge presque sûrement.

Démonstration. D'après le théorème précédent, il suffit de montrer que S_n converge en probabilité. Il suffit donc de montrer que S_n converge dans L^2. Comme L^2 est complet, il suffit de montrer que S_n est une suite de Cauchy dans L^2. Soient n, p entiers :

$$\|S_{n+p} - S_n\|_2^2 = \mathbb{E}(\sum_{k=n+1}^{n+p} X_{n+k})^2 = \operatorname{Var} \sum_{k=n+1}^{n+p} X_{n+k} = \sum_{k=n+1}^{n+p} \operatorname{Var} X_{n+k}$$

Soit $\varepsilon > 0$. On peut choisir N tel que $\sum_{k>N} \operatorname{Var} X_k \leq \varepsilon^2$ car la série de terme général $\operatorname{Var} X_k$ converge. Ainsi, pour tout $n \geq N$ et tout p entier, on a $\|S_{n+p} - S_n\|_2 \leq \varepsilon$; comme $\varepsilon > 0$ est quelconque, cela montre bien que S_n est une suite de Cauchy dans L^2, ce qui achève la preuve. □

Théorème des trois séries

Théorème 7. *Soit $X_1, \dots, X_n$ une suite de variables aléatoires indépendantes.*

Soit c un réel strictement positif quelconque : on note $X_k^{(c)} = X_k \mathbb{1}_{\{|X_k| \leq c\}}$.

Alors, la série de terme général X_n converge presque sûrement si et seulement si les séries de terme général $\operatorname{Var} X_k^{(c)}$, $\mathbb{P}(|X_k| \geq c)$ et $\mathbb{E} X_k^{(c)}$ convergent.

Démonstration. Montrons d'abord que la condition est suffisante. La série de terme général $\mathbb{P}(|X_k| \geq c)$ converge, donc d'après le premier lemme de Borel Cantelli

$$\mathbb{P}(\overline{\lim_{n\to+\infty}} \{|X_k| \geq c\}) = 0,$$

soit

$$\mathbb{P}(\underline{\lim_{n\to+\infty}} \{|X_k| < c\}) = 1,$$

ou encore

$$\mathbb{P}\left(\exists N; k \geq N \Longrightarrow |X_k| < c\right) = 1,$$

ce qui implique que presque sûrement, à partir d'un certain rang, $X_k = X_k^{(c)}$, ce qui signifie que les deux séries sont de même nature. Bien sûr, comme la série de terme général $\mathbb{E}X_k^{(c)}$ converge, la série de terme général $X_k^{(c)}$ et la série de terme général $X_k^{(c)} - \mathbb{E}X_k^{(c)}$ sont de même nature, or cette dernière série converge presque sûrement d'après le théorème 6, ce qui achève la preuve.

Réciproquement, supposons que la série de terme général S_n converge presque sûrement : cela implique que

$$\mathbb{P}(|X_n| > c \text{ infiniment souvent}) = 0,$$

ce qui entraîne que la série de terme général $\mathbb{P}(|X_n| > c)$ d'après le deuxième lemme de Borel-Cantelli. À partir d'un certain rang, X_n et X_n^c sont égales, donc la série de terme général X_n^c converge également presque sûrement. Maintenant, raisonnons par l'absurde et supposons que la série de terme général $\operatorname{Var} X_k^c$ diverge. Alors

$$\frac{\sum_{k=1}^{n} X_k^c}{\sqrt{\sum_{k=1}^{n} \operatorname{Var} X_k^c}} \to 0 \quad \text{p.s.}$$

Posons

$$Z_n = \frac{\sum_{k=1}^{n} (X_k^c - \mathbb{E}X_k^c)}{\sqrt{\sum_{k=1}^{n} \operatorname{Var} X_k^c}}$$

et $\alpha_n = \frac{\sum_{k=1}^{n} \mathbb{E}X_k^c}{\sqrt{\sum_{k=1}^{n} \operatorname{Var} X_k^c}}$ D'après le corollaire 1, $Z_n \Longrightarrow \mathcal{N}(0,1)$, ce qui contredit que $Z_n + \alpha_n$ converge presque sûrement vers 0. En effet, $Z_n \Longrightarrow \mathcal{N}(0,1)$ entraîne que $\varphi_{Z_n}(1)$ tend $\exp(-1/2)$ alors que $Z_n + \alpha_n \to 0$ entraîne que $\varphi_{Z_n}(1) \sim e^{-i\alpha_n}$, ce qui est clairement incompatible en comparant les modules. Il y a une contradiction : on en déduit que la série de terme général $\operatorname{Var} X_k^c$ converge. On peut maintenant appliquer le théorème 6 : la série de terme général $X_k^c - \mathbb{E}X_k^c$ converge presque sûrement. Comme la série de terme général X_k^c converge presque sûrement, on en déduit que la série de terme général $\mathbb{E}X_k^c$ converge.

□

1.3 Exercices sur les sommes de variables indépendantes

1. Faire la preuve du TCL qui a été laissée en exercice.
2. Soit $(X_\lambda)_{\lambda>0}$ une famille de variables aléatoires telles que pour tout $\lambda > 0$, X_λ suive une loi de Poisson $\mathcal{P}(\lambda)$. Montrer que la suite
$$\frac{X_\lambda - \lambda}{\sqrt{\lambda}}$$
converge faiblement vers la loi $\mathcal{N}(0,1)$ lorsque λ tend vers l'infini.
3. (a) Soit $(U_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. Pour $x > 0$, on pose
$$X_n^x = \sum_{k=1}^n \mathbb{1}_{\{U_k \leq \frac{x}{n}\}}.$$
Montrer que $(X_n^x)_{n\geq 1}$ converge en loi et déterminer la loi limite.
   (b) Soit $(u_n)_{n\geq 0}$ une suite croissante, de limite $\ell \in \mathbb{R}$. Pour $x > 0$, on pose
$$f(x) = \sum_{k=0}^{+\infty} e^{-x}\frac{x^k}{k!}u_k.$$
Montrer que f est une fonction croissante et déterminer sa limite en $+\infty$.
4. Soit X_n une suite de variables aléatoires convergeant en loi vers une variable aléatoire X. On pose $\varphi_n = \varphi_{X_n}$. Montrer que la famille $(\varphi_n)_{n\geq 1}$ est uniformément équicontinue, c'est à dire que
$$\forall \varepsilon > 0 \quad \exists \eta > 0 \quad \forall n \geq 1 \quad \forall (x,y) \in \mathbb{R}^2 \quad |x-y| \leq \eta \Longrightarrow |\varphi_n(x) - \varphi_n(y)| \leq \varepsilon.$$
5. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires convergeant en loi vers une variable aléatoire X, des suites de réels $(a_n)_{n\geq 1}$ et $(b_n)_{n\geq 1}$ convergeant respectivement vers les réels a et b. Montrer que la suite de variables aléatoires $(a_nX_n+b_n)_{n\geq 1}$ converge en loi vers la variable aléatoire $aX + b$.
6. *Une preuve probabiliste de la formule de Stirling*
Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires indépendantes suivant la loi exponentielle de paramètre 1. On pose
$$S_n = \sum_{k=0}^n X_k.$$

(a) Montrer que $\frac{S_n-n}{\sqrt{n}}$ converge en loi vers la loi $\mathcal{N}(0,1)$.

(b) Montrer que S_n suit la loi $\Gamma(n+1,1)$. En déduire que la densité de $\frac{S_n-n}{\sqrt{n}}$ s'écrit $g_n(x)=a_nh_n(x)$, avec

$$a_n=\frac{n^{n+1/2}e^{-n}\sqrt{2\pi}}{\Gamma(n+1)}$$

et

$$h_n(x)=a_n\mathbb{1}_{[-\sqrt{n},+\infty[}\frac{1}{\sqrt{2\pi}}e^{-\sqrt{n}x}(1+\frac{x}{\sqrt{n}})^n.$$

(c) Montrer que

$$\lim_{n\to+\infty}\int_0^1 g_n(x)\ dx=\int_0^1\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\ dx$$

(d) Soit Ψ la fonction définie sur $[-1/2,1/2]$ par

$$\forall x\in[-1/2,+1/2]\quad \ln(1+x)=x-\frac{x^2}{2}+x^3\Psi(x).$$

Montrer que

$$x\ge-\sqrt{n}\Longrightarrow h_n(x)-\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}(e^{\frac{x^3}{n^{1/2}}\psi(\frac{x}{\sqrt{n}})}-1)$$

En déduire que $h_n(x)$ converge uniformément vers $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ sur tout compact.

(e) Montrer que

$$\lim_{n\to+\infty}\int_0^1 h_n(x)\ dx=\int_0^1\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\ dx.$$

(f) En déduire la formule de Stirling :

$$n!\sim\sqrt{2\pi}n^{n+1/2}e^{-n}.$$

7. Soit $(X_n)_{n\ge1}$ une suite de variables aléatoires indépendantes telles que pour tout n, X_n suive une loi binomiale $\mathcal{B}(n,\frac{1}{n^{1,01789}})$. Déterminer l'ensemble des valeurs d'adhérences de la suite (X_n).

8. Soit $(X_n)_{n\ge2}$ une suite de variables aléatoires identiquement distribuées telle qu'il existe $\alpha>0$ avec $\mathbb{E}\exp(\alpha|X_1|)<+\infty$.
Montrer que la série de terme général $\frac{X_n}{n(\ln n)^3}$ converge.

9. *Séries de Rademacher*
   Soit $(a_n)_{n\geq 1}$ une suite de réels positifs. Soit X_n une suite de variables aléatoires indépendantes suivant la loi uniforme sur $[-1,1]$. Déterminer des conditions nécessaires et suffisantes pour avoir presque sûrement
   – convergence de la série de terme général $a_n X_n$
   – absolue convergence de la série de terme général $a_n X_n$.
10. Reprendre l'exercice précédent dans le cas où les X_n sont des variables aléatoires indépendantes suivant la loi normale $\mathcal{N}(0,1)$.
11. *Séries à termes positifs*
    Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes à termes positifs. Soit $c > 0$ quelconque. Montrer que la série de terme général X_n converge si et seulement si les séries de termes général $\mathbb{P}(X_n > c)$ et $\mathbb{E}X_n \mathbb{1}_{\{X_n \leq c\}}$ convergent.
12. Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires indépendantes suivant la loi uniforme sur $[0, 2\pi]$.
    Montrer que le rayon de convergence de la série entière
    $$z \mapsto \sum_{n=0}^{+\infty} \cos(X_n) z^n$$
    est presque sûrement 1.
13. *Séries de variables symétriques*
    Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes symétriques. Soit $c > 0$ quelconque. Montrer que la série de terme général X_n converge si et seulement si les séries de termes général $P(X_n > c)$ et $\mathrm{Var}[X_n \mathbb{1}_{\{X_n \leq c\}}]$ convergent.
14. Reprendre l'exercice 9. dans le cas où les X_n sont des variables aléatoires indépendantes suivant de Cauchy $\mathcal{C}(0,1)$.
15. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes identiquement distribuées. Montrez qu'il existe une suite $(a_n)_{n\geq 1}$ de réels non nuls tels que la série de terme général $a_n X_n$ converge presque sûrement
16. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires de même loi non dégénéré. Soit $(a_n)_{n\geq 1}$ une suite réelle. On suppose que $(a_n X_n)_{n\geq 1}$ tend presque sûrement vers zéro. Montrer que $(a_n)_{n\geq 1}$ tend vers zéro.

Chapitre 2

Vecteurs gaussiens

2.1 Quelques rappels sur la matrice de covariance

Si $X = (X_1, \dots, X_n)$ est un vecteur aléatoire dont toutes les composantes admettent un moment d'ordre deux, on convient de dire que le vecteur a un moment d'ordre deux et on appelle matrice de covariance de X la matrice $n \times n$ dont les coefficients sont $(\mathrm{Covar}(X_i, X_j))_{1 \le i,j \le n}$.

Théorème 8. *Si $X = (X_1, \dots, X_n)$ est un vecteur aléatoire admettant un moment d'ordre deux , la matrice de covariance de X est la matrice dans la base canonique de l'application bilinéaire positive*

$$\begin{array}{rcl} \mathbb{R}^n \times \mathbb{R}^n & \to & \mathbb{R} \\ (a, b) & \mapsto & \mathrm{Covar}(\langle X, a \rangle, \langle X, b \rangle) \end{array}$$

C'est une matrice symétrique positive.

Démonstration. À X fixé, l'application $X \mapsto \langle X, a \rangle$ est une application linéaire. Comme on a déjà montré que Covar était une forme bilinéaire symétrique positive, il s'ensuit que l'application considérée ici est une forme bilinéaire symétrique positive. Cette application envoie le couple (e_i, e_j) sur $\mathrm{Covar}(\langle X, e_i \rangle, \langle X, e_j \rangle) = \mathrm{Covar}(X_i, X_j)$. La matrice d'une forme bilinéaire symétrique positive est une matrice symétrique positive. □

Théorème 9. *Soit $X = (X_1, \dots, X_n)$ est un vecteur aléatoire admettant un moment d'ordre deux et de matrice de covariance C_X et d'espérance m_X. Soit A une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$, et b un vecteur de $\mathbb{R}^p$. Alors $Y = AX + c$ admet $C_Y = AC_XA^*$ comme matrice de covariance et l'espérance de Y vaut $Am_X + c$.*

Démonstration.

$$\begin{aligned}
\mathbb{E}Y_i &= \mathbb{E}(\sum_{k=1}^{n} a_{i,k}X_k + c_i) \\
&= \sum_{k=1}^{n} a_{i,k}\mathbb{E}X_k + c_i \\
&= \sum_{k=1}^{n} a_{i,k}m_k + c_i \\
&= (Am + c)_i
\end{aligned}$$

$$\begin{aligned}
\mathrm{Covar}(\langle Y,a\rangle,\langle Y,b\rangle) &= \mathbb{E}\,\mathrm{Covar}(\langle AX+c,a\rangle,\langle AX+c,b\rangle) \\
&= \mathrm{Covar}(\langle AX,a\rangle,\langle AX,b\rangle) \\
&= \mathrm{Covar}(\langle X,A^*a\rangle,\langle X,A^*b\rangle) \\
&= \langle C_X A^*a, A^*b\rangle \\
&= \langle AC_X A^*a, b\rangle
\end{aligned}$$

□

Définition: On dit qu'un vecteur aléatoire $X \in \mathbb{R}^d$ est gaussien si pour tout $a \in \mathbb{R}^d$ la variable aléatoire $\langle X, a\rangle$ est gaussienne .

2.2 Image affine d'un vecteur gaussien

Théorème 10. *L'image d'un vecteur gaussien X d'espérance m_X et de matrice de covariance C_X par une application affine $x \mapsto Ax + b$ est un vecteur gaussien d'espérance $m_Y = Am_X + b$ et de matrice de covariance $C_Y = AC_XA^*$.*

Démonstration. On pose $Y = AX + b$. Soit $a \in \mathbb{R}^d$. $\langle Y,a\rangle = \langle AX,a\rangle + \langle b,a\rangle = \langle X,A^*a\rangle + \langle b,a\rangle$. Comme X est un vecteur gaussien $\langle X,A^*a\rangle$ est une variable aléatoire gaussienne. Quand on ajoute une constante à variable aléatoire gaussienne, on obtient une variable aléatoire gaussienne. Ainsi, pour tout A, $\langle Y,a\rangle$ est une variable aléatoire gaussienne, donc Y est un vecteur gaussien. L'expression de l'espérance et de la covariance est une conséquence du théorème 9. □

Corollaire 2. *Si $X = (X_1,\dots,X_d)$ est gaussien, alors pour tout $I \in \{1,\dots,d\}$, le vecteur $(X_i)_{i\in I}$ est gaussien*

Démonstration. $(X_i)_{i\in I}$ est l'image de $X = (X_1, \dots, X_d)$ par l'application linéaire

$$\begin{array}{rcl} \mathbb{R}^{\{1,\dots,d\}} & \to & \mathbb{R}^I \\ (x_i)_{i\in\{1,\dots,d\}} & \mapsto & (x_i)_{i\in D} \end{array}$$

□

2.3 Exemple fondamental

Théorème 11. *Soit $X_1, \dots, X_d$ d variables aléatoires gaussiennes indépendantes. Alors $X = (X_1, \dots, X_d)$ est un vecteur gaussien.*

Démonstration. Pour $k \in \{1, \dots d\}$, on pose $S_k = \sum_{i=1}^k a_X X_k$. On montre par récurrence sur k que S_k est une variable aléatoire gaussienne. Pour $k = 1$, $S_1 = a_1 X_1$: quand on multiplie une variable gaussienne par une constante, on a une variable aléatoire gaussienne. Supposons acquis que S_k est gaussienne : on a $S_{k+1} = S_k + a_{k+1}X_{k+1}$. $a_{k+1}X_{k+1}$ est une variable gaussienne indépendante de S_k, car S_k est $\sigma(X_1, \dots X_k)$-mesurable. Or, on sait que la somme de deux variables aléatoires gaussiennes indépendantes est une variable aléatoire gaussienne, donc S_{k+1} est une variable aléatoire gaussienne.

