



Cours de Master 1ère année

Année 2006-2007

PROBABILITÉS
et
MODÉLISATIONS STOCHASTIQUES

Table des matières

1	Variables aléatoires gaussiennes	5
1.1	Variabes aléatoires gaussiennes à valeurs réelles	5
1.2	Espérance et covariance de variables aléatoires à valeurs vectorielles	8
	Définitions et notations	8
	Propriétés	9
1.3	Vecteurs aléatoires gaussiens	10
1.4	Convergence vers la loi normale	14
1.5	Loi du chi-deux	15
1.6	Une application à la statistique	19
1.7	Exercices	23
2	Conditionnement	35
2.1	Espérance conditionnelle	35
2.2	Loi conditionnelle	43
2.3	Le cas gaussien	49
2.4	Introduction aux martingales	52
2.5	Exercices	58
3	Chaînes de Markov	73
3.1	Définitions et exemples	73
3.2	Chaîne canonique. Temps d'arrêt	77
3.3	Potentiel. États récurrents, états transients	82
3.4	Mesure invariante et convergence	92
3.5	Exercices	104
4	Simulation	115
4.1	Introduction	115
4.2	Procédés de simulations de variables aléatoires à valeurs réelles	117
	Simulations de variables aléatoires courantes	117
	Procédés généraux de simulation	120

Chapitre 1

Variables aléatoires gaussiennes

Les variables aléatoires gaussiennes apparaissent naturellement comme limite de sommes re-normalisées, de v.a. indépendantes ; pour plus de précisions on peut se reporter à l'énoncé du théorème central limite dans le §4. Ainsi la somme cumulée de petites fluctuations au niveau microscopique donne naissance à une fluctuation macroscopique gaussienne. En plus de cette propriété de "normalité", les v.a. gaussiennes à valeurs multidimensionnelles sont très utilisées dans la modélisation de phénomènes physiques, car elles se prêtent extrêmement bien au calcul.

1.1 Définitions et propriétés des variables aléatoires gaussiennes à valeurs réelles

Définition 1.1 Une variable aléatoire à valeurs réelles X est dite gaussienne réduite et centrée si sa loi de probabilité admet la densité :

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad x \in \mathbb{R}.$$

On note $\mathcal{N}_1(0, 1)$ cette loi. Rappelons que

$$\mathbb{E}[f(X)] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) \exp\left(-\frac{x^2}{2}\right) dx,$$

pour toute fonction borélienne bornée ou positive. En particulier,

$$\int_{\mathbb{R}} \exp\left(-\frac{x^2}{2}\right) dx = \sqrt{2\pi}.$$

Remarque : On introduit la fonction d'erreur erf définie par $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt$. Si X suit une loi $\mathcal{N}_1(0, 1)$ alors $\mathbb{P}(|X| \leq \sqrt{2}x) = \operatorname{erf}(x)$.

Proposition 1.1 Soit X une v.a.r. gaussienne réduite et centrée.

1) Pour tout $z \in \mathbb{C}$, $\mathbb{E}[|e^{zX}|] < \infty$ et

$$\mathbb{E}[\exp zX] = \exp z^2/2. \quad (1.1)$$

En particulier

$$\mathbb{E}[\exp(itX)] = e^{-t^2/2}, \quad \forall t \in \mathbb{R}. \quad (1.2)$$

2) Pour tout $n \in \mathbb{N}$, on a

$$\mathbb{E}(X^n) = \begin{cases} 0 & \text{si } n \text{ est impair,} \\ \frac{(2m)!}{m!2^m}, & \text{si } n \text{ est pair, } n = 2m. \end{cases} \quad (1.3)$$

Démonstration : On vérifie que l'intégrale $\int_{\mathbb{R}} \exp(zx - \frac{1}{2}x^2)dx$ est absolument convergente pour tout z complexe. Par conséquent la quantité $\varphi(z) = \mathbb{E}(e^{zX})$ est bien définie et,

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(zx - \frac{1}{2}x^2)dx.$$

Supposons z réel. On écrit $zx - \frac{1}{2}x^2 = -\frac{1}{2}(x - z)^2 + \frac{z^2}{2}$ et l'on fait le changement de variable $y = x - z$ dans l'intégrale, il vient $\varphi(z) = e^{z^2/2}$. Remarquons que φ et $z \rightarrow e^{z^2/2}$ sont deux fonctions entières ; puisque ces deux fonctions coïncident sur \mathbb{R} , elles sont égales sur \mathbb{C} . En particulier, si $z = it$ avec $t \in \mathbb{R}$, on a : $\mathbb{E}(\exp(itX)) = e^{-t^2/2}$, $t \in \mathbb{R}$.

Soit $n \geq 1$. Sachant que $\lim_{|x| \rightarrow \infty} \frac{|x|^n}{\text{ch}(x)} = 0$, il existe une constante c_n telle que

$$|x|^n \leq c_n(e^x + e^{-x}), \quad \forall x \in \mathbb{R}. \quad (1.4)$$

Puisque $\mathbb{E}(\exp(aX))$ existe pour tout réel a , on déduit de (1.4) que $\mathbb{E}(|X|^n) < \infty$. Par conséquent, $\mathbb{E}(X^n)$ existe pour tout $n \geq 1$.

Par ailleurs, en utilisant le développement en série entière de $x \rightarrow e^{itx}$, on peut affirmer que, presque sûrement,

$$e^{itX} = \lim_{n \rightarrow \infty} S_n, \quad S_n = \sum_{k=0}^n \frac{(it)^k}{k!} X^k.$$

Mais $|S_n| \leq Y$ avec

$$Y = \sum_{k=0}^{\infty} \frac{|t|^k |X|^k}{k!} = e^{|tX|} \leq e^{tX} + e^{-tX}.$$

En utilisant à nouveau le fait que $\mathbb{E}(\exp(aX)) < \infty$, on en déduit que Y est intégrable. Une application du théorème de Lebesgue, conduit à,

$$\mathbb{E}(\exp(itX)) = \mathbb{E} \left(\sum_{n \geq 0} \frac{(itX)^n}{n!} \right) = \sum_{n \geq 0} \frac{i^n t^n}{n!} \mathbb{E}(X^n).$$

Par identification on en déduit (1.3). □

Notons en particulier $\mathbb{E}(X) = 0$, $\text{Var}(X) = 1$. Ce qui justifie le terme “réduit” et “centrée”.

Définition 1.2 Une variable aléatoire réelle est dite gaussienne s'il existe une v.a. X gaussienne réduite et centrée, et deux réels a et b tels que $Y = aX + b$.

On peut identifier a et b à l'aide de l'espérance et la variance de Y , plus précisément,

$$\mathbb{E}(Y) = b, \quad \text{Var}Y = a^2 \times \text{Var}X = a^2.$$

Posons $\sigma = |a|$ et $m = b$, et supposons $a \geq 0$, alors

$$Y = \sigma X + m, \quad X \text{ de loi gaussienne réduite et centrée.} \quad (1.5)$$

Si $a < 0$, on écrit $Y = (-a)(-X) + b$, on peut se ramener au cas précédent en observant que $-X$ suit une loi gaussienne réduite et centrée.

La relation précédente permet d'exprimer X à l'aide de Y . En effet, soient Y une v.a. gaussienne, $m = \mathbb{E}(Y)$ et $\sigma = \sqrt{\text{Var}Y}$. On suppose $\sigma > 0$. Alors la v.a. $Y_* = \frac{Y-m}{\sigma}$ est une v.a. de loi gaussienne réduite et centrée et

$$Y = \sigma Y_* + m. \quad (1.6)$$

On note $\mathcal{N}_1(m, \sigma^2)$ la loi d'une v.a. de loi gaussienne de moyenne m et de variance σ^2 . Un calcul aisé montre :

$$\frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{(x-m)^2}{2\sigma^2} \text{ est la densité de } \mathcal{N}_1(m, \sigma^2). \quad (1.7)$$

Soit Y de loi $\mathcal{N}_1(m, \sigma^2)$, on déduit de (1.6)

$$\mathbb{E}[\exp(itY)] = e^{itm} \mathbb{E}[\exp i(t\sigma)Y_*].$$

Une application directe de (1.1) et (1.2) conduit à,

$$\mathbb{E}[\exp(itY)] = \exp\left(itm - \frac{t^2}{2}\sigma^2\right), \quad t \in \mathbb{R}. \quad (1.8)$$

En utilisant l'injectivité de la transformée de Fourier, on montre que si X est une v.a. telle que la fonction caractéristique (c'est-à-dire la fonction $t \rightarrow \mathbb{E}(\exp(itX))$) est égale à $t \rightarrow \exp(ita - t^2b^2)$ où $a \in \mathbb{R}$ et $b \in \mathbb{R}$, X est une v.a. gaussienne.

Proposition 1.2 Soient Y_1 et Y_2 deux v.a. gaussiennes, indépendantes, Y_1 de loi $\mathcal{N}_1(m_1, \sigma_1^2)$, Y_2 de loi $\mathcal{N}_1(m_2, \sigma_2^2)$. Alors $Y_1 + Y_2$ est une v.a. gaussienne de loi $\mathcal{N}_1(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

Démonstration : Les v.a. Y_1 et Y_2 étant indépendantes, on a pour tout t réel :

$$\varphi(t) = \mathbb{E}[\exp(it(Y_1 + Y_2))] = \mathbb{E}[\exp(itY_1)] \mathbb{E}[\exp(itY_2)].$$

De plus Y_1 et Y_2 sont gaussiennes, d'après (1.8), on a :

$$\begin{aligned} \varphi(t) &= \exp\left(itm_1 - \frac{t^2\sigma_1^2}{2}\right) \exp\left(itm_2 - \frac{t^2\sigma_2^2}{2}\right) \\ &= \exp\left(it(m_1 + m_2) - \frac{t^2(\sigma_1^2 + \sigma_2^2)}{2}\right). \end{aligned}$$

□

Remarques

- 1) Le résultat peut être faux si l'on ne suppose plus Y_1 et Y_2 indépendantes.
 2) Plaçons-nous sous les hypothèses de la proposition 1.2. Si l'on sait que $Y_1 + Y_2$ est gaussienne, de loi $\mathcal{N}_1(m, \sigma^2)$, il est aisé d'identifier les deux paramètres :

$$m = \mathbb{E}(Y_1 + Y_2) = \mathbb{E}(Y_1) + \mathbb{E}(Y_2) = m_1 + m_2,$$

$$\sigma^2 = \text{Var}(Y_1 + Y_2) = \text{Var}Y_1 + \text{Var}Y_2 = \sigma_1^2 + \sigma_2^2.$$

La deuxième égalité a lieu car Y_1 et Y_2 sont deux v.a. indépendantes.

- 3) Le résultat se généralise sans difficulté au cas de n v.a. : soient Y_1, Y_2, \dots, Y_n , n v.a.r., indépendantes, Y_i de loi $\mathcal{N}_1(m_i, \sigma_i^2)$, $1 \leq i \leq n$, alors

$$Y_1 + Y_2 + \dots + Y_n \text{ suit une loi } \mathcal{N}_1 \left(\sum_{i=1}^n m_i, \sum_{i=1}^n \sigma_i^2 \right). \quad (1.9)$$

1.2 Espérance et covariance de variables aléatoires à valeurs vectorielles

Avant de considérer les vecteurs gaussiens, il est bon de rappeler les définitions et propriétés des v.a. à valeurs vectorielles.

Définitions et notations

- 1) Si A est une matrice d'ordre, A^* désigne la matrice transposée. En particulier si $x \in \mathbb{R}^n$ est considéré comme un vecteur unicolonne, x^* est une matrice uniligne. Si x et y sont deux vecteurs de \mathbb{R}^n , leur produit scalaire est noté :

$$\langle x, y \rangle = x^* y = y^* x = \sum_{i=1}^n x_i y_i, \quad x^* = (x_1, \dots, x_n), \quad y^* = (y_1, \dots, y_n).$$

- 2) Une v.a. X à valeurs dans \mathbb{R}^n est la donnée de n v.a. à valeurs réelles X_1, X_2, \dots, X_n . On note X la matrice **unicolonne** de coordonnées X_1, X_2, \dots, X_n :

$$X^* = (X_1, X_2, \dots, X_n) \text{ ou encore } X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_n \end{pmatrix}. \quad (1.10)$$

1.2. ESPÉRANCE ET COVARIANCE DE VARIABLES ALÉATOIRES À VALEURS VECTORIELLES

Pour d'évidentes raisons typographiques, on choisira la première écriture.

