Théorie des probabilités et processus aléatoires

Ecole Centrale de Lyon

24 mars 2023

Table des matières

Pour compléter le cours, on pourra consulter :

Valérie Girardin et Nikolaos Limnios : Probabilités en vue des applications, tomes I et II, Vuibert 2008.

(ouvrage disponible à la bibliothèque en plusieurs exemplaires ainsi que sur scholarvox par un lien à partir du catalogue de la bibliothèque).

Chapitre 1

Fonctions caractéristiques et indépendance de vecteurs aléatoires.

Sommaire

1.1 Transformée de Fourier et fonction caractéristique

Définition 1.1 (*Transformée de Fourier*)

1. Si $f \in L^1(\mathbb{R}^d)$, sa transformée de Fourier $\hat{f}$ est une fonction définie sur $\mathbb{R}^d$ par

$$\hat{f}(t) = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle} f(x)dx,\ t \in \mathbb{R}^d,\ \text{où } \langle t,x\rangle = \sum_{i=1}^{d} t_i x_i.$$

2. Si μ est une mesure finie sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, sa transformée de Fourier est la fonction définie sur $\mathbb{R}^d$ par

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

Remarque : Si μ admet une densité par rapport à la mesure de Lebesgue alors sa transformée de Fourier est celle de sa densité.

Proposition 1.1

Si $f \in L^1(\mathbb{R}^d)$, $\hat{f}$ est uniformément continue.

Théorème 1.1 (*Formule d'inversion de Fourier*)

Si f et $\hat{f}$ sont intégrables, alors f est continue et

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{f}(t) e^{-i\langle t,x\rangle} dt \text{ pour } \lambda\text{-presque tout } x \in \mathbb{R}^d.$$

Proposition 1.2

Si μ est une mesure finie sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ de transformée de Fourier $\hat{\mu}$, alors

$$\mu(]x_1, y_1[\times \ldots \times]x_d, y_d[) = \lim_{t \to +\infty} \int_{-t}^{t} \ldots \int_{-t}^{t} \prod_{k=1}^{d} \frac{e^{-iu_k x_k} - e^{-iu_k y_k}}{iu_k} \hat{\mu}(u_1, \ldots, u_d) du_1 \ldots du_d.$$

Définition 1.2

Soit X un vecteur aléatoire de dimension d. La fonction caractéristique de X est la transformée de Fourier φ_X de sa loi, soit

$$\varphi_X(t) = \mathbb{E}\left[e^{i\langle t, X\rangle}\right], \ t \in \mathbb{R}^d,$$

soit, par la formule de transfert,

$$\varphi_X(t) = \int_{\mathbb{R}^d} e^{i\langle t, x\rangle} \mathbb{P}_X(dx).$$

Proposition 1.3 (*Propriétés de la fonction caractéristique*)

1. *$\varphi_X(0) = 1$ et $|\varphi_X(t)| \leq 1$ pour tout $t \in \mathbb{R}^d$.*
2. *La fonction caractéristique caractérise la loi : deux vecteurs aléatoires de même fonction caractéristique ont même loi.*
3. *Si X a une densité f_X alors $\varphi_X = \hat{f}_X$.*
4. *$\varphi_{-X}(t) = \varphi_X(-t) = \overline{\varphi_X(t)}$ pour tout t. On en déduit X symétrique $\Leftrightarrow$ φ_X paire $\Leftrightarrow$ φ_X réelle.*
5. *Soit X un vecteur aléatoire de $\mathbb{R}^d$ et soient $A \in \mathcal{M}_{d' \times d}(\mathbb{R})$ et $b \in \mathbb{R}^{d'}$. Alors pour tout $t \in \mathbb{R}^{d'}$,*
$$\varphi_{AX+b}(t) = e^{i\langle t, b\rangle} \varphi_X({}^t\!At).$$
Si $d = d' = 1$, $\varphi_{aX+b}(t) = e^{itb} \varphi_X(at)$.

Proposition 1.4

Si φ_X est intégrable, alors X a une densité par rapport à la mesure de Lebesgue donnée par

$$f_X(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle t, x\rangle} \varphi_X(t) dt \ \lambda\text{-p.p}$$

Exemple : Calculer la fonction caractéristique d'une v.a. binomiale de paramètres n et p.

Exemple : Calcul de la fonction caractéristique d'une loi normale : montrer que la fonction

$$\varphi_X(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{itx} e^{-\frac{1}{2}x^2} dx, \ x \in \mathbb{R},$$

est dérivable, calculer sa dérivée et en déduire qu'elle satisfait une équation différentielle ordinaire. La résoudre pour déterminer la fonction caractéristique d'une v.a. gaussienne centrée réduite. En déduire la fonction caractéristique d'une $\mathcal{N}(m, \sigma^2)$.

Théorème 1.2

Soit X une v.a. réelle.

1. *Si X^n est intégrable alors $\varphi_X \in C^n(\mathbb{R})$ et $\varphi_X^{(m)}(t) = i^m \mathbb{E}\left[X^m e^{itX}\right]$ pour tout $m \in \{0, ..., n\}$. En particulier, $\varphi_X^{(m)}(0) = i^m \mathbb{E}[X^m]$ pour tout $m \in \{0, ..., n\}$.*
2. *Si φ_X est k fois dérivable en 0 ($k \geq 2$), alors X admet des moments jusqu'à l'ordre $2\left[\frac{k}{2}\right]$. Ils sont donnés par $\varphi_X^{(m)}(0) = i^m \mathbb{E}[X^m]$ pour tout $1 \leq m \leq 2\left[\frac{k}{2}\right]$.*

Démonstration : Quelques éléments.

1. La proposition est vraie en $m = 0$ pour tout $t \in \mathbb{R}$. Elle se démontre ensuite par récurrence en dérivant sous le signe somme.
2. On remarque que $2\left[\frac{k}{2}\right] = k$ si k est pair. Sinon, c'est l'entier précédent. On admet ce résultat.

Remarque : Si $X = (X_1, ..., X_d)$, $\mathbb{E}[X_i X_j] = -\dfrac{\partial^2 \varphi_X}{\partial t_i \partial t_j}(0)$ quand cela existe.

Exemple : Calcul des moments d'une loi normale. On eut montrer que si $X \rightsquigarrow \mathcal{N}(0, \sigma^2)$ alors

$$\mathbb{E}\left[X^{2k+1}\right] = 0 \text{ pour tout } k \geq 0$$

et

$$\mathbb{E}\left[X^{2k}\right] = \frac{(2k)!}{k!}\frac{\sigma^{2k}}{2^k}.$$

1.2 Vecteurs aléatoires indépendants

1.2.1 Caractérisation des vecteurs aléatoires indépendants

Théorème 1.3

Soient X_1, ...,X_n, n vecteurs aléatoires de dimensions respectives d_1,...,d_n. Les propositions suivantes sont équivalentes.

1. *X_1, ..., X_n sont indépendantes.*
2. *$\forall\ h_i : \mathbb{R}^{d_i} \to \mathbb{R}$ boréliennes bornées pour $i = 1..n$,*
$$\mathbb{E}\left[\prod_{i=1}^n h_i(X_i)\right] = \prod_{i=1}^n \mathbb{E}[h_i(X_i)].$$
3. *$\varphi_{(X_1,...,X_n)}(t_1, ..., t_n) = \varphi_{X_1}(t_1)...\varphi_{X_n}(t_n)$ for all $(t_1, ..., t_n) \in \mathbb{R}^{d_1} \times ... \times \mathbb{R}^{d_n}$.*

Remarque : L'expression dans 2. reste vraie dès que les intégrales ont un sens et que l'on est dans les conditions d'application d'un théorème de Fubini.

Démonstration :

□ On montre l'équivalence entre 1. et 2. pour deux variables aléatoires, le passage à n se faisant par récurrence. Soient donc X et Y deux v.a. de $\mathbb{R}^d$ et $\mathbb{R}^{d'}$ respectivement et soient φ et ψ deux fonctions boréliennes bornées définies sur $\mathbb{R}^d$ et $\mathbb{R}^{d'}$ respectivement.

$$\mathbb{E}[\varphi(X)\psi(Y)] = \int_{\mathbb{R}^d \times \mathbb{R}^{d'}} \varphi(x)\psi(y)\mathbb{P}_{(X,Y)}(dx, dy).$$

Comme les v.a. sont indépendantes, $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$, et par Fubini,

$$\int_{\mathbb{R}^d \times \mathbb{R}^{d'}} \varphi(x)\psi(y)\mathbb{P}_X \otimes \mathbb{P}_Y(dx, dy) = \int_{\mathbb{R}^d} \varphi(x)\mathbb{P}_X(dx) \int_{\mathbb{R}^{d'}} \psi(y)\mathbb{P}_Y(dy).$$

On en déduit bien

$$\mathbb{E}[\varphi(X)\psi(Y)] = \mathbb{E}[\varphi(X)]\mathbb{E}[\psi(Y)].$$

Réciproquement, on remarque que, pour $A \in \mathcal{B}(\mathbb{R}^d)$ et $B \in \mathcal{B}(\mathbb{R}^{d'})$,

$$\mathbb{P}[X \in A, Y \in B] = \mathbb{E}\left[\mathbf{1}_{[X\in A, Y \in B]}\right] = \mathbb{E}\left[\mathbf{1}_{[X\in A]}\mathbf{1}_{[Y\in B]}\right] = \mathbb{E}[\mathbf{1}_A(X)\mathbf{1}_B(Y)]$$

et en appliquant 2. à $\varphi = \mathbf{1}_A$ et $\psi = \mathbf{1}_B$, il vient

$$\mathbb{E}[\mathbf{1}_A(X)\mathbf{1}_B(Y)] = \mathbb{E}[\mathbf{1}_A(X)]\mathbb{E}[\mathbf{1}_B(Y)]$$

soit

$$\mathbb{P}[X \in A, Y \in B] = \mathbb{P}[X \in A]\mathbb{P}[X \in B].$$

□ 2. ⇒ 3. pour $(t_1, ..., t_n)$ fixé avec avec $h_i(x) = e^{i\langle t_i, x\rangle}$ pour $x \in \mathbb{R}^{d_i}$ pour tout $i \in \{1, ..., n\}$.

□ 3. ⇒ 1. se montre en utilisant la formule d'inversion et Fubini. C'est assez technique.

Théorème 1.4

Soient X_1,...,X_n des vecteurs aléatoires de dimensions d_1,...,d_n respectivement.

1. *Si les vecteurs X_1, ..., X_n admettent des densités f_{X_1},...,f_{X_n} et sont indépendants alors $(X_1, ..., X_n)$ admet une densité définie par*

$$f_{(X_1,...,X_n)}(x_1, ..., x_n) = f_{X_1}(x_1)...f_{X_n}(x_n),\ \forall\ (x_1, ..., x_n) \in \mathbb{R}^{d_1} \times ... \times \mathbb{R}^{d_n}. \quad (1.1)$$

2. *Si $(X_1, ..., X_n)$ admet une densité alors les X_i admettent aussi une densité. Dans ce cas, les v.a. X_1,..., X_n sont indépendantes si et seulement si (1.1) est satisfaite.*

Proposition 1.5

Soient X et Y deux vecteurs aléatoires indépendants de carré intégrable. Alors ils sont décorrélés.

Question : Justifier ce résultat avec les théorèmes et propositions précédents.

1.2.2 Convolution de mesures et somme de vecteurs aléatoires indépendants

Dans ce chapitre, on voit que la loi de la somme de v.a. indépendantes est la convolution de leurs lois.

Définition 1.3

Soient μ et ν deux mesures sur $(E, \mathcal{B}(E))$ espace vectoriel normé muni de sa tribu borélienne. La convolution ou produit de convolution de μ et ν, notée $\mu * \nu$ est la mesure image de $\mu \otimes \nu$ par l'application

$$\begin{array}{rccc} T: & E \times E & \to & E \\ & (x, y) & \to & x + y. \end{array}$$

Autrement dit, $\mu * \nu = \mu \otimes \nu \circ T^{-1}$ est définie sur $\mathcal{B}(E)$ et pour tout $B \in \mathcal{B}(E)$,

$$\mu * \nu(B) = \mu \otimes \nu(\{(x, y) \in E^2/\ x + y \in B\}.$$

Proposition 1.6

(i) Pour toute fonction h borélienne positive sur E,

$$\int_E h\, d(\mu * \nu) = \int_{E^2} h(x+y)\mu \otimes \nu(dx, dy). \tag{1.2}$$

(ii) Pour toute fonction h borélienne h est $\mu * \nu$-intégrable si et seulement si $h \circ T$ est $\mu \otimes \nu$-intégrable sur E^2, et dans ce cas (1.2) est satisfaite.

Démonstration : Il suffit d'appliquer la formule de transfert.

Proposition 1.7

On suppose que $E = \mathbb{R}^d$ et que μ et ν sont respectivement de densités f et g par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$, f et g étant intégrables. Alors $\mu * \nu$ admet pour densité $f * g$ par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ et $f * g$ est intégrable.

Démonstration : (à compléter). On montre que

$$\int_{\mathbb{R}^d} h\, d(\mu * \nu) = \int_{\mathbb{R}^d} h(f * g) d\lambda$$

pour toute fonction h borélienne bornée. Soit donc h une telle fonction. On utilise (1.2) puis le théorème de Fubini et enfin l'absolue continuité de μ et ν par rapport à la mesure de Lebesgue. On obtient donc

$$\int_{\mathbb{R}^d} h\, d(\mu * \nu) = \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} h(x+y) f(x) dx \right) g(y) dy.$$

Reste alors à faire le changement de variable $x \to z = x + y$ puis à appliquer à nouveau le théorème de Fubini. On obtient donc

$$\int_{\mathbb{R}^d} h\, d(\mu * \nu) = \int_{\mathbb{R}^d} h(z) \left(\int_{\mathbb{R}^d} f(z-y) g(y) dy \right) dz$$

ce qu'il fallait démontrer.

Proposition 1.8 (*Propriétés du produit de convolution*)

(i) Si μ et ν sont des probabilités, $\mu * \nu$ également.

(ii) Si μ et ν sont discrètes alors $\mu * \nu$ également. Plus précisément, si

$$\mu = \sum_k \mu_k \delta_{x_k} \quad \text{et} \quad \nu = \sum_k \nu_k \delta_{y_k}$$

alors

$$\mu * \nu = \sum_k \sum_l \mu_k \nu_l \delta_{x_k + y_l}.$$

En particulier, si $x_k = y_k = k$ pour tout k, alors

$$\mu * \nu = \sum_k \left(\sum_{l=0}^{k} \mu_l \nu_{k-l} \right) \delta_k.$$

Démonstration : A faire en exercice.

(i) Montrer que $\mu * \nu(E) = 1$ si $\mu(E) = \nu(E) = 1$.

(ii) On admet
$$\mu * \nu = \sum_k \sum_l \mu_k \nu_l \delta_{x_k} * \delta_{y_l}$$
et on calcule seulement, pour $x, y \in E$, $\delta_x * \delta_y(B)$ pour un $B \in \mathcal{B}(E)$. Enfin, on appliquera au cas entier (i.e. lorsque $x_k = y_k = k$ pour tout k).

Théorème 1.5

La transformée de Fourier du produit de convolution de deux fonctions ou deux mesures est le produit des transformées de Fourier.

Démonstration : Soient μ et ν deux mesures boréliennes finies. Pour tout $t \in \mathbb{R}^d$,
$$\widehat{\mu * \nu}(t) = \int_{\mathbb{R}^d} e^{i\langle t,z\rangle}(\mu * \nu)(dz) = \int_{\mathbb{R}^{2d}} e^{i\langle t,x+y\rangle}\mu(dx)\nu(dy).$$
On applique alors Fubini, il vient :
$$\widehat{\mu * \nu}(t) = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle}\mu(dx) \int_{\mathbb{R}^d} e^{i\langle t,y\rangle}\nu(dy) = \hat{\mu}(t)\hat{\nu}(t).$$

Théorème 1.6

Soient X_1,..., X_n n vecteurs aléatoires indépendants à valeurs dans $\mathbb{R}^d$. Alors

1. $\mathbb{P}_{X_1+...+X_n} = \mathbb{P}_{X_1} * ... * \mathbb{P}_{X_n}$.
2. *Si X_1, ..., X_n ont respectivement pour densité f_1,..., f_n alors $X_1 + ... + X_n$ a pour densité $f_1 * ... * f_n$.*
3. $\varphi_{X_1+...+X_n}(t) = \varphi_{X_1}(t)...\varphi_{X_n}(t)$ *pour tout $t \in \mathbb{R}^d$.*

Démonstration : A faire en exercice.

Question : Soient X et Y deux v.a. exponentielles indépendantes de paramètres λ et μ respectivement, $\lambda \neq \mu$. Calculer la densité de $X + Y$.

Question : Soient X et Y deux lois normales indépendantes de paramètres (m, σ^2) et (m', σ'^2) respectivement. En utilisant les fonctions caractéristiques, déterminer la loi de $X + Y$.

1.3 Vecteurs gaussiens

Définition 1.4

Un vecteur aléatoire $X = (X_1, ..., X_d)^t$ est dit gaussien si toute combinaison linéaire de ses composantes est gaussienne.

Remarque :

(i) Si X est un vecteur gaussien, alors toutes ses composantes sont gaussiennes.

(ii) Pour toute fonction $f : \mathbb{R}^d \to \mathbb{R}^{d'}$ affine, si X vecteur gaussien, alors $f(X)$ gaussien.

(iii) Notations usuelles : l'espérance d'un vecteur gaussien est notée $M = \mathbb{E}[X] \in \mathbb{R}^d$, sa matrice de covariance $\Gamma = \mathbf{Cov}[X] \in \mathcal{M}_d(\mathbb{R})$. On note $X \rightsquigarrow \mathcal{N}_d(M, \Gamma)$.

Exemple :

(i) $X = a \in \mathbb{R}^d$, $X \rightsquigarrow \mathcal{N}_d(a, 0)$.

(ii) Si $X \in \mathbb{R}^d$ a toutes ses composantes gaussiennes et indépendantes, X est gaussien.

Théorème 1.7

La fonction caractéristique φ d'un vecteur gaussien de loi $\mathcal{N}_d(M,\Gamma)$ est définie par

$$\varphi(t) = e^{i\langle M,t\rangle} e^{-\frac{1}{2}\langle \Gamma t,t\rangle} \text{ pour tout } t \in \mathbb{R}^d.$$

Démonstration : A faire en exercice. On remarquera que

$$\varphi_X(t) = \varphi_{\langle t,X\rangle}(1)$$

et que si X est gaussien, alors pour $t \in \mathbb{R}^d$, $\langle t, X\rangle$ est une v.a. gaussienne dont on calculera l'espérance et la variance.

Proposition 1.9

Soit $X \rightsquigarrow \mathcal{N}_d(M,\Gamma)$ et soit $Y = AX + B$, avec $A \in \mathcal{M}_{d'\times d}(\mathbb{R})$ et $B \in \mathbb{R}^{d'}$. Alors $Y \rightsquigarrow \mathcal{N}_{d'}(AM + B, A\Gamma A^t)$.

Théorème 1.8

Soit $X = (X_1, ..., X_d)^t \rightsquigarrow \mathcal{N}_d(M,\Gamma)$. Les composantes X_1,..., X_d sont mutuellement indépendantes si et seulement si elles sont deux à deux décorrélées, i.e. si Γ est diagonale.

Démonstration : On a montré l'une des implications proposition 1.5. Réciproquement, si Γ est diagonale, alors pour tout $t \in \mathbb{R}^d$,

$$\varphi_X(t) = e^{i\langle M,t\rangle} e^{-\frac{1}{2}\langle \Gamma t,t\rangle} = \prod_{i=1}^d e^{im_i t_i} \prod_{i=1}^d e^{-\frac{1}{2}\mathbf{Var}[X_i] t_i^2} = \prod_{i=1}^d \varphi_{X_i}(t_i),$$

donc les X_i sont indépendants.