Comme $\langle X, a\rangle = S_d$ quel que soit a , on en déduit que X est un vecteur gaussien. □

2.4 Lois gaussiennes

Théorème 12. *Soit C une matrice symétrique positive $d \times d$ et $m \in \mathbb{R}^d$. Alors, on peut construire un vecteur gaussien admettant m comme espérance et C comme matrice de covariance.*

Démonstration. Comme C est symétrique,on peut la diagonaliser avec une matrice de passage orthogonale. Comme elle est positive, les valeurs propres sont positives. Ainsi, on peut trouver une matrice O orthogonale et des réels positifs $\lambda_1, \dots, \lambda_d$ tels que

$$C = O \begin{pmatrix} \lambda_1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \lambda_2 & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \lambda_{d-1} & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & \lambda_d \end{pmatrix} O^*$$

Posons

$$A = O \begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \sqrt{\lambda_{d-1}} & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & \sqrt{\lambda_d} \end{pmatrix} O^*$$

et soient $(X_1, \dots, X_d)$ d variables aléatoires indépendantes suivant la loi $\mathcal{N}(0,1)$: l'espérance de X est $m_X = 0$ sa matrice de covariance $C_X = \mathrm{Id}_{\mathbb{R}^d}$. D'après le théorème 11, X est gaussien donc, d'après le théorème 9, $Y = AX + m$ est un vecteur gaussien d'espérance $A.0 + m = m$ et de covariance $AC_X A^* = AA^* = C$. □

Théorème 13. *Soit X et Y deux vecteurs gaussiens ayant même espérance et même matrice de covariance. Alors X et Y ont même loi.*

Démonstration. Soit $a \in \mathbb{R}^d$. On pose $V = \langle X, a\rangle$ et $W = \langle Y, a\rangle$. L'espérance de V est $\langle m_X, a\rangle$ et la covariance de V est $\mathrm{Covar}(\langle X, a\rangle, \langle X, a\rangle) = \langle C_X a, a\rangle$. Comme X est gaussien, V est gaussienne, donc $V \sim \mathcal{N}(\langle m_X, a\rangle, \langle C_X a, a\rangle)$. De même $W \sim \mathcal{N}(\langle m_Y, a\rangle, \langle C_Y a, a\rangle)$. Comme $m_X = m_Y$ et $C_X = C_Y$, on en déduit que $V = \langle X, a\rangle$ et $W = \langle Y, a\rangle$ ont même loi. Ainsi, pour tout a, on a $\mathbb{E} \exp(i\langle X, a\rangle) = \mathbb{E} \exp(i\langle Y, a\rangle)$. X et Y ont donc même fonction caractéristique, or la fonction caractéristique caractérise la loi, donc X et Y ont même loi. □

Définition: Soit $m \in \mathbb{R}^d$ et C une matrice $d \times d$ symétrique positive. On note $\mathcal{N}(m, C)$ la loi commune à tous les vecteurs gaussiens admettant m comme espérance et C comme matrice de covariance.

La pertinence de cette définition est assurée par les théorèmes 12 et 13

2.5 Lois gaussiennes et indépendance

Théorème 14. *Soit $d_1, \dots, d_n$ n entiers positifs de somme d. Soit $C_1, \dots C_n$ n matrices symétriques positives, et $m_1, \dots m_n$ n vecteurs C_i étant de taille d_i. Alors*

$$\mathcal{N}(m_1, C_1) \otimes \mathcal{N}(m_2, C_2) \cdots \otimes \mathcal{N}(m_n, C_n) = \mathcal{N}(m, C),$$

avec

$$m = \begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix} \quad \textit{et } C = \begin{pmatrix} C_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & C_n \end{pmatrix}$$

Démonstration. Soit $Y_1, \ldots, Y_n$ n vecteurs gaussiens indépendants, avec $Y_1 \sim \mathcal{N}(m_i, C_i)$. On va d'abord montrer que $X = (Y_1, \ldots, Y_n)$ est gaussien. Soit $a \in \mathbb{R}^n$. On peut écrire $a = (a_1, \ldots, a_n)$, avec a_i de taille d_i. On a

$$\langle X, a \rangle = \sum_{k=1}^{n} \langle Y_k, a_k \rangle.$$

Comme Y_k est gaussien, chaque variable aléatoire $\langle Y_k, a_k \rangle$ est une variable aléatoire gaussienne. Comme les Y_k sont indépendants, les variables aléatoires $\langle Y_k, a_k \rangle$ sont indépendantes. Or on sait qu'une somme de variables aléatoires gaussiennes indépendantes est une variable aléatoire gaussienne, donc $\langle X, a \rangle$ est une variable aléatoire gaussienne. Il s'ensuit que X est un vecteur gaussien. L'expression de l'espérance et de la matrice de covariance ne pose pas de problème, puisque des variables aléatoires indépendantes ne sont pas corrélées. □

Théorème 15. *Soit $d_1, \ldots, d_n$ n entiers positifs de somme d. On suppose que $X = (X_1, \ldots, X_n)$ est un vecteur gaussien dont la matrice de covariance est diagonale par blocs :*

$$C_X = \begin{pmatrix} C_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & C_n \end{pmatrix}$$

Alors, si l'on pose $Y_1 = (X_1, \ldots, X_{d_1})$, $Y_2 = (X_{d_1+1}, \ldots, X_{d_1+d_2})$, $Y_n = (X_{d_1+d_2+\cdots+d_{n-1}+1}, \ldots, X_{d_1+d_2+\cdots+d_n})$, les vecteurs $Y_1, \ldots Y_n$ sont des vecteurs gaussiens indépendants.

Démonstration. On voit que X a même espérance et même matrice de covariance que le vecteur aléatoire considéré au théorème précédent. Comme X et le vecteur aléatoire considéré au théorème précédent sont tous deux gaussiens, ils ont tous deux la même loi, donc la loi de X est $\mathcal{N}(m_1, C_1) \otimes \mathcal{N}(m_2, C_2) \cdots \otimes \mathcal{N}(m_n, C_n)$, ce qui signifie que les Y_i sont indépendants et que pour tout i, on a $Y_i \sim \mathcal{N}(m_i, C_i)$. □

En particulier, on a le corollaire suivant :

Corollaire 3. *Si le vecteur gaussien $X = (X_1, \ldots, X_d)$ a une matrice de covariance dont tous les termes non-diagonaux sont nuls, alors $X_1, \ldots, X_d$ sont des variables aléatoires indépendantes.*

2.6 Lois gaussiennes à densité

Théorème 16. *Soit C une matrice symétrique définie positive et $m \in \mathbb{R}^d$. La loi sur $\mathbb{R}^d$ $\mathcal{N}(m, C)$ admet comme densité par rapport à la mesure de Lebesgue la fonction*

$$f_{m,C}(y) = \frac{1}{(2\pi)^{d/2}} \frac{1}{\sqrt{\det C}} \exp(-\frac{1}{2}\langle C^{-1}(y-m), y-m\rangle).$$

Démonstration. On reprend les notations de la preuve du théorème 12. On a $X \sim \mathcal{N}(0, \mathrm{Id})$ et $Y \sim \mathcal{N}(m, C)$. Par rapport au cas général, on gagne le fait que C est définie positive, ce qui implique que les λ_i sont strictement positifs, et donc que A est inversible. Comme X est composé de n variables aléatoires indépendantes à densité, la densité de X est le produit des densités, soit

$$f_X(x) = \prod_{k=1}^{d} \frac{1}{\sqrt{2\pi}} \exp(-\frac{x_k^2}{2}) = \frac{1}{(2\pi)^{d/2}} \exp(-\frac{1}{2}\langle x, x\rangle).$$

D'après le théorème de C^1-difféomorphisme, $Y = AX + m$ admet comme densité

$$\frac{1}{\det A} f_X(A^{-1}(y-m)) = \frac{1}{\det A} \frac{1}{(2\pi)^{d/2}} \exp(-\frac{1}{2}\langle A^{-1}(y-m), A^{-1}(y-m)\rangle).$$

Or

$$\begin{aligned}
\langle A^{-1}(y-m), A^{-1}(y-m)\rangle &= \langle (A^{-1})^* A^{-1}(y-m), y-m\rangle \\
&= \langle (A^{-1})^* A^{-1}(y-m), y-m\rangle \\
&= \langle (AA^*)^{-1}(y-m), y-m\rangle \\
&= \langle C^{-1}(y-m), y-m\rangle
\end{aligned}$$

D'autre part, $\det C = \det AA^* = (\det A)^2$, donc $\det A = \sqrt{\det C}$. On en déduit que la densité de Y (c'est à dire de la loi de Y) est

$$f_{m,C}(y) = \frac{1}{(2\pi)^{d/2}} \frac{1}{\sqrt{\det C}} \exp(-\frac{1}{2}\langle C^{-1}(y-m), y-m\rangle).$$

$\square$

2.7 Fonction caractéristique des vecteurs gaussiens

Théorème 17. *En dimension quelconque, la fonction caractéristique de la loi normale* $\mathcal{N}(m, C)$ *est*

$$x \mapsto \exp(i\langle x, m\rangle \exp(-\frac{1}{2}\langle Cx, x\rangle)$$

Démonstration.

$$\mathbb{E}\exp(i\langle X, t\rangle) = \mathbb{E}\exp(iY) = \varphi_Y(1),$$

où $Y = \langle X, t\rangle$. comme X est gaussien de covariance C et d'espérance m, Y est gaussien de covariance $\langle Cx, x\rangle$ et d'espérance $\langle x, m\rangle$. On déduit donc le résultat de la formule précédente. □

2.8 Théorème de la limite centrale en dimension d

Théorème 18. *Soit* $X_1, \dots, X_n$ *une suite i.i.d. de vecteurs aléatoires à valeurs dans* $\mathbb{R}^d$*. On suppose que* $\mathbb{E}\|X_1\|^2 < +\infty$*. On note alors* m *l'espérance et* C *la matrice des covariances. Alors*

$$\frac{(X_1 + \dots X_n) - nm}{\sqrt{n}} \Longrightarrow \mathcal{N}(0, C).$$

En fait, le théorème 18 peut être vu comme une conséquence du théorème 2.

Lemme 1. *Soit* $(X_n)_{n\geq 1}$ *une suite de vecteurs aléatoires à valeurs dans* $\mathbb{R}^d$ *et* X *un vecteur aléatoire à valeurs dans* $\mathbb{R}^d$*. Si pour tout* $a \in \mathbb{R}^d$*,* $\langle X_n, a\rangle$ *converge en loi vers* $\langle X, a\rangle$*, alors* X_n *converge en loi vers* X*.*

Démonstration. Soit $a \in \mathbb{R}^d$. On pose $Y_n = \langle X_n, a\rangle$. On a

$$\varphi_{X_n}(a) = \mathbb{E}e^{i\langle X_n, a\rangle} = \mathbb{E}e^{iY_n} = \varphi_{Y_n}(1)$$

Par hypothèse, Y_n converge en loi vers $\langle X, a\rangle$, donc $\varphi_{Y_n}(1)$ converge vers $\varphi_{\langle X,a\rangle}(1) = \mathbb{E}e^{i\langle X,a\rangle} = \varphi_X(a)$. Donc $\varphi_{X_n}(a)$ converge vers $\varphi_X(a)$. Comme c'est vrai pour tout $a \in \mathbb{R}^d$, le théorème de Lévy nous dit que $(X_n)_{n\geq 1}$ converge en loi vers X. □

On peut maintenant prouver le théorème 18.

Démonstration. Soit X un vecteur aléatoire suivant $\mathcal{N}(0, C)$. On pose

$$S_n = (X_1 + \dots X_n) - nm = \sum_{k=1}^{n}(X_k - m).$$

D'après le lemme précédent, il nous suffit de montrer que pour tout $a \in \mathbb{R}^d$ $\langle \frac{S_n}{\sqrt{n}}, a\rangle$ converge en loi vers $\langle X, a\rangle$.

Fixons $a \in \mathbb{R}^d$ et posons $Y_n = \langle X_n - m, a\rangle = a^*(X_n - m)$. Les Y_n sont des variables aléatoires indépendantes qui suivent toutes la même loi, la loi image de P_{X_1} par $x \mapsto \langle x - m, a\rangle = a^*(x - m)$. D'après le théorème 9, leur espérance est 0 et leur variance $a^*Ca = \langle Ca, a\rangle$. Ainsi, d'après le théorème 2, la suite $\frac{\sum_{k=1}^n Y_k}{\sqrt{n}}$ converge en loi vers $\mathcal{N}(0, \langle Ca, a\rangle)$.

Mais

$$\frac{\sum_{k=1}^n Y_k}{\sqrt{n}} = \langle \frac{S_n}{\sqrt{n}}, a\rangle,$$

et, encore d'après le théorème 9, la loi de $\langle X, a\rangle$ est précisément $\mathcal{N}(0, \langle Ca, a\rangle)$. $\langle \frac{S_n}{\sqrt{n}}, a\rangle$ converge donc en loi vers $\langle X, a\rangle$, ce qui achève la preuve.

□

2.9 Exercices sur les vecteurs gaussiens

1. Soient U et V deux variables aléatoires gaussiennes centrées et f une fonction croissante. Montrer que

$$\mathbb{E}\|U\|^2 < \mathbb{E}\|V\|^2 \Longrightarrow \mathbb{E}f(\|U\|) \leq \mathbb{E}f(\|V\|)$$

2. Soit X, Y deux variables aléatoires indépendantes suivant $\mathcal{N}(0, 1)$. On pose

$$U = \begin{pmatrix} 2X \\ 2X \end{pmatrix} \text{ et } V = \begin{pmatrix} 2X \\ \sqrt{5}Y \end{pmatrix}.$$

Montrer que U et V sont gaussiens centrés, puis que l'on a

$$\mathbb{E}\|U\|_2^2 < \mathbb{E}\|V\|_2^2,$$

tandis que

$$\mathbb{E}\|U\|_2^4 > \mathbb{E}\|V\|_2^4.$$

Comparer avec le résultat de l'exercice précédent.

3. Soit X un vecteur gaussien centré de matrice de covariance C. Montrer que

$$\mathbb{E}\|X\|_2^2 = \mathrm{Tr}C.$$

4. (a) Montrer qu'il existe un vecteur gaussien centré dont la matrice de covariance est
$$C = \begin{pmatrix} 5 & 0 & 3 \\ 0 & 6 & 0 \\ 3 & 0 & 2 \end{pmatrix}$$
   (b) Soit X un tel vecteur. Déterminer
$$\sup_{a\in\mathbb{R}^3} \frac{\mathbb{E}|\langle X, a\rangle|}{\|a\|_2}.$$
5. Montrer qu'on peut trouver des variables aléatoires X_1, X_2, X_3 telles que pour tout $i \in \{1,2,3\}$ $X_i \sim \mathcal{N}(0,3)$ et que $\mathrm{Covar}(X_1,X_2) = \mathrm{Covar}(X_1,X_3) = \mathrm{Covar}(X_3,X_2) = 1$. Calculer la variance de $X_1 + 2X_2 + 3X_3$.
6. (a) Soit X et Y deux variables aléatoires indépendantes suivant la loi $\mathcal{N}(0,1)$. Calculer la matrice de covariance du couple $(X, \alpha X+\beta Y)$
   (b) Soit (Z,T) un vecteur gaussien centré, avec $\mathrm{Var}\, Z = 1$. Montrer qu'il existe α et β tels que la matrice de covariance de (Z,T) soit
$$\begin{pmatrix} 1 & \alpha \\ \alpha & \alpha^2+\beta^2 \end{pmatrix}.$$
   En déduire que $\mathbb{E}Z^2T^2 = 3\alpha^2+\beta^2 = \mathrm{Var}\, Z + 2\,\mathrm{Covar}(Z,T)^2$,
   (c) Soit (Z,T) est un vecteur gaussien centré quelconque – on ne suppose plus que $\mathrm{Var}\, Z = 1$ –, montrer
$$\mathbb{E}Z^2T^2 = \mathrm{Var}\, Z\, \mathrm{Var}\, T + 2\,\mathrm{Covar}(Z,T)^2.$$
   (d) Soit (Z,T) un vecteur gaussien de matrice de covariance
$$\begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}.$$
   Calculer $\mathbb{E}Z^2T^2$.
   (e) Soit X un vecteur gaussien centré de matrice de covariance C. Montrer que
$$\mathbb{E}\|X\|^4 = 2\mathrm{Tr}C^*C + (\mathrm{Tr}C)^2.$$
7. Soit X un vecteur gaussien dans $\mathbb{R}^n$ de loi $\mathcal{N}(0,C)$ et α est un réel vérifiant $\alpha < \rho(C)^{-1}$, où $\rho(C)$ est le rayon spectral de C. Montrer que
$$\mathbb{E}\exp(\frac{\alpha}{2}\|X\|_2^2) = \prod_i (1-\alpha\lambda_i)^{-1/2}.$$
   où les λ_i sont les valeurs propres de C.