3) Soit X une v.a. de coordonnées X_1, X_2, \dots, X_n , on note $\mathbf{E}(X)$ le vecteur unicolonne de coordonnées $\mathbf{E}(X_1), \mathbf{E}(X_2), \dots, \mathbf{E}(X_n)$:

$$\mathbf{E}(X)^* = (\mathbf{E}(X_1), \mathbf{E}(X_2), \dots, \mathbf{E}(X_n)). \quad (1.11)$$

On suppose bien sûr que chaque v.a. X_i admet une espérance. **4)** Soit X (resp. Y) une v.a. à valeurs dans \mathbb{R}^n (resp. \mathbb{R}^m), $K_{X,Y}$ est la matrice d'ordre $n \times m$ (n lignes et m colonnes) définie par

$$K_{X,Y} = \mathbf{E}[(X - \mathbf{E}(X))(Y - \mathbf{E}(Y))^*]. \quad (1.12)$$

Remarquons que $X - \mathbf{E}(X)$ est une matrice $n \times 1$ et $(Y - \mathbf{E}(Y))^*$ une matrice $1 \times m$, le produit $(X - \mathbf{E}(X))(Y - \mathbf{E}(Y))^*$ est une matrice $n \times m$. Soit $K_{X,Y}(i, j)$ l'élément de $K_{X,Y}$ situé à la i ème ligne et j ème colonne, alors :

$$K_{X,Y}(i, j) = \text{Cov}(X_i, Y_j); \quad 1 \leq i \leq n, \quad 1 \leq j \leq m, \quad (1.13)$$

où $\text{Cov}(U, V) = \mathbf{E}(UV) - \mathbf{E}(U)\mathbf{E}(V)$. Rappelons que si U et V sont deux v.a. à valeurs réelles,

$$\text{Cov}(U, V) = \mathbf{E}[(U - \mathbf{E}(U))(V - \mathbf{E}(V))] = \mathbf{E}[UV] - \mathbf{E}[U]\mathbf{E}[V].$$

Si l'on choisit $Y = X$, la matrice $K_{X,X}$ est appelée matrice de covariance de X , on note pour simplifier $K_X = K_{X,X}$. Cette matrice est carrée d'ordre n , et :

$$K_X = \mathbf{E}[(X - \mathbf{E}(X))(X - \mathbf{E}(X))^*]. \quad (1.14)$$

$$K_X(i, j) = \text{Cov}(X_i, X_j); \quad 1 \leq i \leq n, \quad 1 \leq j \leq n. \quad (1.15)$$

Nous allons à présent donner quelques propriétés utiles en pratique. La démonstration en est laissée au lecteur.

Propriétés

1) Soient X une v.a. à valeurs \mathbb{R}^n , A une matrice d'ordre $m \times n$, u un vecteur de \mathbb{R}^m . Alors

$$\mathbf{E}[u + AX] = u + A\mathbf{E}(X). \quad (1.16)$$

2) Comme dans le cas unidimensionnel, on a deux formules pour calculer la covariance :

$$K_{X,Y} = \mathbf{E}(XY^*) - \mathbf{E}(X)\mathbf{E}(Y)^*. \quad (1.17)$$

En particulier,

$$K_X = \mathbf{E}(XX^*) - \mathbf{E}(X)\mathbf{E}(X)^*. \quad (1.18)$$

3) Soit A une matrice d'ordre $m \times n$, X une v.a. à valeurs dans \mathbb{R}^n et $u \in \mathbb{R}^m$, alors

$$K_{u+AX} = K_{AX} = AK_X A^*. \quad (1.19)$$

1.3 Vecteurs aléatoires gaussiens

Nous conservons bien sûr les notations de la section précédente, les v.a. vectorielles seront représentées par des matrices unicolonnes.

Définition 1.3 Une v.a. X à valeurs dans \mathbb{R}^n , est dite *gaussienne* (on dit aussi que X est un *vecteur aléatoire gaussien*) si et seulement si pour tout $\lambda \in \mathbb{R}^n$, $\lambda^* = (\lambda_1, \dots, \lambda_n)$,

$$\langle \lambda, X \rangle = \lambda^* X = \sum_{i=1}^n \lambda_i X_i \text{ est une v.a. gaussienne réelle.} \quad (1.20)$$

Remarques :

- 1) Il est clair que si X est un vecteur gaussien, alors chaque composante X_i de X est une v.a.r. gaussienne.
- 2) L'exemple clé de vecteur gaussien est celui où X_1, \dots, X_n sont n v.a. gaussiennes, indépendantes. Il suffit en effet d'appliquer directement (1.9). Attention l'hypothèse d'indépendance est essentielle. Il est facile de construire un exemple où X_1 et X_2 sont deux v.a.r. gaussiennes telles que (X_1, X_2) ne soit pas un vecteur gaussien.

Proposition 1.3 Soit K la matrice de covariance d'un vecteur gaussien X . Alors, pour tout $u \in \mathbb{R}^n$,

$$\mathbb{E}(\exp(i \langle u, X \rangle)) = e^{i \langle u, \mathbb{E}(X) \rangle - \frac{1}{2} u^* K u}, \quad (1.21)$$

où u est représenté sous la forme d'une matrice unicolonne.

Démonstration : La v.a. $Y = \langle u, X \rangle = u^* X$ est une v.a. gaussienne et $\mathbb{E}(Y) = \langle u, \mathbb{E}(X) \rangle$. De plus d'après (1.19) : $\text{Var} Y = u^* K u$. Par conséquent, une application directe de (1.8) conduit à : $\mathbb{E}[\exp(i \langle u, X \rangle)] = e^{i \mathbb{E}(Y) - \frac{1}{2} \text{Var} Y}$. \square

Remarques :

- 1) Si $n = 1$, on retrouve exactement la formulation de (1.8). On peut remarquer d'ailleurs que pour établir la proposition 1.3, on se ramène au cas unidimensionnel.
- 2) La proposition 1.3 signifie que la loi d'un vecteur gaussien X est caractérisée par sa moyenne m et sa matrice de covariance K : si X et Y sont deux vecteurs gaussiens, ayant même moyenne et même matrice de covariance, ils ont même loi. Attention cette propriété concerne les vecteurs gaussiens, elle n'est pas vraie en général, il n'y a aucune raison pour que deux v.a. à valeurs réelles qui ont même moyenne et même variance aient même loi.

Proposition 1.4 Soient X un vecteur gaussien à valeurs dans \mathbb{R}^n , m sa moyenne et K_X sa matrice de covariance, A une matrice $p \times n$ (p lignes et n colonnes) et z un vecteur de \mathbb{R}^p . On pose $Y = AX + z$. Alors Y est un vecteur gaussien et

$$\mathbb{E}(Y) = z + Am, \quad K_Y = AK_X A^*, \quad (1.22)$$

K_Y désignant la matrice de covariance de Y .

Démonstration : Soit $u \in \mathbb{R}^p$ représenté sous forme unicolonne. Alors

$$u^*Y = u^*z + u^*AX = u^*z + v^*X,$$

où l'on a posé $v = A^*u$.

X étant un vecteur gaussien, alors v^*X est une v.a.r. gaussienne. Y est bien un vecteur gaussien. De plus, $\mathbb{E}(Y) = z + A\mathbb{E}(X) = z + Am$; $K_Y = AK_XA^*$. \square

Proposition 1.5 *La matrice de covariance d'un vecteur aléatoire est une matrice symétrique et positive.*

Démonstration : Soit K la matrice de covariance d'un vecteur aléatoire X à valeurs dans \mathbb{R}^n . Puisque $K(i, j) = \text{Cov}(X_i, X_j)$, il est clair que K est symétrique : $K(i, j) = K(j, i)$ pour tout i et j de $\{1, 2, \dots, n\}$. Montrons que K est une matrice positive : $u^*Ku \geq 0$, pour tout vecteur u de \mathbb{R}^n . En effet, posons $Y = u^*X$, alors $\text{Var}Y = u^*Ku \geq 0$. \square

Théorème 1.1 *Soient m un vecteur de \mathbb{R}^n et Γ une matrice carrée symétrique positive d'ordre n . Alors il existe un vecteur gaussien d'espérance m et de matrice de covariance Γ .*

Afin de prouver le théorème on commence par établir un résultat d'algèbre linéaire :

Lemme 1.1 *Soit Γ une matrice d'ordre n , symétrique et positive. Alors il existe une matrice A , carré d'ordre n telle que $\Gamma = AA^*$.*

Démonstration : (Lemme). Γ étant symétrique, il existe une matrice U orthogonale ($UU^* = U^*U = \text{Id}$) telle que $D_1 = U^*\Gamma U$ soit une matrice diagonale. On note $\lambda_1, \lambda_2, \dots, \lambda_n$ les éléments situés sur la diagonale de D_1 . Γ étant de plus positive, $\lambda_i \geq 0$ pour tout $i \in \{1, 2, \dots, n\}$. Soit D la matrice diagonale, dont les éléments situés sur la diagonale sont $\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_n}$. On pose $A = UD$. Rappelons que $U^{-1} = U^*$, donc $\Gamma = UD_1U^*$. Mais $D_1 = D^2 = DD^*$, d'où $\Gamma = UDD^*U^* = UD(UD)^* = AA^*$. \square

Démonstration : (Théorème). Soient Y_1, Y_2, \dots, Y_n , n v.a. gaussiennes réelles, indépendantes, équidistribuées et de lois $\mathcal{N}_1(0, 1)$. On pose $Y^* = (Y_1, \dots, Y_n)$. Y est un vecteur gaussien. On définit X par :

$$X = m + AY, \tag{1.23}$$

où A est une matrice $n \times n$, telle que $\Gamma = AA^*$. D'après la proposition 1.4, X est un vecteur gaussien

$$\mathbb{E}(X) = m + \mathbb{E}(AY) = m + A\mathbb{E}(Y) = m,$$

$$K_X = AK_YA^* = A\text{Id}A^* = AA^* = \Gamma.$$

\square

Notation.

$\mathcal{N}_n(m, \Gamma)$ désigne la loi d'un vecteur gaussien dans \mathbb{R}^n de moyenne m et de matrice de covariance Γ .

Remarque.

Lorsque $n = 1$, la formule (1.23) est analogue à celle que l'on a vue en dimension 1 : $X = m + \sigma Y$ avec $Y \sim \mathcal{N}_1(0, 1)$ et σ l'écart-type de X .

Il est bon de retenir le procédé pour engendrer un vecteur gaussien de moyenne m et de matrice de covariance Γ : on considère

- A une "racine carrée positive" de Γ , c'est-à-dire une matrice carrée telle que $\Gamma = AA^*$.
- Y_1, Y_2, \dots, Y_n n v.a.r., indépendantes de loi gaussienne réduite et centrée.

Alors

$$X = m + AY \text{ suit la loi } \mathcal{N}_n(m, \Gamma). \quad (1.24)$$

Théorème 1.2 Soit X un vecteur gaussien, $X^* = (X_1, \dots, X_n)$. Alors les v.a. X_1, X_2, \dots, X_n sont indépendantes si et seulement si la matrice de covariance K de X est diagonale.

Démonstration : Il est clair que si X_1, \dots, X_n sont indépendantes alors $K(i, j) = \text{cov}(X_i, X_j) = 0$ si $i \neq j$. Donc K est diagonale.

Étudions la réciproque. Nous avons montré :

$$\mathbb{E}(\exp(i \langle u, X \rangle)) = e^{i \langle u, \mathbb{E}(X) \rangle - \frac{1}{2} u^* K u}, \quad u \in \mathbb{R}^n. \quad (1.25)$$

Mais K étant diagonale, on a :

$$u^* K u = \sum_{l=1}^n u_l^2 K(l, l) = \sum_{l=1}^n u_l^2 \text{Var} X_l. \quad (1.26)$$

Notons ϕ_{X_l} la fonction caractéristique de X_l : $\phi_{X_l}(s) = \mathbb{E}[e^{isX_l}]$, $s \in \mathbb{R}$. Si l'on choisit u tel que $u_i = 0$, pour tout $i \neq l$, (1.25) et (1.26) impliquent

$$\phi_{X_l}(u_l) = \mathbb{E}(\exp(iu_l X_l)) = e^{iu_l \mathbb{E}(X_l) - \frac{1}{2} u_l^2 \text{Var} X_l}.$$

On peut donc ré-écrire (1.25) sous la forme suivante :

$$\phi_{X_1}(u_1) \dots \phi_{X_n}(u_n) = \mathbb{E} \left[\exp \left(i \sum_{l=1}^n u_l X_l \right) \right] = \mathbb{E}[\exp(i \langle u, X \rangle)],$$

pour tout $u = (u_1, u_2, \dots, u_n)$. Ce qui signifie que les v.a. X_1, X_2, \dots, X_n sont indépendantes. \square

Théorème 1.3 Soit X un vecteur gaussien à valeurs dans \mathbb{R}^n , de matrice de covariance K .

- 1) X admet une densité si et seulement si K est inversible.
 2) Si K est inversible, la densité de X est donnée par :

$$f(x) = \frac{1}{(2\pi)^{n/2}} \frac{1}{(\det K)^{1/2}} \exp - \left(\frac{1}{2} (x - m)^* K^{-1} (x - m) \right) \quad (1.27)$$

avec $m = \mathbb{E}(X)$.

Démonstration : 1) a) On commence par établir un résultat préliminaire. Soit H un sous-espace vectoriel de \mathbb{R}^n , de dimension strictement plus petite que n . Si ξ est une v.a. à densité alors $\mathbb{P}(\xi \in H) = 0$. En effet soit H' un hyperplan contenant H . Quitte à changer les coordonnées on peut supposer que $H' = \{(x_1, x_2, \dots, x_n); x_n = 0\}$. Notons φ la densité de ξ , on a :

$$\mathbb{P}(\xi \in H) = \mathbb{P}(\xi \in H') = \int_{\mathbb{R}^n} \varphi(x_1, x_2, \dots, x_n) \mathbf{1}\{x_n = 0\} dx_1 dx_2 \dots dx_n = 0.$$

b) On a vu (cf. formule (1.24)) que X a même loi que $m + AY$, où $AA^* = K$ et Y est un vecteur dont toutes les composantes admettent des densités. Si A n'est pas inversible, A considéré comme application linéaire a une image H strictement incluse dans \mathbb{R}^n . Donc $AY \in H$. Raisonnons par l'absurde : si X admet une densité, $X - m$ aussi; d'après ce qui précède on a $\mathbb{P}(X - m \in H) = 0$ et $\mathbb{P}(X - m \in H) = \mathbb{P}(AY \in H) = 1$. Par conséquent X n'a pas de densité.