Corollaire 1.1

Si $X \rightsquigarrow \mathcal{N}_d(0, I_d)$ alors X a pour densité

$$f_X(x) = \frac{1}{\sqrt{2\pi}^d} e^{-\frac{1}{2}\sum_{i=1}^d x_i^2}, \quad \text{pour } x = (x_1, ..., x_n) \in \mathbb{R}^d.$$

Démonstration : La densité est le produit des densité des d v.a. gausiennes centrées réduites.

Remarque : Des v.a. gausiennes ne forment pas nécessairement un vecteur gaussien, même si elles sont décorrélées. Elles ne sont alors pas indépendantes. Voici un exemple à faire à titre d'exercice.

On considère les v.a. indépendantes X_1 gausienne centrée réduite et ε de Bernoulli qui prend les valeurs -1 et 1 avec probabilité 1/2. On définit alors $X_2 = \varepsilon X_1$. On montre que X_2 est une v.a. gaussienne centrée réduite et que X_1 et X_2 sont décorrélées. Mais on montre ensuite que X_1 et X_2 ne sont pas indépendantes et que (X_1, X_2) n'est pas un vecteur gaussien.

Théorème 1.9

Soit $X \rightsquigarrow \mathcal{N}_d(M,\Gamma)$ avec $\Gamma \neq 0$. Soit r le rang de Γ. Alors il existe $B \in \mathcal{M}_{d\times r}(\mathbb{R})$ et $Y \rightsquigarrow \mathcal{N}_r(0, I_r)$ tels que $X = BY + M$ et $BB^t = \Gamma$.

Démonstration : C'est un résultat d'algèbre et une transformation affine d'un vecteur gaussien. Dans les exercices, on construit B explicitement sur un exemple. La démonstration du résultat n'est que la généralisation évidente de cette preuve.

Corollaire 1.2

Soit $X \rightsquigarrow \mathcal{N}_d(M, \Gamma)$. *$X$ admet une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ si et seulement si Γ est inversible. Dans ce cas, elle s'écrit*

$$f(x) = \frac{1}{\sqrt{2\pi}^d \sqrt{\det \Gamma}} e^{-\frac{1}{2}\langle \Gamma^{-1}(x-M), x-M \rangle} \quad \textit{pour } x \in \mathbb{R}^d.$$

Remarque : Si X a pour densité f et si $Y = AX + B$ avec A inversible, alors Y a pour densité la fonction

$$y \mapsto \frac{1}{|\det A|} f(A^{-1}(y - B))$$

Démonstration : (du corollaire). $X = BY + M$ avec $Y \rightsquigarrow \mathcal{N}_d(0, I_d)$ et $\Gamma = BB^t$. Si Γ est inversible, le rang de Γ est plein, B est carrée et $\det B \neq 0$ donc B est inversible. Calculer alors la densité de X à partir de la densité de Y.

Les vecteurs gaussiens sont très utilisés en statistique, notamment à cause du théorème limite central. On rappelle à titre d'exemple ici le théorème de Cochran : si $(X_1, ..., X_n)$ est un échantillon de la loi $\mathcal{N}(m, \sigma^2)$ (des v.a. indépendantes et toutes distribuées selon la loi $\mathcal{N}(m, \sigma^2)$), alors

$$\overline{X} = \frac{1}{n} X_i \text{ et } \Sigma_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

sont indépendantes et

$$\overline{X} \rightsquigarrow \mathcal{N}(m, \frac{\sigma^2}{n}) \text{ et } \frac{n\Sigma_n^2}{\sigma^2} \rightsquigarrow \chi^2_{n-1}.$$

Ces résultats sont utilisés de façon asymptotique pour les grands échantillons gaussiens.

1.4 Exercices

Exercice 1.1. [*Calcul de fonctions caractéristiques usuelles*] Calculer la fonction caractéristique de la v.a. Y suivant :

1. une loi de Poisson $\mathcal{P}(\lambda)$,
2. une loi exponentielle $\mathcal{E}(\lambda)$,
3. une loi uniforme sur $[-a, a]$.

Exercice 1.2. [*Calcul de fonctions caractéristiques-théorème d'inversion*]
Soit la fonction définie sur $\mathbb{R}$ par $f(y) = \alpha/2 e^{-\alpha|y|}$. Calculer la transformée de Fourier de f et en déduire la fonction caractéristique d'une v.a de loi de Cauchy $\mathcal{C}(\alpha)$.

Exercice 1.3. [*Propriétés de fonctions caractéristiques*]

1. Soit X une v.a.r. Montrer que $\varphi_X(u)$ est à valeurs réelles si et seulement si X a une loi symétrique, i.e $\mathbb{P}_X = \mathbb{P}_{-X}$.
2. Montrer que si X et Y sont des variables aléatoires indépendantes et de même loi, alors $Z = X - Y$ a une loi symétrique.

Exercice 1.4. [*Loi de la transformée linéaire d'un vecteur*]

1. Soit X un vecteur aléatoire de dimension n, de densité f. Soit $A : \mathbb{R}^n \to \mathbb{R}^n$, une application linéaire inversible et $Y = AX$ un vecteur aléatoire. Déterminer la densité de Y.
2. Soit $X = (X_1, X_2)$ un vecteur de $\mathbb{R}^2$ de densité f. Déduire de la question précédente la densité du vecteur $(X_1 + X_2, X_1 - X_2)$. Si X_1 et X_2 sont indépendantes, quelle forme prend la densité de $X_1 + X_2$?
3. Soient $X_1, \ldots, X_n$ des v.a. i.i.d. de loi $\mathcal{E}(\lambda)$. On pose $S_k = X_1 + \cdots + X_k$, pour $1 \leq k \leq n$. Calculer la densité du vecteur $(S_1, S_2, \ldots, S_n)$.

Exercice 1.5. [*Somme de v.a indépendantes*]
Soient X et Y deux v.a.r indépendantes.

1. Montrer que si $X \rightsquigarrow \mathcal{P}(\lambda)$ et $Y \rightsquigarrow \mathcal{P}(\mu)$, alors $X + Y \rightsquigarrow \mathcal{P}(\lambda + \mu)$.
2. Montrer que si $X \rightsquigarrow \mathcal{N}(m_1, \sigma_1^2)$ et $Y \rightsquigarrow \mathcal{N}(m_2, \sigma_2^2)$, alors $X + Y \rightsquigarrow \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

Exercice 1.6.
Soit $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ un vecteur gaussien centré de matrice de variance $\Gamma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, où ρ est un réel tel que $-1 < \rho < 1$.

1. Soit $U = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$ et $V = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. On pose $Y = \begin{pmatrix} < U, X > \\ < V, X > \end{pmatrix}$. Montrer que Y est un vecteur gaussien dont on précisera la loi.
2. Les variables aléatoires $< U, X >$ et $< V, X >$ sont-elles indépendantes ?
3. De manière générale, soit $U = \begin{pmatrix} a \\ b \end{pmatrix}$ et $V = \begin{pmatrix} c \\ d \end{pmatrix}$.
   a) Quelle est la loi de $< U, X >$?
   b) Montrer que $\begin{pmatrix} < U, X > \\ < V, X > \end{pmatrix}$ est un vecteur gaussien.
   c) Quelle est la covariance de $< U, X >$ et $< V, X >$?
   d) A quelle condition nécessaire et suffisante les variables $< U, X >$ et $< V, X >$ sont-elles indépendantes ?
   e) Donner un exemple de vecteurs U et V tels que $< U, X >$ et $< V, X >$ soient des variables aléatoires indépendantes.

Exercice 1.7. [*Fonctions d'un vecteur gaussien*]
Soient $a \in \mathbb{R}^* \setminus \{\pm 1\}$ et $X = (X_0, ..., X_d)'$ un $(d+1)$-vecteur gaussien tel que $X_i \rightsquigarrow \mathcal{N}(0, 1)$ pour tout i et

$$\Gamma_X = \left\{ \begin{array}{cccccc} 1 & a & a & \ldots & \ldots & a \\ a & 1 & a^2 & \ldots & \ldots & a^2 \\ a & a^2 & 1 & a^2 & & a^2 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ a & a^2 & & a^2 & 1 & a^2 \\ a & a^2 & \ldots & \ldots & a^2 & 1 \end{array} \right\}$$

1. Soit $Y = (X_0, Y_1, ..., Y_d)'$ tel que
$$X_i = aX_0 + \sqrt{1-a^2}Y_i, i = 1, \dots, d.$$
   Déterminer la loi du vecteur aléatoire Y.
2. Exprimer X en fonction de Y.
3. (a) Déterminer la loi de $S_d = (X_1 + \dots + X_d)/d$, $d \geq 1$.
   (b) On pose $Z = S_d/X_0$. Déterminer la loi de $\sqrt{d}(Z-a)/\sqrt{1-a^2}$ et déduire celle de Z.

Exercice 1.8. [*Décomposition d'un vecteur gaussien*]

Soit X un vecteur gaussien centré de matrice de covariance $\Gamma = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix}$. Montrer qu'il existe une vecteur Y bi-dimensionnel centré réduit tel que $X = BY$ où B vérifie $BB' = \Gamma$.

Chapitre 2

Suites aléatoires

Sommaire

2.1 Autonomie : Limites supérieures et inférieures

2.1.1 Cas des suites réelles

Définition 2.1

Soit $(x_n)_{n\in\mathbb{N}}$ une suite réelle. On définit la limite supérieure de la suite $(x_n)_{n\in\mathbb{N}}$ par

$$\overline{\lim_{n\to+\infty}}\, x_n = \lim_{n\to+\infty} \sup_{m\geq n} x_m$$

et sa limite inférieure par

$$\underline{\lim_{n\to+\infty}}\, x_n = \lim_{n\to+\infty} \inf_{m\geq n} x_m.$$

Ces limites existent dans $\overline{\mathbb{R}}$ car les suites

$$\left(\sup_{m\geq n} x_m\right)_{n\in\mathbb{N}} \text{ et } \left(\inf_{m\geq n} x_m\right)_{n\in\mathbb{N}}$$

sont monotones (respectivement décroissante et croissante).

Si la suite n'est ni majorée ni minorée

$$\overline{\lim_{n\to+\infty}}\, x_n = +\infty \text{ et } \underline{\lim_{n\to+\infty}}\, x_n = -\infty.$$

Si la suite est majorée mais pas minorée

$$\overline{\lim_{n\to+\infty}}\, x_n =\in \overline{\mathbb{R}} \text{ et } \underline{\lim_{n\to+\infty}}\, x_n = -\infty.$$

Si la suite est minorée mais pas majorée

$$\overline{\lim_{n\to+\infty}} x_n = +\infty \text{ et } \underline{\lim_{n\to+\infty}} x_n \in \mathbb{R}.$$

Proposition 2.1

On a

$$\underline{\lim_{n\to+\infty}} x_n \leq \overline{\lim_{n\to+\infty}} x_n.$$

Si la suite est bornée, les limites supérieure et inférieure sont réelles et ces deux valeurs sont les valeurs d'adhérence maximale et minimale de la suite. En outre, toujours dans le cas borné, la suite converge si et seulement si les deux limites coïncident et dans ce cas

$$\lim_{n\to+\infty} x_n = \underline{\lim_{n\to+\infty}} x_n = \overline{\lim_{n\to+\infty}} x_n.$$

2.1.2 Cas des fonctions réelles

Si $(f_n)_{n\in\mathbb{N}}$ est une suite de fonctions réelles définie sur un ensemble Ω, on définit les limites supérieure et inférieure de la suite ponctuellement :

$$\left(\overline{\lim_{n\to+\infty}} f_n\right)(x) = \overline{\lim_{n\to+\infty}} f_n(x), \text{ pour tout } x \in \Omega$$

et

$$\left(\underline{\lim_{n\to+\infty}} f_n\right)(x) = \underline{\lim_{n\to+\infty}} f_n(x), \text{ pour tout } x \in \Omega.$$

L'intérêt de ces limites est qu'elles sont définies même lorsqu'il n'y a pas convergence et elles sont souvent utilisées pour montrer des étapes intermédiaires dans les théorèmes de convergence. Voici par exemple un résultat dont on se sert pour démontrer le théorème de convergence dominée de Lebesgue.

Théorème 2.1 (*Lemme de Fatou*)

Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré et soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions boréliennes positives sur $(\Omega, \mathcal{F})$. Alors

$$\int_\Omega \underline{\lim_{n\to+\infty}} f_n d\mu \leq \underline{\lim_{n\to+\infty}} \int_\Omega f_n d\mu.$$

2.1.3 Cas des ensembles

Soit $(A_n)_{n\in\mathbb{N}}$ une suite de sous-ensembles d'un ensemble non vide Ω. On peut définir une notion de limite de la suite $(A_n)_{n\in\mathbb{N}}$ lorsqu'elle est monotone de la façon suivante :

- Si la suite $(A_n)_{n\in\mathbb{N}}$ est croissante i.e. si $A_n \subset A_{n+1}$ pour tout n, on définit

$$\lim_{n\to+\infty} A_n = \bigcup_{n\in\mathbb{N}} A_n;$$

- Si la suite $(A_n)_{n\in\mathbb{N}}$ est décroissante i.e. si $A_{n+1} \subset A_n$ pour tout n, on définit

$$\lim_{n\to+\infty} A_n = \bigcap_{n\in\mathbb{N}} A_n.$$

On ne cherche pas à généraliser la notion de limite à d'autres cas, mais on définit des limites supérieure et inférieure de n'importe quelle suite d'ensembles $(A_n)_{n\in\mathbb{N}}$ selon le même principe que celui utilisé pour les suites réelles.

Définition 2.2

Soit $(A_n)_{n\in\mathbb{N}}$ une suite de sous-ensembles d'un ensemble non vide Ω. On définit les limites inférieure et supérieure de la suite par

$$\limsup_{n\to+\infty} A_n = \bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} A_m \text{ et } \liminf_{n\to+\infty} A_n = \bigcup_{n\in\mathbb{N}} \bigcap_{m\geq n} A_m.$$

On remarque que

$$\left(\bigcup_{m\geq n} A_m\right)_{n\in\mathbb{N}} \quad \text{est une suite décroissante et}$$

$$\left(\bigcap_{m\geq n} A_m\right)_{n\in\mathbb{N}} \quad \text{est une suite croissante}$$

ce qui justifie l'emploi du terme de limite.

En outre, on vérifie aisément que, comme pour le cas des suites réelles, les limites supérieure et inférieure coïncident avec la notion de limite dans le cas monotone, i.e. si $(A_n)_{n\in\mathbb{N}}$ est monotone

$$\limsup_{n\to+\infty} A_n = \liminf_{n\to+\infty} A_n = \lim_{n\to+\infty} A_n.$$

Proposition 2.2

On a toujours

$$\liminf_{n\to+\infty} A_n \subset \limsup_{n\to+\infty} A_n.$$

Démonstration : Soit $x \in \liminf_n A_n$. Alors il existe n_0 tel que pour tout $m \geq n_0$, $x \in A_m$, autrement dit x est dans tous les A_m à partir d'un certain rang. Pour un point y, être dans la limite supérieure de la suite signifie que pour tout n, il existe $m \geq n$ tel que $x \in A_m$. Donc cela signifie que y est dans une suite infinie de A_m. Il est clair que si x est dans tous les A_m à partir d'un certain rang, x est dans une suite infinie de A_m et donc l'inclusion est vérifiée.

Pour tester la maitrise des notions introduites dans ce paragraphe, on pourra faire en autonomie les exercices suivants.

Exemple : On suppose que $(A_n)_{n\in\mathbb{N}}$ est une suite de sous-ensembles deux à deux disjoints d'un ensemble non vide Ω. Calculer $\limsup_n A_n$ et $\liminf_n A_n$. On pourra commencer par l'exemple

$$A_n = [n, n+1[, \text{ pour tout } n \in \mathbb{N}.$$

Exemple : Soit $(x_n)_{n\in\mathbb{N}}$ une suite de réels strictement positifs. Pour tout $n \in \mathbb{N}$, on définit $A_n = [0, x_n[$. Calculer les limites inférieures et supérieures de la suite $(A_n)_{n\in\mathbb{N}}$ dans les cas suivants :

1. La suite $(x_n)_{n\in\mathbb{N}}$ converge, plus précisément on étudiera uniquement les exemples :

$$(a)\ x_n = 1 + \frac{1}{n} \quad (b)\ x_n = 1 - \frac{1}{n} \quad (c)\ x_{2n} = 1 + \frac{1}{2n} \text{ et } x_{2n+1} = 1 - \frac{1}{2n+1}.$$

2.2 Suites infinies d'événements

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

Pour commencer, nous rappelons quelques résultats sur les familles indépendantes (cf. fin chapitre 1).

Proposition 2.3 (*Principe d'associativité pour les tribus indépendantes*)

Si $(\mathcal{G}_n)_{n\in\mathbb{N}}$ est une suite de tribus indépendantes, alors pour tous I_1, $I_2 \subset \mathbb{N}$ tels que $I_1 \cap I_2 = \varnothing$, les suites $(\mathcal{G}_n)_{n\in I_1}$ et $(\mathcal{G}_n)_{n\in I_2}$ sont indépendantes, c'est-à-dire que les tribus

$$\sigma\left(\bigcup_{n\in I_1} \mathcal{G}_n\right) \text{ et } \sigma\left(\bigcup_{n\in I_2} \mathcal{G}_n\right)$$

sont indépendantes.

Proposition 2.4

Si $\mathcal{A}$ et $\mathcal{B}$ sont deux algèbres de Boole indépendantes, alors $\sigma(\mathcal{A})$ et $\sigma(\mathcal{B})$ le sont également.

Voici maintenant un nouvel outil, la tribu des événements asymptotiques.

Définition 2.3 (*Tribu queue ou tribu asymptotique*)

1. Soit $(\mathcal{G}_n)_{n\in\mathbb{N}}$ une suite de sous-tribus de $\mathcal{F}$. La tribu

$$\mathcal{L} = \bigcap_{n\geq 0} \sigma\left(\bigcup_{m\geq n} \mathcal{G}_m\right)$$

est appelée *tribu asymptotique* de la suite $(\mathcal{G}_n)_{n\in\mathbb{N}}$.

2. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. La *tribu asymptotique ou tribu queue* de la suite est la tribu asymptotique de la suite $(\sigma(X_n))_{n\in\mathbb{N}}$.

Proposition 2.5

(i) Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements. Pour tout n, on pose $\mathcal{G}_n = \sigma(\{A_n\})$ et on note $\mathcal{L}$ la tribu asymptotique de cette suite. Alors $\limsup_n A_n$, $\liminf_n A_n \in \mathcal{L}$.

(ii) Soient $(\mathcal{G}_n)_{n\in\mathbb{N}}$ une suite de tribus et $\mathcal{L}$ sa tribu asymptotique. Pour tout n, on pose

$$\mathcal{F}_n = \sigma\left(\bigcup_{k\geq n} \mathcal{G}_k\right).$$

Soit alors $(A_n)_{n\in\mathbb{N}}$ une suite d'événements tels que $A_n \in \mathcal{F}_n$ pour tout n. On a également $\limsup_n A_n$, $\liminf_n A_n \in \mathcal{L}$.

Démonstration : On remarque que, pour tout $N \in \mathbb{N}$,

$$\limsup_n A_n = \bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} A_m = \bigcap_{n\geq N} \bigcup_{m\geq n} A_m$$

puisque la suite

$$\left(\bigcup_{m\geq n} A_m\right)_{n\in\mathbb{N}}$$

est décroissante. On en déduit que $\limsup_n A_n \in \mathcal{F}_N$ pour tout N et donc $\limsup_n A_n \in \mathcal{L}$ par définition. On procède de même pour la limite inférieure.

Théorème 2.2 (*Loi 0 − 1 de Kolmogorov*)

Soit $(\mathcal{G}_n)_{n\in\mathbb{N}}$ une suite de sous-tribus de $\mathcal{F}$ indépendantes. Alors les événements de sa tribu asymptotique sont de probabilité 0 ou 1.