8. (a) Soit X un vecteur gaussien dans $\mathbb{R}^n$ de loi $\mathcal{N}(0, I_n)$ et O une matrice orthogonale. Montrer que $Y = OX$ a la même loi que X.

   (b) On appelle loi du chi-deux à n degrés de liberté et on note $\chi^2(n)$ la loi de $\|X\|_2^2$. Montrer que $\chi^2(1)$ admet la densité

$$x \mapsto \frac{1}{\sqrt{2\pi}} x^{-1/2} e^{-\frac{x}{2}} \mathbb{1}_{\mathbb{R}_+}(x)$$

   En déduire que $\chi^2(n)$ est une loi à densité.

9. Soit A la matrice d'un projecteur orthogonal de rang r et X une variable aléatoire suivant la loi $\mathcal{N}(0, A)$. Montrer que $\|X\|_2^2$ suit la loi $\chi^2(r)$.

10. Soit X une variable aléatoire suivant la loi $\mathcal{N}(0,1)$. Montrer que $Y = (-1)^{\mathbb{1}_{\{|X|>1\}}} X$ suit la loi $\mathcal{N}(0,1)$, mais que (X, Y) n'est pas gaussien.

Chapitre 3

Espérance conditionnelle

3.1 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

$$A = \underset{i \in B}{\cup} A_i,$$

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

On a alors

$$\begin{aligned}
\mathbb{E}X\mathbb{1}_A &= \mathbb{E}X\sum_{i=1}^{N} \mathbb{1}_{i \in B}\mathbb{1}_{A_i} \\
&= \sum_{i=1}^{N} \mathbb{E}\mathbb{1}_{i \in B}X\mathbb{1}_{A_i} \\
&= \sum_{i=1}^{N} \mathbb{1}_{i \in B}\mathbb{E}X\mathbb{1}_{A_i}
\end{aligned}$$

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

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

Mais

$$\mathbb{E}X'1\!\!1_{A_i} = \sum_{j=1}^{N} \frac{\mathbb{E}X1\!\!1_{A_j}}{\mathbb{P}(A_j)}\mathbb{E}1\!\!1_{A_i}1\!\!1_{A_j} = \mathbb{E}X1\!\!1_{A_i}.$$

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

$$\mathbb{E}X1\!\!1_A = \mathbb{E}X'1\!\!1_A \tag{3.1}$$

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 (3.1) 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.

3.2 construction

Lemme 2. *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*

$$\mathbb{E}X'1\!\!1_A \le \mathbb{E}Y'1\!\!1_A \tag{3.2}$$

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

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

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

Ainsi $\mathbb{E}(X' - Y')1\!\!1_A \le 0$. Mais $(X' - Y')1\!\!1_A$ est positive, donc $(X' - Y')1\!\!1_A = 0$ 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

$$\forall f \in V_2 \quad \langle f - Pf, \mathbb{1}_A \rangle = 0, \tag{3.3}$$

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

$$\mathbb{E} f = \mathbb{E} Pf. \tag{3.4}$$

L'équation (3.3) 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 $P(\{F \in A\}\Delta\{G \in A\}) \le \mathbb{P}(F \neq G) = 0$. Ainsi

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

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

Ainsi, si $f \in L^p$, il est licite de noter $1\!\!1_{\{f\in A\}}$ 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 $f = f.1 = f(1\!\!1_{f>0} + 1\!\!1_{f=0} + 1\!\!1_{f<0}) = f(1\!\!1_{f>0} + 1\!\!1_{f<0}) = f^+ - f^-$, où $f^+ = f1\!\!1_{f>0}$ et $f^- = -f1\!\!1_{f>0}$. Il est facile de voir que $f^+ = f1\!\!1_{f>0}$ et $f^+ = f1\!\!1_{f>0}$

Démontrons maintenant l'analogue du lemme 3.1 : si f est un élément positif de L^2, on a

$$\mathbb{E}f1\!\!1_{Pf<0} = \mathbb{E}Pf1\!\!1_{Pf<0}$$

Mais $f1\!\!1_{Pf<0}$ est une classe de fonctions positives, donc $\mathbb{E}f1\!\!1_{Pf<0} \ge 0$, tandis que $Pf1\!\!1_{Pf<0}$ est une classe de fonctions négatives, donc $\mathbb{E}Pf1\!\!1_{Pf<0} \le 0$, finalement $Pf1\!\!1_{Pf<0} = 0$, 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

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

Mais Pf^+ est positive, donc $\|Pf^+\|_1 = \mathbb{E}Pf^+$. En utilisant (3.4), il vient que $\|Pf^+\|_1 = \mathbb{E}f^+$. De la même manière $\|Pf^-\|_1 = \mathbb{E}f^-$. Finalement, on a

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

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 :

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

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^∞,

$$\mathbb{E}gf = \mathbb{E}g\tilde{P}f \tag{3.5}$$

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

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

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.

3.2.1 Propriétés

D'après le lemme 2, on voit donc que les propriétés :Y $\mathcal{A}$-mesurable et

$$\forall A \in \mathcal{A} \ \mathbb{E}X\mathbb{1}_A = \mathbb{E}Y\mathbb{1}_A \tag{3.6}$$

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 :

- $\mathbb{E}[aX + bY|\mathcal{A}] = a\mathbb{E}[X|\mathcal{A}] + b\mathbb{E}[Y|\mathcal{A}]$
- $X \leq Y$ p.s. $\Longrightarrow \mathbb{E}[X|\mathcal{A}] \leq \mathbb{E}[Y|\mathcal{A}]$ p.s.

– $X \geq 0$ p.s. $\Longrightarrow \mathbb{E}[X|\mathcal{A}] \geq 0$ 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 :
– $\mathbb{E}\mathbb{E}[X|\mathcal{A}] = \mathbb{E}X$
– $\mathbb{E}[\mathbb{E}[X|\mathcal{A}]|\mathcal{A}] = \mathbb{E}[X|\mathcal{A}]$
– $|\mathbb{E}[X|\mathcal{A}]| \leq \mathbb{E}[|X|\ |\mathcal{A}]$

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

Théorème 19. *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'espérange 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 20. *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 $Y_n = Y\mathbb{1}_{|Y|\le n}$. 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 $|XY_n| \le |XY|$. 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 21. *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 22. *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. □

3.2.2 Inégalité de Jensen

Théorème 23. *Soit X une variable aléatoire intégrable à valeurs dans l'intervalle I. Soit f une fonction convexe de I dans $\mathbb{R}$. On suppose que $f(X)$ est intégrable. Alors, pour toute tribu $\mathcal{A}$, on a*

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

Démonstration. Soit Φ l'ensemble des fonctions affines φ telles que

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

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

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

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

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

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

Mais pour tout x $\sup\{\varphi(x); \varphi \in \Phi\} = f(x)$ (il suffit de considérer l'application affine tangente à droite (ou à gauche) au point x), et donc finalement en appliquant cette propriété au point $\mathbb{E}[X|\mathcal{A}]$, on a

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

□

Exercice : une grosse erreur est cachée dans cette preuve. La trouverez-vous ?

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

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

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

$$\mathbb{E}[Y|X] = f(X) \tag{3.7}$$

On écrit souvent $\mathbb{E}[Y|X = x] = f(x)$ pour signifier que l'application f vérifie l'équation (3.7). 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.

Théorème 24. *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 $\mathbb{R}^n$:

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

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

3.3 Exercices sur l'espérance conditionnelle

1. Soient X et Y deux variables aléatoires intégrables indépendantes. Calculer $\mathbb{E}[(1+X)(1+Y)|X]$.
2. Soient $X_1,\dots,X_n$ des variables aléatoires indépendantes identiquement distribuées admettant un moment d'ordre 1. Calculer
$$\mathbb{E}[X_1|(X_1+X_2+\dots X_n)].$$
3. 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.
4. 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 $\mathbb{E}M_n=O(\frac{\ln n}{n})$.
   (a) Montrer que pour toute variable aléatoire réelle X à valeurs dans $[0,1]$ et tout h réel positif, on a
$$\mathbb{E}X\le h+\mathbb{P}(X\ge h).$$
   (b) Pour $i\in\{1,\dots,n\}$, on note
$$A_i(h)=\{\exists j\in\{1,\dots n\}X_j\ge X_i+h\}\cap\bigcap_{1\le j\le n,i\ne j}\{X_j\notin]X_i,X_i+h[\}.$$
   Montrer que
$$\{M_n\ge h\}=\bigcup_{i=1}^{n}A_i(h).$$
   (c) On pose
$$Y_i(h)=\prod_{j\ne i}1\!\!1_{]X_i,X_i+h[^c}(X_j).$$
   Montrer que $1\!\!1_{A_i}(h)\le 1\!\!1_{\{X_i\le 1-h\}}Y_i(h)$.
   (d) Montrer que
$$\mathbb{E}[Y_i(h)|X_i]=\max(X_i,1-h)^{n-1}$$

(e) En déduire que $\mathbb{P}(A_i(h)) \leq (1-h)^n$.

(f) Conclure.

5. (a) 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 $n \in \mathbb{N}$ $\mathbb{E}1\!\!1_{\{N=n\}}Y = \mathbb{E}1\!\!1_{\{N=n\}}X$

   (b) 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 $\mathbb{E}[U|T]$.

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

7. 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

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

8. 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

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

9. Soit

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

   (a) Montrer qu'on peut construire un vecteur gaussien centré (X,Y,Z) admettant M comme matrice de covariance.

   (b) Calculer $\mathbb{E}[X|Y,Z]$.

10. Soit
$$M = \begin{pmatrix} 3 & 0 & 2 \\ 0 & 2 & 1 \\ 2 & 1 & 2 \end{pmatrix}$$
   (a) Montrer qu'on peut construire un vecteur gaussien centré (X, Y, Z) admettant M comme matrice de covariance.
   (b) Calculer $\mathbb{E}[X|Y, Z]$.
   (c) 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
$$\sigma^2 = \langle Mv, v\rangle \text{ avec } v = (1, \frac{2}{3}, -\frac{4}{3})'.$$

Chapitre 4

Martingales

4.1 Définitions

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.

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 1. *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 $\mathbb{E}[X_n] \geq \mathbb{E}[\mathbb{E}[X_{n+1}|\mathcal{F}_n]] = \mathbb{E}[X_{n+1}]$. □

4.2 Premières inégalités

4.2.1 Martingales et fonctions convexes

Théorème 1. *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.

4.2.2 Inégalité de Kolmogorov

Théorème 2. *Soit $(X_n)_{n\geq 0}$ une sous-martingale. Pour tout $\alpha > 0$, on a*

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

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

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

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

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

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

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

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

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

ce qui donne le résultat voulu. □

4.3 Convergence des martingales de carré intégrable

Théorème 3. *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

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

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

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

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

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

Ainsi

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

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

□

4.4 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 2. *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 $t \in \mathbb{R}$ $\{T \leq t\} \in \mathcal{F}_T$. Soit $n \in \mathbb{N}$: Si on note i la partie entière de t, on a

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

□

Théorème 4 (Théorème de Hunt)**.** *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. Il s'agit de montrer que pour tout événement A $\mathcal{F}_S$-mesurable, on a $\mathbb{E}X_T 1\!\!1_A \geq \mathbb{E}X_S 1\!\!1_A$. Soit M un entier déterministe tel que l'on ait $S \leq T \leq M$. Posons $\Delta_k = X_k - X_{k-1}$, avec $X_{-1} = 0$. Notons, que comme (X_n) est une sous-martingale $\mathbb{E}1\!\!1_B \Delta_k$ est positif pour tout ensemble B $\mathcal{F}_{k-1}$-mesurable. On a

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

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 $(\mathcal{F}_n)_{n\ge 0}$ une filtration et $(X_n)_{n\ge 0}$ une 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

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

Théorème 5 (Théorème d'arrêt)**.** *Soit $(\mathcal{F}_n)_{n\ge 0}$ une filtration et $(X_n)_{n\ge 0}$ une sous-martingale adaptée à la filtration $(\mathcal{F}_n)_{n\ge 0}$. Soit T un temps d'arrêt adapté à la filtration $(\mathcal{F}_n)_{n\ge 0}$. Alors la suite $(X_{n\wedge T})_{n\ge 0}$ est une sous-martingale adaptée à la filtration $(\mathcal{F}_n)_{n\ge 0}$.*

Démonstration. Soit A un événement $\mathcal{F}_n$-mesurable et n un entier. On a

$$\begin{aligned}
\mathbb{E}\mathbb{1}_A X_{(n+1)\wedge T} &= \mathbb{E}\mathbb{1}_A X_{(n+1)\wedge T}\mathbb{1}_{\{T\le n\}} + \mathbb{E}\mathbb{1}_A X_{n+1\wedge T}\mathbb{1}_{\{T>n\}} \\
&= \mathbb{E}X_{n\wedge T}\mathbb{1}_A\mathbb{1}_{\{T\le n\}} + \mathbb{E}X_{(n+1)\wedge T}\mathbb{1}_A\mathbb{1}_{\{T>n\}}
\end{aligned}$$

Pour l'instant, on a juste utilisé que quand $T \le n$, 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 $A \cap \{T > n\} \cap \{n\wedge T \le p\}$ est dans $\mathcal{F}_p$. Si $n > p$, l'intersection est l'ensemble vide, donc est $\mathcal{F}_p$-mesurable. Si $n \le p$, alors $\{n \wedge T \le p\} = \Omega$, donc

$$A \cap \{T > n\} \cap \{n \wedge T \le p\} = A \cap \{T > n\} \in \mathcal{F}_n \subset \mathcal{F}_p.$$

Ainsi, en appliquant le théorème de Hunt aux temps d'arrêts $(n+1)\wedge T$ et $n \wedge T$, on a

$$\mathbb{E}X_{(n+1)\wedge T}\mathbb{1}_A\mathbb{1}_{\{T>n\}} \ge \mathbb{E}X_{n\wedge T}\mathbb{1}_A\mathbb{1}_{\{T>n\}}$$

D'où

$$\begin{aligned}\mathbb{E}1\!\!1_A X_{(n+1)\wedge T} &= \mathbb{E}X_{n\wedge T}1\!\!1_A 1\!\!1_{\{T\le n\}} + \mathbb{E}X_{(n+1)\wedge T}1\!\!1_A 1\!\!1_{\{T>n\}} \\ &\ge \mathbb{E}X_{n\wedge T}1\!\!1_A 1\!\!1_{\{T\le n\}} + \mathbb{E}X_{n\wedge T}1\!\!1_A 1\!\!1_{\{T>n\}} \\ &\ge \mathbb{E}X_{n\wedge T}1\!\!1_A\end{aligned}$$

ce qui achève la preuve.

□

On en déduit facilement les deux résultats suivants :

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

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

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

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

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

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

Lemme 4. *Soit $(\mathcal{F}_n)_{n\ge 0}$ une filtration et $(A_n)_{n\ge 0}$ une suite d'événement adaptée à la filtration $(\mathcal{F}_n)_{n\ge 0}$. Soit T un temps d'arrêt adapté à la filtration $(\mathcal{F}_n)_{n\ge 0}$. Alors, la variable aléatoire S définie par*

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

est un temps d'arrêt.

Démonstration. Il suffit de remarquer que pour tout n

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

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

4.5 Convergence L^1 des martingales

4.5.1 Théorème des traversées montantes

Théorème 6. *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 4, 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 $\tau_0 \leq \sigma_1 \leq \tau_1 \leq \sigma_2 \leq \dots \leq \sigma_n \leq \tau_n$.