Lorsque A est inversible, l'application $y \rightarrow m + Ay$ étant bijective et de classe \mathcal{C}^1 , la v.a. $m + AY$ admet une densité.

Puisque $AA^* = K$, $\det A \det A^* = (\det A)^2 = \det K$, on a l'équivalence :

$$A \text{ inversible} \iff K \text{ est inversible.}$$

2) Soit Y un vecteur gaussien centré de matrice de covariance identité. Il est clair que Y admet pour densité :

$$g(y) = \frac{1}{(2\pi)^{n/2}} \exp - \frac{1}{2} \langle y, y \rangle .$$

Posons $X' = AY + m$ donc $Y = A^{-1}(X' - m)$. Le jacobien de $x \rightarrow A^{-1}(x - m)$ est A^{-1} , son déterminant vaut $\det A^{-1} = \frac{1}{\det A}$. Mais $(\det A)^2 = \det K$ et $K^{-1} = (AA^*)^{-1} = (A^*)^{-1}A^{-1}$. De plus,

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

Donc X' admet pour densité la fonction f définie par (1.27). Puisque X et X' ont même loi, X admet pour densité f . \square

1.4 Convergence vers la loi normale

On rappelle la loi forte des grands nombres :

Théorème 1.4 Soit $(X_n; n \geq 1)$ une suite de v.a. à valeurs dans \mathbb{R}^k , indépendantes, de même loi et telles que $\mathbb{E}[|X_1|] < \infty$, $\mathbb{E}[X_1] = m \in \mathbb{R}^k$. Alors

$$\frac{X_1 + X_2 + \dots + X_n}{n}(\omega) \longrightarrow m, \quad \text{quand } n \rightarrow \infty, \quad \text{pour tout } \omega \in \Omega$$

sauf sur un ensemble de probabilité nulle.

Le théorème de la limite centrale précise un peu plus cette convergence :

Théorème 1.5 Soit $(X_n; n \geq 1)$ une suite de v.a. à valeurs dans \mathbb{R}^k , indépendantes et de même loi. On suppose que ces v.a. admettent un moment d'ordre 2. On note m leur espérance commune et Γ la covariance de X_i . Alors

$$\frac{X_1 + X_2 + \dots + X_n - nm}{\sqrt{n}}$$

converge en loi, lorsque $n \rightarrow \infty$ vers une v.a. de loi $\mathcal{N}_k(0, \Gamma)$.

Démonstration : Pour démontrer la convergence en loi d'une suite de variables aléatoires T_n vers une variable aléatoire T il suffit de montrer la convergence des fonctions caractéristiques (voir Appendice).

On pose $T_n = \frac{X_1 + X_2 + \dots + X_n - nm}{\sqrt{n}}$.

On calcule alors la fonction caractéristique, qui peut s'écrire comme le produit de n fonctions caractéristiques car les variables aléatoires X_i sont indépendantes.

$$\Phi_{T_n}(t) = \mathbb{E}[\exp i \langle t, T_n \rangle] = e^{-i \langle t, \sqrt{n}m \rangle} \mathbb{E} \left[\exp i \left\langle \frac{t}{\sqrt{n}}, X_1 \right\rangle \right]^n.$$

Par une simple translation, on se ramène à considérer des variables aléatoires centrées.

On pose $\tilde{X}_1 = X_1 - m$. Alors $\Phi_{T_n}(t) = \mathbb{E} \left(\exp \frac{i}{\sqrt{n}} \langle t, \tilde{X}_1 \rangle \right)^n$. Comme \tilde{X}_1 admet un moment d'ordre 2, sa fonction caractéristique, c'est-à-dire sa transformée de Fourier, admet un développement de Taylor d'ordre 2,

$$\Phi_{\tilde{X}_1}(t) = 1 - \frac{1}{2} t^* K t + o(|t|^2).$$

Ainsi, quand $n \rightarrow \infty$,

$$\Phi_{T_n}(t) = \left(1 - \left(\frac{t}{\sqrt{n}} \right)^* K \left(\frac{t}{\sqrt{n}} \right) + o \left(\frac{|t|^2}{n} \right) \right)^n \longrightarrow \exp -\frac{1}{2} t^* K t.$$

□

Corollaire 1.1 Soit $(X_n; n \geq 1)$ une suite de v.a.r. de Bernoulli, indépendantes et de même loi, $\mathbb{P}(X_n = 1) = p$, $\mathbb{P}(X_n = 0) = 1 - p = q$. Alors

$$\frac{X_1 + \dots + X_n - np}{\sqrt{npq}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}_1(0, 1).$$

Remarque.

En pratique dès que $np > 15$, on approxime la loi de $\frac{X_1 + \dots + X_n - np}{\sqrt{npq}}$ par $\mathcal{N}_1(0, 1)$. Notons que $X_1 + \dots + X_n$ suit $\mathcal{B}(n, p)$.

1.5 Loi du chi-deux

Pour tout $\lambda > 0$ et $a > 0$, on note $\gamma(\lambda, a)$ la loi sur \mathbb{R} de densité :

$$\frac{x^{\lambda-1}}{a^\lambda} \frac{1}{\Gamma(\lambda)} \left(\exp - \frac{x}{a} \right) \mathbb{1}_{\{x>0\}}, \quad (1.28)$$

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

Le but de cette section est de montrer qu'il existe un lien étroit entre les lois gamma et gaussiennes. On commence par un résultat préliminaire.

Proposition 1.6 1) Si X et Y sont deux v.a. indépendantes, la loi de X (resp. Y) a pour densité f (resp. g) alors $X + Y$ a pour densité $f * g$ où

$$f * g(x) = \int_{\mathbb{R}} f(x-t)g(t)dt = \int_{\mathbb{R}} g(x-t)f(t)dt; \quad x \in \mathbb{R}.$$

2) En particulier si, X a pour loi $\gamma(\lambda, a)$ et Y a pour distribution $\gamma(\mu, a)$ alors $X + Y$ suit la loi $\gamma(\lambda + \mu, a)$.

3) Si X a pour loi $\gamma(\lambda, a)$ et $\rho > 0$ alors ρX suit une loi $\gamma(\lambda, a\rho)$.

Démonstration : 1) Le premier résultat est classique, on refait brièvement la démonstration. Soit φ une fonction borélienne positive et $\Delta = \mathbb{E}[\varphi(X + Y)]$. Puisque le couple (X, Y) a pour densité $f(x)g(y)$, on a,

$$\Delta = \iint_{\mathbb{R}^2} \varphi(x + y)f(x)g(y) dx dy.$$

On fait le changement de variables : $u = x, v = x + y$, d'où

$$\Delta = \iint_{\mathbb{R}^2} \varphi(v)f(u)g(v - u)dudv = \int_{\mathbb{R}} \varphi(v)(f * g)(v)dv.$$

2) Etudions le cas particulier où $X \sim \gamma(\lambda, a)$ et $Y \sim \gamma(\mu, a)$. $X + Y$ a pour densité h , avec

$$h(x) = \int_0^x \frac{1}{a^\lambda}(x - t)^{\lambda-1} \frac{1}{\Gamma(\lambda)} e^{-\frac{x-t}{a}} \frac{1}{a^\mu} t^{\mu-1} \frac{1}{\Gamma(\mu)} e^{-\frac{t}{a}} dt; \quad x \geq 0,$$

et $h(x) = 0$ si $x < 0$. On en déduit,

$$h(x) = c \left(\int_0^x (x - t)^{\lambda-1} t^{\mu-1} dt \right) e^{-x/a}, \quad \text{avec } c = \frac{1}{a^{\lambda+\mu}} \frac{1}{\Gamma(\lambda)\Gamma(\mu)}. \quad (1.29)$$

On fait le changement de variable $t = ux$, on obtient

$$h(x) = c' x^{\lambda+\mu-1} e^{-x/a}; \quad c' = c \int_0^1 (1 - u)^{\lambda-1} u^{\mu-1} du. \quad (1.30)$$

Mais on sait que h est une densité, d'où

$$\int_{\mathbb{R}} h(x)dx = c' \int_0^\infty x^{\lambda+\mu-1} e^{-x/a} dx = c' a^{\lambda+\mu} \Gamma(\lambda + \mu) = 1. \quad (1.31)$$

On a montré que $X + Y$ suit une loi $\gamma(\lambda + \mu, a)$.

3) Si $X \sim \gamma(\lambda, a)$ et $\rho > 0$, alors un changement de variable immédiat montre que ρX suit une loi $\gamma(\lambda, a\rho)$. \square

Remarque.

En utilisant (1.29), (1.30) et (1.31) on a montré :

$$\int_0^1 (1 - t)^{\lambda-1} t^{\mu-1} dt = \frac{\Gamma(\lambda)\Gamma(\mu)}{\Gamma(\lambda + \mu)}; \quad \lambda > 0, \quad \mu > 0. \quad (1.32)$$

Définition 1.4 On appelle loi du chi-deux à n degrés de liberté, la loi de v.a. $\sum_{i=1}^n X_i^2$, où les v.a. X_1, X_2, \dots, X_n sont indépendantes et ont pour loi commune la loi gaussienne réduite et centrée. On note $\chi^2(n)$ cette distribution.

Proposition 1.7 La loi du $\chi^2(n)$ coïncide avec la loi $\gamma(n/2, 2)$. En particulier la loi de $\sum_{i=1}^n X_i^2$ est $\gamma(n/2, 2)$, où les v.a. X_1, X_2, \dots, X_n sont indépendantes et de loi $\mathcal{N}_1(0, 1)$.

Démonstration : 1) On commence par calculer la loi de X_1^2 . Posons $\Delta = \mathbb{E}[f(X_1^2)]$ où f est une fonction borélienne positive. On a :

$$\Delta = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x^2) \exp -\frac{x^2}{2} dx = \int_0^\infty f(x^2) \exp -\frac{x^2}{2} dx.$$

On fait le changement de variable $y = x^2$, il vient

$$\Delta = \frac{1}{\sqrt{2\pi}} \int_0^\infty f(y)y^{-1/2} \exp(-y/2)dy.$$

Donc $\chi^2(1) = \gamma(1/2, 2)$.

2) Posons $\xi_n = \sum_{i=1}^n X_i^2$. Montrons par récurrence sur $n \geq 1$: $\xi_n \sim \gamma(n/2, 2)$.

On a vu à l'étape précédente que $\xi_1 \sim \gamma(1/2, 2)$, la propriété est réalisée pour $n = 1$. Explicitons $n \implies n + 1$. On remarque que X_{n+1}^2 est indépendante de ξ_n , $\xi_n \sim \gamma(n/2, 2)$ et $X_{n+1}^2 \sim \gamma(1/2, 2)$. D'après le 2) de la proposition 1.6, $\xi_{n+1} \sim \gamma(n/2 + 1/2, 2) = \gamma(\frac{n+1}{2}, 2)$. \square

On s'intéresse à une suite de n variables aléatoires gaussiennes. On introduit tout d'abord la notion de moyenne et de variance empirique.

Soient X_1, X_2, \dots, X_n , n variables aléatoires indépendantes et de même loi à valeurs réelles. On définit deux nouvelles v.a. \bar{X}_n et S_n^2 la **moyenne** et la **variance empirique** :

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n); \quad S_n^2 = \frac{1}{n-1} \left(\sum_{k=1}^n (X_k - \bar{X})^2 \right). \quad (1.33)$$

En fait, la renormalisation étonnante de la variance empirique provient du fait que la moyenne et la variance empirique sont des estimateurs sans biais de l'espérance et de la variance de la loi de X_1 : si $m \in \mathbb{R}$ est l'espérance de X_1 et σ^2 sa variance, on a (exercice) : $\mathbb{E}[\bar{X}_n] = m$ et $\mathbb{E}[S_n^2] = \sigma^2$. Par ce procédé on peut donc obtenir une bonne approximation de la variance d'une variable aléatoire par la variance empirique. Il est alors important de connaître l'erreur commise dans un problème donné en prenant à la place de la variance (inconnue) la variance empirique (qui est aléatoire). On s'intéresse donc à la loi de S_n^2 qui précisera sa distribution autour de son espérance.

Théorème 1.6 Soient X_1, X_2, \dots, X_n , n v.a. réelles, indépendantes, de même loi gaussienne $\mathcal{N}_1(m, \sigma^2)$. Alors

1) \bar{X}_n et S_n^2 sont indépendantes.

2) \bar{X}_n suit une loi $\mathcal{N}_1(m, \frac{\sigma^2}{n})$ et $\frac{n-1}{\sigma^2} S_n^2$ a pour loi $\chi^2(n-1)$.

Démonstration : 1) La première étape consiste à vérifier que l'on peut se ramener au cas où $m = 0$ et $\sigma = 1$. On pose

$$X'_i = \frac{X_i - m}{\sigma} \iff X_i = \sigma X'_i + m \quad 1 \leq i \leq n. \quad (1.34)$$

Les v.a. X'_1, X'_2, \dots, X'_n sont indépendantes, chacune de loi $\mathcal{N}_1(0, 1)$. On note \bar{X}'_n et $S_n'^2$ la moyenne et la variance empirique de cette nouvelle suite. On déduit de (1.34),

$$\bar{X}_n = \sigma \bar{X}'_n + m. \quad (1.35)$$

Par conséquent, $X_i - \bar{X}_n = \sigma(X'_i - \bar{X}'_n)$. On en déduit

$$S_n^2 = \sigma^2 S_n'^2. \quad (1.36)$$

Donc si le théorème 1.6 est réalisé lorsque $m = 0$ et $\sigma = 1$, d'après (1.35) et (1.36) ce théorème l'est encore dans le cas général.