Démonstration : Pour tout n, on pose

$$\mathcal{F}_n = \sigma\left(\bigcup_{k\geq n} \mathcal{G}_k\right).$$

Comme les tribus de la suite sont indépendantes, pour tout n, $\mathcal{G}_n$ est indépendante de $\mathcal{F}_{n+k}$ pour tout $k \geq 1$, et donc de

$$\bigcap_{k\geq n+1} \mathcal{F}_k = \mathcal{L}.$$

On en déduit que, pour tout n, $\mathcal{F}_n$ est également indépendante de $\mathcal{L}$ et donc finalement, $\mathcal{L} = \bigcap_n \mathcal{F}_n$ est indépendante d'elle-même. Donc tous les événements de $\mathcal{L}$ sont indépendants d'eux-mêmes et on a vu que les événements indépendants d'eux-mêmes sont les événements de probabilité 0 ou 1.

Exemple :

1. Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements indépendants, $\limsup_n A_n$ et $\liminf_n A_n$ sont des événements négligeables ou presque certains. En outre
   - $\mathbb{P}[\liminf_n A_n] = 1$ signifie que, presque sûrement il existe un n tel que pour tout $k \geq n$, A_k est réalisé, c'est-à-dire, presque sûrement tous les A_n sont réalisés à partir d'un certain rang.
   - De même, on peut monter que $\mathbb{P}[\limsup_n A_n] = 1$ signifie que, presque sûrement, une infinité des A_n se réalise.
   - Que signifient $\mathbb{P}[\liminf_n A_n] = 0$ et $\mathbb{P}[\limsup_n A_n] = 0$? Quelles sont finalement les situations possibles ?

   Application : On joue à pile ou face. Montrer que, presque sûrement, on a une infinité de "pile" et une infinité de "face".
2. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. indépendantes. Alors soit $\sum_n X_n$ converge presque sûrement soit $\sum_n X_n$ diverge presque sûrement. En effet, on définit la tribu asymptotique de la suite comme dans la définition 2.3-2 et on remarque que

   $$\sum_{n\geq 0} X_n \text{ converge si et seulement si, pour tout } N,\ \sum_{n\geq N} X_n \text{ converge.}$$

   Or

   $$\left[\sum_{n\geq N} X_n \text{ converge}\right] \in \mathcal{F}_N$$

   donc l'événement $[\sum_n X_n \text{ converge}] \in \mathcal{F}_N$ pour tout N et donc $[\sum_n X_n \text{ converge}] \in \mathcal{L}$. On en déduit que la probabilité de cet événement est 0 ou 1, ce qu'il fallait démontrer.

Théorème 2.3 (*Lemme de Borel-Cantelli*)

Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements de $\mathcal{F}$.

1. *Si $\sum_n \mathbb{P}[A_n] < +\infty$ alors $\mathbb{P}[\limsup_n A_n] = 0$ (presque sûrement, seul un nombre fini de A_n se réalise).*
2. *Si $\sum_n \mathbb{P}[A_n] = +\infty$ et si les $(A_n)_{n\in\mathbb{N}}$ sont indépendants, alors $\mathbb{P}[\limsup_n A_n] = 1$ (presque sûrement une infinité de A_n se réalise).*

Démonstration :

1. On a
$$\mathbb{P}\left[\limsup_n A_n\right] = \lim_{n\to+\infty} \mathbb{P}\left[\bigcup_{m\geq n} A_m\right] \leq \lim_{n\to+\infty} \sum_{m\geq n} \mathbb{P}[A_m].$$
Or si la série converge, la queue de la série tend vers 0, et donc la probabilité de la limite supérieure est nulle.
2. On montre en fait que
$$\mathbb{P}\left[\left(\limsup_n A_n\right)^c\right] = \mathbb{P}\left[\liminf_n A_n^c\right] = 0.$$
En effet,
$$\mathbb{P}\left[\liminf_n A_n^c\right] = \lim_{n\to+\infty} \mathbb{P}\left[\bigcap_{m\geq n} A_m^c\right] = \lim_{n\to+\infty} \lim_{N\to+\infty} \prod_{k=n}^{N} \mathbb{P}[A_k^c]$$
car les événements sont indépendants. Or
$$\prod_{k=n}^{N}(1-\mathbb{P}[A_k]) \leq \prod_{k=n}^{N} \mathrm{e}^{-\mathbb{P}[A_k]} = \mathrm{e}^{-\sum_{k=n}^{N}\mathbb{P}[A_k]}$$
et cette dernière quantité tend vers 0 quand N tend vers l'infini parce que la série des probabilités diverge. On a donc bien montré le résultat souhaité.

2.3 Convergence de suites de variables aléatoires

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

2.3.1 Différents types de convergence

Définition 2.4

Soient $(X_n)_{n\in\mathbb{N}}$ et X une suite de v.a. et une v.a. définies sur $(\Omega, \mathcal{F})$ et à valeurs dans $(E, \mathcal{B}(E))$ un espace vectoriel normé muni de sa tribu borélienne. On dit que

- ☐ $(X_n)_{n\in\mathbb{N}}$ converge vers X presque sûrement si $\mathbb{P}[\lim_{n\to+\infty} X_n = X] = 1$.
- ☐ $(X_n)_{n\in\mathbb{N}}$ converge vers X en probabilités si pour tout $\varepsilon > 0$, $\mathbb{P}[|X_n - X| > \varepsilon] \to 0$ quand $n\to+\infty$.
- ☐ $(X_n)_{n\in\mathbb{N}}$ converge vers X en loi si $\mathbb{E}[h(X_n)] \to \mathbb{E}[h(X)]$ pour toute fonction h continue bornée sur $(E, \mathcal{B}(E))$ et à valeurs dans $\mathbb{R}$.
- ☐ $(X_n)_{n\in\mathbb{N}}$ converge vers X dans $L^p(\Omega, \mathcal{F}, \mathbb{P})$ (pour un $p\in[1,+\infty]$) si toutes les v.a. X_n et X sont $L^p(\Omega, \mathcal{F}, \mathbb{P})$ et $||X_n - X||_p \to 0$ quand $n\to+\infty$

Remarque : Si $(E, \mathcal{E}, \mu)$ est un espace mesuré et si l'on considère des fonctions mesurables à valeurs dans un espace vectoriel normé de norme $|\cdot|$ muni de sa tribu borélienne $(F, \mathcal{B}(F))$, alors

- Si $1 \leq p < +\infty$, la norme $||\cdot||_p$ est définie par
$$||f||_p = \left(\int_\Omega |f|^p d\mu\right)^{\frac{1}{p}}, \text{ pour tout } f \in L^p(E, \mathcal{E}, \mu).$$
- L'espace $L^\infty(E, \mathcal{E}, \mu)$ est l'espace des fonctions mesurables presque sûrement bornées et la norme L^∞ est la plus petite des constantes qui majore presque sûrement la norme dans F, appelée aussi le supremum essentiel.
- Les espaces L^p sont des espaces de Banach.
- Pour tous $f \in L^p$ et $g \in L^q$ avec $1/p + 1/q = 1$, $fg \in L^1$ et on a l'inégalité de Hölder
$$||fg||_1 \leq ||f||_p ||g||_q.$$
Dans le cas $p = q = 2$, cette inégalité est connue sous le nom d'inégalité de Cauchy-Schwarz.
- Pour tout $f,\ g \in L^2$, on définit
$$\langle f, g\rangle_2 = \int_E fg d\mu.$$
Cette forme bilinéaire est un produit scalaire sur L^2 qui est un espace de Hilbert.
- Si $\mu(E) < +\infty$ alors $L^\infty \subset L^p \subset L^1$ pour tout $p \in]1, +\infty[$.
- Si $\mu(E) = 1$ alors $||f||_p \leq ||f||_{p'}$ pour tout $p' \geq p$ dès que $f \in L^{p'}$.
- Si $\mu(E) = 1$ et si $f \in L^p$ alors on a l'inégalité de Markov suivante :
$$\mu(|f| > \varepsilon) \leq \frac{1}{\varepsilon^p}\mathbb{E}\left[|f|^p\right] \text{ pour tout } \varepsilon > 0.$$
- Soit $f \geq 0$ presque partout, une fonction intégrable. Alors pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que pour tout $A \in \mathcal{E}$,
$$\mu(A) < \delta \Rightarrow \int_A f d\mu < \varepsilon.$$

Remarque : Dans les espaces L^p, on identifie les fonctions qui sont égales presque partout.

2.3.2 Critères de convergence

Proposition 2.6

La suite $(X_n)_{n\in\mathbb{N}}$ converge presque sûrement si et seulement si
$$\mathbb{P}\left[\limsup_{n\to+\infty}[|X_n - X| > \varepsilon]\right] = 0 \textit{ pour tout } \varepsilon > 0.$$

Démonstration : $X_n \to X$ si et seulement si pour tout $\varepsilon > 0$, il existe $n \in \mathbb{N}$ tel que pour tout $m \geq n$, $|X_m - X| \leq \varepsilon$. Autrement dit
$$[X_n \to X \text{ quand } n \to +\infty] = \bigcap_{\varepsilon>0}\bigcup_{n\in\mathbb{N}}\bigcap_{m\geq n}[|X_m - X| \leq \varepsilon] = \bigcap_{\varepsilon>0}\liminf_{n\to+\infty}[|X_n - X| \leq \varepsilon].$$
Les ensembles $\liminf_n[|X_n - X| \leq \varepsilon]$ décroissent quand ε décroit vers 0. Donc
$$\mathbb{P}\left[X_n \to X\right] = 1 \text{ si et seulement si } \mathbb{P}\left[\liminf_n[|X_n - X| \leq \varepsilon]\right] = 1 \text{ pour tout } \varepsilon > 0,$$
d'où le résultat annoncé en passant au complémentaire.

Théorème 2.4 (***Convergence en loi***)

Soient $(X_n)_{n\in\mathbb{N}}$ et X une suite de v.a. et une v.a. définies sur $(\Omega, \mathcal{F})$ et à valeurs dans $(E, \mathcal{B}(E))$ un espace vectoriel normé muni de sa tribu borélienne. Les propositions suivantes sont équivalentes :

(i) $(X_n)_{n\in\mathbb{N}}$ *converge vers* X *en loi.*

(ii) *Pour toute fonction* h *uniformément continue bornée de* $(E, \mathcal{B}(E))$ *dans* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $\mathbb{E}[h(X_n)] \to \mathbb{E}[h(X)]$ *quand* $n \to +\infty$.

(iii) $\overline{\lim}_{n\to+\infty} \mathbb{P}[X_n \in F] \leq \mathbb{P}[X \in F]$ *pour tout ensemble* F *fermé.*

(iv) $\underline{\lim}_{n\to+\infty} \mathbb{P}[X_n \in O] \geq \mathbb{P}[X \in O]$ *pour tout ensemble* O *ouvert.*

(v) *Fonctions de répartition : dans le cas où* $E = \mathbb{R}$, $F_{X_n}(x) \to F_X(x)$ *en tout point de continuité* x *de* F_X.

(vi) *Fonctions caractéristiques : dans le cas où* $E = \mathbb{R}^d$, $\varphi_{X_n}(t) \to \varphi_X(t)$ *pour tout* $t \in \mathbb{R}^d$.

Démonstration : On fait une preuve partielle du théorème.

(i)⇒(ii) Evident.

(ii)⇒(iii) Soit F un fermé. Pour tout $k \in \mathbb{N}^*$, on définit

$$F_k = \left\{ x \in E / \ d(x, F) \leq \frac{1}{k} \right\} \text{ où } d(x, F) = \inf_{y\in F} |x - y| \text{ est la distance de } x \text{ à } F \text{ dans } E.$$

Soit g définie sur $\mathbb{R}^+$ par $g(t) = \mathbf{1}_{[0,1]}(t)(1 - t)$ uniformément continue sur $\mathbb{R}^+$. Pour tout k, on définit alors sur E la fonction f_k par $f_k(x) = g(d(x, F)k)$. Alors f_k est uniformément continue car d est Lipschitzienne. De plus, $\mathbf{1}_F \leq f_k \leq \mathbf{1}_{F_k}$. On en déduit que, pour tout k,

$$\mathbb{P}[X_n \in F] = \mathbb{E}[\mathbf{1}_F(X_n)] \leq \mathbb{E}[f_k(X_n)] \text{ pour tout } n,$$

et donc en passant à la limite supérieure et en utilisant (ii),

$$\limsup_{n\to+\infty} \mathbb{P}[X_n \in F] \leq \mathbb{E}[f_k(X)].$$

Mais, en utilisant la seconde inégalité, $\mathbb{E}[f_k(X)] \leq \mathbb{P}[X \in F_k]$, d'où

$$\limsup_{n\to+\infty} \mathbb{P}[X_n \in F] \leq \mathbb{P}[X \in F_k] \text{ pour tout } k.$$

Comme $(F_k)_k$ est une suite décroissante d'ensembles, $[X \in F_k]$ est une suite décroissante d'événements et on obtient

$$\limsup_{n\to+\infty} \mathbb{P}[X_n \in F] \leq \mathbb{P}\left[\bigcap_k [X \in F_k]\right] = \mathbb{P}\left[X \in \bigcap_k F_k\right].$$

enfin, comme F est fermé,

$$F = \bigcap_k F_k$$

ce qui conclut la preuve.

(iii)⇔(iv) Par passage au complémentaire.

(iii),(iv)⇒(v) Soit $t \in \mathbb{R}$ fixé. Pour tout n, $F_{X_n}(t) = \mathbb{P}[X \in]-\infty, t]]$ donc en utilisant (iii) on obtient $\overline{\lim}_n F_{X_n}(t) \leq F_X(t)$. D'autre part, $\mathbb{P}[X_n \in]-\infty, t]] \geq \mathbb{P}[X_n \in]-\infty, t[\,]$ pour tout n et on en déduit $\underline{\lim}_n F_{X_n}(t) \geq \mathbb{P}[X \in]-\infty, t[\,]$ en utilisant (iv). Mais pour tout $s < t$,

$$\mathbb{P}[X \in]-\infty, t[\,] \geq \mathbb{P}[X \in]-\infty, s]] = F_X(s),$$

i.e. $\underline{\lim}_n F_{X_n}(t) \geq F_X(s)$ pour tout $s < t$. On en déduit donc que si F_X est continue en t, $\underline{\lim}_n F_{X_n}(t) \geq F_X(t)$, ce qu'il fallait démontrer pour conclure.

...⇒(i) Pour remonter de (ii), (iii) ou (iv), ou (v) vers (i), il faut utiliser des arguments de densité que l'on admet ici.

(ii)⇒(vi) Evident.

(vi)⇒(i) Théorème de Paul Lévy ci-dessous.

Théorème 2.5 (*Théorème de Paul Lévy*)

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de vecteurs aléatoires. Si la suite des fonctions caractéristiques $(\varphi_{X_n})_{n\in\mathbb{N}}$ converge simplement vers une fonction g continue en 0 alors g est fonction caractéristique d'un vecteur aléatoire X et $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X.

Nous admettons la preuve de ce résultat qui repose sur la convergence étroite des mesures et sur le lien entre cette convergence et la convergence des transformées de Fourier des mesures.

Théorème 2.6

Soient $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ deux suites de v.a. convergeant respectivement en loi vers X et Y.

1. *Si $Y = c$ presque sûrement, avec c constante, alors le couple $((X_n, Y_n))_{n\in\mathbb{N}}$ converge en loi vers $(X, Y) = (X, c)$.*
2. *Si X_n et Y_n sont indépendantes pour tout n alors $((X_n, Y_n))_{n\in\mathbb{N}}$ converge en loi vers un couple de v.a. $(\overline{X}, \overline{Y})$ indépendantes et telles que $\mathcal{L}(\overline{X}) = \mathcal{L}(X)$ et $\mathcal{L}(\overline{Y}) = \mathcal{L}(Y)$.*

Démonstration :

1. A faire en exercice (section 2.5.2).
2. Pour tout n, si X_n et Y_n sont indépendantes, on a, pour tout (t, s), $\varphi_{(X_n,Y_n)}(t, s) = \varphi_{X_n}(t)\varphi_{Y_n}(s) \to \varphi_X(t)\varphi_Y(s)$, d'où le résultat.

Remarque : On voit donc que la convergence en loi des marginales n'entraîne pas la convergence en loi des vecteurs en général. En revanche, la convergence en loi des vecteurs entraîne bien la convergence en loi des coordonnées. Par exemple $\varphi_{X_n}(t) = \varphi_{(X_n,Y_n)}(t, 0)$.

Proposition 2.7

La suite $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X dans $\mathbb{R}^d$ si et seulement si, pour tout $a \in \mathbb{R}^d$, $(\langle a, X_n\rangle)_{n\in\mathbb{N}}$ converge en loi vers $\langle a, X\rangle$.

Démonstration : Si $(X_n)_n$ converge en loi vers X, alors comme $x \to \langle a, x\rangle$ est une application continue, $x \to h(\langle a, x\rangle)$ est une fonction continue bornée pour tout h continue bomée et $\mathbb{E}[h(\langle a, X_n\rangle)] \to \mathbb{E}[h(\langle a, X\rangle)]$ quand n tend vers $+\infty$, donc $(\langle a, X_n\rangle)_{n\in\mathbb{N}}$ converge bien en loi vers $\langle a, X\rangle$.

Réciproquement, on remarque que pour tout t et pour tout n $\varphi_{X_n}(t) = \varphi_{\langle t,X_n\rangle}(1)$ et donc la convergence en loi des produits entraîne bien la convergence en loi de X_n.

Remarque : Avec cette proposition, on voit bien que la convergence en loi des coordonnées ne suffit pas pour avoir la convergence en loi du vecteur : il faut la convergence en loi e toutes les combinaisons linéaires des coordonnées pour avoir la convergence en loi du vecteur.

2.3.3 Liens entre les convergences des v.a.

Théorème 2.7

$(X_n)_{n\in\mathbb{N}}$ et X sont respectivement une suite de v.a. et une v.a. définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans un espace vectoriel normé E. On a :

1. *Si $(X_n)_{n\in\mathbb{N}}$ converge vers X dans $L^{p'}$ alors $(X_n)_{n\in\mathbb{N}}$ converge vers X dans L^p pour $1 \leq p \leq p' \leq \infty$.*
2. *Si $(X_n)_{n\in\mathbb{N}}$ converge vers X dans un L^p pour $1 \leq p < +\infty$, alors $(X_n)_{n\in\mathbb{N}}$ converge vers X en probabilités.*
3. *Si $(X_n)_{n\in\mathbb{N}}$ converge vers X presque surement alors $(X_n)_{n\in\mathbb{N}}$ converge vers X en probabilités.*
4. *Si $(X_n)_{n\in\mathbb{N}}$ converge vers X en probabilités alors $(X_n)_{n\in\mathbb{N}}$ converge vers X en loi.*
5. *Si $(X_n)_{n\in\mathbb{N}}$ converge vers X dans L^∞ alors $(X_n)_{n\in\mathbb{N}}$ converge vers X presque surement.*
6. *Si $(X_n)_{n\in\mathbb{N}}$ converge vers X en probabilités alors il existe une sous-suite de $(X_n)_{n\in\mathbb{N}}$ qui converge vers X presque sûrement.*
7. *Si $(X_n)_{n\in\mathbb{N}}$ converge vers X dans un L^p pour $1 \leq p < +\infty$, alors il existe une sous-suite de $(X_n)_{n\in\mathbb{N}}$ qui converge vers X presque sûrement.*

Question : Faire un schéma pour représenter ces convergences.