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. □

4.5.2 Le théorème de convergence de Doob

Théorème 7. *Soit $(X_n)_{n\geq 1}$ une sous-martingale. Si $K = \sup_{n\geq 1} \mathbb{E}|X_n| < +\infty$, alors la suite $(X_n)_{n\geq 1}$ converge presque surement vers une variable aléatoire X qui est telle que $\mathbb{E}|X| \leq K$.*

Démonstration. 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

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

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

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

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 $\mathbb{E}|X| = \mathbb{E} \underline{\lim}_{n\to+\infty} |X_n| \leq \underline{\lim}_{n\to+\infty} \mathbb{E}|X_n| \leq K$. Ainsi, $|X|$ est intégrable, ce qui implique qu'elle prend presque sûrement ses valeurs dans $\mathbb{R}$.

□

4.6 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 8. *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 8 (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 3, C_S est $\mathcal{F}_S$-mesurable et $\mathbb{E}[C_S|\mathcal{F}_S] = C_S$. Finalement, $\mathbb{E}[X_T|\mathcal{F}_S] \geq M_S + C_S = X_S$. □

4.7 Exercices sur les martingales

1. 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 $X_{n+1} = U_{n+1} + (1 - U_{n+1})X_n^2$.
   (a) Montrer que la suite $(X_n)_{n\geq 0}$ est une sous-martingale.
   (b) On suppose maintenant que $a \in [0, 1]$.
      i. Montrer que $(X_n)_{n\geq 0}$ converge presque sûrement vers une variable aléatoire X^*.
      ii. Donner la valeur de X^*·
2. Soit $(X_n)_{n\geq 1}$ 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
$$S_n = \sum_{k=1}^{n} X_i.$$
   (a) On pose $Y_n = S_n^3 - 3nS_n$. Montrer que $(Y_n)_{n\geq 1}$ est une martingale par rapport à la filtration $(\mathcal{F}_n)_{n\geq 1}$.
   (b) 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 $(\mathcal{F}_n)_{n\geq 1}$?

3. 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.
   (a) Montrer que $(X_n)_{n\geq 1}$ est une martingale.
   (b) Montrer que pour tout réel a les suites $(X_n - a)^+_{n\geq 1}$ et $(X_n - a)^-_{n\geq 1}$ sont des martingales.
   (c) En déduire que pour $n > p \geq 1$, X_n est presque sûrement supérieur ou égal à a sur l'événement $\{X_p \geq a\}$.
   (d) En déduire que $(X_n)_{n\geq 1}$ est presque sûrement constante.
4. 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
$$Y_n^t = e^{tS_n - n\varphi(t)}.$$
   (a) Montrer que $(Y_n^t)_{n\geq 1}$ est une martingale par rapport à la filtration naturelle associée aux $(X_n)_{n\geq 1}$.
   (b) On suppose désormais que t est non-nul. Montrer que $\varphi(t/2) < \varphi(t)/2$.
   (c) En déduire que $(Y_n^t)_{n\geq 1}$ converge presque sûrement vers 0.
   (d) Retrouver ce résultat à partir de la loi forte des grands nombres.
5. Soit $(\mathcal{F}_n)_{n\geq 1}$ une filtration de l'espace $(\Omega, \mathcal{F}, P)$. 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 $Y_n = \mathbb{E}[Z|\mathcal{F}_n]$.
   (a) Montrer que $(Y_n)_{n\geq 1}$ est une martingale par rapport à la filtration $(\mathcal{F}_n)_{n\geq 1}$.
   (b) Démontrer qu'il existe une variable aléatoire Y intégrable vers laquelle $(Y_n)_{n\geq 1}$ converge presque sûrement.
   (c) Soit W une variable aléatoire intégrable et (T_n) une suite bornée dans L^1. Montrer que
$$\lim_{a\to+\infty} \sup_{n\geq 1} \mathbb{E}|W|\mathbb{1}_{\{|T_n|\geq a\}} = 0.$$
   (d) Démontrer que la suite (Y_n) est équiintégrable sous $\mathbb{P}$, c'est à dire que
$$\lim_{a\to+\infty} \sup_{n\geq 1} \int_{\mathbb{R}\setminus[-a,a]} |x| \, d\mathbb{P}_{Y_n}(x) = 0.$$
   (e) En déduire que Y_n converge en norme 1 vers Y.

(f) Montrer que $Y = \mathbb{E}[Z|\mathcal{F}_\infty]$.

6. 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.

7. *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 $i \in \{1, \ldots, n-1\}$ 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 $k \in \{0, \ldots, n-1\}$ et tout $\theta \in]0, 1[$.
Mon pion part d'un point $i_0 \in \{1, \ldots, n-1\}$; 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 $X_0, \ldots, X_n$.
   (a) Soit g une fonction concave supérieure à f. Montrer que $(g(X_n))_{n\in\mathbb{N}}$ est une surmartingale.
   (b) Soit T un temps d'arrêt fini p.s. Montrer que
   $$\mathbb{E}(f(X_T)) \leq \mathbb{E}(g(X_T)) \leq g(i_0).$$
   (c) 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
   $$\mathbb{E}(f(X_T)) \leq \Psi(i_0).$$
   (d) 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 $[s, t]$.
   (e) 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 $\mathbb{E}(f(X_{T_{\mathrm{opt}}}))$ en fonction de $f(A)$ et $f(B)$, et en déduire que

$$\mathbb{E}(f(X_{T_{\mathrm{opt}}})) = \Psi(i_0).$$

Chapitre 5

Loi d'un processus

5.1 Loi d'un processus

Définition: Un processus stochastique est une famille $(X_t)_{t\in T}$ 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. Dans ce cours, nous ne considérerons que des processus indexés par un ensemble dénombrable. Aussi tous les théorèmes que nous donnerons seront-ils énoncés dans le cas où $T = \mathbb{N}$, à charge pour le lecteur sérieux de faire les modifications qui s'imposent...et surtout de consulter des livres plus à même de satisfaire sa curiosité.

Concrètement pour nous, un processus stochastique ne sera rien d'autre qu'un famille $X = (X_n, n \in \mathbb{N})$ de variables aléatoires définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$

$$\forall n \in \mathbb{N}, X_n : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

Définition: On appelle trajectoire de X tout élément $X(\omega) = (X_n(\omega), n \in \mathbb{N}), \omega \in \Omega$.

Définition: On définit la tribu borélienne sur $\mathbb{R}^{\mathbb{N}}$, notée $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$, comme étant la plus petite tribu qui rend mesurable les projections $\Pi_i : \mathbb{R}^{\mathbb{N}} \longrightarrow \mathbb{R}, \omega = (\omega_n)_{n\in\mathbb{N}} \longmapsto \omega_i$.

Définition: Pour toute partie non vide S et S' de $\mathbb{N}$ 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: Étant donné un processus $X = (X_n, n \in \mathbb{N})$ où

$$\forall n \in \mathbb{N}, X_n : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

$X : \omega \longmapsto X(\omega)$ est une application mesurable de Ω dans l'espace des trajectoires $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$. L'image de $\mathbb{P}$ par X s'appelle loi du processus X, notée $\mathbb{P}_X$.

Notation : $\mathcal{P}^*(\mathbb{N}), \mathcal{F}(\mathbb{N}), \mathcal{D}(\mathbb{N})$ désignent respectivement l'ensemble des parties non vides, l'ensemble des parties finies non vides et l'ensemble des parties dénombrables non vides de $\mathbb{N}$.

Il est clair que $\mathcal{F}(\mathbb{N}) \subseteq \mathcal{D}(\mathbb{N}) \subseteq \mathcal{P}^*(\mathbb{N})$.

Définition: On appelle processus extrait d'un processus $X = (X_n, n \in \mathbb{N})$ tout processus $X_S = (X_n, n \in S), S \in \mathcal{P}^*(\mathbb{N})$.

Définition: On appelle loi de dimension finie d'un processus $X = (X_n, n \in \mathbb{N})$ la loi de tout processus extrait $X_S, S \in \mathcal{F}(\mathbb{N})$.

Proposition 3. *Les lois de dimension finie d'un processus $X = (X_n, n \in \mathbb{N})$ où*

$$\forall n \in \mathbb{N}, X_n : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

sont les images de la loi $\mathbb{P}_X$ de X par les projections de $\mathbb{R}^{\mathbb{N}}$ sur les espaces-produits finis $\mathbb{R}^S, S \in \mathcal{F}(\mathbb{N})$.

Théorème 9. *Deux processus stochastiques $X = (X_n, n \in \mathbb{N})$ et $X' = (X'_n, n \in \mathbb{N})$ où*

$$\forall n \in \mathbb{N}, X_n : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

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 $P'_{X'}$ ont les mêmes images par projection sur les espaces-produits de dimension finie $(\mathbb{R}^S, \mathcal{B}(\mathbb{R}^S))$. On considère

$$\mathcal{C} = \{\prod_{n \in \mathbb{N}} A_n, \forall n, A_n \in \mathcal{B}(\mathbb{R}), \text{ et } \exists N, \forall n \geq N A_n = \mathbb{R}\}$$

$\mathbb{P}_X$ et $P'_{X'}$ 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 $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$, donc X et X' ont la même loi. $\square$

Définition: On appelle processus canonique associé à un processus stochastique $X = (X_n, n \in \mathbb{N})$ où $\forall n \in \mathbb{N}, X_n : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ le processus $\Pi = (\Pi_n, n \in \mathbb{N})$ formé par les projections Π_n de $\mathbb{R}^{\mathbb{N}}$ sur $\mathbb{R}$, qui à $X = (X_k, k \in \mathbb{N})$ associe X_n, sa n-ième composante.

Théorème 10. *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. $\Pi : (X_n)_{n\in\mathbb{N}} \longmapsto (X_n(\omega), \omega \in \Omega)_{n\in\mathbb{N}}$ est l'application identité de $\mathbb{R}^{\mathbb{N}}$. Donc l'image de $\mathbb{P}_X$ par Π est $\mathbb{P}_X$. □

5.2 Théorème d'existence de Kolmogorov (admis)

Définition: On se place sur $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$. On dit que $(Q_S, S \in \mathcal{F}(\mathbb{N}))$ où pour tout S, Q_S désigne une probabilité sur $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$, est un système projectif de lois si pour tout S, S' de $\mathcal{F}(\mathbb{N})$ tels que $S \supseteq S'$, $Q_{S'}$ est l'image de Q_S par $\Pi^S_{S'}$

Théorème 11. *Théorème d'existence de Kolmogorov (admis).*
On se place sur $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$. Pour toute partie $S \in \mathcal{F}(\mathbb{N})$, soit Q_S une mesure de probabilité sur $(\mathbb{R}^S, \mathcal{B}(\mathbb{R}^S))$.

1. *pour qu'il existe un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et pour tout $n \in \mathbb{N}$ une variable aléatoire $X_n : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que $(Q_S, S \in \mathcal{F}(\mathbb{N}))$ soit l'ensemble des lois de dimension finie du processus $X = (X_n, n \in \mathbb{N})$*
2. *ou pour qu'il existe une mesure de probabilité Q sur $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ dont l'image par Π_S soit Q_S quelle que soit la partie finie, non vide S de $\mathbb{N}$,*
3. *il faut et il suffit que $(Q_S, S \in \mathcal{F}(\mathbb{N}))$ soit un système projectif de lois.*

Exemple: 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

$$\forall B \in \mathcal{B}(\mathbb{R}^n), \mathbb{P}_{n+1}(B \times \mathbb{R}) = \mathbb{P}_n(B)(*).$$

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 $S = \{s_1, \ldots, s_p\}$, on note $\mathbb{P}_S$ l'image de $\mathbb{P}_{\max(S)}$ par la projection $\Pi_S^{\{1,\ldots,\max(S)\}}$. 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 $\Pi_S^{\{1,\ldots,n\}}$, ce qui se fait facilement par récurrence.
Exemple: 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 $\forall n \geq 1, \mathbb{P}_{X_n} = \mu_n$. En effet, si pour tout $n \geq 1$, on pose $\mathbb{P}_n = \mu_1 \otimes \cdots \otimes \mu_n$, la condition $(*)$ est satisfaite puisque $\forall n \geq 1, \forall B \in \mathcal{B}(\mathbb{R}^n)$,

$$\mathbb{P}_{n+1}(B\times\mathbb{R}) = \mu_1\otimes\cdots\otimes\mu_n\otimes\mu_{n+1}(B\times\mathbb{R}) = \mu_1\otimes\cdots\otimes\mu_n(B)\mu_{n+1}(\mathbb{R}) = \mathbb{P}_n(B)$$

D'après l'exemple précédent, il existe une suite $(X_n, n \geq 1)$ de variables aléatoires simultanées telle que

$$\forall n \geq 1, \mathbb{P}_{(X_1,\ldots,X_n)} = \mu_1 \otimes \cdots \otimes \mu_n$$

On en déduit que $\forall n \geq 1, \mathbb{P}_{X_1} = \mu_1, \ldots, \mathbb{P}_{X_n} = \mu_n$ et que les variables aléatoires réelles $X_1, \ldots, X_n$ sont indépendantes.

5.3 Processus réels stationnaires

Définition: Un processus stochastique réel $(X_n, n \in \mathbb{N})$ est dit stationnaire si quels que soient les entiers $d \geq 1, n_1, \ldots, n_d$ choisis dans $\mathbb{N}$, tels que $n_1 < \cdots < n_d$, les vecteurs aléatoires réels d-dimensionnels $(X_{n_1}, \ldots, X_{n_d})$ et $(X_{n_1+1}, \ldots, X_{n_d+1})$ suivent la même loi.

Il en résulte évidemment que $(X_{n_1}, \ldots, X_{n_d})$ et $(X_{n_1+h}, \ldots, X_{n_d+h})$ 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 12. *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 $\mathbb{P}$.

– $\forall A \in \mathcal{F}, \mathbb{P}(T(A)) = \mathbb{P}(T^{-1}(T(A))) = \mathbb{P}(A)$.

$\square$

Théorème 13. *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 $\mathcal{F}$. $\square$

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 14. *Pour qu'un processus réel* $X = (X_n, n \in \mathbb{N})$ *soit stationnaire, il faut et il suffit que l'opérateur de translation* θ *conserve sa loi* $\mathbb{P}_X$.

Démonstration. D'après le théorème précédent, θ conserve $\mathbb{P}_X$ si et seulement si pour tout cylindre mesurable à base finie C de $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}})), \mathbb{P}_X(\theta^{-1}(C)) = \mathbb{P}_X(C)$. Un tel cylindre s'écrivant sous la forme $\Pi^{-1}_{(s_1,\ldots,s_n)}(B)$,où $n \in \mathbb{N}^*, s_1, \ldots, s_n \in \mathbb{N}$ tels que $s_1 < \cdots < s_n$, et $B \in \mathcal{B}(\mathbb{R}^n)$, la condition nécessaire et suffisante s'écrit : $\forall n \in \mathbb{N}^*, \forall s_1 < \cdots < s_n \in \mathbb{N}, \forall B \in \mathcal{B}(\mathbb{R}^n)$,

$$\begin{aligned}
\mathbb{P}_{(X_{s_1+1},\ldots,X_{s_n+1})}(B) &= \mathbb{P}_X(\Pi^{-1}_{(s_1+1,\ldots,s_n+1)}(B)) \\
&= \mathbb{P}_X(\theta^{-1}(\Pi^{-1}_{(s_1,\ldots,s_n)}(B))) \\
&= \mathbb{P}_X(\Pi^{-1}_{(s_1,\ldots,s_n)}(B)) \\
&= \mathbb{P}_{(X_{s_1},\ldots,X_{s_n})}(B).
\end{aligned}$$

Elle équivaut à l'égalité des lois $\mathbb{P}_{(X_{s_1+1},\ldots,X_{s_n+1})}$ et $\mathbb{P}_{(X_{s_1},\ldots,X_{s_n})}$, quels que soient $n \in \mathbb{N}^*$, et $s_1, \ldots, s_n \in \mathbb{N}$ avec $s_1 < \cdots < s_n$ ce qui est la définition de la stationnarité du processus X. $\square$

Les processus stationnaires fournissent ainsi un exemple fondamental de transformation conservant la mesure.