2) On suppose que chaque v.a. X_i est de loi $\mathcal{N}_1(0, 1)$.

Soit u_1 le vecteur de \mathbb{R}^n de coordonnées $\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}$. Ce vecteur est de norme 1, il existe donc u_2, u_3, \dots, u_n de \mathbb{R}^n tels que $\mathcal{B} = \{u_1, u_2, \dots, u_n\}$ soit une base orthonormée de \mathbb{R}^n . Soit A la matrice carrée d'ordre n , dont les colonnes sont u_1, u_2, \dots, u_n . Par construction A est une matrice orthogonale : $AA^* = A^*A = \text{Id}$. On note $Y = A^*X$, et $Y^* = (Y_1, Y_2, \dots, Y_n)$. D'après la proposition 1.4, Y est un vecteur gaussien centré de matrice de covariance K_Y :

$$K_Y = A^*K_X(A^*)^* = A^*\text{Id}.A = A^*A = \text{Id},$$

les v.a. X_1, X_2, \dots, X_n étant indépendantes, réduites et centrées la matrice de covariance K_X de X est l'identité. Puisque la première ligne de A^* est $u_1^* = \left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}\right)$, on a :

$$Y_1 = \frac{1}{\sqrt{n}}(X_1 + X_2 + \dots + X_n) = \sqrt{n}\bar{X}_n. \quad (1.37)$$

On va exprimer S_n^2 à l'aide de Y .

$$\begin{aligned} (n-1)S_n^2 &= \sum_{k=1}^n (X_k - \bar{X}_n)^2 = \sum_{k=1}^n (X_k^2 - 2X_k\bar{X}_n + \bar{X}_n^2) \\ &= \left(\sum_{k=1}^n X_k^2\right) - 2\bar{X}_n \left(\sum_{k=1}^n X_k\right) + n\bar{X}_n^2. \end{aligned}$$

Par conséquent,

$$(n-1)S_n^2 = \left(\sum_{k=1}^n X_k^2\right) - 2\bar{X}_n(n\bar{X}_n) + n\bar{X}_n^2 = \left(\sum_{k=1}^n X_k^2\right) - n\bar{X}_n^2. \quad (1.38)$$

Mais $Y = A^*X$, A^* est une matrice orthogonale, elle conserve la norme : $\sum_{k=1}^n Y_k^2 = \sum_{k=1}^n X_k^2$. On déduit de (1.37) et (1.38) : $(n-1)S_n^2 = \sum_{k=1}^n Y_k^2 - Y_1^2 = \sum_{k=2}^n Y_k^2$. Y est un vecteur gaussien de matrice de covariance identité, les v.a. Y_1, Y_2, \dots, Y_n sont donc indépendantes et ont pour loi $\mathcal{N}_1(0, 1)$, on en déduit l'indépendance de \bar{X}_n et S_n^2 . De plus $\bar{X}_n \sim \mathcal{N}_1(0, 1/n)$, et d'après la proposition 1.7, la loi de S_n^2 est $\chi^2(n-1)$. \square

1.6 Une application à la statistique

Nous terminerons ce chapitre en donnant une application du théorème 1.6 à la théorie des tests statistiques.

On s'intéresse à l'approximation d'une loi discrète sur $E = \{1, 2, \dots, k\}$, notée $p = (p_1, p_2, \dots, p_k)$:

$$p_i > 0, \quad \forall i \in \{1, 2, \dots, k\} \quad \text{et} \quad \sum_{i=1}^k p_i = 1. \quad (1.39)$$

On considère Y_1, Y_2, \dots, Y_n , un échantillon de loi p ; ce qui signifie que les v.a. Y_1, Y_2, \dots, Y_n sont indépendantes et ont même loi p . On introduit :

$$N_n(i) = \#\{j; 1 \leq j \leq n, Y_j = i\} = \sum_{j=1}^n \mathbf{1}_{\{Y_j=i\}}; \quad i \in E. \quad (1.40)$$

$N_n(i)$ est le nombre de fois où $Y_j = i$, lorsque j varie de 1 à n , $\frac{N_n(i)}{n}$ est donc la fréquence d'apparition de i .

Proposition 1.8 *Soit \sqrt{p} le vecteur de \mathbb{R}^k de coordonnées $(\sqrt{p_1}, \sqrt{p_2}, \dots, \sqrt{p_k})$. On note*

$$T_n := \left(\frac{N_n(1) - np_1}{\sqrt{np_1}}, \frac{N_n(2) - np_2}{\sqrt{np_2}}, \dots, \frac{N_n(k) - np_k}{\sqrt{np_k}} \right); \quad n \geq 1.$$

Alors T_n converge en loi, lorsque $n \rightarrow \infty$, vers une v.a. de loi gaussienne $\mathcal{N}_k(0, \Gamma)$ avec $\Gamma = \text{Id} - \sqrt{p}(\sqrt{p})^$.*

Démonstration : Soit (e_1, e_2, \dots, e_k) la base canonique de \mathbb{R}^k . On introduit :

$$X_j = \sum_{l=1}^k \mathbf{1}_{\{Y_j=l\}} e_l; \quad 1 \leq j \leq n. \quad (1.41)$$

$(X_j; j \geq 1)$ est une famille de vecteurs aléatoires, indépendants, de même loi, et à valeurs dans \mathbb{R}^k . Les X_j ont un moment d'ordre 2, d'après le théorème central limite,

$$Z_n = \frac{1}{\sqrt{n}} \left(\sum_{j=1}^n X_j - n\mathbb{E}(X) \right), \quad \text{converge en loi vers } \mathcal{N}_k(0, A),$$

où A désigne la matrice de covariance de X_1 . D'après (1.41),

$$Z_n = \frac{1}{\sqrt{n}} \left(\sum_{l=1}^k \left\{ \sum_{j=1}^n \mathbf{1}_{\{Y_j=l\}} - np_l \right\} e_l \right) = \frac{1}{\sqrt{n}} \left(\sum_{l=1}^k (N_n(l) - np_l) e_l \right).$$

Donc

$$Z_n^* = \left(\frac{N_n(1) - np_1}{\sqrt{n}}, \frac{N_n(2) - np_2}{\sqrt{n}}, \dots, \frac{N_n(k) - np_k}{\sqrt{n}} \right).$$

Il reste donc à évaluer la matrice $A = (a_{i,j})$.

$$a_{i,j} = \text{cov}(\mathbf{1}_{\{Y_1=i\}}, \mathbf{1}_{\{Y_1=j\}}) = \mathbb{P}(Y_1 = i, Y_1 = j) - \mathbb{P}(Y_1 = i)\mathbb{P}(Y_1 = j).$$

D'où $a_{i,j} = p_i\delta_{i,j} - p_i p_j$, avec $\delta_{i,j} = \mathbf{1}_{\{i=j\}}$. Soit u un vecteur de \mathbb{R}^k , $u^* = (u_1, u_2, \dots, u_k)$. On lui associe $v \in \mathbb{R}^k$, $v^* = (u_1/\sqrt{p_1}, u_2/\sqrt{p_2}, \dots, u_k/\sqrt{p_k})$. On a :

$$\langle u, T_n \rangle = u^* T_n = v^* Z_n = \langle v, Z_n \rangle.$$

Soit Φ_n la fonction caractéristique de T_n ; alors

$$\Phi_n(u) = \mathbb{E}[\exp(i \langle u, T_n \rangle)] = \mathbb{E}[\exp(i \langle v, Z_n \rangle)].$$

Sachant que Z_n converge en loi, vers $\mathcal{N}_k(0, A)$,

$$\lim_{n \rightarrow \infty} \Phi_n(u) = \exp -\frac{1}{2} v^* A v.$$

Un calcul immédiat conduit à :

$$\begin{aligned} v^* A v &= \sum_{i=1}^k a_{ii} v_i^2 + \sum_{1 \leq i \neq j \leq k} a_{i,j} v_i v_j = \sum_{i=1}^k (1 - p_i) u_i^2 - \sum_{1 \leq i \neq j \leq k} \sqrt{p_i} \sqrt{p_j} u_i u_j \\ &= \sum_{i=1}^k u_i^2 - \left(\sum_{i=1}^k \sqrt{p_i} u_i \right) \left(\sum_{j=1}^k \sqrt{p_j} u_j \right) = u^* \Gamma u. \end{aligned}$$

□

Considérons alors

$$d_n^2 = \|T_n\|^2 = \sum_{i=1}^k \left(\frac{N_n(i) - np_i}{\sqrt{np_i}} \right)^2 = \sum_{i=1}^k \frac{n}{p_i} \left(\frac{N_n(i)}{n} - p_i \right)^2. \quad (1.42)$$

$\frac{d_n^2}{n}$ correspond au carré d'une pseudo-distance dans \mathbb{R}^k , entre la fréquence empirique $\left(\frac{N_n(i)}{n}; 1 \leq i \leq k \right)$ et $(p_i; 1 \leq i \leq k)$.

Proposition 1.9 d_n^2 converge en loi, lorsque $n \rightarrow \infty$, vers la loi $\chi^2(k-1)$.

Avant de démontrer ce résultat, introduisons un lemme préliminaire :

Lemme 1.2 Soit $\Gamma = \text{Id.} - (\sqrt{p})(\sqrt{p})^*$. Alors il existe une matrice orthogonale B telle que

$$B\Gamma B^* = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}$$

Démonstration : Soit u un vecteur de \mathbb{R}^k , alors

$$\sqrt{p}(\sqrt{p})^*u = \langle \sqrt{p}, u \rangle \sqrt{p}.$$

Par conséquent $u \rightarrow \sqrt{p}(\sqrt{p})^*u$ est l'opérateur de projection orthogonale sur le vecteur normé \sqrt{p} . Donc Γ est la projection orthogonale sur F , F désignant le sous-espace vectoriel de \mathbb{R}^k orthogonal à \sqrt{p} : $F = \{v \in \mathbb{R}^k \mid \langle v, \sqrt{p} \rangle = 0\}$. Il existe une base orthonormée \mathcal{B}_1 pour laquelle Γ soit représenté par une matrice diagonale dont la diagonale est formée uniquement de 1 sauf le dernier élément qui est nul. On choisit B^* , la matrice de changement de base, de la base canonique à \mathcal{B}_1 . B est une matrice orthogonale, le lemme 1.2 est alors vérifié. \square

Démonstration : (Proposition 1.9)

Puisque $x \in \mathbb{R}^k \rightarrow \|x\|^2$ est continue, d'après la proposition 1.8, la v.a.r d_n^2 converge en loi vers $\|Z\|^2$, Z de loi $\mathcal{N}_k(0, \Gamma)$.

Posons $Z' = BZ$, où B est la matrice définie par le lemme 1.2. Puisque B est une matrice orthogonale, $\|Z'\|^2 = \|Z\|^2$. De plus Z' est un vecteur gaussien centré, de matrice de covariance $K_{Z'} = B\Gamma B^*$. Ainsi la loi de $\|Z'\|^2$ est $\chi^2(k-1)$. \square

Application : le test du chi-deux

La théorie des tests consiste à formuler des hypothèses particulières sur les paramètres ou sur les lois qui interviennent dans les problèmes étudiés, puis à apporter un jugement sur ces

hypothèses. Ce jugement est basé d'une part sur les résultats obtenus sur un ou plusieurs échantillons extraits de la population concernée et d'autre part, sur l'acceptation d'un certain *risque* dans la prise de décision. Le test du chi-deux est un test de conformité de deux distributions, ou encore un *test d'ajustement* entre une distribution théorique et une distribution expérimentale.

On veut tester si une v.a. Y a pour loi p , c'est-à-dire si

$$\mathbb{P}(Y = i) = p_i; \quad 1 \leq i \leq k.$$

On définit deux hypothèses possibles :

$$H_0 : Y \text{ a pour loi } p, \quad H_1 : Y \text{ n'a pas la loi } p,$$

et on aimerait choisir entre les deux suivant les données fournies par un échantillon. Pour établir la crédibilité de l'hypothèse H_0 , des règles très précises doivent être énoncées pour permettre de conclure au rejet ou à l'acceptation de H_0 . Cependant pour des événements dans lesquels intervient le hasard, il est impossible de prendre une bonne décision sans risque de se tromper. Il faut donc mettre en oeuvre une règle conduisant à rejeter H_0 , si elle est vraie, que dans une faible proportion des cas. Cette décision a un caractère probabiliste, toute décision comporte un risque qu'elle soit erronée. Ce risque, noté α , qui est le risque de rejeter à tort l'hypothèse H_0 alors qu'elle est vraie et qui favorise donc l'hypothèse H_1 s'appelle *seuil de signification* ou *risque de première espèce*.

$$\alpha = \mathbb{P}\{\text{rejeter } H_0 \mid H_0 \text{ vraie}\} = \mathbb{P}\{\text{choisir } H_1 \mid H_0 \text{ vraie}\}.$$

La probabilité α , appelée aussi *risque du client*, est choisie a priori par l'utilisateur. Les valeurs les plus utilisées pour α sont 0,05 et 0,01.

La règle de décision comporte un risque β ou *risque de deuxième espèce*, c'est le risque de ne pas rejeter H_0 alors que H_1 est vraie :

$$\beta = \mathbb{P}\{\text{ne pas rejeter } H_0 \mid H_1 \text{ vraie}\}.$$

Cette probabilité, appelée également *risque du fournisseur* (celui de voir, par exemple, une bonne production refusée), dépend de l'hypothèse alternative H_1 .