Démonstration :

1. Vu dans une section précédente. Repose sur les inégalités entre les normes obtenues grâce aux inégalités de Hölder.
2. On suppose que $(X_n)_n$ converge vers X dans L^p et on utilise les inégalités de Markov de la remarque à la suite de la definition 2.4 : pour tout $\varepsilon > 0$ et pour tout n,
$$\mathbb{P}\left[|X_n - X| > \varepsilon\right] \leq \frac{1}{\varepsilon^p}\mathbb{E}\left[|X_n - X|^p\right].$$
3. On suppose $X_n \to X$ presque sûrement. Soit $\varepsilon > 0$ fixé. On a
$$\mathbb{P}\left[|X_n - X| > \varepsilon\right] = \int_\Omega \mathbf{1}_{[|X_n - X| > \varepsilon]} d\mathbb{P}.$$
Comme $X_n \to X$ presque sûrement, la fonction indicatrice est nulle pour n assez grand en presque tout point de Ω (donc converge p.s. vers 0). En outre, la fonction indicatrice est majorée par la constante 1 qui est intégrable et donc, par le théorème de convergence dominée,
$$\mathbb{P}\left[|X_n - X| > \varepsilon\right] \to 0 \text{ quand } n \to +\infty.$$
4. On suppose que $X_n \to X$ en probabilité. Soit h une fonction uniformément continue bornée. Alors, pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que pour tous $(x, y) \in E^2$ tels que $|x - y| < \eta$, $|h(x) - h(y)| < \varepsilon$. Par ailleurs, pour tout $\eta > 0$,
$$\begin{aligned} \left|\mathbb{E}\left[h(X_n)\right] - \mathbb{E}\left[h(X)\right]\right| &\leq \int_\Omega |h(X_n) - h(X)| d\mathbb{P} \\ &= \int_{[|X_n - X| < \eta]} |h(X_n) - h(X)| d\mathbb{P} + \int_{[|X_n - X| \geq \eta]} |h(X_n) - h(X)| d\mathbb{P}. \end{aligned}$$
Donc pour le η défini plus haut, on obtient
$$\left|\mathbb{E}\left[h(X_n)\right] - \mathbb{E}\left[h(X)\right]\right| \leq \varepsilon\mathbb{P}\left[|X_n - X| < \eta\right] + 2\|h\|_\infty \mathbb{P}\left[|X_n - X| \geq \eta\right].$$

On majore $\mathbb{P}[|X_n - X| < \eta]$ par 1 et pour ε et η fixés, on définit n_0 tel que, pour tout, $n \geq n_0$,

$$\mathbb{P}[|X_n - X| \geq \eta] \leq \frac{\varepsilon}{2\|h\|_\infty}.$$

On a donc montré, pour tout $\varepsilon > 0$, il existe n_0 tel que pour tout $n \geq n_0$,

$$|\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| \leq 2\varepsilon$$

ce qui conclut la preuve.

5. Evident.
6. On suppose $X_n \to X$ en probabilité et on fixe $\varepsilon > 0$. Pour α quelconque, il existe n_0 tel que pour tout $n \geq n_0$
$$\mathbb{P}[|X_n - X| > \varepsilon] \leq \alpha.$$
Alors, on peut construire une suite *croissante* d'entiers $(n_k)_{k\in\mathbb{N}^*}$ (de sorte que $(X_{n_k})_k$ soit une suite extraite de $(X_n)_{n\in\mathbb{N}}$) telle que pour tout k
$$\mathbb{P}[|X_{n_k} - X| > \varepsilon] \leq \frac{1}{k^2}.$$
On a alors en particulier
$$\sum_{k\geq 1} \mathbb{P}[|X_{n_k} - X| > \varepsilon] \leq \sum_{k\geq 1} \frac{1}{k^2} < +\infty$$
et donc, d'après le lemme de Borel-Cantelli Théorème 2.3,
$$\mathbb{P}\left[\limsup_{k\to+\infty}[|X_{n_k} - X| > \varepsilon]\right] = 0$$
ce qui, d'après le critère de convergence Proposition 2.6, équivaut à $X_{n_k} \to X$ presque sûrement. On a donc bien une sous-suite qui converge presque sûrement.
7. Il suffit d'appliquer les propositions 2. et 6. ci-dessus.

Proposition 2.8

Si $(X_n)_n$ converge en loi vers une constante c alors $(X_n)_n$ converge en probabilité vers c.

Démonstration : Soit $\varepsilon > 0$. $|X_n - c| \geq \varepsilon$ si et seulement si $[X_n \in F]$ où $F =]-\infty, c-\varepsilon] \cup [c+\varepsilon, +\infty[$ est fermé. Donc, comme $(X_n)_n$ converge vers c en loi,

$$\overline{\lim_{n\to+\infty}}\ \mathbb{P}[|X_n - c| \geq \varepsilon] \leq \mathbb{P}[c \in F] = 0,$$

et $(X_n)_n$ converge vers X en probabilité.

Théorème 2.8

[Convergence dominée pour la convergence en probabilité] Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. qui converge en probabilité vers une v.a. X pour un $p \geq 1$. S'il existe $Y \in L^p$ telle que $|X_n| \leq |Y|$ presque sûrement pour tout n alors $(X_n)_n$ converge vers X dans L^p.

Démonstration : Comme $(X_n)_{n\in\mathbb{N}}$ converge en probabilité vers X, il existe une sous-suite qui converge presque sûrement vers X et on en déduit $|X| \leq |Y|$ presque sûrement et donc $X \in L^p$. On peut donc remplacer X_n par $X_n - X$ et X par 0 et comme la convergence L^p est la convergence L^1 de la pième puissance, on est ramené à démontrer le résultat suivant : si $(X_n)_n$ converge vers 0 en probabilité et si la suite est dominée par $|Y|$ pour $Y \in L^1$, alors $(X_n)_n$ vers 0 dans L^1, i.e. $\mathbb{E}[|X_n|] \to 0$ quand $n \to 0$. Pour tout n et pour tout $\varepsilon > 0$, on a

$$\begin{aligned} \int_\Omega |X_n| d\mathbb{P} &= \int_{[|X_n|>\varepsilon]} |X_n| d\mathbb{P} + \int_{[|X_n|\leq\varepsilon]} |X_n| d\mathbb{P} \\ &\leq \int_{[|X_n|>\varepsilon]} |Y| d\mathbb{P} + \varepsilon \mathbb{P}[|X_n| \leq \varepsilon]. \end{aligned}$$

On majore le second terme par ε. Pour le premier, on utilise la dernière propriété de la remarque qui fait suite à la définition 2.4 en vérifiant qu'elle implique que si $(A_n)_n$ est une suite d'événements tels que $\mathbb{P}[A_n] \to 0$ quand $n \to +\infty$, alors

$$\int_{A_n} Z d\mathbb{P} \to 0 \text{ quand } n \to +\infty$$

pour toute fonction Z presque sûrement positive intégrable. On l' applique à $A_n = [|X_n| > \varepsilon]$ dont la probabilité converge bien vers 0 quand n tend vers l'infini et $Z = |Y|$, on en déduit que le premier terme tend vers 0 quand n tend vers $+\infty$, puis que $\mathbb{E}[|X_n|] \leq 2\varepsilon$ pour n assez grand et donc $\lim_{n\to+\infty} \mathbb{E}[|X_n|] = 0$.

2.4 Théorèmes limites

2.4.1 Lois des grands nombres

Théorème 2.9 (*Loi faible des grands nombres*)

Soit $(X_n)_n$ une suite de v.a. i.i.d. intégrables. Alors

$$\frac{X_1 + \dots + X_n}{n} \to \mathbb{E}[X_1] \text{ en probabilité et dans } L^1.$$

Remarque : En 1A, on a montré une version L^2 pour des v.a. décorrélées à variances bornées dans laquelle on obtient la convergence en probabilité et dans L^2.

Théorème 2.10 (*Loi forte des grands nombres*)

Soit $(X_n)_n$ une suite de v.a. i.i.d. intégrables. Alors

$$\frac{X_1 + \dots + X_n}{n} \to \mathbb{E}[X_1] \text{ presque sûrement.}$$

En outre, si la suite est dans L^2, la convergence a lieu dans L^2 également.

2.4.2 Théorème limite central

Théorème 2.11 (*Théorème limite central*)

Si $(X_n)_n$ est une suite de v.a. i.i.d. de $\mathbb{R}^d$ de carré intégrable alors

$$\sqrt{n}\left(\frac{X_1 + \dots + X_n}{n} - \mathbb{E}[X_1]\right) \to \mathcal{N}_d(0, \mathbf{Cov}[X_1]).$$

Démonstration : On montre la convergence des fonctions caractéristiques.

- ☐ dans $\mathbb{R}$. On pose $\overline{X}_n = \dfrac{X_1 + ... + X_n}{n}$, on suppose $\mathbb{E}[X_1] = 0$ et on pose $\mathbf{Var}[X_1] = \sigma^2$. Alors, pour tout t,

$$\varphi_{\sqrt{n}\overline{X}_n}(t) = \varphi_{X_1+\ldots+X_n}\left(\frac{t}{\sqrt{n}}\right) = \varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)^n = \left(1 + 0 - \frac{1}{2}\frac{t^2}{n}\sigma^2 + o\Big(\frac{1}{n}\Big)\right)^n.$$

  Cette quantité converge[1] vers $\exp\left(-\frac{1}{2}\sigma^2 t^2\right)$ qui est bien la valeur en t de la fonction caractéristique d'une v.a. $\mathbb{N}(0, \sigma^2)$ donc le résultat est démontré.

- ☐ dans $\mathbb{R}^d$. On utilise l'égalité $\varphi_{\sqrt{n}\overline{X}_n}(t) = \varphi_{\langle t, \sqrt{n}\overline{X}_n\rangle}(1)$ pour tout $t \in \mathbb{R}^d$ et on conclut en utilisant le résultat déjà démontré dans $\mathbb{R}$ (à faire en exercice).

2.5 Exercices

2.5.1 Calculs asymptotiques

Exercice 2.9.
Soit (X_n) une suite aléatoire i.i.d. On pose $S_n = X_1 + \cdots + X_n$. Montrer que

1. $\sup_n X_n = +\infty$ p.s ou $\sup_n X_n < +\infty$ p.s.
2. $\dfrac{S_n}{n}$ diverge ou converge p.s.

Exercice 2.10. [*Une marche aléatoire simple*]
Un joueur lance une pièce une infinité de fois. Soit $p \in]0, 1[$, $p \neq 1/2$, la probabilité d'obtenir "face" à un lancer. Soit S_n la v.a. égale au nombre de face obtenu diminué du nombre de "pile", en n lancers. On pose $S_0 = 0$.

1. Calculer $\mathbb{P}[S_n = 0]$ selon la parité de n.
2. Soit $N = Card\{n \in \mathbb{N}^* : S_n = 0\}$ le nombre d'instants où la trajectoire coupe l'axe des temps. Montrer que l'espérance de N est finie.
3. Calculer $\mathbb{P}[\limsup(S_n = 0)]$.

Exercice 2.11. [*Apparition d'un mot dans un schéma de Bernoulli*]
On cherche la probabilité d'apparition d'une chaîne de piles et faces fixée dans une suite infinie de lancers indépendants d'une pièce équilibrée ou non. On fixe donc un mot de longueur k avec l piles et $k - l$ faces. On note X_1, X_2, ... les résultats des lancers et pour tout $n \geq 1$ et A_n l'événement "le mot apparaît dans la séquence X_{n+1},...,X_{n+k}".

1. Calculer la probabilité de A_n.
2. Montrer que presque sûrement le mot fixé apparaît une infinité de fois.

1. Ce résultat asymptotique est aussi valable pour des nombres complexes.

2.5.2 Convergence des variables aléatoires

Exercice 2.12. [*Théorème de Slutsky*]
On suppose que la suite de v.a. $(X_n)_n$ converge en loi vers une v.a. X et que la suite $(Y_n)_n$ converge en loi vers C, avec C constante presque sûrement.

1. Montrer que $(Y_n)_n$ converge en probabilité vers C (Proposition 2.8).
2. En déduire que le couple $((X_n, Y_n))_n$ converge en loi vers (X, C). Indication : On pourra utiliser le point ii du théorème 2.4 .
3. En déduire que

$$X_n + Y_n \xrightarrow{\mathcal{L}} X + C, \quad X_n Y_n \xrightarrow{\mathcal{L}} XC, \quad X_n/Y_n \xrightarrow{\mathcal{L}} X/C, \ (si\ C \neq 0).$$

2.5.3 Les théorèmes de convergence

Exercice 2.13.
Soit $(X_n)_n$ une suite aléatoire i.i.d. de loi P de fonction caractéristique ϕ. Posons $Y_n = (X_n + Y_{n-1})/2$ pour tout $n \geq 1$, avec $Y_0 = X_0/2$.

1. (a) Exprimer Y_n en fonction de X_i pour $0 \leq i \leq n$.
   (b) Calculer la fonction caractéristique ϕ_n de Y_n en fonction de ϕ et de n.
   (c) Etudier la convergence en loi de la suite $(Y_n)_n$ lorsque P est la loi de Cauchy $\mathcal{C}s(1)$.
2. Etudier la convergence de $(\sum_{i=1}^n X_i/n)_n$ et $(\sum_{i=1}^n X_i/\sqrt{n})_n$ pour la loi P ci-dessus.

Exercice 2.14. Démontrer le théorème 2.8 pour $p = 1$.
Exercice 2.15. Soit $(\xi_i, i \geq 1)$ une suite de v.a.i.i.d de loi $\mathcal{N}(0, 1)$. Pour $n \geq 1$, on pose

$$X_0 = 0, X_n = \theta X_{n-1} + \xi_n,$$

où $\theta \in \mathbb{R}$.

1. Calculer la fonction caractéristique de X_n. En déduire la loi de X_n.
2. On suppose que $|\theta| < 1$. Montrez que X_n converge en loi vers une variable X dont on précisera la loi.
3. On suppose que $|\theta| > 1$. Déterminer la loi de $\theta^{-n} X_n$.

Chapitre 3

Chaines de Markov

Sommaire

Les chaînes de Markov sont des processus à temps discret ayant deux caractéristiques :

1. l'ensemble des états pris par le processus est dénombrable.
2. Conditionnellement au présent, le futur est indépendant du passé.

Ces caractéristiques (ainsi que l'homogénéité — cf. plus loin) permettent une étude poussée de ces processus.

3.1 Généralités sur les chaînes de Markov.

Soit $X = (X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$. On suppose que tous les X_n prennent leurs valeurs dans un ensemble E dénombrable. On peut alors considérer que $E = \mathbb{N}$ et nous noterons les éléments de E comme des entiers avec les notations i, j, k,...

Définition 3.1

Soit $X = (X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires entières. On dit que X est une *chaîne de Markov* si, pour tout $n \in \mathbb{N}$ et pour tout $(i_0, ..., i_n, j) \in \mathbb{N}^{n+2}$,

$$\mathbb{P}\left[X_{n+1} = j | X_0 = i_0, ..., X_n = i_n\right] = \mathbb{P}\left[X_{n+1} = j | X_n = i_n\right].$$

Une chaîne de Markov est donc caractérisée par ses *probabilités de transition* $\mathbb{P}\left[X_{n+1} = j | X_n = i\right]$. En particulier :

Définition 3.2

On dit qu'une chaîne de Markov $X = (X_n)_{n\in\mathbb{N}}$ est *homogène* si, pour tout $n \in \mathbb{N}$ et pour tout $(i,j) \in \mathbb{N}^2$, $\mathbb{P}[X_{n+1} = j | X_n = i]$ ne dépend pas de n.

On note en général pour une chaîne homogène, la probabilité de passer de l'état i dans l'état j

$$p_{ij} = \mathbb{P}[X_{n+1} = j | X_n = i].$$

La « matrice » $P = (p_{ij})_{i,j\in\mathbb{N}}$ est alors appelée la *matrice de transition* de la chaîne. Elle est caractérisée par le fait que tous ses coefficients sont positifs et que, pour tout $i \in \mathbb{N}$,

$$\sum_{j\in\mathbb{N}} p_{ij} = 1.$$

Une telle matrice s'appelle une matrice stochastique.

On notera $p_{ij}^{(n)} = \mathbb{P}[X_n = j | X_0 = i]$ la probabilité de passer de l'état i à j en n étapes.

3.2 Exemples.

3.2.1 La chaîne à deux états.

Déterminer la matrice stochastique associée à cette chaîne et réalisez-en une représentation sous forme de graphe dont les sommets représentent les états et les arcs orientés et pondérés les probabilités de transition.

3.2.2 La marche aléatoire sur $\mathbb{Z}$.

Elle décrit la position d'un mobile qui se déplace, indépendamment de sa position, d'un pas vers la droite avec probabilité p et d'un pas vers la gauche avec probabilité q et reste sur place avec une probabilité $1 - p - q$.

3.2.3 Le modèle de diffusion d'Ehrenfest.

C'est un modèle discret de diffusion d'un gaz. On suppose que a boules numérotées de 1 à a sont réparties dans deux urnes notées A et B. A chaque étape, on tire au hasard un numéro de 1 à a avec probabilité uniforme, puis on change d'urne la boule qui a été tirée. Modéliser la suite $(X_n)_{n\geq 1}$ des nombres de boules successifs qu'il y a dans l'urne A.

3.2.4 La ruine du joueur.

Les joueurs A et B jouent à pile ou face. Il y a a euros en jeu en tout. Un euro passe de A à B avec probabilité p et de B à A avec probabilité $q = 1 - p$. Le jeu s'arrête lorsque l'un des joueurs est ruiné. On note X_n la fortune de A à l'issue de la n-ième partie.

3.2.5 Le modèle bonus-malus.

Une compagnie d'assurance classe les niveaux de bonus-malus de ses clients selon l'échelle 0,1,2,..., l'état 0 étant le plus avantageux pour le client. On note X_n le niveau de malus d'un client l'année n. L'année $n+1$, le niveau de malus est calculé en respectant les règles suivantes :

1) le niveau de malus ne peut baisser ; 2) le niveau de malus augmente du nombre d'accidents par an ; 3) ce nombre suit tous les ans une loi de Poisson de paramètre λ indépendante du nombre d'accidents les années passées.

3.3 Propriétés des chaînes de Markov homogènes.

Dans cette section $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov homogène. On note π^n la *loi ou distribution* de X_n ; c'est le vecteur ligne (éventuellement infini) de $[0,1]^E$, défini, pour tout $j \in E$, par

$$\pi_j^n = \mathbb{P}[X_n = j]$$

et satisfaisant donc $\sum_{j\in E} \pi_j^n = 1$.

Proposition 3.1

Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov de matrice de transition P. Alors, pour tout $n \in \mathbb{N}$,

$$\mathbb{P}[X_{n+1} = j] = \sum_{i\in E} p_{ij}\mathbb{P}[X_n = i], \ \forall\ j \in E.$$

Démonstration : On utilise la formule des probabilités totales :

$$\mathbb{P}[X_{n+1} = j] = \sum_{i\in E} \mathbb{P}[X_{n+1} = j|X_n = i]\mathbb{P}[X_n = i] = \sum_{i\in E} p_{ij}\mathbb{P}[X_n = i].$$

Avec la notation pour la loi de la chaîne, l'égalité de la proposition précédente s'écrit

$$\pi^{n+1} = \pi^n P.$$

De cette égalité, on déduit par récurrence que

$$\pi^n = \pi^0 P^n.$$

Plus précisément, on a la propriété suivante :

Proposition 3.2 (*Propriété de Chapman-Kolmogorov*)

Soit X une chaîne de Markov homogène de matrice de transition P. Pour tout n, on note $P^{(n)}$ la matrice de transition de la chaîne en n étapes. Alors

$$P^{(n)} = P^n$$

Démonstration : On le montre par récurrence : soient $i,\ j \in E$. Pour tout $n \in \mathbb{N}$, la probabilité de passer de i à j entre les instants n et $n+2$ est

$$p_{ij}^{(2)} = \mathbb{P}[X_{n+2} = j|X_n = i] = \sum_{k\in E} \frac{\mathbb{P}[X_{n+2} = j, X_{n+1} = k, X_n = i]}{\mathbb{P}[X_n = i]}$$

$$= \sum_{k\in E} \mathbb{P}[X_{n+2} = j|X_{n+1} = k, X_n = i]\mathbb{P}[X_{n+1} = k|X_n = i] = \sum_{k\in E} p_{kj}p_{ik} = (P^2)_{ij}.$$

Proposition 3.3 (*Loi de dimension finie d'une chaîne*)

Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov de matrice de transition P. Alors, la loi de dimension finie de la chaîne est donnée, pour tout $m \geq 1$ et pour tout $(i_0, ..., i_m) \in E^{m+1}$, par

$$\mathbb{P}[X_0 = i_0, ..., X_m = i_m] = \mathbb{P}[X_0 = i_0] \prod_{k=0}^{m-1} p_{i_k i_{k+1}}.$$

Démonstration : On a

$$\begin{aligned}\mathbb{P}[X_0 = i_0, ..., X_m = i_m] &= \mathbb{P}[X_m = i_m | X_0 = i_0, ..., X_{m-1} = i_{m-1}]\mathbb{P}[X_0 = i_0, ..., X_{m-1} = i_{m-1}] \\ &= p_{i_{m-1} i_m}\mathbb{P}[X_0 = i_0, ..., X_{m-1} = i_{m-1}]\end{aligned}$$

et le résultat souhaité s'obtient par récurrence.