Théorème 15. *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.
1. Soit $n \in \mathbb{N}^*$, $s_1 < \cdots < s_n$ dans $\mathbb{N}$. On a :

$$\begin{array}{rcl} T : (\Omega, \mathcal{F}, \mathbb{P}) & \rightarrow & (\Omega, \mathcal{F}, \mathbb{P}) \\ (\xi \circ T^{s_1}, \ldots, \xi \circ T^{s_n}) : (\Omega, \mathcal{F}, \mathbb{P}) & \rightarrow & (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)) \end{array}$$

   puis
$$(\xi \circ T^{s_1+1}, \ldots, \xi \circ T^{s_n+1}) : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$$

   par composition. Alors la loi de $(\xi \circ T^{s_1+1}, \ldots, \xi \circ T^{s_n+1})$ est l'image de $\mathbb{P}$ par $(\xi \circ T^{s_1}, \ldots, \xi \circ T^{s_n})$ (car T conserve la mesure), qui est la loi de $(\xi \circ T^{s_1}, \ldots, \xi \circ T^{s_n})$. $(\xi \circ T^n, n \geq 1)$ est donc stationnaire par définition.
2. Même démonstration en remplaçant $s_1, \ldots, s_n \in \mathbb{N}$ par $s_1, \ldots, s_n \in \mathbb{Z}$.

□

5.4 Processus gaussiens

5.4.1 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 4. *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.

□

5.4.2 Condition d'existence

Théorème 16. *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. □

5.4.3 Processus gaussiens stationnaires

Théorème 17. *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. □

5.5 Exercices sur les processus

1. 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.
2. 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.
3. Soit φ une application mesurable de $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ et $(X_n)_{n\geq 1}$ un processus stationnaire. Démontrer que le processus $(Y_n)_{n\geq 1}$ défini par $Y_n = \varphi(X_n, X_{n+1}, \dots) = \varphi(\theta^{n-1} \circ X)$ est stationnaire.
4. 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.

5. 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.
6. 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 $[0, 1]$.
7. 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 $x \mapsto e^{2i\pi x}$?
8. 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.

Chapitre 6

Chaînes de Markov

6.1 Dynamique markovienne

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 $(\Omega, \mathcal{F}, P)$. 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 $(i,j) \in S \times S$ $p_{i,j} = \nu(j)$.

Propriétés :

1. Si $(X_n)_{n\geq 0}$ est une chaîne de Markov de matrice de passage P et que $x_0, \dots, x_{n-1}$ sont tels que $\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots X_{n-1} = x_{n-1})$, alors $\mathbb{P}(X_n = x_n | X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}) = p_{x_{n-1}, x_n}$. Autrement dit,
$$\mathbb{P}(X_n = x_n | X_0, \dots, X_{n-1}) = p_{X_{n-1}, x_n}. \quad (6.1)$$
Cela signifie que toute l'information que $X_0, \dots, X_{n-1}$ peuvent nous apporter sur X_n est comprise dans X_n.
2. (6.1) implique que $\mathbb{P}(X_n = x_n | X_{n-1}) = p_{X_{n-1}, x_n}$

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 suites 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.

Lemme 5. *Soit S un ensemble fini ou dénombrable, ν une loi sur S et χ une mesure sur $S^S = \mathcal{F}(S,S)$.*

Soit $(f_n)_{n\geq 1}$ une suite de v.a.i.i.d. 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.

6.2 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

$$\forall i \in S \quad \sum_{j\in S} p_{i,j} = 1.$$

6.2.1 Existence des chaînes de Markov

Théorème 18. *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

$$\chi_P = \underset{i\in S}{\otimes} \mu_i,$$

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 $(X_n)_{n\ge 0}$ 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 $(X_k)_{k\ge 0}$ 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 $n \ge 1$ 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 9. *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*

$$\mathbb{P}^\nu = \int \mathbb{P}^i \, d\nu \tag{6.2}$$

c'est à dire que pour tout borélien A de $S^{\mathbb{N}}$, on a

$$\mathbb{P}^\nu(A) = \int \mathbb{P}^i(A) \, d\nu \tag{6.3}$$

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 $\mathbb{P}_X$.

6.2.2 Puissances des matrices stochastiques

Théorème 19. *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

$$\begin{aligned}
\mu_{n+1}(j) &= \mathbb{P}^\nu(X_{n+1} = j) \\
&= \sum_{i \in S} \mathbb{P}^\nu(X_n = i, X_{n+1} = j) \\
&= \sum_{i \in S} \mathbb{P}^\nu(X_n = i) p_{i,j} \\
&= \sum_{i \in S} \mu_n(i) p_{i,j} \\
&= (\mu_n M)(j)
\end{aligned}$$

□

6.2.3 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

$$\mathbb{P}^i(X_n = x_n) = \sum_{(x_0,\dots x_{n-1})\in S^n} \mathbb{P}^i(X_0 = x_0, X_1 = x_1, \dots X_{n-1} = x_{n-1}, X_n = x_n). \tag{6.4}$$

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

$$\mathbb{P}^i(\exists n > 0;\ X_n = j) = \mathbb{P}^i(\underset{n\geq 1}{\cup}\ \{X_n = j\}),$$

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.

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 6. *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 $d(i) = d(j)$. □

Définition: Si une chaîne irréductible a ses états de période 1, on dit qu'elle est apériodique.

Les lemme suivant se révélera très utile par la suite

Lemme 7. *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 10. *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 11. *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 $N = \max(N(x), x \in S) + \mathrm{diam}(G)$. □

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 $\mathbb{P}^x(X_1 = x) = 1$.

6.3 Propriété de Markov

Théorème 20. *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

$$\begin{aligned}
\mathbb{E}1\!\!1_A 1\!\!1_{\{X_{p+.}\in B\}} &= \sum_i \mathbb{E}1\!\!1_A 1\!\!1_{\{X_p=i\}} 1\!\!1_{\{X_{p+.}\in B\}} \\
&= \sum_i \mathbb{P}(A, X_p = i, X_{p+.} \in B) \\
&= \sum_i \mathbb{P}(A, X_p = i)\mathbb{P}^i(X \in B) \\
&= \sum_i \mathbb{E}[1\!\!1_{A\cap\{X_p=i\}}\mathbb{P}^i(X \in B)] \\
&= \sum_i \mathbb{E}[1\!\!1_{A\cap\{X_p=i\}} f_B(X_p)] \\
&= \mathbb{E}[1\!\!1_A f_B(X_p)]
\end{aligned}$$

□

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

$$\begin{aligned}
&\mathbb{P}((X_0, \dots X_{n-1}) \in A, X_n = i, (X_{n+1}, \dots, X_{n+p}) \in B) \\
= \; &\mathbb{P}((X_0, \dots X_{n-1}) \in A, X_n = i)\mathbb{P}^i((X_1, \dots, X_p \in B)).
\end{aligned}$$

Corollaire 12. *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 :

$$\begin{aligned}\mathbb{P}^i(X \in B) &= \mathbb{P}^i(X \in \Theta^{-1}(B)) \\ &= \mathbb{P}^i(\Theta(X) \in B) \\ &= \sum_{i:P^i(X_1=j)>0} \mathbb{P}^i(X_1 = j)\mathbb{P}^j(X \in B) \\ &= \sum_{i:p_{i,j}>0} p_{i,j}\mathbb{P}^j(X \in B)\end{aligned}$$

□

6.4 Exercices sur les chaînes de Markov

1. *Chaîne à deux états.* Soit $\{X_n : n \geq 0\}$ une chaîne de Markov à valeurs dans $\{0, 1\}$ et de probabilité de transition :

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix},\ 0 \leq \alpha, \beta \leq 1.$$

   (a) 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 $(\alpha, \beta) \neq (0, 0)$.

   (b) Vérifier que pour toute loi initiale μ, on a

$$\mathbb{P}^\mu(X_n = 0) = \frac{\beta}{\alpha+\beta} + (1-\alpha-\beta)^n\left(\mu(0) - \frac{\beta}{\alpha+\beta}\right).$$

   (c) 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 $(\alpha, \beta) \neq (1, 1)$.

(d) (*Mesure stationnaire*) Prouver que, pour tout $n \in \mathbb{N}$,

$$\mathbb{P}^{\nu}(X_n \in A) = \nu(A).$$

2. *Représentation canonique et simulation des chaînes de Markov.*
   (a) 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.
   (b) 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.
   (c) 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.
   (d) Application. Comment simuler une chaîne de Markov homogène de matrice de transition $P = (p_{i,j})$? Ecrire un algorithme explicite si

   $$P = \begin{pmatrix} 0.25 & 0.5 & 0.25 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{pmatrix}.$$

3. *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 $\mathbb{P}(X_n = a) = \mathbb{P}(T \leq n)$.

4. *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

$$f(x) = 1 \text{ pour } x \in A \text{ et } f(x) = (Qf)(x) \text{ pour } x \not inA.$$

5. *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.

6. 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 $\mathbb{P}(B \cap \{Y \notin A\}) = 0$.

7. *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.

   (a) 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.

   (b) 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

$$pu_{n+1} - u_n + qu_{n-1} = 0.$$

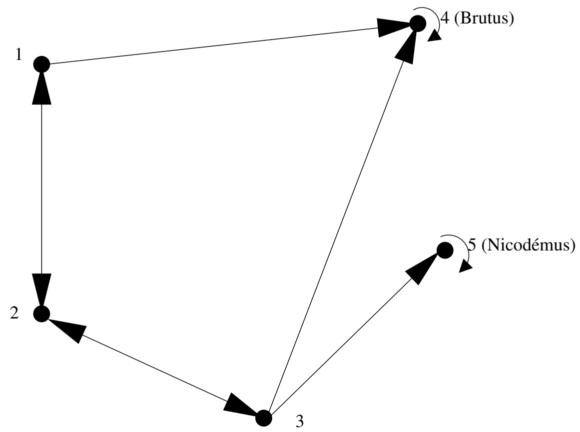
Que valent u_a et u_{-b} ?

(c) Résoudre l'équation de récurrence et en déduire

$$p_g = \frac{(\frac{q}{p})^b - 1}{(\frac{q}{p})^{a+b} - 1}. \tag{6.5}$$

8. *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 (6.5), le tout étant de le justifier...

   On suggère de poser $T' = \inf\{n; S_n \le -b\}$ et, pour tout $a > 0$, $U_a = \inf\{n; S_n \ge a\}$ et $G^a = \{U_a \le T'\}$.
9. *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 ?



10. 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 ?
11. Soit a et b des entiers supérieurs ou égaux à 2, $(D_n)_{n\ge 1}$ 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 ?

12. *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 $\mathbb{E}Y = \int_0^{+\infty} \mathbb{P}(Y > t)\ dt$.

13. *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 $\mathbb{E}^1 T, \mathbb{E}^2 T, \mathbb{E}^3 T$.

14. *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 :

$$X_{n+1} = \sum_{k=1}^{2N} \mathbb{1}_{\{Y_{n+1,k} \leq X_n\}},$$

où $(Y_{n,k})_{n\geq 1, k\in\{1,\dots,2N\}}$ est une suite de variables aléatoires indépendantes suivant la loi uniforme sur l'ensemble fini $\{1,\dots,2N\}$. X_0 est indépendante des $(Y_{n,k})$.

(a) Montrer que X_n est une chaîne de Markov à valeurs dans $E = \{0,\dots,2N\}$.

(b) 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.

(c) Montrer que $(X_n)_{n\geq 0}$ converge presque sûrement vers une variable aléatoire X_∞.

(d) Déterminer la loi de X_∞ en fonction de la loi de X_0.

15. *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 $f(0) = f(1) = 0$, $f(2) = 1$. Montrer que $(f(X_n))$ n'est pas une chaîne de Markov.

16. *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).

Chapitre 7

Récurrence et mesures invariantes

7.1 Temps d'arrêt et propriété de Markov forte

Théorème 21. *Soit S un ensemble dénombrable, $(X_k)_{k\ge 0}$ 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 $(X_{T+k})_{k\ge 0}$ 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. Comme être une chaîne de Markov est une propriété de la loi, on peut supposer que $(X_n)_{n\ge 0}$ est obtenue par le procédé décrit plus haut : $X_{n+1} = f_{n+1}(X_n)$ où $(f_n)_{n\ge 1}$ est une suite de variables aléatoires indépendantes de loi χ_M indépendante de X_0.

Posons $Y_k = X_{T+k}$ si $T < +\infty$ et $Y_k = X_k$ sinon. De même posons $g_k = f_{T+k}$ si $T < +\infty$ et $g_k = f_k$ sinon. Il est facile de voir que $(Y_k)_{k\ge 0}$ vérifie la récurrence $Y_{n+1} = g_{n+1}(Y_n)$

Soit p un entier et B un borélien de $\mathcal{F}(S,S)^{\mathbb{N}_*}$. On a

$$\begin{aligned}
\mathbb{P}(T = p, A, g \in B) &= \mathbb{P}(T = p, A, f_{p+.} \in B)\\
&= \mathbb{P}(T = p, A)\mathbb{P}(f_{p+.} \in B)\\
&= \mathbb{P}(T = p, A)\mathbb{P}(f_{p+.} \in B)
\end{aligned}$$

car comme l'événement $\{T = p\} \cap A$ est $\sigma(X_0, f_1, \dots, f_p)$-mesurable, il est indépendant de l'événement $\{f_{p+.} \in B\}$ qui est $\sigma(f_{p+1}, f_{p+2}, \dots)$-mesurable.

Maintenant la loi de $f_{p+.}$ est la même loi que celle de f : c'est $\chi_M^{\otimes \mathbb{N}^*}$. On a donc $\mathbb{P}(T=p, A, g \in B) = \mathbb{P}(T=p, A)\mathbb{P}(f \in B)$. En faisant la somme pour p variant de 1 à $+\infty$, on obtient

$$\mathbb{P}(T < +\infty, A, g \in B) = \mathbb{P}(T < +\infty, A)\mathbb{P}(f \in B) \tag{7.1}$$

Soit maintenant B' un Borélien de S^N et A' un événement $\mathcal{F}_T$-mesurable. On a

$$\{T < +\infty, X_T = i, X_{T+.} \in B'\} = \{T < +\infty, X_T = i; G_i(g) \in B'\},$$

où l'on a posé

$$G_i((h_n)_{n\geq 1}) = (i, h_1(i), h_2 \circ h_1(i), h_3 \circ h_2 \circ h_1(i), \dots)..$$

En appliquant l'égalité (7.1) précédemment démontrée avec $A = A' \cap \{X_T = i\}$ et $B = G_i^{-1}(B')$, on obtient

$$\begin{aligned} \mathbb{P}(T < +\infty, X_T = i, X_{T+.} \in B', A') &= \mathbb{P}(T < +\infty, X_T = i, A')\mathbb{P}(f \in G_i^{-1}(B')) \\ &= \mathbb{P}(T < +\infty, X_T = i, A')\mathbb{P}^i(X \in B') \end{aligned}$$

□

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 13. *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 :

$$\overline{\mathbb{P}}(X_{X_{T+.}} \in B | \mathcal{F}_T) = f_B(X_T), \text{ avec } f_B(x) = \mathbb{P}^x(X \in B).$$

En particulier, sous $\overline{\mathbb{P}}$, $X_{T+.}$ est une chaîne de Markov de loi initiale la loi de X_T sous $\overline{\mathbb{P}}$.

Ces résultats sont souvent employés dans le cas où $\mathbb{P}(T < \infty) = 1$, puisqu'alors $\mathbb{P} = \overline{\mathbb{P}}$.