Élaboration d'un test et démarche à suivre :

- 1) formuler de façon précise les hypothèses H_0 et H_1 ,
- 2) fixer, avant l'expérience, le risque α de première espèce, c'est-à-dire le risque de rejeter l'hypothèse H_0 alors qu'elle est vraie,
- 3) choisir la statistique la mieux adaptée en fonction des caractéristiques de la population étudiée,
- 4) déterminer la région critique ou région de rejet de l'hypothèse H_0 au profit de l'hypothèse H_1 et en déduire la règle de décision,
- 5) calculer effectivement la valeur numérique de la variable de décision en utilisant les données de l'échantillon,
- 6) donner les conclusions du test.

Dans le cas qui nous intéresse (le cadre du test du chi-deux), la variable de décision est la pseudo-distance entre les fréquences empiriques et p . En effet, on suppose que l'on peut reproduire des copies indépendantes de $Y : Y_1, Y_2, \dots, Y_n$, qui sera notre échantillon.

On définit alors d_n^2 comme dans (1.42). Si d_n^2 est grand, on choisira de préférence l'hypothèse H_1 , par contre, si d_n^2 est petit, les fréquences empiriques sont proches de la probabilité attendue p et donc on choisira l'hypothèse H_0 . Il faut donc trouver un seuil θ_α , dépendant du risque α , qui définira la région de rejet.

$$\mathbb{P}(d_n^2 > \theta_\alpha) = \alpha.$$

En appliquant la proposition 1.9, on obtient, pour n suffisamment grand

$$\mathbb{P}(d_n^2 > \theta_\alpha) \sim \mathbb{P}(\xi > \theta_\alpha),$$

où ξ suit une loi $\chi^2(k-1)$. Le test du chi-deux consiste donc à définir le seuil θ_α par l'équation

$$\mathbb{P}(\xi > \theta_\alpha) = \alpha.$$

Dans la pratique, n assez grand veut dire $np_i \geq 10$ pour tout $i \in \{1, 2, \dots, k\}$.

1.7 Exercices

1.1. Soit $I := \int_0^\infty e^{-x^2/2} dx$. Montrer que I est une intégrale convergente. Calculer I^2 en passant en coordonnées polaires. Que vaut I ? En déduire la valeur de $\Gamma(1/2)$.

1.2. Loi gamma. Pour tout $a, b > 0$, on appelle loi $\gamma(a, b)$ la loi de densité :

$$\frac{1}{\Gamma(a)} x^{a-1} b^a e^{-bx} \mathbf{1}_{\{x>0\}}.$$

i) Calculer la transformée de Laplace de cette loi, c'est-à-dire :

$$\frac{1}{\Gamma(a)} \int_0^\infty e^{-\lambda x} x^{a-1} b^a e^{-bx} dx.$$

Démontrer que l'espérance de cette loi vaut a/b et sa variance a/b^2 .

ii) Vérifier que $\gamma(a, b) * \gamma(a', b) = \gamma(a + a', b)$. En déduire que :

$$\int_0^1 x^{a-1} (1-x)^{a'-1} dx = \frac{\Gamma(a)\Gamma(a')}{\Gamma(a+a')}$$

iii) *injectivité de la transformée de Laplace* Soient ξ, η deux variables aléatoires positives. On suppose que, pour tout λ réel positif :

$$\mathbb{E}(e^{-\lambda\xi}) = \mathbb{E}(e^{-\lambda\eta}).$$

a) On note $\varphi(z) = \mathbb{E}(e^{z\xi})$ et $\psi(z) = \mathbb{E}(e^{z\eta})$. Montrer que φ et ψ sont bien définies, continues sur $\{z \in \mathbb{C} : \operatorname{Re}(z) \leq 0\}$ et holomorphes sur $\{z \in \mathbb{C} : \operatorname{Re}(z) < 0\}$.

b) Montrer que φ et ψ coïncident sur \mathbb{R}_- . En déduire que ces fonctions coïncident sur $\{z \in \mathbb{C} : \operatorname{Re}(z) < 0\}$, et donc que, pour tout $t \in \mathbb{R}$:

$$\mathbb{E}(e^{it\xi}) = \mathbb{E}(e^{it\eta}).$$

c) Montrer que ξ et η ont la même loi.

iv) *Loi de chi-deux.* Soient X_1, X_2, \dots, X_n , n variables aléatoires réelles indépendantes gaussiennes standard. Montrer que :

$$Z := \sum_{j=1}^n X_j^2 \sim \gamma(n/2, 1/2).$$

(On dit que Z suit une loi de χ^2 à n degrés de liberté.)

1.3. Loi de chi-deux décentrée. Soient X_1, X_2, \dots, X_n , n variables aléatoires réelles gaussiennes, indépendantes, telles que $X_j \sim \mathcal{N}(m_j, 1)$.

i) Montrer que la fonction caractéristique de $Z := \sum_{j=1}^n X_j^2$ est :

$$\mathbb{E}(e^{itZ}) = \frac{1}{(1 - 2it)^{n/2}} \exp\left(\frac{itr}{1 - 2it}\right),$$

où $r := \sum_{j=1}^n m_j^2$. (On dit que Z suit une loi de χ^2 à décentrée n degrés de liberté, de paramètre de décentrage r , $\chi^2(n, r)$).

ii) Calculer $\mathbb{E}(Z)$ et $\operatorname{Var}(Z)$.

1.4. Loi beta. Soient $a, b > 0$. Désignons par Z_a, Z_b deux variables aléatoires indépendantes de lois respectivement $\gamma(a, 1)$, $\gamma(b, 1)$.

i) Trouver la densité du vecteur aléatoire

$$(U, V) := \left(Z_a + Z_b, \frac{Z_a}{Z_a + Z_b} \right).$$

ii) Les variables aléatoires U et V sont-elles indépendantes ? (On dit que V suit une loi $\mathfrak{B}(a, b)$.)

1.5. Loi de Cauchy et loi d'arcsinus.

i) Soient X et Y deux variables aléatoires indépendantes de même loi gaussienne standard. Montrer que la variable aléatoire $C = X/Y$ suit une loi de Cauchy.

ii) Soit V une variable aléatoire de loi $\mathfrak{B}(1/2, 1/2)$ (loi d'arcsinus). Démontrer que la variable aléatoire $1/V$ a la même loi que la variable aléatoire $1 + C^2$, où C est de loi de Cauchy.

1.6. Soit $E \sim \gamma(1, 1)$ et $V \sim \mathfrak{B}(1/2, 1/2)$ deux variables indépendantes (E est de loi $\mathcal{E}(1)$). Montrer que :

$$2EV \sim G^2,$$

où G est une variable aléatoire gaussienne standard. On pourra observer que la relation précédente équivaut à : pour tout $a \geq 0$:

$$\sqrt{\frac{2}{\pi}} \int_a^\infty e^{-t^2/2} dt = \mathbb{E} \left[\exp -\frac{a^2}{2V} \right].$$

1.7. Loi de Fisher-Snedecor. Soient $X \sim \gamma(a, b)$ et $Y \sim \gamma(c, d)$ deux variables aléatoires indépendantes.

i) Trouver la loi de la variable aléatoire X/Y .

ii) Supposons que $a = n/2$, $c = m/2$, $b = d = 1/2$. Trouver la loi de :

$$R := \frac{X/n}{Y/m}$$

(On dit que R suit une loi de Fisher-Snedecor à n et m degrés de liberté).

iii) Calculer $\mathbb{E}(R)$ et $\text{Var}(R)$ (discussion suivant les valeurs de n et m).

1.8. Loi de Student. Soient $X \sim \mathcal{N}(0, 1)$ et $Y \sim \chi^2(n)$ deux variables aléatoires indépendantes.

i) Trouver la loi de la variable aléatoire :

$$T := \frac{X}{\sqrt{Y/n}}$$

(On dit que T suit une loi de Student à n degrés de liberté).

ii) Calculer $\mathbb{E}(T)$ et $\text{Var}(T)$.

1.9. On note par g_n la densité d'une variable aléatoire de loi $\chi^2(n)$:

$$g_n(x) := \frac{1}{2^{n/2} \Gamma(n/2)} x^{n/2-1} e^{-x/2} \mathbf{1}_{\{x \geq 0\}}.$$

i) Étudier la régularité de g_n à l'origine et le graphe de g_n .

ii) Pour quelle valeur de x , $g_n(x)$ est-elle maximum ?

iii) Étudier le comportement de ce maximum quand $n \uparrow \infty$.

1.10. Approximation de la fonction de répartition de la loi gaussienne standard. Soit

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

i) Vérifier que, pour tout $x > 0$:

$$\frac{1}{x\sqrt{2\pi}} e^{-x^2/2} = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} (1 + 1/t^2) dt.$$

ii) En déduire, pour $x > 0$:

$$1 - \Phi(x) \leq \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}.$$

iii) Démontrer que :

$$\frac{1}{\sqrt{2\pi}} \left(\frac{1}{x} - \frac{1}{x^3} \right) e^{-x^2/2} = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} \left(1 - \frac{3}{t^4} \right) dt.$$

En déduire que, pour $x > 0$:

$$\frac{1}{\sqrt{2\pi}} \left(\frac{1}{x} - \frac{1}{x^3} \right) e^{-x^2/2} \leq 1 - \Phi(x) \leq \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}.$$

iv) Démontrer que :

$$1 - \Phi(x) \sim \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}, \text{ lorsque } x \uparrow \infty.$$

1.11. Soit $X_1, X_2, \dots, X_n, \dots$ une suite de variables aléatoires réelles gaussiennes indépendantes, de même loi $\mathcal{N}(m, \sigma^2)$.

i) Soit

$$S_n := \frac{X_1 + \dots + X_n - nm}{n^\alpha}, \quad 1/2 < \alpha \leq 1.$$

Montrer que S_n est une variable aléatoire gaussienne et calculer son espérance et sa variance.

ii) Utiliser l'exercice précédent pour majorer $\mathbb{P}(|S_n| > \varepsilon)$.

iii) En déduire que $S_n \rightarrow 0$ p.s.

1.12. Soient X et Y deux variables aléatoires indépendantes de même loi.

i) Montrer que si X et Y sont deux variables gaussiennes standard, alors $X + Y$ et $X - Y$ sont indépendantes.

ii) *Théorème de Bernstein.* Réciproquement, on suppose que X et Y sont de carré intégrable et que $X + Y$ et $X - Y$ sont indépendantes. On veut montrer que X et Y sont deux variables gaussiennes. Pour cela :

a) Montrer qu'on peut supposer que X et Y sont centrées, de variance 1.

b) Montrer que φ , la fonction caractéristique commune de X et de Y , satisfait l'égalité $\varphi(2t) = \varphi(t)^3 \varphi(-t)$. En déduire que φ ne s'annule nulle part.

c) On pose $\psi(t) := \varphi(t)/\varphi(-t)$. Montrer que $\psi(2t) = \psi(t)^2$ et que $\psi(t) = 1 + o(t^2)$, lorsque $t \downarrow 0$. En déduire que, pour tout t , $\psi(t) = 1$ et que $\varphi(t) = \varphi(t/2)^4$. Conclure.

1.13. Soit X une variable aléatoire réelle gaussienne standard et ε une autre variable aléatoire indépendante de X , prenant seulement les valeurs ± 1 .

i) Montrer que $Y := \varepsilon X$ est une variable aléatoire réelle gaussienne standard.

ii) Calculer $\text{Cov}(X, Y)$.

iii) La variable aléatoire $X + Y$ est-elle gaussienne ?

1.14. Soit X une variable aléatoire réelle gaussienne standard. Pour tout $a > 0$ on note :

$$Y_a := -X \mathbf{1}_{\{|X| \leq a\}} + X \mathbf{1}_{\{|X| > a\}}.$$

- i) Montrer que Y_a est une variable aléatoire réelle gaussienne standard.
- ii) Démontrer que le couple (X, Y_a) n'est pas gaussien.
- iii) Y-a-t il une valeur de a telle que la matrice de covariance de (X, Y_a) soit l'identité?

1.15. Soit (X_1, X_2) un couple gaussien tel que :

$$\mathbb{E}(X_1) = \mathbb{E}(X_2) = 0, \quad \mathbb{E}(X_1^2) = \mathbb{E}(X_2^2) = 1, \quad \mathbb{E}(X_1 X_2) = \rho.$$

- i) Montrer que $|\rho| \leq 1$.
- ii) Calculer la densité du couple (X_1, X_2) .
- iii) Retrouver, à partir du résultat de ii) les lois marginales de X_1 et X_2 , ainsi que les quantités $\mathbb{E}(X_1)$, $\mathbb{E}(X_2)$, $\mathbb{E}(X_1^2)$, $\mathbb{E}(X_2^2)$, $\mathbb{E}(X_1 X_2)$.
- iv) Donner une condition nécessaire et suffisante pour que X_1 et X_2 soient indépendantes.
- v) Que se passe-t-il quand $|\rho| = 1$?

1.16. Soient X et Y deux variables aléatoires réelles, indépendantes, de même loi, centrées et de variance 1. On suppose que la variable aléatoire $(X + Y)/\sqrt{2}$ a la même loi que X et Y .