3.4 Irréductibilité, probabilités stationnaires.

On considère des chaînes à états dans $E = \mathbb{N}$.

3.4.1 Classification des états d'une chaîne.

Définition 3.3

Soient $i, j \in E$. On dit que i conduit à j et on note $i \rightsquigarrow j$ s'il existe n tel que $p_{ij}^{(n)} > 0$.

Proposition 3.4

C'est une relation réflexive ($i \rightsquigarrow i$), et transitive ($i \rightsquigarrow j$ et $j \rightsquigarrow k$, alors $i \rightsquigarrow k$) sur les états.

Démonstration : $p_{ii}^{(0)} = 1$ donc $i \rightsquigarrow i$.

Si $i \rightsquigarrow j$ et $j \rightsquigarrow k$, alors $\exists\, n,\ m \geq 0$ tels que $p_{ij}^{(n)} > 0$ et $p_{jk}^{(m)} > 0$. Il vient $p_{ik}^{(n+m)} = \sum_{l\geq 0} p_{il}^{(n)} p_{lk}^{(m)} \geq p_{ij}^{(n)} p_{jk}^{(m)} > 0$ donc $i \rightsquigarrow k$.

Définition 3.4

Soient $i, j \in E$. On dit que i et j communiquent et on note $i \leftrightsquigarrow j$ si $i \rightsquigarrow j$ et $j \rightsquigarrow i$.

Proposition 3.5

C'est une relation d'équivalence.

Démonstration : Elle est symétrique par définition, transitive et réflexive d'après la proposition précédente.

On définit ainsi des classes d'équivalence sur E, ce qui nous donne une partition des états.

Exemple : Classification des états des chaînes à deux état.

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

Définition 3.5

On dit qu'une chaîne est irréductible s'il existe une seule classe d'équivalence pour la relation.

La notion d'irréductibilité est lié à la matrice de transition de la chaîne et non à la loi initiale. Pour montrer qu'une chaine de Markov est irréductible, il peut être utile de tracer un graphe orienté dont les sommets sont les états de la chaîne et où une arête représente une transition possible (l'arête (i,j) existe uniquement si $p_{ij} > 0$). La chaîne est alors irréductible si et seulement s'il existe un chemin fermé passant au moins une fois par tous les états de la chaîne.

3.4.2 Probabilités stationnaires.

L'une des problématiques les plus courantes concernant les chaînes de Markov homogènes consiste à déterminer leurs *distributions ou (probabilités ou mesures) invariantes (ou stationnaires)* et à étudier la convergence éventuelle de la chaîne vers ces distributions.

Définition 3.6

Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov homogène de matrice de transition P. On dit que la distribution $\pi = (\pi_i)_{i\in E}$ est *invariante* pour la chaîne si

$$\pi = \pi P.$$

Remarque : Du fait que P est une matrice stochastique, la valeur 1 est une valeur propre de P associée aux vecteurs propres constants. Les distributions invariantes, si elles existent, sont les transposées des vecteurs propres de P^t associés à la valeur propre 1.

Proposition 3.6

1. *Si l'espace d'états est fini, il existe toujours au moins une mesure stationnaire.*
2. *Si la chaîne est irréductible, il existe au plus une mesure stationnaire.*

Remarque : Le point 1 est une conséquence de théorème de Perron-Frobenius.

Corollaire 3.1

Si la chaîne est irréductible sur un espace d'états fini, il existe une unique mesure stationnaire.

Lorsque la distribution stationnaire est unique, on s'intéresse au comportement asymptotique de la chaîne, en particulier, on cherche les conditions sous lesquelles la loi d'une chaîne converge vers la distribution stationnaire.

Proposition 3.7

Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov à états finis $\{1,...,N\}$. On suppose que, pour tout $i \in \{1,...,N\}$, $(\mathbb{P}[X_n = i])_{n\in\mathbb{N}}$ converge et on note π_i sa limite. Alors $\pi = (\pi_1,...,\pi_N)$ est une distribution invariante pour la chaîne.

Démonstration : Comme l'ensemble des distributions de probabilité sur $\{1,...,N\}$ est fermé dans $\mathbb{R}^N$, π est une distribution de probabilité. D'autre part, on a $\pi^{n+1} = \pi^n P$ pour tout n et comme l'application $\pi \to \pi P$ est linéaire de $\mathbb{R}^N$ dans lui-même, elle est continue, et donc π est un point fixe de cette application, soit $\pi = \pi P$.

Remarque : Pour avoir un résultat analogue pour un espace d'état E dénombrable, il faut se donner une topologie sur $\mathbb{R}^E$ qui assure

— la fermeture de l'espace des distributions sur E dans $\mathbb{R}^E$,

— la continuité de l'opérateur $\pi \to \pi P$ de l'espace des distributions sur E dans lui-même.

Par conséquent les limites en loi de la chaîne, si elles existent, sont les distributions invariantes de la chaîne. Réciproquement, il est clair que si π est une probabilité invariante alors une chaîne admettant π comme distribution initiale est stationnaire. Par contre, on n'est pas assuré de la convergence de toute chaîne vers la distribution invariante même si elle existe et est unique.

Dans la suite, nous cherchons à déterminer les distributions stationnaires des chaînes de Markov homogènes et nous étudions la convergence en loi de ces chaînes.

3.5 Décomposition canonique des chaînes de Markov.

3.5.1 Récurrence/transience.

Définitions 1. Soit X une chaîne de Markov sur E et soit $i \in E$.

— On note τ_i le premier passage de la chaîne à l'état i, i.e.

$$\tau_i = \inf\{n > 0/\ X_n = i\}.$$

La variable aléatoire entière τ_i s'appelle un *temps d'arrêt* cf Chapitre 4.

— Le nombre de passages en i de la chaîne est noté N_i, $N_i = \sum_{n \geq 0} \mathbf{1}_{[X_n = i]}$.

Dans la suite, on note $\mathbf{P}_i$ la probabilité conditionnée par le fait que la chaîne soit initialement dans l'état i. L'espérance relative à cette probabilité est notée $\mathbb{E}_i$.

Proposition 3.8

Pour tout $i \in E$ et pour tout $j \in E$

$$\mathbb{E}_i[N_j] = \sum_{n \geq 0} p_{ij}^{(n)}.$$

Démonstration : Comme

$$N_j = \sum_{n \geq 0} \mathbf{1}_{[X_n = j]}, \text{ d'où } \mathbb{E}_i[N_j] = \sum_{n \geq 0} \mathbb{E}_i \left[\mathbf{1}_{[X_n = j]}\right].$$

Mais $\mathbb{E}_i \left[\mathbf{1}_{[X_n = j]}\right] = \mathbf{P}_i[X_n = j] = p_{ij}^{(n)}$.

Définitions 2. Soit $i \in E$. On dit que

— i est récurrent si $\mathbf{P}_i[\tau_i < +\infty] = 1$ (partant de i, la chaîne repasse en i presque sûrement) ;

— i est transient si $\mathbf{P}_i[\tau_i < +\infty] < 1$.

Théorème 3.1

Soit $i \in E$. Les propositions suivantes sont équivalentes :

1. *L'état i est récurrent ;*
2. $\mathbf{P}_i[N_i = +\infty] = 1$;
3. $\mathbb{E}_i[N_i] = +\infty$.

Démonstration : La démonstration de ce théorème repose sur le résultat suivant $N_i \rightsquigarrow \mathcal{G}(1 - \mathbf{P}_i[\tau_i < +\infty])$. De plus

$$\mathbf{P}_i[N_i = +\infty] = \lim_{k\to+\infty} \mathbf{P}_i[N_i > k] = \lim_{k\to+\infty} (\mathbf{P}_i[\tau_i < +\infty])^k$$

D'où la première équivalence. On obtient aussi facilement l'équivalence entre 1) et 3).

Théorème 3.2

Si i est récurrent et $i \rightsquigarrow j$ alors j est récurrent et $j \rightsquigarrow i$.

Démonstration : Si j ne conduit pas à i alors comme il est possible d'aller de i en j, la récurrence de i est mise en défaut puisqu'avec probabilité non nulle on peut aller de i à j et on ne retournera jamais en i. Donc $j \rightsquigarrow i$.

Montrons qu'alors j est récurrent. On note n_{ij} et n_{ji} les entiers tels que

$$p_{ij}^{(n_{ij})} > 0 \text{ et } p_{ji}^{(n_{ji})} > 0.$$

On remarque que, pour tout $n \geq 1$,

$$p_{jj}^{(n_{ij}+n_{ji}+n)} \geq p_{ji}^{(n_{ji})} p_{ii}^{(n)} p_{ij}^{(n_{ij})}.$$

On en déduit que

$$\sum_{n\geq 0} p_{jj}^{(n_{ij}+n_{ji}+n)} \geq p_{ji}^{(n_{ji})} p_{ij}^{(n_{ij})} \sum_{n\geq 0} p_{ii}^{(n)}.$$

Finalement, comme i est récurrent et par la proposition 3.8 et le théorème 3.1 appliqué deux fois, l'état j est récurrent.

Corollaire 3.2 (*nature des états d'une chaîne irréductible*)

Dans une chaîne irréductible, tous les états sont de même nature.

On affine la classification des états récurrents avec la définition suivante.

Définitions 3. Soit i un état. On dit que

— i est récurrent positif si $\mathbb{E}_i[\tau_i] < +\infty$;

— i est récurrent nul si $\mathbb{E}_i[\tau_i] = +\infty$.

3.5.2 Lien avec les mesures invariantes.

Théorème 3.3

1. *Une chaîne n'ayant que des états transients n'admet pas de probabilité invariante.*
2. *S'il existe un état récurrent positif alors il existe une probabilité invariante.*
3. *Si X est une chaîne récurrente irréductible alors la probabilité invariante si elle existe est unique et $\pi_i = 1/\mathbb{E}_i[\tau_i]$.*

Théorème 3.4

Si X admet une probabilité invariante π, alors tout $j \in E$ tel que $\pi_j > 0$ est récurrent

Théorème 3.5

Soit X une chaîne irréductible. Alors les propositions suivantes sont équivalentes :

1. *il existe un état récurrent positif ;*
2. *il existe une probabilité invariante ;*
3. *tous les états sont récurrents positifs.*

Exemple : On considère la marche aléatoire sur $\mathbb{Z}$ de l'exemple 3.2.2.

— Si $p = q$ alors la chaîne est récurrente nulle.

— Si $p \neq q$ alors la marche aléatoire est transiente.

On le démontre dans le cas $p + q = 1$ soit $q = 1 - p$.

1. Monter que la chaîne est irréductible.
2. Montrer que pour tout $n \geq 0$, on a : $p_{0,0}^{(2n+1)} = 0$ et $p_{0,0}^{(2n)} = \binom{2n}{n} p^n (1-p)^n$.
3. En utilisant la formule de Sterling $n\,! \sim \sqrt{2\pi n}\left(\frac{n}{\mathrm{e}}\right)^n$, trouver un équivalent de $p_{0,0}^{(2n)}$ et en déduire que la marche aléatoire est récurrente si $p = 1/2$ et transiente sinon.
4. Lever l'indécision sur le type de récurrence dans le cas $p = 1/2$.

3.6 Deux théorèmes ergodiques.

Les théorèmes ergodiques sont les théorèmes qui permettent de caractériser le comportement asymptotique de moyenne d'une chaine de Markov à l'aide de la probabilité invariante lorsque celle-ci est unique ou de conclure à la convergence en loi d'une chaîne de Markov .

On donne ici pour une chaine à états dénombrables un théorème sur la convergence des moyennes ainsi qu'un deuxième résultat sur la convergence en loi de la chaine. Ce dernier résultat nécessitant l'introduction de la notion d'apériodicité comme condition nécessaire et suffisante.

Les résultats de cette partie sont issus du polycopié de cours de Nizar Touzi.

3.6.1 Un théorème ergodique pour les chaînes irréductible et récurrente positive.

Théorème 3.6 (*Théorème ergodique*)

Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov irréductible et récurrente positive sur E. Alors il existe une unique distribution invariante ν et, pour toute fonction $g : E \to \mathbb{R}$ positive ou telle que $\int |g| d\nu = \sum_{k\in E} |g(k)| \nu(k) < \infty$, on a :

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{n} g(X_i) \to \int g d\nu \ ps.$$

quelle que soit la loi initiale π^0 de X_0.

3.6.2 Convergence en loi des chaines de Markov

Le théorème ergodique implique que pour une chaine de Markov irréductible récurrente positive de loi invariante ν alors

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{P}\left[X_k = i\right] = \frac{1}{n}\sum_{k=1}^{n}\pi^k(i) \to \nu(i),\ \mathbf{P}_j - \text{p.s.}$$

quelque soit $i, j \in E$.

On s'intéresse maintenant à la convergence des $\pi^n(i)$ et non plus simplement à la convergence des moyennes de Cesaro. Ceci nécessite l'hypothèse d'apériodicité.

Définition 3.7

- Soit $i \in E$. On appelle période de l'état i le pgcd (plus grand commun diviseur) des n tels que $p_{ii}^{(n)} > 0$. On le note $d(i)$.
- Si $d(i) = 1$, on dit que i est apériodique.

Proposition 3.9

L'apériodicité est une caractéristique de classe.

Corollaire 3.3 (*chaîne apériodique*)

S'il existe un état apériodique dans une chaîne irréductible alors tous les états sont apériodiques et on dit alors que la chaîne est apériodique.

Lemme 3.1

Pour tout $i \in E$ les assertions suivantes sont équivalentes

1. $d(i) = 1$
2. *il existe $n(i)$ tel que $p^n(i,i) > 0$ pour tout $n \geq n(i)$.*

Démonstration : 2) implique 1) est triviale.

On montre l'autre implication. Soit des entiers $n_1, n_2, \ldots, n_k$ tel que $p^{n_l}(i,i) > 0$ et $PGCD(n_1, \ldots, n_k) = 1$. Le théorème de Bezout assure l'existence de $q_1, \ldots, q_k \in \mathbb{Z}$ tels que

$$\sum_{l=1}^{k} q_l n_l = 1 = a(i) - b(i) \text{ où } a(i) = \sum_{l=1}^{k} q_l^+ n_l \text{ et } b_i = \sum_{l=1}^{k} q_l^- n_l.$$

On pose $n(i) = b(i)^2 - 1 = (b(i) - 1)b(i) + b(i) - 1$. Ainsi pour tout $n \geq n(i)$ la division euclidienne de n par $b(x)$ s'écrit

$$\begin{aligned} n &= db(i) + r \\ &= (d - r)b(i) + ra(i) \\ &= (d - r)\sum_{l=1}^{k} q_l^+ n_l + r\sum_{l=1}^{k} q_l^- n_l. \end{aligned}$$

avec $d \geq r$. L'égalité de Chapman-Kolomorov implique donc que $p^n(i,i) > 0$.

Théorème 3.7

Si X est une chaîne irréductible, récurrente positive, apériodique d'unique loi invariante ν alors $(X_n)_{n\in\mathbb{N}}$ converge en loi vers la probabilité invariante.

La preuve de ce résultat repose sur la proposition suivante

Proposition 3.10

Soient X^1 et X^2 deux chaines de Markov indépendantes de même matrice de transition P irréductible et apériodique. Alors la chaine produit $Y := (X^1, X^2)$ est irréductible apériodique. Si de plus P est récurrente positive, il en est de même pour Y.

Démonstration : Soit$(i_1, i_2, j_1, j_2) \in E^4$. La matrice de transition de la chaine Y est donnée par

$$(P \otimes P)(i_1, i_2, i_1, i_2) = P(i_1, i_1)P(i_2, i_2).$$

Comme P est irréductible, il existe deux entiers m_1 et m_2 tels que $P^{m_k}(i_k, j_k) > 0, k = 1, 2$. Par ailleurs on peut appliquer le lemme 3.1 à j_1 et j_2. On en déduit que pour $n \geq n(j_1)+n(j_2)+m_1+m_2$, $P^n(i_k, j_k) \geq P^{m_k}(i_k, j_k)P^{n-m_k}(j_k, j_k) > 0$, et

$$(P \otimes P)^n(i_1, i_2, j_1, j_2) \geq P^n(i_1, j_1)P^n(i_2, j_2) > 0.$$

Ceci implique l'irréductibilité de la chaine Y. On en déduit l'apériodicité en remarquant que $(P \otimes P)^n(i_1, i_2, i_1, i_2) > 0$ et $(P \otimes P)^{n+1}(i_1, i_2, i_1, i_2) > 0$ pour $n \geq n(i_1) + n(i_2)$ et donc que $d(i_1, i_2) = 1$. Enfin si P est récurrente positive elle admet une unique loi invariante ν et on vérifie que $\nu \otimes \nu$ est invariante pour Y. D'où le résultat avec le théorème 3.5.

démonstration du théorème 3.7 Soient X^1 et X^2 deux chaînes indépendantes de même matrice de transition que X. On introduit la chaîne produit $Y := (X^1, X^2)$. Soit $T := \inf\{n \geq 0, X_n^1 = X_n^2\}$. On peut remarquer que même si les deux chaînes ne sont pas issues de la même loi initiale, après le temps T elles ont la même distribution. En remarquant que

$$\begin{aligned}
\mathbb{P}\left[X_n^1 = i\right] &= \mathbb{P}\left[X_n^1 = i, T > n\right] + \mathbb{P}\left[X_n^1 = i, T \leq n\right] \\
&= \mathbb{P}\left[X_n^1 = i, T > n\right] + \mathbb{P}\left[X_n^2 = i, T \leq n\right] \\
&\leq \mathbb{P}\left[T > n\right] + \mathbb{P}\left[X_n^2 = i\right].
\end{aligned}$$

Ainsi on obtient que $\mathbb{P}\left[X_n^1 = i\right] - \mathbb{P}\left[X_n^2 = i\right] \leq \mathbb{P}\left[T > n\right]$. En inversant les rôles il en découle que $\left|\mathbb{P}\left[X_n^1 = i\right] - \mathbb{P}\left[X_n^2 = i\right]\right| \leq \mathbb{P}\left[T > n\right]$ et si X_0^2 est distribué selon la loi invariante ν, on a

$$\left|\mathbb{P}\left[X_n^1 = i\right] - \nu(i)\right| \leq \mathbb{P}\left[T > n\right].$$

En passant à la limite on a que $\mathbb{P}\left[T > n\right] \to \mathbb{P}\left[T = \infty\right]$ par convergence monotone. De plus d'après le résultat précédent Y est récurrente positive. Par conséquent , $T = \inf_{i \in E} \tau_{i,i} < \infty$ p.s. où $\tau_{i,i}$ est le premier temps de retour de la chaîne Y au point (i, i).

3.7 Exercices

Exercice 3.1.[Ruine du joueur] On reprend le modèle de l'exemple 3.2.4.

1. Rappeler l'expression de X_n ,$n \geq 0$ et la matrice de transition. La chaine est-elle irreductible ?
2. Déterminer les états de la chaine.
3. Existe-t-il une probabilité invariante ? Est-elle unique ? Pourquoi ?