Démonstration. Soit $A \in \mathcal{F}_T$. En partitionnant l'espace, on a

$$\begin{aligned}
\mathbb{P}(A, T < +\infty, X_{T+.} \in B) &= \sum_{i \in S} \mathbb{P}(A, T < +\infty, X_T = i, X_{T+.} \in B) \\
&= \sum_{i \in S} \mathbb{P}(T < +\infty, X_T = i, A)\mathbb{P}^i(X \in B),
\end{aligned}$$

où la dernière égalité vient de la propriété de Markov forte que l'on a démontrée. Ainsi,

$$\begin{aligned}
\mathbb{P}(A, T < +\infty, X_{T+.} \in B) &= \sum_{i \in S} \mathbb{P}(T < +\infty, X_T = i, A) f_B(i) \\
&= \sum_{i \in S} \mathbb{E}[\mathbb{1}_{\{T<+\infty, X_T=i, A\}} f_B(i)] \\
&= \sum_{i \in S} \mathbb{E}[\mathbb{1}_{\{T<+\infty, X_T=i, A\}} f_B(X_T)] \\
&= \mathbb{E}[\sum_{i \in S} \mathbb{1}_{\{T<+\infty, X_T=i, A\}} f_B(X_T)] \\
&= \mathbb{E}[\mathbb{1}_{\{T<+\infty, A\}} f_B(X_T)].
\end{aligned}$$

(On a utilisé le théorème de Tonelli pour échanger la somme et l'espérance.) Soit, en divisant par $\mathbb{P}(T < +\infty)$.

$$\overline{\mathbb{P}}(A, X_{T+.} \in B) = \overline{\mathbb{E}}[\mathbb{1}_A f_B(X_T)]$$

Comme $f_B(X_T)$ est $\mathcal{F}_T$-mesurable, il découle que

$$\overline{\mathbb{P}}(X_{T+.} \in B | \mathcal{F}_T) = f_B(X_T).$$

□

7.2 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 22. *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. En particulier $\mathbb{E}N_i=+\infty$.*

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

$$\mathbb{P}^i(N_i \ge k+1) = \mathbb{P}^i(N_i \ge k+1, T_i<+\infty) = \mathbb{P}^i(T_i<+\infty)\mathbb{P}^i(\sum_{k=1}^{+\infty} \mathbb{1}_{\{i\}}(X_{T+i} \ge k | T_i<+\infty).$$

Or T_i est un temps d'arrêt. Donc, d'après la propriété de Markov forte, sachant $T_i<+\infty$, $(X_{T+i})_{i\ge 0}$ 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 $(X_i)_{i\ge 0}$. On en déduit

$$\mathbb{P}^i(N_i \ge k+1) = \mathbb{P}^i(T_i<+\infty)\mathbb{P}^i(N_i\ge k).$$

Par récurrence, on en déduit

$$\mathbb{P}^i(N_i \ge k) = \mathbb{P}^i(T_i<+\infty)^k$$

D'après le théorème de continuité séquencielle décroissante, on a $P(N_i=+\infty)=\lim_{k\to+\infty}\mathbb{P}^i(N_i\ge k)$. Cette limite vaut donc 0 si i est transient, 1 si i est récurrent. Pour $k\ge 1$, on a

$$\begin{aligned}\mathbb{P}^i(1+N_i=k) &= \mathbb{P}^i(1+N_i\ge k)-P(N_i\ge k)\\ &= \mathbb{P}^i(T_i<+\infty)^{k-1}-\mathbb{P}^i(T_i<+\infty)^k\\ &= \mathbb{P}^i(T_i<+\infty)^{k-1}(1-\mathbb{P}^i(T_i<+\infty)),\end{aligned}$$

ce qui montre que $1+N_i$ suit bien une loi géométrique de paramètre $1-\mathbb{P}^i(T_i<+\infty)$. De plus

$$\mathbb{E}^i N_i = \sum_{k=1}^{+\infty}\mathbb{P}^i(N_i\ge k) = \sum_{k=1}^{+\infty} = \mathbb{P}^i(T_i<+\infty)^k = \frac{\mathbb{P}^i(T_i<+\infty)}{1-\mathbb{P}^i(T_i<+\infty)} < +\infty.$$

□

Corollaire 14. *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 15. *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 16. *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 $\{T_j < +\infty, \forall k > T_j X_k \neq i\} \subset \{N_j < +\infty\}$, donc comme i est récurrent, $\mathbb{P}^i(N_i < +\infty) = 0$, donc $\mathbb{P}^j(T_i = +\infty) = 0$. 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*.

7.3 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 $\forall n \geq 0 \mu M^n = \mu$. 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

$$\forall i, j \in S \quad \mu(i) m_{i,j} = \mu(j) m_{j,i}.$$

Théorème 23. *Soit* $(X_n)_{n\geq 0}$ *une chaîne de Markov de loi initiale* ν *réversible sous l'action de* M. *Alors*

$$\forall n \geq 1 (X_0, X_1, \dots X_n) \text{ et } (X_n, X_{n-1}, \dots, X_0) \text{ ont même loi sous } \mathbb{P}^\nu.$$

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 24. *Si la matrice de transition* M *est irréductible et admet une probabilité invariante, alors les chaînes de Markov associées à* M *sont récurrentes.*

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 $\lim\limits_{n\to+\infty} \mu(i)m_{i,j}^{(n)} = 0$ 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. $\square$

Théorème 25. *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 $(\mu_n)_{n\geq 0}$ 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 $(\mu I - M^n))_{n\geq 0}$, est bornée, il s'ensuit que toute valeur d'adhérence de $(\mu_n)_{n\geq 0}$ est laissée fixe par M. Comme $\mathcal{M}(S)$ est compacte, $(\mu_n)_{n\geq 0}$ a au moins une valeur d'adhérence donc M au moins une mesure invariante. $\square$

Corollaire 17. *Une chaîne de Markov irréductible dont l'espace d'états est fini est récurrente.*

7.4 Théorème de la probabilité stationnaire

Théorème 26. *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 8. *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 9. *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 $(g_n)_{n\geq 1}$ est une suite de v.a.i.i.d de loi χ_M.

D'après le lemme 5, (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 $(X_n)_{n\geq 0}$ et $(Y_n)_{n\geq 0}$ sont indépendants.

Ainsi, d'après le lemme 8, (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 10, 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 $((Z_n)_{n\geq 0} = ((X_n, Y_n))_{n\geq 0}$ est une chaîne de Markov irréductible. Comme $M \otimes M$ admet $\mu \otimes \mu$ comme mesure invariante, la dynamique est donc récurrente : $(Z_n)_{n\geq 0}$ 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 $n \geq \tau$, 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 $(X_n)_{n\geq 0}$ et $(Y_n)_{n\geq 0}$ ayant la même matrice de transition irréductibles, telles que $(X_n, Y_n)_{n\geq 0}$ ne soit pas irréductible et que $(X_n, Y_n)_{n\geq 0}$ ne coupe jamais la diagonale. Donner deux exemples d'un tel phénomène, l'un avec S fini, l'autre avec S infini.

7.5 Théorème ergodique des chaînes de Markov

Théorème 27. *Soit $(X_n)_{n\geq 0}$ une chaîne de Markov irréductible et récurrente. Pour tout $x \in S$, on a*

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{k=0}^{n-1} 1\!\!1_{\{x\}}(X_k) \to \frac{1}{\mathbb{E}^x T_x}.$$

Démonstration. Pour tout $k \geq 1$, posons $T^k = \inf\{n > 0, \sum_{k=1}^{n} 1\!\!1_{\{x\}}(X_k) \geq k\}$. Les T^k sont des temps d'arrêt adaptés à la filtration naturelle engendrée par $(X_n)_{n\geq 0}$. Comme la chaîne est récurrente, 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 :

$$\begin{aligned}
\mathbb{P}^x(A, T^{k+1} - T_k > n) &= \mathbb{P}^x(A, \cap_{j=T^k+1}^{T^k+n} 1\!\!1_{\{x\}}(X_j) = 0) \\
&= \mathbb{P}^x(A)\mathbb{P}^x(\cap_{j=1}^{n} 1\!\!1_{\{x\}}(X_j) = 0) \\
&= \mathbb{P}^x(A)\mathbb{P}^x(T^1 > n)
\end{aligned}$$

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 $\mathbb{E}^x T^x$. Lorsque la loi initiale n'est pas la masse de Dirac en x, il suffit de remarquer que la suite $\frac{T^n}{n}$ a le même comportement asymptotique que la suite $\frac{T^n - T^1}{n-1}$ et appliquer à nouveau la propriété de Markov forte : la loi sous $\mathbb{P}$ de $(\frac{T^n - T^1}{n-1})_{n\ge 2}$ est la loi sous $\mathbb{P}^x$ de $(\frac{T^n}{n})_{n\ge 1}$, d'où le résultat. Posons $S_n = \sum_{k=0}^{n-1} 1\!\!1_{\{x\}}(X_k)$. Un instant de réflexion montre que $T^{S_n} \le n < T^{S_n+1}$. On en déduit

$$\frac{T^{S_n}}{S_n} \le \frac{n}{S_n} < \frac{T^{S_n+1}}{S_n+1}\frac{S_n+1}{S_n}$$

Comme 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. □

Théorème 28. *Soit $(X_n)_{n\ge 0}$ une chaîne de Markov irréductible admettant une probabilité invariante μ. Pour tout $x \in S$, on a*

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} 1\!\!1_{\{x\}}(X_k) = \frac{1}{\mathbb{E}^x T_x} = \mu(x).$$

Démonstration. Une chaîne de Markov irréductible admettant une probabilité invariante est toujours récurrente. Le théorème ergodique des chaînes de Markov s'applique donc, et on a pour toute loi initiale ν :

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} 1\!\!1_{\{x\}}(X_k) = \frac{1}{\mathbb{E}^x T_x}\mathbb{P}^\nu\text{p.s.}$$

Comme $|\frac{1}{n}\sum_{k=0}^{n-1} 1\!\!1_{\{x\}}(X_k)| \le 1$, le théorème de convergence dominée s'applique et on a

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{P}^\nu(X_k = x) = \frac{1}{\mathbb{E}^x T_x}.$$

Si l'on prend pour ν la mesure invariante μ, on a pour tout $k \ge 0$ $\mathbb{P}^\nu(X_k = x) = \mu(x)$. On en déduit que $\frac{1}{\mathbb{E}^x T_x} = \mu(x)$, ce qui achève la preuve. □

Corollaire 18. *Une chaîne de Markov irréductible a au plus une probabilité invariante.*

Remarque : ce résultat admet une réciproque, qui ne sera pas démontrée dans ce cours : si une chaîne de Markov irréductible apériodique vérifie

$\mathbb{E}^x T_x < +\infty$ (ou de manière équivalente $\frac{1}{\mathbb{E}^x T_x} > 0$, ce qui fait qu'on dit de telles chaînes qu'elles sont récurrentes positives), alors la chaîne admet une mesure invariante. 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.

7.6 Exercices sur la récurrence et les mesures invariantes

1. *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

   $$T_{k+1} = T_k + 1\!\!1_{\{X_{T_k}\notin A\}} \inf\{n \geq 0, X_{T_k+n} \neq X_{T_k}\},$$

   avec la convention $0.\infty = 0$.

   (a) Montrer que les $(T_k)_{k\in\mathbb{N}}$ sont des temps d'arrêt pour $(X_n)_{n\in\mathbb{N}}$, finis presque sûrement.

   (b) 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.

2. *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 J une partie de E. On observe cette chaîne de Markov seulement lors de ses passages par J, et on note Y_m la $m^{\text{ième}}$ observation. Plus formellement, pour $m \geq 1$, on note

   $$T_m = \inf\left\{n \geq 1 + T_{m-1} \;\middle|\; X_n \in J\right\} \text{ et } T_0 = \inf\left\{n \geq 0 \;\middle|\; X_n \in J\right\}.$$

   On suppose que $\forall m \geq 0$, $T_m < +\infty$, et on pose $Y_m = X_{T_m}$.

   (a) Montrer que T_m est un temps d'arrêt, 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.

   (b) Montrer que $(Y_n)_{n\in\mathbb{N}}$ est une chaîne de Markov homogène.

3. 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”.

4. *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.
   (a) Quelle est sa matrice de transition P ? La chaîne est-elle irréductible ? apériodique ?
   (b) On suppose que d est impair. Montrer que $(X_n)_{n\geq 0}$ converge en loi et préciser la limite.
   (c) 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.
5. *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 $x = (x(1), \ldots, x(N)) \in (\mathbb{Z}/2\mathbb{Z})^N$, 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. $(\delta_1, \ldots, \delta_N)$ 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 $\{1, \ldots, N\}$. $(V_n)_{n\geq 1}$ est indépendante de X_0.
   (a) Montrer que $(X_n)_{n\geq 0}$ est une chaîne de Markov.
   (b) 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.

(c) 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 $U * \gamma = U$.

(d) Montrer que si X_0 suit la loi U, alors $X_0(1), X_0(2), \dots X_0(N)$ sont indépendantes.

(e) 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 $(Y_n)_{n\geq 0}$. .

(f) Retrouver ce dernier résultat par un calcul direct.

6. *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.

7. *Un modèle de prédiction météo (!)* On suppose que le temps qu'il fera demain depend des deux jours précédents. On suppose que :

   $\mathbb{P}($ il pleut demain $|$ il a plu hier et aujourd'hui$) = 0,7$
   $\mathbb{P}($ il pleut demain $|$ il a plu aujourd'hui mais pas hier$) = 0,5$
   $\mathbb{P}($ il pleut demain $|$ il a plu hier mais pas aujourd'hui$) = 0,4$
   $\mathbb{P}($ il pleut demain $|$ il n'a pas plu ni hier, ni aujourd'hui$) = 0,2$

   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 ?

8. *Chaîne de Markov réversible*

   (a) 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.

(b) Trouver rapidement la probabilité stationnaire de la marche aléatoire symétrique sur les sommets de l'hypercube de dimension d.

(c) 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.)

9. *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.

Annexe A

Théorème de Levy

A.1 Rappels sur la convergence en loi

On dit qu'une suite $(\mu_n)_{n\geq 1}$ de mesures de probabilités sur $\mathbb{R}^d$ converge faiblement vers la mesure de probabilité μ lorsque pour toute fonction f continue bornée de $\mathbb{R}^d$ dans $\mathbb{R}$, on a

$$\lim_{n\to+\infty} \int_{\mathbb{R}^d} f\ d\mu_n = \int_{\mathbb{R}^d} f\ d\mu.$$

Par extension, on dit qu'une suite de variables aléatoires X_n converge en loi vers la variable aléatoire X (ou vers la loi μ) si la suite de mesures P_{X_n} converge faiblement vers P_X (ou vers la loi μ).

Ainsi, dire que X_n converge en loi vers X signifie que pour toute fonction continue bornée, $\mathbb{E}f(X_n)$ converge vers $\mathbb{E}f(X)$.

Rappel : si μ et ν sont deux mesures telles que pour toute fonction f continue bornée de $\mathbb{R}^d$ dans $\mathbb{R}$, on a $\int_{\mathbb{R}^d} f\ d\mu = \int_{\mathbb{R}^d} f\ d\nu$, alors $\mu = \nu$.

On rappelle deux théorèmes très utiles dont la preuve peut être trouvée dans le cours de licence.

Théorème 25 (Théorème de Portmanteau)**.** *Les propositions suivantes sont équivalentes*

1. *μ_n converge faiblement vers μ.*
2. *Pour toute fonction f uniformément continue bornée de $\mathbb{R}^d$ dans $\mathbb{R}$, on a*

$$\lim_{n\to+\infty} \int_{\mathbb{R}^d} f\ d\mu_n = \int_{\mathbb{R}^d} f\ d\mu.$$

3. *Pour tout fermé F, $\mu(F) \geq \overline{\lim}_{n\to+\infty} \mu_n(F)$.*

4. *Pour tout ouvert O, $\mu(O) \leq \varliminf_{n\to+\infty} \mu_n(O)$.*
5. *Pour tout borélien A dont la frontière ∂A vérifie $\mu(\partial A) = 0$, on a $\lim_{n\to+\infty} \mu_n(A) = \mu(A)$.*

Théorème 26. *Si une suite $(\mu_n)_{n\geq 1}$ de mesures de probabilités sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est telle que pour toute fonction f continue positive à support compact de $\mathbb{R}^d$ dans $\mathbb{R}$, on a*

$$\lim_{n\to+\infty} \int_{\mathbb{R}^d} f \ d\mu_n = \int_{\mathbb{R}^d} f \ d\mu,$$

alors μ_n converge faiblement vers μ.

Théorème 29. *Soit X_n une suite de variables aléatoires réelles, X une autre variable aléatoire. Alors X_n converge en loi vers X si et seulement si pour tout point x où F_X est continue, $F_{X_n}(x)$ tend vers $F(x)$ lorsque n tend vers l'infini.*

A.2 Tension

Définition: On dit qu'une famille $\mathcal{M}$ de mesures de probabilités sur $\mathbb{R}^d$ est tendue si pour tout $\varepsilon > 0$, il existe un compact K de $\mathbb{R}^d$ tel que pour tout $\mu \in \mathcal{M}$, on a $\mu(K^c) \leq \varepsilon$.
Exemple: Une famille constitué d'une unique mesure μ sur $\mathbb{R}^d$ est tendue.