- i) Soit $\varphi(t) := \mathbb{E}(e^{itX})$ la fonction caractéristique de X . Montrer que $\varphi'(0) = 0$, $\varphi''(0) = -1$.
- ii) Démontrer que, pour tout t , $\varphi^2(t/\sqrt{2}) = \varphi(t)$. En déduire que, pour tout t , $\varphi(t) \neq 0$.
- iii) Montrer que, pour tout t et tout $n \in \mathbb{N}$,

$$\frac{\log \varphi(t)}{t^2} = \frac{\log \varphi(t/2^{n/2})}{(t/2^{n/2})^2}.$$

En déduire que X et Y sont des variables aléatoires gaussiennes standard.

1.17. Loi de Rayleigh. Soit (X, Y) un couple gaussien centré, de matrice de covariance égale à l'identité. Soit $R = \sqrt{X^2 + Y^2}$ et $Q = Y/X$. Montrer que R et Q sont deux variables aléatoires. Quelles sont les lois de R et Q ?

1.18. Soit (X, Y) un couple de variables aléatoires de densité :

$$f(x, y) := \lambda \exp \left[-\frac{1}{2}(2x^2 - 2xy + y^2) \right].$$

- i) Que vaut λ ? Trouver les lois marginales de X et de Y .
- ii) Montrer que (X, Y) est un vecteur gaussien centré. Quelle est sa matrice de covariance ?
- iii) Les variables X et Y sont-elles indépendantes ? non-corrélées ?

1.19. Soit (X, Y) un couple gaussien. On suppose que X et Y sont centrées, de variance 1 et on note par ρ le coefficient de corrélation du couple. Montrer que :

$$\mathbb{P}(XY > 0) = \frac{1}{2} + \frac{1}{\pi} \arcsin \rho.$$

Calculer $\mathbb{P}(XY < 0)$. Soit $(X, Y, Z)^*$ un vecteur gaussien centré de matrice de covariance

$$\begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_3 & 1 \end{pmatrix}.$$

Utiliser les résultats précédents pour calculer $\mathbb{P}(X > 0, Y > 0, Z > 0)$.

1.20. Soit (X, Y) un couple gaussien. On suppose que X et Y sont centrées, de variance 1 et le coefficient de corrélation du couple vaut ρ . Montrer que :

$$\mathbb{E}[\max\{X, Y\}] = \sqrt{\frac{1-\rho}{\pi}}.$$

Que vaut $\mathbb{E}[\max\{X, Y\}^2]$?

1.21. i) Montrer qu'il existe un triplet gaussien (X_1, X_2, X_3) tel que :

$$\mathbb{E}(X_1) = \mathbb{E}(X_2) = \mathbb{E}(X_3) = 0, \quad \mathbb{E}(X_1^2) = \mathbb{E}(X_2^2) = \mathbb{E}(X_3^2) = 1,$$

$$\mathbb{E}(X_1X_2) = \mathbb{E}(X_1X_3) = \mathbb{E}(X_2X_3) = 1/2.$$

ii) Quelle est la loi de $X_1 - X_2 + 2X_3$?

iii) Montrer que, pour tout $a \in \mathbb{R}$, $(X_1 + aX_2, X_1 - X_2)$ est un couple gaussien. Existe-t-il un a tel que $(X_1 + aX_2$ et $X_1 - X_2)$ soient indépendantes ?

iv) Calculer la fonction caractéristique et la densité de (X_1, X_2, X_3) .

1.22. Théorème de Cochran. Soit $X = (X_1, \dots, X_n)^*$ un vecteur gaussien centré de covariance identité. Soit A une matrice orthogonale $n \times n$. Définissons $Y := AX$.

i) Quelle est la loi de $\|Y\|^2 := \sum_{j=1}^n Y_j^2$?

ii) On considère une décomposition en somme directe orthogonale de \mathbb{R}^n , c'est-à-dire $\mathbb{R}^n = \bigoplus_{j=1}^p E_j$, E_j orthogonal à E_k pour $j \neq k$. Soit Π_{E_j} la projection orthogonale de \mathbb{R}^n sur E_j et on note $X_{E_j} := \Pi_{E_j}(X)$. Montrer que les variables aléatoires $X_{E_j}, j = 1, \dots, p$, sont indépendantes et que $\|X_{E_j}\|^2$ suit une loi de χ^2 à r_j degrés de liberté, avec $r_j := \dim E_j$.

1.23. Soit X un vecteur gaussien de dimension n , centré, de matrice de covariance K (de type $n \times n$). Soit A une matrice $m \times n$ et B une matrice $p \times n$. On définit $Y := AX, Z := BX$. Montrer que Y et Z sont indépendantes si et seulement si $AKB^* = 0$.

1.24. Soit X un vecteur gaussien de dimension n , centré, de matrice de covariance K inversible. Montrer que $X^*K^{-1}X \sim \chi^2(n)$.

1.25. Soit

$$K = \begin{pmatrix} 1 & 1 & -1 \\ 1 & 2 & 0 \\ -1 & 0 & 3 \end{pmatrix}.$$

- i) Montrer qu'il existe un triplet gaussien (X, Y, Z) , centré et de matrice de covariance K . Calculer la densité de ce triplet.
 ii) Trouver la loi de $U := X + 2Y - Z$.
 iii) Montrer que $Y^2/2 + Z^2/3 \sim \chi^2(2)$.
 iv) Montrer que $(X + Y, Y - Z)$ est un couple gaussien. Calculer sa densité.

1.26. Théorème de Lévy-Cramer. Soient X et Y deux variables aléatoires réelles indépendantes telles que $X + Y$ soit une variable gaussienne. On veut alors montrer que X et Y sont des variables gaussiennes.

i) Soit ξ une variable aléatoire réelle telle que $h_\xi(u) := \mathbb{E}[e^{u\xi^2}] < \infty$, pour $u > 0$. On note $\psi_\xi(z) := \mathbb{E}[e^{z\xi}]$, $z \in \mathbb{C}$.

a) Montrer que

$$|xz| \leq ux^2 + \frac{|z|^2}{u}, \quad u > 0, \quad x \in \mathbb{R}, \quad z \in \mathbb{C}.$$

b) Montrer que ψ_ξ est entière (holomorphe sur \mathbb{C}).

c) Si ψ_ξ ne s'annule pas et si $\ln |\psi_\xi(z)| \geq c + c'|z|^2$, pour deux constantes c, c' , montrer que

$$\psi_\xi(z) = \exp\left(bz + \frac{az^2}{2}\right), \quad b \in \mathbb{R}, \quad a \geq 0.$$

On utilisera le théorème de Liouville : si f est entière telle que $|\operatorname{Re}f(z)| \leq c_1 + c_2|z|^2$, pour tout $z \in \mathbb{C}$, alors f est un polynôme de degré inférieur ou égal à 2.

d) En déduire que ξ est une variable aléatoire gaussienne.

ii)

a) On peut supposer que les médianes de X et Y sont nulles.

b) Montrer que, pour tout $t \geq 0$,

$$\mathbb{P}(|X + Y| > t) \geq \mathbb{P}(X > t, Y > 0) + \mathbb{P}(X < -t, Y < 0) + \mathbb{P}(|X| > t, Y = 0),$$

$$\mathbb{P}(X > t, Y > 0) \geq \frac{1}{2}\mathbb{P}(X > t) - \mathbb{P}(X > t, Y = 0),$$

$$\mathbb{P}(X < -t, Y < 0) \geq \frac{1}{2}\mathbb{P}(X < -t) - \mathbb{P}(X < -t, Y = 0),$$

$$\mathbb{P}(|X + Y| > t) \geq \frac{1}{2}\mathbb{P}(|X| \geq t).$$

c) Soit φ une fonction borélienne, positive et on note $\Phi(x) := a + \int_0^x \varphi(t)dt$, $x \geq 0$. Montrer que, pour toute variable aléatoire $\eta \geq 0$,

$$\mathbb{E}\Phi(\eta) = a + \int_0^\infty \varphi(t)\mathbb{P}(\eta > t)dt.$$

d) En déduire que, pour toute variable aléatoire ξ ,

$$\mathbb{E}[e^{u\xi^2}] = 1 + 2u \int_0^\infty te^{ut^2} \mathbb{P}(|\xi| > t) dt.$$

- e) En utilisant **ii.b)** et le point précédent, montrer que $h_X(u) \leq 2h_{X+Y}(u)$, $u > 0$.
 f) En déduire qu'il existe $u_0 > 0$, tel que, pour $0 < u < u_0$, $h_X(u), h_Y(u) < \infty$.
 g) Démontrer le théorème de Lévy-Cramer.

1.27. Soit $X \sim \mathcal{N}_1(m, \sigma^2)$ telle que $\mathbb{P}(X \geq 3) = 0.8413$ et $\mathbb{P}(X \geq 9) = 0.0228$. Calculer m et σ (on donne $\Phi(1) = 0.8413$ et $\Phi(2) = 0.9772$).

1.28. Un ordinateur sort les chiffres au hasard de la façon suivante :

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 32 & 22 & 23 & 31 & 21 & 23 & 28 & 25 & 18 & 27 \end{pmatrix}.$$

La machine fournit-elle des nombres équi-répartis (la table de $\chi^2(9)$ donne $\theta_{0.05} = 16.92$) ?

1.29. *Loi log-normale.* Soit X une variable aléatoire telle que $\ln X$ est une variable aléatoire gaussienne standard. Trouver la densité de X et calculer ses moments $\alpha_n := \mathbb{E}(X^n)$, $n \in \mathbb{N}^*$.

1.30. On considère la fonction :

$$h(x) := \frac{1}{\sqrt{2\pi e}} x^3 \mathbf{1}_{[-1,1]}(x), \quad x \in \mathbb{R}.$$

et on note par g la densité gaussienne standard. On définit la fonction :

$$f(x, y) := g(x)g(y) + h(x)h(y).$$

Démontrer que f est une densité de probabilité sur \mathbb{R}^2 qui n'est pas gaussienne, mais que les densités marginales sont gaussiennes.

1.31. Soit $a \in \mathbb{R}$, $|a| \leq 1$ et on note la fonction de répartition gaussienne réduite par :

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt = \int_{-\infty}^x g(t) dt, \quad x \in \mathbb{R}$$

i) Montrer que la fonction :

$$F(x, y) := \Phi(x) \Phi(y) [1 + a(1 - \Phi(x))(1 - \Phi(y))], \quad (x, y) \in \mathbb{R}^2$$

est une fonction de répartition d'un couple aléatoire ayant les fonctions de répartition marginales $\Phi(x)$ et $\Phi(y)$. Démontrer que, si $a \neq 0$, le couple n'est pas gaussien.

ii) Montrer que la fonction :

$$f(x, y) := g(x)g(y)[1 + a(2\Phi(x) - 1)(2\Phi(y) - 1)], \quad (x, y) \in \mathbb{R}^2$$

est une fonction de densité d'un couple aléatoire ayant les marginales $g(x)$ et $g(y)$. Démontrer que, si $a \neq 0$, le couple n'est pas gaussien.

1.32. *Promenade aléatoire dans \mathbb{R} et vecteurs gaussiens.* Partant de l'origine, un individu choisit une direction (à droite ou à gauche) au hasard, puis il fait un pas de 1 mètre dans cette direction, et recommence ainsi indéfiniment. On pose $S_0 = 0$ et soit S_k sa position après le k -ième pas. On note

$$X_k := S_k - S_{k-1}, k \geq 1 \quad Z_n := S_n / \sqrt{n}.$$

Montrer que, pour $0 \leq s \leq t$, $Z_n^{(2)} := (Z_{[ns]}, Z_{[nt]})$, converge en loi, quand $n \rightarrow \infty$ vers un vecteur gaussien centré $B^{(2)} := (B_s, B_t)$ de matrice de covariance

$$\begin{pmatrix} s & s \\ s & t \end{pmatrix}.$$

Énoncer et démontrer un résultat pour des vecteurs de taille $r \geq 2$.

1.33. Soient $g_i(x, y)$ $i = 1, 2$ deux densités gaussiennes bidimensionnelles centrées de matrice de covariance

$$\begin{pmatrix} 1 & \rho_i \\ \rho_i & 1 \end{pmatrix}, i = 1, 2.$$

On définit :

$$f(x, y) := c_1 g_1(x, y) + c_2 g_2(x, y)$$

où c_1, c_2 sont arbitraires $c_1, c_2 \geq 0$ et $c_1 + c_2 = 1$.

i) Montrer que f est la densité d'un couple (X, Y) et que, si $\rho_1 \neq \rho_2$, alors le couple n'est pas gaussien.

ii) Démontrer que les variables X et Y sont des gaussiennes standard avec le coefficient de corrélation $\rho := c_1 \rho_1 + c_2 \rho_2$.

iii) Les variables X et Y sont-elles non-corrélées ? indépendantes ?

1.34. Soit (X, Y) un couple aléatoire de densité

$$f(x, y) := \frac{1}{2\pi\sqrt{3}} \left[\exp\left(-\frac{2}{3}(x^2 + xy + y^2)\right) + \exp\left(-\frac{2}{3}(x^2 - xy + y^2)\right) \right],$$

où $(x, y) \in \mathbb{R}^2$.

i) Montrer que le couple (X, Y) n'est pas gaussien.

ii) Montrer que X et Y sont des variables aléatoires gaussiennes standard.

iii) Les variables X et Y sont-elles non-corrélées ? indépendantes ?

1.35. Soient ξ_1 et ξ_2 deux variables aléatoires indépendantes de même loi gaussienne standard. On considère les vecteurs aléatoires (X_1, X_2) et (Y_1, Y_2) définis respectivement par :

$$X_1 := \xi_1 + \xi_2, X_2 := 2\xi_1 + \xi_2$$

et

$$Y_1 := \sqrt{2}\xi_1, Y_2 := (3/\sqrt{2})\xi_1 + (1/\sqrt{2})\xi_2.$$

- i) Calculer les espérances et les matrices de covariance des deux couples (X_1, X_2) et (Y_1, Y_2) .
 ii) Montrer que ces deux couples sont des couples gaussiens. Que peut-on remarquer ?