Exercice 3.2.[Bruit qui court] Un message pouvant prendre 2 formes (*oui* ou *non*) est transmis à travers n intermédiaires. On suppose que chaque intermédiaire transmet le message de façon correcte avec une probabilité $p \in]0,1[$ ou le déforme en son contraire avec une probabilité $1-p$. Les intermédiaires sont indépendants.

1. Modéliser cette situation par une chaîne de Markov à 2 états.
2. Calculer la probabilité que l'information transmise par le n-ième intermédiaire soit conforme à l'information initiale. (*indication : remarquer que* $(1,1)$ *et* $(1,-1)$ *sont vecteurs propres de* P *et diagonaliser* P)
3. Que se passe-t-il lorsque $n \to +\infty$?

Exercice 3.3.[Maladie contagieuse] Une maladie se contracte avec une probabilité $0,05$. Quand on l'a contractée, on peut soit en guérir soit avoir des séquelles irréversibles. Ces séquelles sont associées à une immunité totale par la suite. Si on guérit avec une probabilité de $0,3$, en revanche, on n'est immunisé que dans 50% des cas. Dans la population, par ailleurs, $1/5$ des gens est naturellement immunisé.

1. Modéliser l'état d'un individu vis à vis de la maladie dans la période de temps $(n, n+1)$ par une chaîne de Markov à 3 états : non immunisé (NI), immunisé sans séquelle (ISS), avec séquelles (S). Donner son graphe, sa loi initiale et sa matrice de transition.
2. Classifier les états.
3. Existe-t-il une probabilité invariante ? Est-elle unique ? Pourquoi ?

Exercice 3.4.[Suites de v.a. et chaînes de Markov] Soient S et C deux ensembles finis ou dénombrables et $f : S \times C \to S$ une application. On considère une v.a. X_0 à valeurs dans S, ainsi qu'une suite de v.a. $(U_n)_{n\geq 0}$ à valeurs dans C. On définit $(X_n)_{n\geq 0}$ par la formule de récurrence :

$$X_{n+1} = f(X_n, U_{n+1})$$

On suppose que les v.a. $(X_0, U_1, ..., U_n, ...)$ sont indépendantes, et les $(U_i)_{i\geq 0}$ identiquement distribuées.

1. Montrer que $(X_n)_{n\geq 0}$ est une chaîne de Markov homogène, de matrice de transition
$$P(x,y) = P(f(x,U_1) = y)$$
2. **Application : marche aléatoire** Soient $X_0, \xi_1, \xi_2, ...$ des v.a. à valeurs dans $\mathbb{N}$ indépendantes. X_0 est de loi μ et les ξ_i sont de même loi. On définit la marche aléatoire $(X_n)_{n\geq 0}$ par :
$$X_n = X_0 + \xi_1 + ... + \xi_n, \ (n \geq 1)$$
Montrer que $(X_n)_{n\geq 0}$ est une chaîne de Markov homogène. Calculer sa matrice de transition.

Exercice 3.5.[Urne d'Ehrenfest] On considère deux urnes qui contiennent en tout N boules. A chaque unité de temps, une boule est prise au hasard parmi toutes les boules et elle est placée dans l'autre urne. Soit X_n le nombre de boules dans l'urne I au bout de n déplacements. On suppose que $X_0 = 1$.

1. Donner la matrice de transition P de la chaîne de Markov.
2. En note $m_n = \mathbb{E}[X_n]$. Montrer que $m_{n+1} = 1 + \frac{N-2}{N} m_n$.
3. En déduire que $m_n = \frac{N}{2} + \left(1 - \frac{2}{N}\right)^n \left(1 - \frac{N}{2}\right)$.

Exercice 3.6. On considère une marche aléatoire symétrique en deux dimensions dont l'espace des états est l'ensemble $\{(i,j) : i = 0,1,2; j = 0,1,2\}$ i.e la probabilité d'aller à droite, à gauche, en bas ou en haut est de $1/4$. De plus, on suppose que les frontières sont réfléchissantes ; c'est-à-dire lorsque le processus effectue une transition qui le ferait sortir de la région définie par l'espace des états, alors il retourne à la dernière position qu'il occupait (sur la frontière).

1. Déterminer la matrice de transition de la chaîne de Markov.
2. Classifier ses états.
3. Existe-t-il une probabilité invariante ? Est-elle unique ? Pourquoi ?
4. Étudier la convergence de la suite.

Exercice 3.7. On considère une chaine de Markov homogène, admettant $(r+1)$ états, numérotés de 0 à r ($r \geq 2$). On suppose que si le processus est dans l'état 0, il a la probabilité $\frac{1}{2}$ d'y rester et la probabilité $\frac{1}{2r}$ de passer dans l'un quelconque des états $1, 2, \cdots, r$. Il a, d'autre part, la probabilité nulle de passer à l'état 0 à partir de l'un des états $1, 2, \cdots, r$. Enfin, s'il est dans l'état i $(1 \leq i \leq r-1)$ (resp. r), il a la possibilité d'y rester avec la probabilité $\frac{1}{2}$, ou d'aller dans $(i+1)$ (resp. 1) avec la probabilité $\frac{1}{2}$.

1. Donner la matrice de transition P pour $r = 2, 3$ et les différentes classes ainsi que leur périodicité.
2. Pour $r = 2$ et $n \geq 1$, déterminer P^n ; trouver $\lim_n P^n$.
3. Pour $r \geq 2$, déterminer $\lim_n P^n$. (Ind. on calculera $\lim_n p_{i,j}^n$)

Exercice 3.8.[Couverture publicitaire] Un fabricant veut fixer le niveau de publicité qu'il fait passer dans un média. Il peut choisir entre une couverture publicitaire élevée (E) et une couverture publicitaire moyenne (M). Les ventes mensuelles sont réparties en trois catégories suivant leur nombre : C_1 (peu de ventes), C_2 (nombre de ventes normal) et C_3 (beaucoup de ventes). On estime que l'évolution de la catégorie des ventes mensuelles au cours du temps peut être représentée par une chaîne de Markov, dont la matrice de transition dépend de la couverture publicitaire. Les deux matrices de transition sont les suivantes :

$$\text{couv. élevée : } \quad P_E = \begin{pmatrix} 0.2 & 0.5 & 0.3 \\ 0.1 & 0.5 & 0.4 \\ 0.1 & 0.2 & 0.7 \end{pmatrix} \qquad \text{et couv. moyenne : } \quad P_M = \begin{pmatrix} 0.6 & 0.4 & 0 \\ 0.4 & 0.5 & 0.1 \\ 0.4 & 0.5 & 0.1 \end{pmatrix}$$

Un mois de vente de la catégorie C_1 (respectivement C_2, et C_3) rapporte environ 9 000 euros (respectivement 12 000 et 18 000 euros). Une forte couverture publicitaire coûte 5 000 euros par mois, alors qu'une couverture publicitaire moyenne ne coûte que 1000 par mois. Calculer le bénéfice moyen du fabricant sur une grande période de temps lorsque la couverture publicitaire est élevée, puis lorsqu'elle est moyenne. Quel est le choix le plus rentable ?

Exercice 3.9.[Marche aléatoire sur $\mathbb{Z}$-Préparation du BE] On considère la marche aléatoire sur $\mathbb{Z}$ de l'exemple 3.2.2.

— Si $p = q$ alors la chaîne est récurrente nulle.

— Si $p \neq q$ alors la marche aléatoire est transiente.

On le démontre dans le cas $p + q = 1$ soit $q = 1 - p$.

1. Monter que la chaîne est irréductible.
2. Montrer que pour tout $n \geq 0$, on a : $p_{0,0}^{(2n+1)} = 0$ et $p_{0,0}^{(2n)} = \binom{2n}{n} p^n (1-p)^n$.
3. En utilisant la formule de Sterling $n\,! \sim \sqrt{2\pi n}\,\left(\frac{n}{\mathrm{e}}\right)^n$, trouver un équivalent de $p_{0,0}^{(2n)}$ et en déduire que la marche aléatoire est récurrente si $p = 1/2$ et transiente sinon.
4. Lever l'indécision sur le type de récurrence dans le cas $p = 1/2$.

Exercice 3.10.[Bonus-Malus-Préparation du BE] On reprend le modèle de l'exemple 3.2.5.

1. Rappeler l'expression de X_n ,$n \geq 0$ et la matrice de transition. La chaine est-elle irréductible ?
2. Déterminer les états de la chaine.
3. Existe-t-il une probabilité invariante ? Si oui est-elle unique ? Pourquoi ?
4. La suite de $(X_n)_{n\geq 0}$ admet-elle une limite ? Si oui en quel sens ?

Chapitre 4

Conditionnement et martingales

Sommaire

4.1 Espérance conditionnelle

On suppose connu partiellement le résultat d'une expérience aléatoire et on cherche à estimer ce résultat au mieux en fonction de cette observation partielle.

4.1.1 Exemple introductif

On suppose connu le nombre de "pile" obtenus lorsque l'on lance deux fois une pièce de monnaie équilibrée et on cherche à estimer au mieux le résultat du premier lancer.

Modélisation : On note X_i le résultat du i-ème lancer avec la convention $X_i = 1$ si le lancer donne "pile" et $X_i = 0$ sinon. On suppose que les lancers sont indépendants. On cherche à estimer X_1 connaissant $X_1 + X_2$.

Question : Donner les valeurs de X_1 dans les cas suivants : $X_1 + X_2 = 0$, $X_1 + X_2 = 2$ et $X_1 + X_2 = 1$. Dans le denier cas on donnera la loi de X_1 sachant $X_1 + X_2 = 1$.

Sur l'événement $[X_1 + X_2 = 1]$, on n'a aucune information supplémentaire et on peut au mieux calculer la moyenne α de X_1.

Question : Combien vaut-elle ?

En choisissant

$$X_1 = \begin{cases} 0 \text{ si } X_1 + X_2 = 0 \\ \alpha \text{ si } X_1 + X_2 = 1 \\ 1 \text{ si } X_1 + X_2 = 2 \end{cases}$$

on a choisi de donner pour chaque valeur prise par $X_1 + X_2$, l'espérance de X_1 sur cet ensemble.

Conclusion : On remarque alors plusieurs faits concernant la nouvelle variable aléatoire que l'on vient de créer :

- elle est constante là où $X_1 + X_2$ est constante (elle est $\sigma(X_1 + X_2)$-mesurable) ;
- elle a la même moyenne que X_1 sur les ensembles où $X_1 + X_2$ est constante ;
- autrement dit, sur ces ensembles, elle est la meilleure approximation constante de X_1 au sens des moindres carrés (i.e. dans L^2).

Ces remarques constituent en fait les caractérisations de l'approximation que l'on cherche à définir. Cette approximation, qui est une variable aléatoire, s'appelle l'espérance conditionnelle de X_1 sachant $X_1 + X_2$ et se note $\mathbb{E}[X_1 | X_1 + X_2]$.

4.1.2 Définition

Cette notion ne concerne que des v.a. intégrables et l'espérance conditionnelle est elle-même une v.a. intégrable.

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

Proposition 4.1

Soit X une v.a. définie sur $(\Omega, \mathcal{F}, \mathbb{P})$ et soit $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. Si X est intégrable, alors il existe une unique v.a. à un ensemble négligeable près Z intégrable $\mathcal{G}$-mesurable telle que :

$$\text{Pour tout } A \in \mathcal{G},\ \int_A Z d\mathbb{P} = \int_A X d\mathbb{P}. \tag{4.1}$$

Remarque : On retrouve les deux caractérisations discutées dans l'exemple précédent : l'espérance conditionnelle de X est mesurable par rapport à l'information qui décrit le conditionnement et la moyenne de X est préservée sur les événements constituant cette information.

Définition 4.1 (*Espérance conditionnelle*)

La v.a. Z de la proposition précédente s'appelle l'espérance conditionnelle de X sachant $\mathcal{G}$. On la note $\mathbb{E}[X | \mathcal{G}]$.

Remarque : Si X est $\mathcal{G}$-mesurable, on a alors $\mathbb{E}[X | G] = X$ presque sûrement. Ce résultat, bien que trivial est extrêmement important.

Démonstration : On commence par supposer $X \geq 0$ $\mathbb{P}$-presque sûrement et on définit l'application

$$\mu(A) = \int_A X d\mathbb{P} \text{ pour tout } A \in \mathcal{G}.$$

Comme X est intégrable, μ est une mesure bornée sur $\mathcal{G}$, et pour tout $A \in \mathcal{G}$, $\mathbb{P}(A) = 0 \Rightarrow \mu(A) = 0$. D'après le théorème de Radon-Nikodym, il existe une unique fonction Z $\mathcal{G}$-mesurable telle que

$$\mu(A) = \int_A Z d\mathbb{P} \text{ pour tout } A \in \mathcal{G}.$$

Dans ce résultat, Z est intégrable positive et $Z = \dfrac{d\mu}{d\mathbb{P}}$ est unique à un ensemble négligeable près.

Si X n'est pas positive presque sûrement, on pose

$$Z = \frac{d\mu^+}{d\mathbb{P}} - \frac{d\mu^-}{d\mathbb{P}}$$

avec

$$\mu^+(A) = \int_A X^+ d\mathbb{P} \text{ et } \mu^-(A) = \int_A X^- d\mathbb{P} \text{ pour tout } A \in \mathcal{G}$$

et le résultat reste vrai.

Définition 4.2 (*Conditionnement par rapport à une variable*)

Soit X une v.a.r. intégrable et soit T une v.a.r. On pose $\mathcal{G} = \sigma(T)$. L'espérance conditionnelle de X sachant T est $\mathbb{E}[X|T] = \mathbb{E}[X|\sigma(T)]$.

Remarque : D'après un résultat du chapitre 2, $\mathbb{E}[X|T] = g(T)$ pour une certaine fonction borélienne g. Dans l'exemple introductif, on remarque

$$\mathbb{E}[X_1|X_1+X_2] = 0 \cdot \mathbf{1}_{[X_1+X_2=0]} + \frac{1}{2} \cdot \mathbf{1}_{[X_1+X_2=1]} + 1 \cdot \mathbf{1}_{[X_1+X_2=2]} = \frac{X_1+X_2}{2}.$$

4.1.3 Caractérisations et exemples

Voici une autre caractérisation possible de $\mathbb{E}[X|\mathcal{G}]$.

Proposition 4.2

Sous les hypothèses de la proposition 4.1, $Z = \mathbb{E}[X|\mathcal{G}]$ si et seulement si Z est $\mathcal{G}$-mesurable, intégrable et si

$$\textit{Pour tout } Y\ \mathcal{G}\textit{-mesurable et t.q. } YX \textit{ intégrable, } \int_\Omega YZ d\mathbb{P} = \int_\Omega YX d\mathbb{P}. \tag{4.2}$$

On pourra montrer ce résultat en deuxième lecture, il reprend les arguments constructifs du chapitre 3 et les résultats de convergence du chapitre 4.

On a une caractérisation dans le cas L^2 qui reprend l'idée que la meilleure approximation constante d'une fonction L^2 est sa moyenne vue dans le cas discret à la section 4.1.1.

Proposition 4.3

Si $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ et si $\mathcal{G}$ est une sous-tribu de $\mathcal{F}$, alors $\mathbb{E}[X|\mathcal{G}]$ est la projection orthogonale de X sur $L^2(\Omega, \mathcal{G}, \mathbb{P})$.

Démonstration : L'espace $L^2(\Omega, \mathcal{G}, \mathbb{P})$ est un sous-espace vectoriel fermé de $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Soit Z la projection orthogonale de X sur $L^2(\Omega, \mathcal{G}, \mathbb{P})$. Alors Z est $\mathcal{G}$-mesurable et de carré intégrable par construction. De plus, pour tout $Y \in L^2(\Omega, \mathcal{G}, \mathbb{P})$,

$$\langle Y, X - Z \rangle_2 = 0,$$

c'es-à-dire

$$\int_\Omega YZ d\mathbb{P} = \int_\Omega YX d\mathbb{P}.$$

Donc Z satisfait bien les conditions de la définition de l'espérance conditionnelle en prenant $Y = \mathbf{1}_A$ pour $A \in \mathcal{G}$.

Proposition 4.4 (*Cas discret*)

Soit X une v.a. intégrable et soit $\mathcal{G}$ une sous-tribu de $\mathcal{F}$.

1. *Si $\mathcal{G} = \{\varnothing, \Omega\}$, alors $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$ presque sûrement.*
2. *Si $\mathcal{G}$ est engendrée par la partition $\mathcal{P} = \{A_i,\ i \in \mathbb{N}\}$ telle que $\mathbb{P}(A_i) \neq 0$ pour tout $i \in \mathbb{N}$, alors*
$$\mathbb{E}[X|\mathcal{G}] = \sum_{i\in\mathbb{N}} \left(\frac{1}{\mathbb{P}[A_i]} \int_{A_i} X d\mathbb{P} \right) \mathbf{1}_{A_i}.$$

Démonstration :

1. $\mathbb{E}[X]$ est $\mathcal{G}$ mesurable car constante et
$$(i)\ \int_{\varnothing} \mathbb{E}[X]d\mathbb{P} = \int_{\varnothing} X d\mathbb{P} \text{ et } (ii)\ \int_{\Omega} \mathbb{E}[X]d\mathbb{P} = \int_{\Omega} X d\mathbb{P}$$
puisque $\mathbb{P}(\Omega) = 1$.
2. On pose
$$Z = \sum_{i \in \mathbb{N}} \left(\frac{1}{\mathbb{P}[A_i]} \int_{A_i} X d\mathbb{P} \right) \mathbf{1}_{A_i}.$$
Alors Z est $\mathcal{G}$-mesurable car constante sur tous les éléments de la partition. En outre, tout élément de $\mathcal{G}$ étant l'union d'un ensemble dénombrable de A_i l'égalité (4.1) aura lieu en tout A de $\mathcal{G}$ si elle a lieu en tout A_i de la partition. Soit donc $i \in \mathbb{N}$. Sur A_i, Z vaut
$$\frac{1}{\mathbb{P}[A_i]} \int_{A_i} X d\mathbb{P}.$$
Donc
$$\int_{A_i} Z d\mathbb{P} = \mathbb{P}[A_i] \times \frac{1}{\mathbb{P}[A_i]} \int_{A_i} X d\mathbb{P} = \int_{A_i} X d\mathbb{P}$$
ce qu'il fallait démontrer.

Proposition 4.5 (*Cas à densité*)

Soient X et T deux variables aléatoires réelles, X étant intégrable. On suppose que (X,T) admet une densité $f : \mathbb{R}^2 \to \mathbb{R}$ et on note f_T la densité marginale de T. Alors $\mathbb{E}[X|T] = g(T)$ avec
$$g(t) = \mathbf{1}_{[f_T \neq 0]}(t) \frac{1}{f_T(t)} \int_{\mathbb{R}} x f(x,t) dx \text{ pour tout } t \in \mathbb{R}.$$

Démonstration : On rappelle que
$$f_T(t) = \int_{\mathbb{R}} f(x,t) dx \text{ pour tout } t \in \mathbb{R}.$$
On pose $Z = g(T)$ avec g définie dans la proposition. Alors g est borélienne (application du théorème de Fubini), donc Z est bien $\sigma(T)$-mesurable. Soit $A \in \sigma(T)$. Alors il existe $B \in \mathcal{B}(\mathbb{R})$ tel que $A = T^{-1}(B)$. On en déduit
$$\int_A Z d\mathbb{P} = \int_{\Omega} \mathbf{1}_{T^{-1}(B)} Z d\mathbb{P} = \int_{\Omega} \mathbf{1}_B(T) g(T) d\mathbb{P}.$$
On utilise alors la définition de g et la formule de transfert. Il vient, après simplification par $f_T(t)$ (un peu délicate mais il suffit de remarquer que $f_T(t) = 0 \Rightarrow f(t,x) = 0$ pour presque tout x),
$$\int_A Z d\mathbb{P} = \int_{\mathbb{R}} \mathbf{1}_B(t) \int_{\mathbb{R}} x f(x,t) dx dt = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}_B(t) x f(x,t) dx dt.$$
Mais en utilisant à nouveau la formule de transfert, on trouve
$$\int_A Z d\mathbb{P} = \int_{\Omega} \mathbf{1}_B(T) X d\mathbb{P} = \int_A X d\mathbb{P}$$
ce qu'il fallait démontrer.