Démonstration. Considérer la suite d'ensemble $A_n = B(0, n)$: La suite A_n^c est décroissante et son intersection est $\varnothing$, donc d'après le théorème de continuité séquentielle décroisante $\mu(A_n)^c$ tend vers $\mu(\mathbb{R}^d)$, ce qui montre que bien que pour n assez grand $\mu(A_n^c)$ ne dépasse pas ε. □

Lemme 10. *La réunion de deux familles tendues est une famille tendue.*

Démonstration. Soit $\mathcal{M}$ et $\mathcal{N}$ deux familles tendues. Soit $\varepsilon > 0$. Comme $\mathcal{M}$ est tendue, il existe un compact K_1 tel que $\forall \mu \in \mathcal{M} \mu(K_1^c) \leq \varepsilon$. De même, il existe un compact K_1 tel que $\forall \mu \in \mathcal{N} \mu(K_2^c) \leq \varepsilon$. Maintenant, si l'on pose $K = K_1 \cup K_2$, K est compact et il est facile de voir que

$$\forall \mu \in \mathcal{M} \cup \mathcal{N} \mu(K^c) \leq \varepsilon.$$

Ainsi $\mathcal{M} \cup \mathcal{N}$. □

Corollaire 19. *Toute famille finie de mesures sur $\mathbb{R}^d$ est tendue.*

Le lemme suivant peut également être utile

Lemme 11. *Pour k entre 1 et d et x on note $\pi_k(x)$ la k-ième composante de x. Soit $\mathcal{M}$ une famille de mesures sur $\mathbb{R}^d$. $\mathcal{M}$ est tendue si et seulement si pour tout k entre 1 et d la famille $\pi_k^{-1}\mathcal{M}$ est tendue.*

Démonstration. – Supposons que $\mathcal{M}$ est tendue. Soit k entre 1 et d et $\varepsilon > 0$. Il existe un compact K tel que pour tout $\mu \in \mathcal{M}$ $\mu(K) \geq 1 - \varepsilon$. Alors, il est clair que $\pi_k K$ est compact est que pour tout $\mu \in \mathcal{M}$, on a

$$\pi_k^{-1}\mu(\pi_k K) = \mu(x \in \mathbb{R}^d : \pi_k(x) \in \pi_k K) \geq \mu(K) \geq 1 - \varepsilon.$$

– Réciproquement, supposons que $pi_k^{-1}\mathcal{M}$ est tendue. Soit $\varepsilon > 0$. Pour tout k entre 1 et d il existe un compact K_d tel que pour tout $\mu \in \mathcal{M}$ $\pi_k^{-1}\mu(K_k) \geq 1 - \varepsilon/d$. Posons $K = K_1 \times K_2 \times \ldots K_d$. On a $K^c = \cup_{i=1}^d \pi_k^{-1}(K_k^c)$, donc

$$\begin{aligned} \mu(K^c) &\leq \sum_{i=1}^d \mu(\pi_k^{-1}(K_k^c)) \\ &= \sum_{i=1}^d \pi_k^{-1}\mu(K_k^c) \\ &\leq \sum_{i=1}^d \varepsilon/d \\ &\leq \varepsilon \end{aligned}$$

□

Définition: On dit qu'une famille $\mathcal{M}$ de mesures de probabilités est relativement compacte si et seulement toute suite d'éléments de $\mathcal{M}$ admet une sous-suite convergente.

Théorème 30 (Théorème de Prohorov)**.** *Toute famille tendue est relativement compacte.*

Afin de ne pas se perdre dans des détails techniques, on va donner la preuve uniquement dans le cas où $d = 1$. On va s'appuyer sur le lemme suivant

Lemme 12 (Théorème de Helly)**.** *De toute suite $(F_n)_{n\geq 1}$ de fonctions de répartition on peut extraire une sous-suite $(F_{n_k})_{k\geq 1}$ telle qu'il existe une fonction F croissante continue à droite avec*

$$F_{n_k}(x) \to F(x)$$

en chaque point de continuité de F.

Démonstration. À l'aide du procédé diagonal d'extraction, on commence par extraire une suite $(n_k)_{k\geq 1}$ telle que $F_{n_k}(x)$ converge en tout point x rationnel. On note $G(x)$ la limite obtenue. C'est une fonction croissante. On définit alors

$$F(x) = \inf\{G(r); r \in \mathbb{Q} \cap]x, +\infty[\}.$$

Il est encore clair que F est croissante. Montrons que F est continue à droite. Soit $x \in \mathbb{R}$ et $\varepsilon > 0$. Par définition de F, il existe $r > x$, avec r rationnel tel que $G(r) < F(x) + \varepsilon$. Maintenant, on a

$$\forall y \in [x, r[\quad F(x) \leq F(y) \leq G(r) < F(x) + \varepsilon,$$

ce qui montre bien que F est continue à droite. Reste à montrer que F_{n_k} converge vers F en chaque point de continuité de F. Soit x un point de continuité de x et $\varepsilon > 0$. On peut trouver η tel que $|F(x) - F(y)| \leq \varepsilon$ en tout y de $[x - \eta, x + \eta]$. Comme $\mathbb{Q}$ est dense dans $\mathbb{R}$, on peut trouver des rationnels r et s tels que $x - \eta \leq r \leq s \leq x + \eta$. On a pour tout $k \geq 1$:

$$F_{n_k}(r) \leq F_{n_k}(x).$$

On en déduit

$$F(r) = \lim_{k\to+\infty} F_{n_k}(r) = \varliminf_{k\to+\infty} F_{n_k}(r) \leq \varliminf_{k\to+\infty} F_{n_k}(x),$$

ce qui implique que pout tout $\varepsilon > 0$

$$F(x) - \varepsilon \leq \varliminf_{k\to+\infty} F_{n_k}(x).$$

On en déduit finalement que

$$F(x) \leq \varliminf_{k\to+\infty} F_{n_k}(x).$$

De la même manière, on montre que

$$\varlimsup_{k\to+\infty} F_{n_k}(x) \leq F(x).$$

Finalement, on a

$$F(x) \leq \varliminf_{k\to+\infty} F_{n_k}(x) \leq \varlimsup_{k\to+\infty} F_{n_k}(x) \leq F(x),$$

ce qui montre bien que

$$\lim_{k\to+\infty} F_{n_k}(x) = F(x).$$

□

On peut maintenant prouver le théorème de Prohorov.

Démonstration. Soit μ_n une suite formées de mesures appartenant à une famille $\mathcal{M}$ tendue. On note F_n la fonction de répartion de μ_n. D'après le théorème de Helly, on peut trouver une suite strictement croissante n_k et une fonction continue à droite croissante telle que $F_{n_k}(x)$ converge vers $F(x)$ en chaque point de continuité de F. Comme F est croissante continue à droite, on sait (voir un cours de théorie de la mesure) qu'il existe une mesure μ telle que pour tous a et b $F(b) - F(a) = \mu(]a, b])$. Par construction, F prend ses valeurs entre 0 et 1, donc $\mu(\Omega) = \lim_{n\to+\infty} \mu(] - n, n]) \le 1$. Il s'agit maintenant de voir que cette mesure est une mesure de probabilité. Soit $\varepsilon > 0$. Comme $\mathcal{M}$ est tendue, il existe un compact K tel que pour tout $\mu \in \mathcal{M}$, on ait $\mu(K) \ge 1 - \varepsilon$. Posons $M = \sup\{|x|; x \in K\}$. Comme l'ensemble des points de discontinuités d'une fonction croissante est au plus dénombrable, il existe $a < -M$ et $b > M$ tels que F soit continue en a et en b. Ainsi pour tout $k \ge 1$, on a

$$\begin{aligned} F_{n_k}(b) - F_{n_k}(a) &= \mu_{n_k}(]a, b]) \\ &\ge \mu_{n_k}([-m, m]) \\ &\ge \mu_{n_k}(k) \\ &\ge 1 - \varepsilon \end{aligned}$$

Comme a et b sont des points de continuité de F, on obtient, en faisant tendre k vers $+\infty$: $F(b) - F(a) \ge 1 - \varepsilon$, ce qui entraîne $\mu(\omega) \ge \mu(]a, b]) \ge 1 - \varepsilon$. Comme ε peut être pris aussi petit que l'on veut, cela entraîne $\mu(\omega) \ge 1$, d'où $\mu(\Omega) = 1$. μ est donc bien une mesure de probabilité dont F est la fonction de répartition. Comme $F_{n_k}(x)$ converge vers $F(x)$ en chaque point de continuité de F, on peut conclure que μ_{n_k} converge en loi vers μ.

□

Corollaire 20. *Soit* $(\mu_n)_{n\ge1}$ *telle que*

– $\{\mu_n; n \ge 1\}$ *est tendue*

– *Toute sous-suite de* $(\mu_n)_{n\ge1}$ *qui converge en loi converge vers* μ.

Alors $(\mu_n)_{n\ge1}$ *converge en loi vers* μ.

Démonstration. Soit f une fonction continue bornée. On doit montrer que la suite $(\int f(x)\ d\mu_n(x))_{n\ge1}$ converge vers $\int f(x)\ d\mu(x)$. Soit $(n_k)_{k\ge1}$ une suite srictement croissante d'entiers quelconque. La suite $(\mu_{n_k})_{k\ge1}$ est une suite d'éléments d'une famille tendue. Je peux donc en extraire une sous-suite convergente $(\mu_{n_{k_i}})_{i\ge1}$. Mais d'après la deuxième hypothèse la limite ne peut être que μ. Il s'ensuit que

$$\lim_{i\to+\infty} \int f(x)\ d\mu_{n_{k_i}} = \int f(x)\ d\mu(x).$$

Ainsi, de chaque sous-suite de la suite $(\int f(x)\ d\mu_n(x))_{n\geq 1}$, on peut extraire une sous-suite qui tend vers $\int f(x)\ d\mu(x)$. Cela signifie que la suite $(\int f(x)\ d\mu_n(x))_{n\geq 1}$ converge vers $\int f(x)\ d\mu(x)$, ce qui achève la preuve. □

A.3 Théorèmes de Levy

On rappelle que la fonction caractéristique φ_μ d'une mesure μ est définie par

$$\varphi_\mu(t) = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle}\ d\mu(x).$$

On rappelle aussi que la fonction caractérique caractérise la loi : si deux mesures de probabilités μ et ν vérifient en tout point t de $\mathbb{R}^d$: $\varphi_\mu(t) = \varphi_\nu(t)$, alors $\mu = \nu$.

On rappelle enfin que la fonction caractéristique d'une loi est uniformément continue sur $\mathbb{R}^d$.

Théorème 31. *Soit μ_n une suite de mesures de probabilités sur $\mathbb{R}^d$. On a les résultats suivants :*

- *Si μ_n converge en loi vers une probabilité μ, alors la suite φ_{μ_n} converge ponctuellement vers φ_μ.*
- *Si la suite φ_{μ_n} converge ponctuellement vers φ_μ, où φ_μ est la loi de la probabilité μ, alors μ_n converge en loi vers μ.*
- *Si la suite φ_{μ_n} converge ponctuellement vers une fonction φ continue en l'origine, alors il existe une mesure de probabilité μ dont la fonction caractéristique est μ et μ_n converge en loi vers μ.*

Démonstration.

Le premier point résulte de la définition de la convergence en loi et du fait que pour tout t, la fonction $x \mapsto e^{i\langle t,x\rangle}$ est continue bornée.

Le deuxième point est une conséquence du troisième, en utilisant le fait que la fonction caractéristique φ_μ est bien continue en l'origine.

Montrons donc le troisième point (sans utiliser le deuxième, bien sûr !). Ce qui nous manque maintenant, c'est la tension de la suite $(\mu_n)_{n\geq 1}$, car si $(\mu_n)_{n\geq 1}$ est tendue, elle admet une valheur d'adhérence μ et il ressort alors clairement du premier point que

- $\varphi_\mu = \varphi$.
- Si une mesure μ_* est valeur d'adhérence de la suite $(\mu_n)_{n\geq 1}$, alors $\varphi_{\mu_*} = \varphi$, donc $\mu_* = \mu$.

Grâce au corollaire précédent, cela implique alors la convergence en loi de μ_n vers μ. Il reste donc à montrer que $(\mu_n)_{n\geq 1}$ est tendue. Pour cela, il suffit de montrer que pour tout k entre 1 et d, la suite de mesures marginales $(\pi_k^{-1}\mu_n)_{n\geq 1}$ est tendue. Notons qu'on a simplement $\varphi_{\pi_k^{-1}\mu_n)}(x) = \varphi_{\mu_n}(0,\dots,0,x,0,\dots,0)$ et la suite de fonctions $(\varphi_{\pi_k^{-1}\mu_n)}(x))_{n\geq 1}$ converge simplement vers la fonction $x \mapsto \varphi((0,\dots,0,x,0,\dots,0)$, qui est continue en 0. Ainsi, on voit qu'on est ramené à démontrer qu'une famille $(\nu_n)_{n\geq 1}$ de mesures de probabilités sur $\mathbb{R}$ dont les fonctions caractéristiques ψ_n convergent simplement vers une fonction ψ continue en 0 est tendue. Soit $n \geq 1$ et $a > 0$

$$\begin{aligned}
\frac{1}{2a}\int_{-a}^{a}(1-\psi_n(t))\,dt &= \frac{1}{2a}\int_{-a}^{a}(1-\psi_n(t))\,dt \\
&= \frac{1}{2a}\int_{-a}^{a}\int(1-e^{itx})\,d\nu_n(x)\,dt \\
&= \int\frac{1}{2a}\int_{-a}^{a}(1-e^{itx})\,dt\,d\nu_n(x) \\
&= \int 1-\frac{\sin ax}{ax}\,d\nu_n(x)
\end{aligned}$$

La fonction $\theta \mapsto 1 - \frac{\sin\theta}{\theta}$ est positive et pour $|\theta| \geq \pi/2$, on a $\frac{\sin\theta}{\theta} \leq \frac{2}{\pi}$, d'où

$$\begin{aligned}
\frac{1}{2a}\int_{-a}^{a}(1-\psi_n(t))\,dt &= \int 1-\frac{\sin ax}{ax}\,d\nu_n(x) \\
&\geq \int \mathbb{1}_{|ax|\geq\pi/2}(1-\frac{2}{\pi}\,d\nu_n(x) \\
&\geq (1-\frac{2}{\pi})\nu_n(\{x\in\mathbb{R} : |x| \geq \frac{\pi}{2a}\})
\end{aligned}$$

Soit maintenant $\varepsilon > 0$: on va montrer qu'il existe un compact qui porte à ε près la charge de chaque ν_n. Comme ψ est continue en 0, il existe $a > 0$ tel que

$$\sup_{t\in[-a,a]}|1-\psi(t)| \leq \frac{\varepsilon}{2}(1-\frac{2}{\pi}),$$

ce qui entraîne

$$\frac{1}{2a}\int_{-a}^{a}(1-\psi(t))\,dt \leq \frac{\varepsilon}{2}(1-\frac{2}{\pi}).$$

D'autre part, d'après le théorème de convergence dominée, $\frac{1}{2a}\int_{-a}^{a}(1-\psi_n(t))\,dt$ converge vers $\frac{1}{2a}\int_{-a}^{a}(1-\psi(t))\,dt$, donc il existe n_0 tel que pour $n > n_0$, on ait

$$\frac{1}{2a}\int_{-a}^{a}(1-\psi(t))\,dt \leq \varepsilon(1-\frac{2}{\pi}),$$

ce qui entraîne $\nu_n(\{x \in \mathbb{R} : |x| \geq \frac{\pi}{2a}\}) \leq \varepsilon$. La famille $(\nu_n)_{n \leq n_0}$ est finie, donc elle est tendue : il existe un compact K tel que pour tout $n \leq n_0$, on ait $\nu_n(K) \geq 1 - \varepsilon$. Enfin $K' = K \cup [-\frac{\pi}{2a}, \frac{\pi}{2a}]$ est compact et pour tout $n \geq 1$, on a $\nu_n(K') \geq 1 - \varepsilon$, ce qui achève la preuve. □

A.4 Exercices

1. *Existence des lois stables d'indice α*
   Soient $Y_{n,k}$ une suite de variables aléatoires indépendantes suivant la loi de Poisson $\mathcal{P}(cn^\alpha / |k|^{1+\alpha})$, où c est une constante positive et α un réel vérifiant $0 < \alpha < 2$. On pose

$$Z_n = \frac{1}{n} \sum_{k=-n^2}^{n^2} kY_{n,k}.$$

   (a) Montrer que la fonction caractéristique de Z_n peut s'écrire $\varphi_{Z_n}(\theta) = \exp(-2c\theta^\alpha u_n(\theta))$, avec

$$u_n(\theta) = \int_0^{n|\theta|} f(\frac{\theta}{n} \ulcorner \frac{nx}{\theta} \urcorner)\, dx,$$

   où $f(x) = \frac{1-\cos x}{x^{1+\alpha}}$.