1.36. Soient ξ_1 et ξ_2 deux variables aléatoires indépendantes de même loi gaussienne standard. On considère le vecteur aléatoire défini par :

$$(X, Y) := \begin{cases} (\xi_1, |\xi_2|), & \text{si } \xi_1 \geq 0 \\ (\xi_1, -|\xi_2|), & \text{si } \xi_1 < 0. \end{cases}$$

Montrer que le couple (X, Y) n'est pas gaussien, mais que X et Y sont des variables aléatoires gaussiennes.

1.37. Soit (X, Y) un couple aléatoire de densité

$$f(x, y) := \begin{cases} [1/(\pi(1 - \rho^2)^{1/2})] \exp[-\frac{1}{2}\rho^{-2}(x^2 - 2\rho xy + y^2)], & \text{si } xy \geq 0 \\ 0, & \text{si } xy < 0, \end{cases}$$

où $|\rho| < 1$.

- i) Montrer que f est une densité de probabilité, mais que le couple (X, Y) n'est pas gaussien.
 ii) Montrer que X et Y sont des variables gaussiennes centrées réduites.

1.38. Soit ξ une variable aléatoire gaussienne réelle d'espérance m et de variance σ^2 et on note par g sa densité. On considère le vecteur aléatoire (X_1, \dots, X_n) de densité

$$f_n(x_1, \dots, x_n) := \left(\prod_{i=1}^n g(x_i) \right) \left(1 + \prod_{j=1}^n (x_j - m)g(x_j) \right), (x_1, \dots, x_n) \in \mathbb{R}^n.$$

i) Calculer

$$\int_{\mathbb{R}} (x - m)g^2(x)dx.$$

- ii) Montrer que f_n est une densité de probabilité.
 iii) On choisit p , $2 \leq p \leq n-1$, composantes du vecteur (X_1, \dots, X_n) , par exemple les premières (X_1, \dots, X_p) . Montrer que la densité f_p de (X_1, \dots, X_p) satisfait :

$$f_p(x_1, \dots, x_p) = g(x_1) \dots g(x_p).$$

iv) En déduire que les variables X_1, \dots, X_p sont indépendantes et gaussiennes.

1.39. Soit (X, Y, Z) un vecteur gaussien de moyenne et de matrice de covariance :

$$m = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad K = \begin{pmatrix} 6 & -3 & -3 \\ -3 & 6 & -3 \\ -3 & -3 & 6 \end{pmatrix};$$

(X, Y, Z) a-t-il une densité ? Quelle est la densité de (X, Y) ? Écrire Z en fonction de X et Y .

1.40. Soit $X : \Omega \rightarrow \mathbb{R}^d$ un vecteur aléatoire de loi P_X tel que toutes ses coordonnées X_1, \dots, X_d sont des variables aléatoires réelles de carrés intégrables. Soit $\Phi_X : \mathbb{R}^d \rightarrow \mathbb{C}$ la fonction caractéristique de X :

$$\Phi_X(t) := \mathbb{E} [e^{i\langle t, X \rangle}] = \int_{\mathbb{R}^d} e^{i\langle t, x \rangle} P_X(dx).$$

Montrer que Φ_X est une fonction deux fois continûment différentiable et on a les développements de Taylor à l'origine :

$$\Phi_X(t) = 1 + i\langle t, \mathbb{E}(X) \rangle - \frac{1}{2} t^* \mathbb{E}(XX^*) t + o(|t|^2)$$

et

$$\ln \Phi_X(t) = i\langle t, \mathbb{E}(X) \rangle - \frac{1}{2} t^* K_X t + o(|t|^2),$$

où $K_X = \mathbb{E}(XX^*) - (\mathbb{E}X)(\mathbb{E}X)^*$ est la matrice de covariance de X .

1.41. Soit $X = (X_1, \dots, X_n)^*$ un vecteur gaussien centré de matrice de covariance l'identité. Soit H le sous-espace vectoriel de $L^2(\Omega)$ engendré par $\{X_1, \dots, X_n\}$, c'est-à-dire $Y \in H$ si et seulement si $\exists t \in \mathbb{R}^n$ tel que $Y = \langle t, X \rangle$. Pour tout Y et Y' de H on note $\langle Y, Y' \rangle_H = \mathbb{E}(YY') = \text{Cov}(Y, Y')$.

i) Montrer que $\langle \cdot, \cdot \rangle_H$ est un produit scalaire sur H .

ii) Soient Z et Z' éléments de H . Montrer que Z et Z' sont indépendantes si et seulement si $\langle Z, Z' \rangle_H = 0$.

iii) Soient Y_1, \dots, Y_p et Z_1, \dots, Z_r , $p + r$ variables aléatoires appartenant à H . Montrer que $(Y_1, \dots, Y_p)^*$ et $(Z_1, \dots, Z_r)^*$ sont deux vecteurs aléatoires indépendants si et seulement si $H_y \subset H_z^\perp$ où H_y et H_z sont les sous-espaces vectoriels de H engendrés par $\{Y_1, \dots, Y_p\}$ et $\{Z_1, \dots, Z_r\}$.

iv) On note $\bar{X} := (X_1 + \dots + X_n)/n$. Montrer que les variables $(X_1 - \bar{X}, \dots, X_n - \bar{X})^*$ et \bar{X} sont indépendantes.

v) En déduire que les variables aléatoires \bar{X} et $W = \max_{1 \leq j \leq n} X_j - \min_{1 \leq j \leq n} X_j$ sont indépendantes.

1.42. Soit $\{X_n : n \geq 1\}$ une suite de vecteurs aléatoires indépendants, de même loi, de carré intégrable et à valeurs dans \mathbb{R}^d . $\{\xi_n : n \geq 1\}$ désigne une suite de variables aléatoires réelles indépendantes, bornées, équidistribuées. On suppose que la suite $\{X_n : n \geq 1\}$ est indépendante de la suite $\{\xi_n : n \geq 1\}$ et que X_1 ou ξ_1 est centrée. On note $Y_n = (\xi_1 X_1 + \dots + \xi_n X_n)/\sqrt{n}$. Montrer que la suite $\{Y_n : n \geq 1\}$ converge en loi lorsque $n \rightarrow \infty$, vers une loi gaussienne que l'on caractérisera.

1.43. Un ordinateur effectue les calculs avec 9 chiffres après la virgule et arrondit tous les nombres à cette précision. On suppose qu'il effectue 10^6 opérations élémentaires et que les erreurs d'arrondi commises à chaque opération sont indépendantes et de loi uniforme sur $[-10^{-9}/2, 10^{-9}/2]$ et que l'erreur sur le résultat final est la somme des erreurs commises à chaque opération. Evaluer la probabilité pour que l'erreur finale soit en valeur absolue, inférieure à $10^{-6}/2$ (on donne $\int_0^{3^{1/2}} (2\pi)^{-1/2} e^{-x^2/2} dx = 0.45837$).

Chapitre 2

Conditionnement

2.1 Espérance conditionnelle

On commence par un rappel sur les espaces de Hilbert. Soit H un espace de Hilbert et F un sous-espace fermé. Pour tout x de H il existe un unique $y \in F$, appelé projection orthogonale de x sur F , vérifiant l'une des conditions équivalentes :

$$\forall z \in F, \quad \langle x - y, z \rangle = 0. \quad (2.1)$$

$$\forall z \in F, \quad \langle x, z \rangle = \langle y, z \rangle. \quad (2.2)$$

$$\forall z \in F, \quad \|x - y\| \leq \|x - z\|. \quad (2.3)$$

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. On supposera que \mathcal{A} est une tribu complète, c'est-à-dire qu'elle contient tous les ensembles \mathbb{P} -négligeables : pour tout A tel que $\mathbb{P}(A) = 0$, alors $A \in \mathcal{A}$. On note par \mathcal{N} l'ensemble des ensembles \mathbb{P} -négligeables.

On rappelle que $L^2(\Omega, \mathcal{A}, \mathbb{P})$ désigne l'ensemble des classes d'équivalence de v.a. : deux v.a. Y et Y' sont égales dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$ si $Y - Y'$ est presque sûrement nulle.

On considère une sous tribu \mathcal{B} de \mathcal{A} . Si $\mathcal{N} \in \mathcal{B}$, alors l'espace $L^2(\Omega, \mathcal{B}, \mathbb{P})$ est inclus dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$. Dans le cas général, on pose $\mathcal{B}' = \mathcal{B} \cup \mathcal{N}$, par définition $\mathcal{B}' \subset \mathcal{A}$. On vérifie facilement que \mathcal{B}' est une tribu. \mathcal{B}' est la plus petite tribu complète contenant \mathcal{B} et incluse dans \mathcal{A} . De plus $L^2(\Omega, \mathcal{B}', \mathbb{P}) \subset L^2(\Omega, \mathcal{A}, \mathbb{P})$. Par abus, on identifiera $L^2(\Omega, \mathcal{B}, \mathbb{P})$ et $L^2(\Omega, \mathcal{B}', \mathbb{P})$.

Proposition 2.1 *Pour toute v.a. X de $L^2(\Omega, \mathcal{A}, \mathbb{P})$, il existe une unique (classe de) v.a. Y telle que :*

$$Y \in L^2(\Omega, \mathcal{B}, \mathbb{P}), \quad (2.4)$$

$$\mathbb{E}[XZ] = \mathbb{E}[YZ], \text{ pour tout } Z \in L^2(\Omega, \mathcal{B}, \mathbb{P}). \quad (2.5)$$

Notation.

On note $Y = \mathbb{E}(X|\mathcal{B})$ ou encore $Y = \mathbb{E}^{\mathcal{B}}(X)$. Donc si $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$, $\mathbb{E}(X|\mathcal{B})$ est caractérisé par les deux propriétés :

$$\mathbb{E}(X|\mathcal{B}) \in L^2(\Omega, \mathcal{B}, \mathbb{P}), \quad (2.6)$$

$$\mathbb{E}[Z\mathbb{E}(X|\mathcal{B})] = \mathbb{E}[ZX]; \quad \forall Z \in L^2(\Omega, \mathcal{B}, \mathbb{P}). \quad (2.7)$$

Démonstration : (Proposition 2.1)

On choisit $H = L^2(\Omega, \mathcal{A}, \mathbb{P})$ et $F = L^2(\Omega, \mathcal{B}, \mathbb{P})$. On utilise le rappel précédent, les propriétés (2.2) et (2.5) étant équivalentes. \square

Exemples :

Exemple 1 : On commence par l'exemple le plus simple : $\mathcal{B} = \{\emptyset, \Omega\}$. Les v.a. \mathcal{B} -mesurables sont les constantes, donc $\mathbb{E}(X|\mathcal{B}) = c$. Si Z est \mathcal{B} -mesurable, $Z = a$, la propriété (2.7) devient :

$$\mathbb{E}[Z \mathbb{E}(X|\mathcal{B})] = \mathbb{E}[ac] = ac = \mathbb{E}[aX] = a\mathbb{E}(X), \quad \forall a \in \mathbb{R}.$$

Par conséquent $c = \mathbb{E}(X)$. En conclusion : $\mathbb{E}(X) = \mathbb{E}(X|\{\emptyset, \Omega\})$.

Exemple 2 : Soit $\{A_1, A_2, \dots, A_n\}$ une partition de Ω , telle que

$$A_i \in \mathcal{A}, \text{ et } 0 < \mathbb{P}(A_i) < 1, \text{ pour tout } 1 \leq i \leq n.$$

On note \mathcal{B} la tribu engendrée par A_1, A_2, \dots, A_n : $\mathcal{B} = \sigma(A_1, A_2, \dots, A_n)$. Puisque les A_i forment une partition, on a l'équivalence :

$$B \in \mathcal{B} \iff B = \emptyset \text{ ou } B = \cup_{j=1}^k A_{i_j}. \quad (2.8)$$

Afin de calculer l'espérance conditionnelle de X sachant \mathcal{B} , il est nécessaire auparavant de caractériser les v.a. \mathcal{B} -mesurables. Le résultat est le suivant :

$$\text{Une v.a. } X \text{ est } \mathcal{B} \text{ - mesurable } \iff X \text{ est constante sur chaque } A_i. \quad (2.9)$$

Il est clair que la condition est suffisante. Il s'agit de montrer le caractère nécessaire : supposons X est \mathcal{B} -mesurable. A_i étant non vide ($\mathbb{P}(A_i) > 0$), on choisit $\omega_i \in A_i$. On note $x_i = X(\omega_i)$. Mais $\{X = x_i\}$ appartient à \mathcal{B} , et a une intersection non vide avec A_i , d'après (2.8), $\{X = x_i\}$ contient A_i . Par conséquent X vaut x_i sur A_i .