4.1.4 Principaux résultats

Il existe un certain nombre de résultats importants qui sont utilisés régulièrement dans le calcul des espérances conditionnelles. Chaque preuve d'un de ces résultats constitue un exercice facile et intéressant, nous laissons au lecteur le soin de rédiger ces preuves. Seuls le théorème 4.1 qui est le plus important pour les applications et le moins évident à manipuler, et le corollaire qui traite le cas des les vecteurs gaussiens sont démontrés.

Un résultat de type "barycentre".

Proposition 4.6

Soit X une v.a. intégrable définie sur $(\Omega, \mathcal{F}, \mathbb{P})$ et soient $\mathcal{G}$ et $\mathcal{G}'$ deux sous-tribus de $\mathcal{F}$ telles que $\mathcal{G}' \subset \mathcal{G}$. Alors

$$\mathbb{E}\left[\mathbb{E}\left[X\,|\mathcal{G}\right]|\mathcal{G}'\right](=\mathbb{E}\left[\mathbb{E}\left[X\,|\mathcal{G}'\right]|\mathcal{G}\right]) = \mathbb{E}\left[X\,|\mathcal{G}'\right] \textit{ presque sûrement.}$$

L'espérance d'une espérance conditionnelle.

Corollaire 4.1

Sous les hypothèses sur X et $\mathcal{G}$ de la proposition précédente, $\mathbb{E}\left[\mathbb{E}\left[X\,|\mathcal{G}\right]\right] = \mathbb{E}\left[X\right]$.

Conditionnement sous hypothèse de mesurabilité.

Proposition 4.7

Soient X et Y deux v.a.r. définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ et soit $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. On suppose XY et Y intégrables. Alors, si X est $\mathcal{G}$-mesurable,

$$\mathbb{E}\left[XY\,|\mathcal{G}\right] = X\mathbb{E}\left[Y\,|\mathcal{G}\right] \textit{ presque sûrement.}$$

Conditionnement sous hypothèse d'indépendance.

Proposition 4.8

Soit $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ et soit $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. On suppose X indépendante de $\mathcal{G}$. Alors

$$\mathbb{E}\left[X\,|\mathcal{G}\right] = \mathbb{E}\left[X\right] \textit{ presque sûrement.}$$

Corollaire 4.2

Soit (X, T) un vecteur gaussien, X et T étant de dimension quelconque. Alors, si $\mathbf{Cov}\left[T\right]$ inversible, on a

$$\mathbb{E}\left[X\,|T\right] = \mathbb{E}\left[X\right] + \mathbf{Cov}\left[X,T\right]\mathbf{Cov}\left[T\right]^{-1}(T - \mathbb{E}\left[T\right]).$$

Démonstration : On suppose pour commencer que X et T sont centrés. Il suffit de montrer que

$$\mathbb{E}\left[X\,|T\right] = \mathbf{Cov}\left[X,T\right]\mathbf{Cov}\left[T\right]^{-1}T.$$

Soit donc $Z = \mathbf{Cov}\left[X,T\right]\mathbf{Cov}\left[T\right]^{-1}T$. Z est linéaire en T donc $\sigma(T)$-mesurable et intégrable. On montre que $\mathbb{E}\left[X\,|T\right] = \mathbb{E}\left[Z\,|T\right]$ car alors on aura $\mathbb{E}\left[X\,|T\right] = \mathbb{E}\left[Z\,|T\right] = Z ps$. On a :

$$\mathbf{Cov}\left[Z,T\right] = \mathbf{Cov}\left[X,T\right]\mathbf{Cov}\left[T\right]^{-1}\mathbf{Cov}\left[T,T\right] = \mathbf{Cov}\left[X,T\right]$$

donc $\mathbf{Cov}[Z-X,T]=0$ et $Z-X$ et T sont décorrélés. Mais $(Z-X,T)$ est un vecteur gaussien (comme transformée linéaire de (X,T)), donc $Z-X$ et T sont indépendants. On en déduit en particulier que

$$\mathbb{E}[Z-X\,|T]=\mathbb{E}[Z-X]=0 \text{ puisque } X \text{ et } T \text{ sont centrés.}$$

Cela implique finalement le résultat cherché $\mathbb{E}[X\,|T]=\mathbb{E}[Z\,|T]=Z$ presque sûrement puisque Z est $\sigma(T)$-mesurable.

Si X et T sont quelconque, on applique le résultat que l'on vient de montrer à $X-\mathbb{E}[X]$ et $T-\mathbb{E}[T]$ centrés pour obtenir le résultat souhaité.

Théorème principal qui combine les résultats précédents.

Théorème 4.1

Soient X et T deux v.a.r. définies sur $(\Omega,\mathcal{F},\mathbb{P})$ et soit $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. Soit $\psi:\mathbb{R}^2\to\mathbb{R}$ une application borélienne. On suppose $\psi(X,T)$ intégrable et $\psi(x,T)$ intégrable pour tout $x\in\mathbb{R}$.

On suppose que X est $\mathcal{G}$-mesurable et que T est indépendante de $\mathcal{G}$. Alors

$$\mathbb{E}[\psi(X,T)\,|\mathcal{G}]=\phi(X) \text{ p.s., où } \phi(x)=\mathbb{E}[\psi(x,T)] \text{ pour tout } x\in\mathbb{R}.$$

Ce théorème donne en fait la seule façon d'écrire proprement et conjointement le fait que X, $\mathcal{G}$-mesurable, n'est pas modifié par le conditionnement et que T, indépendante de $\mathcal{G}$, doit être intégrée.

Démonstration : On pose $Z=\phi(X)$. Alors ϕ est borélienne par Fubini et donc Z est $\sigma(X)$-mesurable donc $\mathcal{G}$-mesurable. Z est également intégrable (toujours par Fubini) et on va montrer pour conclure la proposition 4.2. Soit donc Y $\mathcal{G}$-mesurable de sorte que YZ et YX soient intégrables. On veut montrer que

$$\int_\Omega YZ d\mathbb{P}=\int_\Omega Y\psi(X,T)d\mathbb{P}.$$

On a, par la formule de transfert,

$$\int_\Omega Y\psi(X,T)d\mathbb{P}=\int_{\mathbb{R}^3} y\psi(x,t)\mathbb{P}_{(X,Y,T)}(dx,dy,dt),$$

mais (X,Y) et T sont indépendants donc $\mathbb{P}_{(X,Y,T)}=\mathbb{P}_{(X,Y)}\otimes\mathbb{P}_T$. On en déduit par Fubini

$$\int_\Omega Y\psi(X,T)d\mathbb{P}=\int_{\mathbb{R}^2} y\int_{\mathbb{R}}\psi(x,t)d\mathbb{P}_T(dt)d\mathbb{P}_{(X,Y)}(dx,dy).$$

Mais pour tout $x\in\mathbb{R}$,

$$\int_{\mathbb{R}}\psi(x,t)d\mathbb{P}_T(dt)=\int_\Omega\psi(x,T)d\mathbb{P}=\mathbb{E}[\psi(x,T)]=\phi(x).$$

Il vient

$$\int_\Omega Y\psi(X,T)d\mathbb{P}=\int_{\mathbb{R}^2} y\phi(x)d\mathbb{P}_{(X,Y)}(dx,dy)=\int_\Omega Y\phi(X)d\mathbb{P},$$

ce qu'il fallait démontrer.

Remarque : L'espérance conditionnelle satisfait la plupart des propriétés de l'espérance (linéarité, convergence monotone, convergence dominée, inégalité de Jensen...).

4.2 Filtrations, temps d'arrêt

4.2.1 Définitions

Dans ce chapitre, on introduit les filtrations de tribus qui permettent de faire dépendre du temps la quantité d'information connue. Les temps d'arrêt sont des temps aléatoires dont la définition s'appuie sur les filtrations pour introduire la notion de non-anticipativité.

Définition 4.3 (*Filtration*)

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. On appelle *filtration de* $\mathcal{F}$ toute suite croissante de sous-tribus de $\mathcal{F}$.

On appelle *espace probabilisé filtré* le quadruplet $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ où $\mathbb{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ est une filtration de $\mathcal{F}$.

Définition 4.4 (*Processus adapté*)

Soit $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\in\mathbb{N}}, \mathbb{P})$ un espace probabilisé filtré. Un *processus* $(\mathcal{F}_n)_{n\in\mathbb{N}}$*-adapté* est une suite de v.a. X_n définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ telles que X_n est $\mathcal{F}_n$-mesurable pour tout $n \in \mathbb{N}$.

Définition 4.5 (*Filtration naturelle*)

Soit $(X_n)_{n\in\mathbb{N}}$ une suite définie sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. On appelle *filtration naturelle* du processus $(X_n)_{n\in\mathbb{N}}$ la plus petite filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ qui rend le processus adapté. Elle est définie par

$$\mathcal{F}_n = \sigma - \{X_k,\ k \leq n\},\ \text{pour tout } n \in \mathbb{N}.$$

Définition 4.6 (*Temps d'arrêt*)

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et soit $\mathbb{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ une filtration de $\mathcal{F}$. Un $\mathbb{F}$*-temps d'arrêt* est une v.a. T à valeurs dans $\overline{\mathbb{N}} = \mathbb{N} \cup \{+\infty\}$ qui vérifie

$$[T \leq n] \in \mathcal{F}_n \text{ pour tout } n \in \mathbb{N}.$$

Voici une définition équivalente des temps d'arrêt qui est couramment utilisée dans le cas discret mais qui n'a pas d'équivalent en temps continu au contraire de la définition ci-dessus.

Proposition 4.9

Sous les conditions de la définition précédente, T est un $\mathbb{F}$-temps d'arrêt si et seulement si

$$[T = n] \in \mathcal{F}_n \textit{ pour tout } n \in \mathbb{N}.$$

Démonstration : On suppose T temps d'arrêt. Alors $[T = 0] = [T \leq 0] \in \mathcal{F}_0$ et pour tout $n \geq 1$

$$[T = n] = [T \leq n] \backslash [T \leq n-1] \in \mathcal{F}_n$$

d'après les propriétés des tribus. Réciproquement,

$$[T \leq n] = \bigcup_{k \leq n} [T = k] \text{ pour tout } n \geq 0,$$

ce qui permet de conclure.

On remarque quand on lit ces définitions, que ces temps d'arrêt sont non anticipatifs, au sens où on n'utilise que les informations sur les événements survenus avant le temps n pour déterminer si le temps d'arrêt est survenu ou non à cet instant-là.

4.2.2 Exemples

Par exemple, parmi les temps aléatoires, on trouve le premier temps (ou le second ...) où un événement particulier se produit. La production ou non de cet événement au temps n est généralement modélisée par une v.a. X_n $\mathcal{F}_n$-mesurable, ce qui assure la non anticipativité.

On suppose maintenant que $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ est un espace probabilisé filtré et que $\mathbb{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$. Les variables aléatoires sont définies sur cet espace.

Exemple :

1. T constante. Alors $[T = n]$ est vide ou égale à Ω quelque soit n et donc $[T = n] \in \mathcal{F}_n$ pour tout n et T est un $\mathbb{F}$-temps d'arrêt.
2. Premier temps d'entrée d'un processus adapté dans un ensemble. Soit $(X_n)_{n\in\mathbb{N}}$ un processus adapté à valeurs dans $(E, \mathcal{B}(E))$un espace vectoriel normé muni de sa tribu borélienne. Soit enfin $B \in \mathcal{B}(E)$. Alors la v.a. T définie par
$$T = \inf\{n \geq 0/\ X_n \in B\} \text{ (et } T = +\infty \text{ si } X_n \notin B \text{ pour tout } n)$$
est un $\mathbb{F}$-temps d'arrêt. On l'appelle le premier temps d'atteinte de B ou premier temps d'entrée dans B.
En effet, pour tout $n \in \mathbb{N}$, $T \leq n$ si et seulement si il existe $k \leq n$ tel que $X_k \in B$, i.e.
$$[T \leq n] = \bigcup_{k=0}^{n} [X_k \in B] \in \mathcal{F}_n.$$
3. Dans ce même contexte, le dernier temps de passage dans B n'est pas un $\mathbb{F}$-temps d'arrêt car, pour savoir s'il s'agit du dernier passage, il faut connaître les événements futurs. Précisément, soit T' ce temps.
$$T' = \sup\{n \geq 0/\ X_n \in B\} \text{ (et } T' = +\infty \text{ si } X_n \notin B \text{ pour tout } n).$$
Alors $T' \leq n$ si et seulement si il existe $k \leq n$ tel que $X_k \in B$ et pour tout $k \geq n+1$, $X_k \notin B$, i.e.
$$[T' \leq n] = \left(\bigcup_{k=0}^{n} [X_k \in B]\right) \cap \left(\bigcap_{k \geq n+1} [X_k \notin B]\right)$$
et cet événement n'est pas dans $\mathcal{F}_n$ en général.

Question : Définir le second temps d'atteinte de B. Est-ce un temps d'arrêt ? Justifier votre réponse.

4.2.3 Propriétés

On suppose toujours que $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ est un espace probabilisé filtré, que $\mathbb{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ et que les variables aléatoires sont définies sur cet espace.

Proposition 4.10 (*Premières propriétés*)

1. *Soient S et T deux temps d'arrêt. Alors, $S + T$, $S \vee T$, $S \wedge T$ (et en particulier, pour une constante $n \in \mathbb{N}$, $T + n$, $T \vee n$, $T \wedge n$) sont des temps d'arrêt.*
2. *Si $(T_n)_{n\in\mathbb{N}}$ est une suite de temps d'arrêt alors $\sup_n T_n$ et $\inf_n T_n$ également*

Cette proposition peut-être démontrée en exercice.

Il est important de pouvoir donner des informations sur une suite à un instant aléatoire. Si $(X_n)_{n\in\mathbb{N}}$ est une suite et T un temps aléatoire tel que $T < +\infty$, on peut définir la valeur de la suite au temps T comme la v.a. X_T définie par

$$(X_T)(\omega) = X_{T(\omega)}(\omega) \text{ pour tout } \omega \in \Omega.$$

Définition 4.7 (*Suite arrêtée*)

Soit $(X_n)_{n\in\mathbb{N}}$ une suite $\mathbb{F}$-adaptée et soit T un $\mathbb{F}$-temps d'arrêt. On appelle suite $(X_n)_{\in\mathbb{N}}$ arrêtée au temps T la suite $(X_{T\wedge n})_{n\in\mathbb{N}}$.

Proposition 4.11

Une suite $\mathbb{F}$-adaptée $(X_n)_{n\in\mathbb{N}}$ arrêtée au $\mathbb{F}$-temps d'arrêt T est une suite $\mathbb{F}$-adaptée.

Démonstration : En effet, on remarque que, pour tout $n \in \mathbb{N}$

$$X_{T\wedge n} = \left(\sum_{k=0}^{n} \mathbf{1}_{[T=k]} + \mathbf{1}_{[T>n]}\right) X_{T\wedge n} = \sum_{k=0}^{n} \mathbf{1}_{[T=k]} X_k + \mathbf{1}_{[T>n]} X_n.$$

Pour tout $k \leq n$, $[T = k] \in \mathcal{F}_n$ et X_k $\mathcal{F}_n$-mesurable. De plus $[T > n] = [T \leq n]^c \in \mathcal{F}_n$ donc $X_{T\wedge n}$ est $\mathcal{F}_n$-mesurable.

Avant de définir les martingales qui synthétisent les éléments introduits dans les deux premières sections de ce chapitre, nous allons encore introduire la notion d'événements antérieurs à un temps aléatoire. Cette notion nous est connue pour un temps déterministe puisque les événements antérieurs à n sont les événements de $\mathcal{F}_n$. On veut donc définir une tribu $\mathcal{F}_T$ pour un temps aléatoire T.

Définition 4.8

On appelle tribu des événements antérieurs à T, v.a. à valeurs dans $\overline{\mathbb{N}}$, la famille

$$\mathcal{F}_T = \{A \in \mathcal{F} / \ \forall n \in \mathbb{N}, A \cap [T \leq n] \in \mathcal{F}_n\}.$$

Proposition 4.12

Si T est un $\mathbb{F}$-temps d'arrêt alors $\mathcal{F}_T$ est une sous-tribu de $\mathcal{F}$.

Proposition 4.13

Soient $(X_n)_{n\in\mathbb{N}}$ une suite $\mathbb{F}$-adaptée à valeur dans $(E, \mathcal{B}(E))$, un espace vectoriel normé muni de sa tribu borélienne, et T un $\mathbb{F}$-temps d'arrêt. Alors la v.a. $X_T \mathbf{1}_{[T<+\infty]}$ est $\mathcal{F}_T$-mesurable.

Démonstration : Soit $B \in \mathcal{B}(E)$. On définit

$$A = [X_T \mathbf{1}_{[T<+\infty]} \in B].$$

On doit montrer que, pour tout $n \in \mathbb{N}$,

$$A \cap [T \leq n] \in \mathcal{F}_n.$$

Or

$$A \cap [T \leq n] = \bigcup_{k=0}^{n} (A \cap [T = k]) = \bigcup_{k=0}^{n} ([X_k \in B] \cap [T = k]) \in \mathcal{F}_n.$$

4.3 Martingales à temps discret

Soit $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, avec $\mathbb{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$, un espace probabilisé filtré. On considère des suites à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

4.3.1 Définition et propriétés

Définition 4.9 (*Martingale*)

Une suite aléatoire $(X_n)_{n\geq 0}$ définie sur $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ est une $\mathbb{F}$-*martingale* si, pour tout $n \geq 0$,

1. X_n et $\mathcal{F}_n$-mesurable,
2. X_n est intégrable,
3. $\mathbb{E}\left[X_{n+1} \,|\mathcal{F}_n\right] = X_n$ presque sûrement.

Définition 4.10 (*Sur- et sous-martingales*)

Une suite définie comme ci-dessus satisfaisant, pour tout n, les hypothèses 1. et 2. et

4. $\mathbb{E}\left[X_{n+1} \,|\mathcal{F}_n\right] \leq X_n$ (resp. $\mathbb{E}\left[X_{n+1} \,|\mathcal{F}_n\right] \geq X_n$) presque sûrement

est appelée une $\mathbb{F}$-*sur-martingale* (resp. $\mathbb{F}$-*sous-martingale*).

Remarque :

(i) Lorsqu'il n'y a pas d'ambigüité sur la filtration concernée, on peut parler plus simplement de martingale.

(ii) Une suite à la fois sur- et sous- martingale relativement à une même filtration est une martingale (toujours relativement à cette filtration).

(iii) Toute $\mathbb{F}$-martingale est martingale relativement à sa filtration naturelle.

(iv) La relation 3. pour tout n dans la définition de la martingale peut être remplacée par

$$\mathbb{E}\left[X_{n+k} \,|\mathcal{F}_n\right] = X_n \text{ presque sûrement, pour tout } n \geq 0 \text{ et pour tout } k \geq 1.$$

Question : Démontrer les deux derniers points ci-dessus.