   (b) Soit $A > 0$ et $\theta > 0$. À l'aide du théorème de convergence dominée, démontrer la convergence de

$$\int_A^{n\theta} f(\frac{\theta}{n} \ulcorner \frac{nx}{\theta} \urcorner)\, dx$$

   lorsque n tend vers l'infini.

   (c) Montrer que pour tout $x > 0$, on a

$$f'(x) = \sum_{k=1}^{+\infty} (-1)^{k-1} \frac{2k - \alpha - 1}{(2k)!} x^{2k-\alpha-2}.$$

   En déduire l'existence d'un réel A tel que f est monotone sur $]0, A[$.

   (d) Montrer que pour un choix approprié de c, la suite de fonctions φ_{Z_n} converge vers $\varphi(\theta) = \exp(-|\theta|^\alpha)$.

   (e) Soit α avec $0 < \alpha \leq 2$. Montrer qu'il existe une mesure m_α dont la fonction caractéristique soit la fonction $\theta \mapsto \exp(-|\theta|^\alpha)$.

(f) Soient $X_1, \dots X_n$ des variables aléatoires indépendantes identiquement distribuées suivant la loi m_α. Montrer qu'il existe λ_n tel que $\lambda_n(X_1 + X_2 + \cdots + X_n)$ suive la loi m_α.

2. Déterminer un réel λ_n tel que la mesure μ_n dont le support est $\{-n, -(n-1), \dots, n-1, n\}$ et vérifiant

$$\forall k \in \{-n, -(n-1), \dots, n-1, n\} \quad \mu(k) = \lambda_n(2n + 1 - 2|k|)$$

soit une mesure de probabilité. Soit X_n suivant la loi μ_n. Montrer que X_n/n converge en loi. (Indication : on pourra déterminer ν_n telle que $\mu_n = \nu_n * \nu_n$.

Annexe B

Indications

B.1 Exercices sur les sommes de variables indépendantes

1. Pas d'indication.
2. Utiliser le premier théorème de Levy.
3. (a) Commencer par déterminer la loi de X_n.

   (b) Pour montrer la croissance, on pourra remarquer que $f(x) = \lim_{n\to+\infty} \mathbb{E}u \circ X_n^x$. Pour déterminer la limite, il est commode de se ramener au cas où $(u_n)_{n\geq 0}$ est à valeurs positives et remarquer qu'alors, pour tout n, il existe un polynôme P_n tel que

   $$\forall x > 0 \quad f(x) \geq (1 - \frac{P_n(x)}{e^x})u_n.$$

4. on pourra utiliser la fonction f_A, continue et affine par morceaux qui vaut 0 sur $]-\infty, A]$, et 1 sur $[A+1, +\infty[$ et montrer que

   $$|\varphi_n(s) - \varphi_n(t)| \leq |s-t|(A+1) + 2\int_{\mathbb{R}} f_A(u)\ d\mathbb{P}_{|X_n|}(u).$$

5. Plusieurs méthodes possibles : utiliser le résultat de l'exercice précédent ou considérer la loi du triplet (a_n, b_n, X_n).
6. (a) On pourra utiliser le théorème de la limite centrale et le résultat de l'exercice 5.

   (b) Se souvenir qu'une loi exponentielle est une loi Gamma.

   (c) Utiliser le théorème de la limite centrale.

   (d) Remarquer que Ψ est bornée.

   (e) Intégrer.

(f) Cela revient à montrer que (a_n) tend vers 1.

7. À k fixé, chercher un équivalent de $\mathbb{P}(X_n = k)$.
8. On peut par exemple montrer que presque sûrement, $|X_n| \le \frac{2}{\alpha}\ln n$ pour n assez grand.
9. On peut commencer par montrer que la série diverge si (a_n) ne tend pas vers zéro ;
10. On peut remarquer que si $Y_n = a_n X_n$ avec (a_n) de limite nulle, alors $\mathbb{E}|Y_n^{(c)}| \sim a_n \mathbb{E}|X_0|$.
11. Remarquer que $(X_n 1\!\!1_{\{X_n \le c\}})^2 \le c X_n 1\!\!1_{\{X_n \le c\}}$.
12. La série diverge pour $z = 1$.
13. Appliquer le théorème des trois séries.
14. Remarquer que $\mathbb{P}(|a_n X_n| \ge 1) = \operatorname{atan} a_n$.
15. Il suffit d'avoir $\mathbb{P}(|a_n X_n| > \frac{1}{n^2}) \le \frac{1}{n^2}$.
16. Dans un premier temps, on pourra remarquer que

$$\overline{\lim_{n\to+\infty}} |a_n X_n| \ge \left(\underline{\lim_{n\to+\infty}} |a_n|\right)\left(\overline{\lim_{n\to+\infty}} |X_n|\right) \ge \alpha \left(\underline{\lim_{n\to+\infty}} |a_n|\right)\left(\overline{\lim_{n\to+\infty}} 1\!\!1_{\{|X_n| \ge \alpha\}}\right)$$

B.2 Exercices sur les vecteurs gaussiens

1. Remarquer que $\frac{\sigma(U)}{\sigma(V)} V$ a même loi que U.
2. On rappelle que si X suit $\mathcal{N}(0,1)$, $\mathbb{E}X^4 = 3$.
3. Trouver une matrice orthogonale O tel que la matrice de covariance de OX soit diagonale.
4. (a) Déterminer le spectre de C.

   (b) On pourra commencer par trouver une constante K telle que

$$\forall a \in \mathbb{R}^3 \quad \mathbb{E}|\langle X, a\rangle| = K(\mathbb{E}|\langle X, a\rangle|_2^2)^{1/2}.$$

5. On pourra montrer qu'il existe un vecteur gaussien dont la matrice de covariance est

$$\begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}$$

6. (a) Relire le cours !

   (b) Se souvenir qu'une matrice de covariance est symétrique positive.

   (c) Si (Z,T) est gaussien $(\frac{Z}{\sigma(Z)},T)$ aussi.
   (d) Utiliser la question précédente.
   (e)
$$\mathbb{E}\|X\|^4 = \mathbb{E}\Big(\sum_{k=1}^{n} X_k^2\Big)^2.$$

7. Trouver une matrice orthogonale O tel que la matrice de covariance de OX soit diagonale.
8. (a) Revoir l'image d'un vecteur gaussien par une transformation affine.
   (b) On pourra utiliser un changement de variables.
9. On pourra diagonaliser A dans une base orthogonale.
10. Pour la première question, on pourra s'intéresser à la fonction de répartition de Y ; pour la deuxième on pourra montrer que $0 < \mathbb{P}(X+Y=0) < 1$.

B.3 Exercices sur l'espérance conditionnelle

1. Revoir les propriétés de l'espérance conditionnelle.
2. 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.
3. 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.
4. (a) $\Omega = \{X < h\} \cup \{X \geq h\}$.
   (b) Un max se réalise toujours en au moins un point.
   (c) L'indicatrice de l'intersection est le produit des indicatrices.
   (d) Écrire $Y_i(h)$ sous la forme $F(X_i, Z)$, avec Z indépendant de X_i.
   (e) Utiliser (c) et (d)
   (f) Prendre $h = \alpha\frac{\ln n}{n}$, avec α bien choisi.
5. (a) Écrire $A = \cup_{n\geq 0} A \cap \{N=n\}$.
   (b) On peut utiliser la question précédente.
6. 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.

7. On peut s'inspirer de la preuve du calcul de $\mathbb{E}[g(X,Y)|X]$ lorsque X et Y sont indépendantes.
8. On peut le voir comme un cas particulier de l'exercice précédent
9. On pourra remarquer que $X + Y - Z = 0$.
10. (a) Calculer les mineurs.
    (b) On pourra écrire X sous la forme $X = \alpha Y + \beta Z + R$, avec R indépendant de $\sigma(Y, Z)$.
    (c) 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.

B.4 Exercices sur les martingales

1. (a) Remarquer que $x \leq (x^2 + 1)/2$.
   (b) i. On pourra remarquer que (X_n) prend ses valeurs dans $[0, 1]$
       ii. Appliquer le théorème de convergence dominée
2. Utiliser les propriétés de l'espérance conditionnelle
3. (a) On rappelle que si $Y \geq 0$ et $\mathbb{E}Y = 0$, alors Y est nulle presque sûrement.
   (b) $x \mapsto (x - a)^+$ est une fonction convexe.
   (c) On pourra montrer que $(X_n - a)^-$ est presque sûrement nulle sur l'événement $\{X_p \geq a\}$.
   (d) Si deux nombres sont distincts, il y a un rationnel entre les deux.
4. (a) Classique.
   (b) Utiliser l'inégalité de Cauchy-Schwartz
   (c) On pourra montrer que $Y_n^t = o(Y_n^{t/2})$.
   (d) 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.
5. (a) Remarquer les conditionnements successifs.
   (b) On pourra montrer que $(Y_n)_{n\geq 1}$ est bornée dans L^1

   (c) On pourra remarquer que

$$\mathbb{E}|W|\mathbb{1}_{\{|T_n \geq a\}} \leq t\mathbb{P}(|T_n| \geq a) + \mathbb{E}|W|\mathbb{1}_{\{|W| \geq t\}}.$$

   (d) On pourra démontrer que

$$\mathbb{E}|Y_n|\mathbb{1}_{\{|Y_n \geq a\}} \leq 2\mathbb{E}|Z|\mathbb{1}_{\{|Y_n \geq a\}}.$$

   (e) Utiliser le (c) ; on pourra remarquer que

$$|Y_n - Y| \leq |Y_n|\mathbb{1}_{\{|Y_n| \geq a\}} + |Y|\mathbb{1}_{\{|Y_n| \geq a\}} + |Y_n - Y|\mathbb{1}_{\{|Y_n| < a\}}.$$

   .

   (f) Traiter d'abord le cas où Z est positive et utiliser un critère d'identification des mesures.

6. On rappelle que $\sin x \leq x$ pour tout $x \geq 0$.
7. (a) $-g$ est convexe.
   (b) On pourra utiliser le théorème d'arrêt.
   (c) Utiliser la question précédente.
   (d) 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 $[s,t]$.
   (e) On pourra montrer successivement

$$\mathbb{E}f(X_T) = \mathbb{E}\Psi(X_T) = \Psi(\mathbb{E}X_T) = \Psi(i_0).$$

B.5 Exercices sur les processus

1. 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 $Y_n = (-1)^n X_n$.

2. On peut considérer les applications

$$\begin{array}{rcl} \psi_n : \mathbb{R}^{n+1} & \to & \mathbb{R}^n \\ (x_1,\dots,x_n) & \mapsto & (x_1+2x_2, x_2+2x_3,\dots,x_{n-1}+2x_n) \end{array}$$

3. On peut considérer les applications

$$\begin{array}{rcl} \psi_n : \mathbb{R}^{\mathbb{N}} & \to & \mathbb{R}^n \\ ((x_n)_{n\geq 1}) & \mapsto & (\varphi(x), \varphi(\theta(x), \dots, \varphi(\theta^{n-1}(n)) \end{array}$$

4. 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.
5. 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.
6. On rappelle que deux mesures sur $\mathbb{R}$ sont égales si elles coïncident sur les ensembles $]-\infty, a]$, où a décrit $\mathbb{R}$.
7. 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 $x \mapsto e^{2i\pi x}$.
8. Remarquer que $\mathbb{E}[Z_i Z_j] = \delta_{i,j}$.

B.6 Exercices sur les chaînes de Markov

1. 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.
2. (a) Revoir le cours.
   (b) Découper l'intervalle $[0,1]$ en morceaux de différentes longueurs.
   (c) Utiliser le (a).
   (d) Recoller les morceaux.
3. On pourra remarquer que $\{X_n = a\} \subset \{T \leq n\}$ et que

$$\{T \leq n\} \backslash \{X_n = a\} \subset \{\exists k < n; X_k = a \text{ et } X_{k+1} \neq a\}.$$

4. 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.
5. 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 $X_{n+1} = g(X_n, Z_{n+1})$.

6. 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, $\mathbb{P}(\cap_{n\geq N}\{X_n = x\}) = 0$.
7. (a) On peut établir que $S'_{n+1} = S'_n + 1\!\!1_{\{S'_n \notin \{-b,a\}\}} X_{n+1}$.
   (b) Comme dans l'exercice 4, on utilisera la propriété de Markov.
   (c) On rappelle que les solutions d'une récurrence linéaire forment un espace vectoriel.
8. On peut montrer que $\{T' = +\infty\} = \cap_{a\geq 1} G^a$.
9. Raisonner comme dans l'exercice 4 et résoudre le système linéaire.
10. Relire les définitions.
11. 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 $b|j$.
12. À t fixé, on pourra remarquer que $\mathbb{P}(1\!\!1_A F((X_n)_{n\geq 1} > t) = \mathbb{P}(A, F((X_n)_{n\geq 1} > t)$ et appliquer la propriété de Markov.
13. 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.
14. (a) 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.
    (b) Se ramener à l'étude d'une somme de variables de Bernoulli.
    (c) 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 .
    (d) On pourra remarquer que la suite $(\mathbb{E}X_n)_{n\geq 1}$ est constante.
15. 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 $\mathbb{P}^a(Y_1 = b, Y_2 = c) > 0$.
16. Commencer par trouver en candidat pour la matrice.

B.7 Exercices sur la récurrence et les mesures invariantes

1. (a) Il y a plusieurs manières de montrer que T_k est un temps d'arrêt. On peut procéder par récurrence ou remarquer que

$$T_k = \inf\{n \geq 0; n \in A \text{ ou } \sum_{i=0}^{n-1} 1\!\!1_{\{} X_i \neq X_{i+1} \geq k\}.$$

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 $\mathbb{P}^{\mu}_{X_{T_k}}$.

(b) 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 $X_{T_{k+1}} = \Psi(X_{T+.})$.

2. (a) On peut remarquer que $T_m = \inf\{n \geq 1; \sum_{k=1}^{n} 1\!\!1_{X_k\in J} \geq m\}$
   (b) S'inspirer de l'exercice précédent.
3. 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.
4. (a) On discutera suivant la parité de d
   (b) On pourra exhiber une mesure invariante.
   (c) Si X_0 est pair, la suite $(X_{2n})_{n\geq 0}$ ne prend que des valeurs "paires".
5. (a) Utiliser la représentation canonique
   (b) On remarque que $Y_{n+1} = Y_n + (-1)^{1\!\!1_{\{X_n(V_{n+1})=1\}}}$.
   (c) Commencer par étudier le cas où γ est une masse de Dirac.
   (d) Calculer la probabilité de chaque N-uplet.
   (e) Utiliser les questions précédentes.
   (f) On pourra montrer que la chaîne est réversible.
6. 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.
7. Utiliser le théorème ergodique des chaînes de Markov
8. (a) C'est dans le cours !
   (b) De manière plus générale, on peut remarquer qu'une marche aléatoire symétrique sur un groupe fini admet toujours une probabilité réversible très simple.

(c) 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.

9. On peut représenter l'état du système au temps n par un vecteur $(x(1), x(2), \ldots, x(2N))$, 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 $1/N^2$.

Annexe C

Problème

Partie I

Soit $(U_n)_{n\geq 1}$ une suite de variables aléatoires suivant la loi uniforme sur $[0,1]$. On définit une suite de variables aléatoires par la donnée de $X_0 \in \mathbb{N}$ 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
$$F(\sigma) = |\{i \in \{1,\dots,N\}; j < i \Longrightarrow \sigma(i) < \sigma(j)\}|.$$

2. On définit une suite (Z_n) par récurrence comme suit :
On pose $Z_0 = N$ et pour $k \geq 1$,
$$Z_k = \min(\sigma(1), \sigma(2), \dots, \sigma(k)) - 1,$$
avec la convention $\sigma(i) = +\infty$ pour $i > n$.
Montrer que $\forall n \in \mathbb{N}^* \quad \forall j \in \{0,\dots,N\}$
$$\mathbb{P}(Z_{n+1} = j | \sigma(1),\dots,\sigma(n)) = \frac{1}{N-n}\mathbb{1}_{\{j<Z_n\}} + \frac{N-n-Z_n}{N-n}\mathbb{1}_{\{j=Z_n\}}.$$
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 $0.\infty = 0$. Déterminer la loi de la suite $(Z_{T_k})_{k\geq 0}$.

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.