Proposition 4.14 (*Propriétés des martingales*)

1. *Si $(X_n)_{n\in\mathbb{N}}$ est une martingale, alors $\mathbb{E}[X_n] = \mathbb{E}[X_0]$ pour tout $n \geq 0$.*
2. *Les constantes sont des martingales par rapport à toute filtration.*
3. *Tout combinaison linéaire de martingales est une martingale (par rapport à une même filtration donnée).*
4. *$(X_n)_{n\in\mathbb{N}}$ est une sous-martingale si et seulement si $(-X_n)_{n\in\mathbb{N}}$ est une sur-martingale.*
5. *Si $(X_n)_{n\in\mathbb{N}}$ est une martingale et φ une application convexe telle que $\varphi(X_n)$ est intégrable pour tout $n \in \mathbb{N}$, alors $(\varphi(X_n))_{n\in\mathbb{N}}$ est une sous-martingale. En particulier, si $(X_n)_{n\in\mathbb{N}}$ est une martingale alors $(|X_n|)_{n\in\mathbb{N}}$ est une sous-martingale et si $(X_n)_{n\in\mathbb{N}}$ est une martingale de carré intégrable alors $(X_n^2)_{n\in\mathbb{N}}$ est une sous-martingale.*
6. *Si $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ sont des sous-martingales alors $(X_n \vee Y_n)_{n\in\mathbb{N}}$ est une sous-martingale.*

Question : Démontrer la proposition précédente, comme exercice d'application sur l'espérance condtionnelle.

Exemple :

1. **Marche aléatoire :** Soit $(\xi_n)_{n\in\mathbb{N}}$ une suite i.i.d. de v.a. d'espérance m. Pour tout $n \in \mathbb{N}$, on pose $S_n = \xi_1 + ... + \xi_n$. $(S_n)_{n\in\mathbb{N}}$ est appelée *marche aléatoire*. C'est un processus adapté à la filtration naturelle de la suite $(\xi_n)_{n\in\mathbb{N}}$ que l'on note $(\mathcal{F}_n)_{n\in\mathbb{N}}$.

   Question :

   (a) Montrer que, pour tout $n \geq 1$, $\mathbb{E}\left[S_{n+1} \left| \mathcal{F}_n \right.\right] = S_n + m$ presque sûrement et en déduire la nature de la suite dans les cas $m = 0$, $m > 0$, $m < 0$.

   (b) Etudier le cas particulier où les ξ_n sont de loi $\mathcal{B}(p)$ en fonction de $p \in [0, 1]$.

2. **Martingale fermée :** Soit X une v.a. intégrable et soit $(\mathcal{F}_n)_{n\in\mathbb{N}}$ une filtration. Pour tout n, on pose $X_n = \mathbb{E}\left[X \left| \mathcal{F}_n \right.\right]$. Montrer alors que $(X_n)_{n\in\mathbb{N}}$ est une martingale.

3. **Changement de probabilité :** Soient $\mathbb{P}$ et $\mathbb{Q}$ deux probabilités sur $(\Omega, \mathcal{F})$ et soir $\mathbb{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ une filtration de $\mathcal{F}$. Pour tout $n \in \mathbb{N}$, on définit $\mathbb{P}_n$ et $\mathbb{Q}_n$ les restrictions de $\mathbb{P}$ et $\mathbb{Q}$ à $\mathcal{F}_n$. On suppose que, pour tout n, $\mathbb{Q}_n << \mathbb{P}_n$ et on note X_n la densité de $\mathbb{Q}_n$ par rapport à $\mathbb{P}_n$. Alors $(X_n)_{n\in\mathbb{N}}$ est une $\mathbb{F}$-martingale.

   Démonstration : La suite $(X_n)_{n\in\mathbb{N}}$ est $\mathbb{F}$-adaptée et intégrable par construction. De plus, pour tout $n \in \mathbb{N}$ et pour tout $A \in \mathcal{F}_n$, comme $\mathcal{F}_n \subset \mathcal{F}_{n+1}$, on a $\mathbb{Q}_n(A) = \mathbb{Q}_{n+1}(A)$, soit

   $$\int_A X_n d\mathbb{P} = \int_A X_{n+1} d\mathbb{P}$$

   i.e. $\mathbb{E}\left[X_n + 1 \left| \mathcal{F}_n \right.\right] = X_n$, ce qu'il fallait démontrer.

4.3.2 Théorème d'arrêt

La propriété de martingale peut s'étendre aux temps d'arrêt bornés.

Proposition 4.15

Soit $(M_n)_{n\in\mathbb{N}}$ une $\mathbb{F}$-martingale. Soient T et S deux $\mathbb{F}$-temps d'arrêt majorés par une constante c presque sûrement et tels que $S \leq T$. Alors

$$\mathbb{E}\left[M_T \left| \mathcal{F}_S \right.\right] = M_S \text{ presque sûrement.}$$

Démonstration : En premier lieu, M_S est bien $\mathcal{F}_S$-mesurable. Soit $B \in \mathcal{F}_S$. D'une part

$$\int_B M_S d\mathbb{P} = \sum_{j=0}^{c} \int_{B\cap[S=j]} M_j d\mathbb{P} = \sum_{j=0}^{c} \int_{B\cap[S=j]} M_c d\mathbb{P} = \int_B M_c d\mathbb{P}$$

car $B \cap [S = j] = (B \cap [S \leq j]) \cap [S > j - 1] \in \mathcal{F}_j$ pour $j \leq c$ et $(M_n)_{n\in\mathbb{N}}$ est une martingale. D'autre part, comme $\mathcal{F}_S \subset \mathcal{F}_T$, $B \in \mathcal{F}_T$ et on a également

$$\int_B M_T d\mathbb{P} = \int_B M_c d\mathbb{P}.$$

On en déduit

$$\int_B M_S d\mathbb{P} = \int_B M_T d\mathbb{P},$$

ce qu'il fallait démontrer.

On déduit de cette proposition le théorème d'arrêt suivant.

Théorème 4.2 (*Théorème d'arrêt*)

Soient $(M_n)_{n\in\mathbb{N}}$ une $\mathbb{F}$-martingale et T un $\mathbb{F}$-temps d'arrêt. Alors $(M_{T\wedge n})_{n\in\mathbb{N}}$ est une $\mathbb{F}$-martingale.

Démonstration : On a vu que $(M_{T\wedge n})_{n\in\mathbb{N}}$ est une suite $\mathbb{F}$-adaptée. En outre,

$$M_{T\wedge n} = \sum_{k=0}^{n} M_k \mathbf{1}_{[T=k]}$$

avec

$$|M_k \mathbf{1}_{[T=k]}| \leq |M_k|$$

pour tout $k \leq n$. Donc $M_{T\wedge n}$ est intégrable. Soit alors $n \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{E}\left[M_{T\wedge(n+1)} \,|\mathcal{F}_n\right] &= \mathbb{E}\left[M_{T\wedge n}\mathbf{1}_{[T\leq n]} + M_{n+1}\mathbf{1}_{[T>n]} \,|\mathcal{F}_n\right] \\
&= \mathbb{E}\left[M_{T\wedge n}\mathbf{1}_{[T\leq n]} \,|\mathcal{F}_n\right] + \mathbb{E}\left[M_{n+1}\mathbf{1}_{[T>n]} \,|\mathcal{F}_n\right].
\end{aligned}$$

La première variable conditionnée est $\mathcal{F}_n$-mesurable donc

$$\mathbb{E}\left[M_{T\wedge n}\mathbf{1}_{[T\leq n]} \,|\mathcal{F}_n\right] = M_{T\wedge n}\mathbf{1}_{[T\leq n]} \text{ presque sûrement,}$$

et comme $[T > n] \in \mathcal{F}_n$,

$$\mathbb{E}\left[M_{n+1}\mathbf{1}_{[T>n]} \,|\mathcal{F}_n\right] = \mathbf{1}_{[T>n]}\mathbb{E}\left[M_{n+1} \,|\mathcal{F}_n\right] = M_n\mathbf{1}_{[T>n]} \text{ p.s.}$$

puisque $(M_n)_{n\in\mathbb{N}}$ est une martingale. D'où, comme sur $[T > n]$, $M_n = M_{T\wedge n}$,

$$\mathbb{E}\left[M_{T\wedge(n+1)} \,|\mathcal{F}_n\right] = M_{T\wedge n}\mathbf{1}_{[T\leq n]} + M_{T\wedge n}\mathbf{1}_{[T>n]} = M_{T\wedge n} \text{ p.s.}$$

ce qu'il fallait démontrer.

4.3.3 Convergence des martingales

Exemple : [Urne de Polya] Ce processus est un exemple de processus de branchement.

Une urne contient r boules rouges et b boules blanches. Soit c un entier positif fixé. A tout temps n, on extrait une boule au hasard dans l'urne et on l'y replace avec c boules de la même couleur. On note X_n la proportion de boules blanches dans l'urne après n tirages.

- ☐ On modélise la suite $(X_n)_{n\in\mathbb{N}}$.
  1. Donner X_0.
  2. Exprimer X_{n+1} en fonction de X_n.
  3. Montrer que $(X_n)_{n\in\mathbb{N}}$ est une martingale (par rapport à sa filtration naturelle).
- ☐ On cherche maintenant le comportement asymptotique de la suite $(X_n)_{n\in\mathbb{N}}$. L'enjeu est ici de voir s'il y a une proportion limite. En fait, on peut montrer (on ne le fera pas ici) que $(X_n)_{n\in\mathbb{N}}$ converge presque sûrement vers une v.a. de loi bêta de paramètres $(b/c, r/c)$. En outre, si $b = r = c$ alors la limite presque sûre suit une loi uniforme sur $[0, 1]$.

L'étude de la convergence des martingales en temps discret est un chapitre important de la théorie des probabilités. Nous donnons ici quelques résultats sans preuve qui donnent néanmoins un aperçu des problématiques et quelques outils pratiques.

Théorème 4.3

Toute martingale bornée dans L^1 converge presque sûrement vers une v.a. intégrable.

Remarque :

. Ce théorème s'applique par exemple à l'urne de Polya car la suite $(X_n)_{n\in\mathbb{N}}$ est bornée dans L^∞ ($0 \leq X_n \leq 1$ presque sûrement pour tout n).

. Ce résultat est bien sûr faux pour une suite quelconque de L^1.

On peut remplacer la condition de borne dans L^1 par la condition d'équi-intagrabilité suivante.

Définition 4.11 (*Suite équi-intégrable*)

Une suite $(X_n)_{n\in\mathbb{N}}$ est dite équi-intégrable si

$$\lim_{c\to+\infty}\sup_{n\in\mathbb{N}}\int_{[|X_n|\geq c]}|X_n|\,d\mathbb{P}=0.$$

Proposition 4.16 (*Critères d'équi-intégrabilité*)

(i) Les suites bornées dans L^p, pour $p>1$ sont équi-intégrables.

(ii) Les suites dominées par une fonction L^1 sont équi-intégrables.

Avec cette condition, on va pouvoir préciser le lien entre la suite et sa limite.

Théorème 4.4

Soit $(M_n)_{n\in\mathbb{N}}$ une martingale.

1. *Si $(M_n)_{n\in\mathbb{N}}$ est équi-intégrable alors $(M_n)_{n\in\mathbb{N}}$ converge presque sûrement et dans L^1 vers une v.a. intégrable M et $M_n=\mathbb{E}[M\,|\mathcal{F}_n]$ pour tout n.*
2. *Si $M_n=\mathbb{E}[Y\,|\mathcal{F}_n]$ pour tout n où $Y\in L^1(\Omega,\mathcal{F},\mathbb{P})$, alors $(M_n)_{n\in\mathbb{N}}$ converge presque sûrement et dans L^1 vers $\mathbb{E}[Y\,|\mathcal{G}]$ où*

$$\mathcal{G}=\sigma\left(\bigcup_{n\in\mathbb{N}}\mathcal{F}_n\right).$$

Enfin, ce dernier théorème aborde la problématique de la convergence L^2 (les convergence L^1 et presque sûres découlaient déjà du théorème précédent et des critères d'équi-intégrabilité).

Théorème 4.5

Si $(M_n)_{n\in\mathbb{N}}$ est une martingale bornée dans L^2 alors $(M_n)_{n\in\mathbb{N}}$ converge dans L^2 vers une v.a. M telle que $M_n=\mathbb{E}[M\,|\mathcal{F}_n]$ presque sûrement pour tout n.

4.4 Exercices

4.4.1 Conditionnement

Exercice 4.1.

Soient X et Y deux v.a. réelles définies sur un même espace probabilisé.

1. On suppose X et Y à valeurs dans $\mathbb{N}$ et X intégrable, calculer $\mathbb{E}[X\,|Y]$.
2. On suppose X et Y indépendantes de loi de Poisson de paramètre respectif λ et μ, calculer $\mathbb{E}[X\,|Y+X]$.
3. On suppose X et Y indépendantes de loi uniforme sur $[0,1]$. On pose $U=\max(X,Y)$ et $V=\min(X,Y)$, calculer $\mathbb{E}[V\,|U]$.

Exercice 4.2. [*Espérance conditionnelle*]

Soient X et Y deux v.a indépendantes, intégrables et de même loi P.

1. Montrer que $\mathbb{E}[X\,|X+Y] = \mathbb{E}[Y\,|X+Y]$ ps.
2. En déduire $\mathbb{E}[X\,|X+Y]$.

Exercice 4.3. [*Vecteurs gaussiens*]

Soit (X,Y) un couple gaussien centré de matrice de covariance $\begin{pmatrix} 4/3 & 1 \\ 1 & 1 \end{pmatrix}$.

Calculer $\mathbb{E}[X\,|Y-X]$ et donner sa loi.

Exercice 4.4. [*Fonction caractéristique et conditionnement*] On suppose que X et Y sont deux v.a telles que $Y \rightsquigarrow \mathcal{N}(m,\alpha^2)$ et $\mathbb{E}\left[e^{itX}|Y\right] = e^{-\sigma^2t^2/2}e^{itY}$.

1. Déterminer la fonction caractéristique de X et en déduire sa loi.
2. Même question pour (X,Y)
3. Montrer que $X-Y$ et Y sont indépendantes
4. Déterminer l'espérance conditionnelle de Y sachant X.

Exercice 4.5. [*Variance conditionnelle*]

Soit X une variable aléatoire réelle de carré intégrable définie sur un espace probabilisé $(\Omega,\mathcal{F},\mathbb{P})$ et soit $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. On pose

$$\mathbf{Var}[X|\mathcal{G}] = \mathbb{E}\left[(X-\mathbb{E}[X\,|\mathcal{G}])^2\,|\mathcal{G}\right].$$

Montrer que

$$\mathbf{Var}[X] = \mathbb{E}[\mathbf{Var}[X|\mathcal{G}]] + \mathbf{Var}[\mathbb{E}[X\,|\mathcal{G}]].$$

Exercice 4.6. [*Somme aléatoire de variables aléatoires*]

Soit $(X_i)_{i\geq 1}$ une suite de v.a. d'espérance μ. Soit N une v.a. à valeurs dans $\mathbb{N}^*$, indépendante des v.a. X_i, d'espérance m. On pose

$$S_N = \sum_{i=1}^{N} X_i.$$

Calculer $\mathbb{E}[S_N\,|N]$ puis en déduire $\mathbb{E}[S_N]$.

4.4.2 Temps d'arrêt

Exercice 4.7. Montrer la proposition 4.10

Exercice 4.8. Soient T et S deux temps d'arrêt. On suppose $T \leq S$. Montrer que $\mathcal{F}_T \subset \mathcal{F}_S$.

Exercice 4.9. Soient T et S deux temps d'arrêt.

1. Montrer que $\mathcal{F}_{T\wedge S} = \mathcal{F}_T \cap \mathcal{F}_S$.
2. Montrer que $\mathcal{F}_{T\vee S} \supset \mathcal{F}_T \bigvee \mathcal{F}_S$ $(= \sigma(\mathcal{F}_T \cup \mathcal{F}_S)$ par définition).

4.4.3 Martingales

Exercice 4.10. [*La martingale du casino*]

Le casino propose un jeu de Pile ou Face avec une pièce équilibrée. A chaque tour le joueur doit, pour pouvoir jouer, faire une mise. S'il perd, il perd la totalité de sa mise et s'il gagne il remporte deux fois sa mise (donc sa mise plus un gain égal à sa mise).

Un joueur décide d'adopter la stratégie suivante : il mise un euro au premier tour. Tant qu'il perd, il double sa mise et dès qu'il gagne il s'arrête de jouer.

1. Proposer une modélisation du jeu à l'aide d'une suite de variables aléatoires $(X_n)_{n\geq 1}$. On précisera les propriétés de la suite et la loi des va.
2. Rappeler pourquoi, presque sûrement, le joueur va gagner.
3. Modéliser le gain du joueur par une martingale arrêtée. On précisera la martingale $(Z_n)_{n\geq 0}$ représentant le gain si le jeu ne s'arrête jamais et T l'instant où le jeu s'arrête.
4. Quel est le gain du joueur quand finit le jeu ? Quel est le gain moyen à tout temps n ?
5. Soit M la mise totale du joueur au cours du jeu. Exprimer M puis calculer son espérance.

Exercice 4.11. [*Autour du théorème d'arrêt*]

Soient $(X_n)_n$ une $\mathbb{F}$-martingale et T un $\mathbb{F}$-temps d'arrêt.

1. On suppose pour cette question que $T \leq c$. Montrer, en utilisant le théorème d'arrêt, que $\mathbb{E}[X_T] = \mathbb{E}[X_S]$ pour $T, S \leq c$ où S est un temps d'arrêt. En déduire que $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.
2. On suppose pour cette question que T est fini p.s., que X_T est intégrable et que $\mathbb{E}\left[|X_n|\mathbf{1}_{[T>n]}\right]$ tend vers 0 quand n tend vers l'infini. Montrer que $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.
   (*Ind : Remarquer que* $X_{T\wedge n} = X_{T\wedge n}\mathbf{1}_{[T\leq n]} + X_{T\wedge n}\mathbf{1}_{[T>n]}$).
3. On suppose que $(X_n)_n$ converge dans L^1 vers une variable X. On définit
   $$X_T = \begin{cases} X_n & \text{sur} \quad [T=n] \\ X & \text{sur} \quad [T=+\infty] \end{cases}$$
   et l'on suppose que X_T est intégrable. Montrer que $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.
4. Soit $(Y_n)_n$ une suite de variables aléatoires intégrables adaptée à $\mathbb{F}$. On suppose que, pour tout temps d'arrêt T borné, $\mathbb{E}[Y_T] = \mathbb{E}[Y_0]$.
   (a) Soient n fixé et $A \in \mathcal{F}_n$. On définit $T = n\mathbf{1}_{A^c} + (n+1)\mathbf{1}_A$. Montrer que T est un $\mathbb{F}$-temps d'arrêt et en déduire que $\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = Y_n$ presque sûrement.
   (b) Montrer que $(Y_n)_n$ est une martingale.

Exercice 4.12. [*Martingales et convergence p.s.*]

Soit $(X_n)_n$ une suite aléatoire i.i.d de loi exponentielle de paramètre 1. Soit $(\varepsilon_n)_n$ une suite aléatoire i.i.d telle $\mathbb{P}[\varepsilon_1 = 1] = \mathbb{P}[\varepsilon_1 = -1] = 1/2$. On suppose que les deux suites sont indépendantes, et l'on note $\mathcal{F}_n$ la tribu engendrée par $(\varepsilon_1, \ldots, \varepsilon_n, X_1, \ldots, X_n)$ pour $n \geq 1$, avec $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Soit $p \in]1, 2[$. On pose

$$M_n = \sum_{i=1}^{n} \varepsilon_i (X_1 + \cdots + X_i)^{-1/p}, \; n > 0.$$

1. Montrer que $(M_n)_n$ est une $\mathbb{F}$-martingale.
2. Montrer que
$$\mathbb{E}\left[(M_n - M_1)^2\right] \leq \int_{\mathbb{R}_+} x^{-2/p}(1 - e^{-x})dx, \; n \in \mathbb{N}.$$
En déduire que $(M_n)_n$ converge p.s..

Exercice 4.13.[*Suite exercice 2.5.3*]

1. On suppose que $|\theta| > 1$. Montrez que $\theta^{-n} X_n$ converge p.s vers une variable X. Donnez une expression explicite pour X et précisez sa loi.
2. On suppose $\theta = 1$. Calculez $Z_n = E(X_n | (X_{n-1}, X_{n+1}))$. Montrez que $Z_n/\sqrt{n}$ converge en loi et trouvez la loi limite.