On peut à présent calculer $\mathbb{E}(X|\mathcal{B})$. D'après ce qui précède, $\mathbb{E}(X|\mathcal{B}) = \sum_{i=1}^n x_i \mathbf{1}_{A_i}$. On prend une v.a. test Z , également \mathcal{B} -mesurable :

$$Z = \sum_{i=1}^n z_i \mathbf{1}_{A_i}.$$

On a :

$$\mathbb{E}[Z \mathbb{E}(X|\mathcal{B})] = \mathbb{E} \left[\sum_{i=1}^n x_i z_i \mathbf{1}_{A_i} \right] = \sum_{i=1}^n x_i z_i \mathbb{P}(A_i), \quad (2.10)$$

et

$$\mathbb{E}[ZX] = \mathbb{E} \left[\sum_{i=1}^n z_i X \mathbf{1}_{A_i} \right] = \sum_{i=1}^n z_i \mathbb{E}(X \mathbf{1}_{A_i}). \quad (2.11)$$

L'égalité entre (2.10) et (2.11) a lieu pour tout z_1, z_2, \dots, z_n , si et seulement si :

$$x_i = \frac{1}{\mathbb{P}(A_i)} \mathbb{E}(X \mathbf{1}_{A_i}) = \mathbb{E}(X|A_i), \quad 1 \leq i \leq n.$$

En conclusion,

$$\mathbb{E}(X|\mathcal{B}) = \sum_{i=1}^n \mathbb{E}(X|A_i)\mathbf{1}_{A_i}. \quad (2.12)$$

Proposition 2.2 1) *L'opérateur d'espérance conditionnelle est linéaire de $L^2(\Omega, \mathcal{A}, \mathbb{P})$, à valeurs dans $L^2(\Omega, \mathcal{B}, \mathbb{P})$:*

$$\mathbb{E}(aX + bY|\mathcal{B}) = a\mathbb{E}(X|\mathcal{B}) + b\mathbb{E}(Y|\mathcal{B}) ; \forall (a, b) \in \mathbb{R}^2. \quad (2.13)$$

2) *Si X appartient à $L^2(\Omega, \mathcal{A}, \mathbb{P})$ et $X \geq 0$ alors $\mathbb{E}(X|\mathcal{B}) \geq 0$ p.s.*

Démonstration : 1) Puisque la projection orthogonale sur un sous espace fermé d'un espace de Hilbert est un opérateur linéaire, il est clair que $X \rightarrow \mathbb{E}(X|\mathcal{B})$ est linéaire.

2) Supposons $X \geq 0$. On pose $A = \{\mathbb{E}(X|\mathcal{B}) < 0\}$. Puisque $\mathbb{E}(X|\mathcal{B})$ est \mathcal{B} -mesurable, alors $A \in \mathcal{B}$. On choisit $Z = \mathbf{1}_A$ et on applique (2.5) :

$$\mathbb{E}(X\mathbf{1}_A) = \mathbb{E}(\mathbb{E}(X|\mathcal{B})\mathbf{1}_A).$$

Mais

$$X\mathbf{1}_A \geq 0 \text{ et } \mathbb{E}(X|\mathcal{B})\mathbf{1}_A \leq 0 \implies \mathbb{E}(X\mathbf{1}_A) \geq 0 \text{ et } \mathbb{E}(\mathbb{E}(X|\mathcal{B})\mathbf{1}_A) \leq 0.$$

Par conséquent $\mathbb{E}(\mathbb{E}(X|\mathcal{B})\mathbf{1}_A) = 0$.

Rappelons que si U est une v.a. de signe constant, d'espérance nulle alors U est presque sûrement nulle. Donc $\mathbb{E}(X|\mathcal{B})\mathbf{1}_A = 0$ p.s. Mais $\mathbb{E}(X|\mathcal{B}) < 0$ sur A . Ainsi $\mathbb{P}(A) = 0$. \square

Remarque.

Si X et Y sont deux v.a. de $L^2(\Omega, \mathcal{A}, \mathbb{P})$ telles que $X \geq Y$, alors

$$\mathbb{E}(X|\mathcal{B}) \geq \mathbb{E}(Y|\mathcal{B}). \quad (2.14)$$

Il suffit en effet de remarquer que $X - Y \geq 0$. Ainsi

$$\mathbb{E}(X - Y|\mathcal{B}) = \mathbb{E}(X|\mathcal{B}) - \mathbb{E}(Y|\mathcal{B}) \geq 0.$$

Le cadre des espaces L^2 est très agréable pour définir l'espérance conditionnelle. Il est toutefois trop restrictif.

Théorème 2.1 *Soit X une v.a. \mathcal{A} -mesurable.*

1) *Si X est intégrable, il existe une unique v.a. (à une égalité presque sûre près) intégrable, notée $\mathbb{E}(X|\mathcal{B})$ telle que*

$$\mathbb{E}(X|\mathcal{B}) \text{ est } \mathcal{B}\text{-mesurable.} \quad (2.15)$$

$$\mathbb{E}(ZX) = \mathbb{E}(Z\mathbb{E}(X|\mathcal{B})), \quad (2.16)$$

pour toute v.a. Z , \mathcal{B} -mesurable bornée.

2) *Si X est positive, il existe une "unique" v.a. $\mathbb{E}(X|\mathcal{B}) \geq 0$ vérifiant (2.15) et (2.16) pour toute v.a. Z , \mathcal{B} -mesurable, positive.*

Démonstration : a) On commence par montrer l'existence de $Y = \mathbb{E}(X|\mathcal{B})$, lorsque $X \geq 0$. On se ramène au cadre L^2 , en tronquant X , on pose,

$$X_n = X \wedge n = \inf\{X, n\} ; n \geq 1.$$

X_n étant une v.a. bornée, X_n appartient à $L^2(\Omega, \mathcal{A}, \mathbb{P})$, on peut définir :

$$Y_n = \mathbb{E}(X_n|\mathcal{B}) ; n \geq 1.$$

Sachant que la suite X_n est croissante et positive, on déduit de (2.14) que Y_n est également croissante et positive. En particulier $(Y_n)_{n \geq 1}$ converge presque sûrement vers Y , Y est à valeurs dans $[0, +\infty]$.

Mais comme Y_n est \mathcal{B} -mesurable, Y l'est aussi ; Y vérifie ainsi (2.15).

On vérifie à présent que Y satisfait (2.16), lorsque Z est une v.a. positive, bornée, \mathcal{B} -mesurable. Les deux suites $(ZY_n ; n \geq 0)$ et $(ZX_n ; n \geq 0)$ sont positives et croissantes. D'après le théorème de convergence croissante,

$$\lim_{n \rightarrow \infty} \mathbb{E}(ZY_n) = \mathbb{E}(ZY) ; \lim_{n \rightarrow \infty} \mathbb{E}(ZX_n) = \mathbb{E}(ZX).$$

Mais $\mathbb{E}(ZY_n) = \mathbb{E}(Z \mathbb{E}(X_n|\mathcal{B})) = \mathbb{E}(ZX_n)$, $n \geq 1$. On passe à la limite, $n \rightarrow \infty$: $\mathbb{E}(ZX) = \mathbb{E}(ZY)$.

Supposons de plus X intégrable. On prend $Z = 1$, d'après (2.16) on a :

$$\mathbb{E}[\mathbb{E}(X|\mathcal{B})] = \mathbb{E}(X) < \infty, \text{ si } X \geq 0 \text{ et } X \in L^1(\Omega, \mathcal{A}, \mathbb{P}). \quad (2.17)$$

b) On montre que $\mathbb{E}(X|\mathcal{B})$ existe lorsque X est intégrable. On va se ramener au cas précédent, pour ce faire on écrit :

$$X = X_+ - X_-$$

où $x_+ = \sup(x, 0)$, et $x_- = \sup(0, -x)$. Puisque X est intégrable, X_+ et X_- le sont aussi. De plus $X_+ \geq 0$ et $X_- \geq 0$. On peut appliquer le résultat de l'étape précédente et (2.17) : il existe deux v.a. Y_1 et Y_2 , positives, \mathcal{B} -mesurables, intégrables telles que

$$\mathbb{E}(ZY_1) = \mathbb{E}(ZX_+), \quad \mathbb{E}(ZY_2) = \mathbb{E}(ZX_-),$$

pour toute v.a. Z positive, bornée, \mathcal{B} -mesurable. On pose $Y = Y_1 - Y_2$. Y est intégrable. Il est aisé de vérifier, grâce à la propriété de linéarité (2.13) que l'on a :

$$\begin{aligned} \mathbb{E}(ZX) &= \mathbb{E}(Z(X_+ - X_-)) = \mathbb{E}(ZX_+) - \mathbb{E}(ZX_-) = \mathbb{E}(ZY_1) - \mathbb{E}(ZY_2) \\ &= \mathbb{E}(Z(Y_1 - Y_2)) = \mathbb{E}(ZY). \end{aligned}$$

c) On établit l'unicité. On commence par considérer le cas où $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$. Supposons qu'il existe une v.a. $Y' \in L^1(\Omega, \mathcal{B}, \mathbb{P})$ vérifiant :

$$\mathbb{E}(Y'Z) = \mathbb{E}(XZ), \quad (2.18)$$

pour toute v.a. Z , \mathcal{B} -mesurable, bornée. On en déduit : $\mathbb{E}(UZ) = 0$, avec $U = Y - Y'$. Soient $a < 0$ et $Z = \mathbf{1}_{\{U < a\}}$. Il est clair que Z est une v.a. \mathcal{B} -mesurable, bornée. Donc

$$0 = \mathbb{E}[UZ] = \mathbb{E}[U\mathbf{1}_{\{U < a\}}] \leq a\mathbb{P}(U < a) \leq 0.$$

On en déduit : $a\mathbb{P}(U < a) = 0$ implique $\mathbb{P}(U < a) = 0$.

On choisit $a = -\frac{1}{n}$, la suite d'ensembles $\{U < -\frac{1}{n}\}$ étant croissante,

$$\lim_{n \rightarrow \infty} \mathbb{P}(U < -\frac{1}{n}) = \mathbb{P}(U < 0).$$

Sachant que $\mathbb{P}(U < -\frac{1}{n}) = 0$, on a montré : $\mathbb{P}(U < 0) = 0$.

Mais $(-U)$ vérifie la même propriété : $\mathbb{E}[(-U)Z] = 0$, donc $\mathbb{P}(-U < 0) = \mathbb{P}(U > 0) = 0$. En conclusion U est presque sûrement nulle, ce qui signifie que $Y = Y'$ p.s.

Le cas où $X \geq 0$, se traite d'une manière analogue.

On part de l'égalité : $\mathbb{E}(YZ) = \mathbb{E}(Y'Z)$ et on choisit $Z = \mathbf{1}_{\{Y \leq a < b < Y'\}}$, on montre $\mathbb{P}(Y < Y') = 0$, puis $\mathbb{P}(Y > Y') = 0$; soit finalement : $Y = Y'$ p.s. \square

Remarque.

a) Lorsque $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$; il est équivalent de dire :

(2.16) a lieu pour toute v.a. Z , \mathcal{B} -mesurable bornée,

(2.16) a lieu pour toute v.a. Z , \mathcal{B} -mesurable bornée et positive.

b) lorsque $X \geq 0$ ou $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, les deux assertions suivantes sont équivalentes :

(2.16) a lieu pour toute v.a. Z , \mathcal{B} -mesurable bornée et positive.

(2.16) a lieu pour toute v.a. $Z = \mathbf{1}_A$, A appartenant à \mathcal{B} .

En ce qui concerne la première équivalence, il suffit de décomposer Z sous la forme : $Z = Z_+ - Z_-$. Pour la seconde on utilise le résultat d'approximation suivant : si Z est une v.a. \mathcal{B} -mesurable, il existe une suite (Z_n) de v.a. étagées qui converge presque sûrement vers Z (une v.a. T est dite étagée si $T = \sum_{i=1}^n t_i \mathbf{1}_{A_i}$ avec $A_i \in \mathcal{B}$).

Cas particulier : σ -algèbre engendrée par une v.a.

Comme l'a montré le cas où $\mathcal{B} = \sigma(A_1, A_2, \dots, A_n)$ il est parfois difficile de caractériser les v.a. \mathcal{B} -mesurables, et donc de calculer une espérance conditionnelle sachant \mathcal{B} . Cette tâche est facilitée lorsque \mathcal{B} est engendrée par une v.a. T . Soit T une application mesurable $T : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E})$. La σ -algèbre engendrée par T , notée $\sigma(T)$ est définie de la manière suivante :

$$\sigma(T) = \{A \in \mathcal{A} ; \exists C \in \mathcal{E}, A = T^{-1}(C)\}. \quad (2.19)$$

Soit X une application mesurable $X : (\Omega, \mathcal{A}) \rightarrow (F, \mathcal{F})$. On dit que X est $\sigma(T)$ -mesurable si $X^{-1}(C) \in \sigma(T)$ pour tout $C \in \mathcal{F}$. Ces applications sont caractérisées de la manière suivante :

$$X \text{ est } \sigma(T)\text{-mesurable} \iff \text{il existe } f : (E, \mathcal{E}) \rightarrow (F, \mathcal{F}) \text{ mesurable, telle que } X = f \circ T. \quad (2.20)$$

Pour les exemples pratiques : $E = \mathbb{R}^d$ et $F = \mathbb{R}$.

Revenons à la définition de $\mathbb{E}(X|\mathcal{B})$, lorsque $\mathcal{B} = \sigma(T)$. On note

$$\mathbb{E}(X|T) = \mathbb{E}(X|\sigma(T)). \quad (2.21)$$

La condition (2.15) est aisée : on cherche une fonction mesurable $f : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que

$$\mathbb{E}(X|T) = f(T) \quad (2.22)$$

$$\mathbb{E}[g(T)X] = \mathbb{E}[g(T)f(T)], \quad (2.23)$$

pour toute fonction mesurable $g : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, bornée positive. On peut se restreindre (voir remarque) à $g = \mathbf{1}_C$, $C \in \mathcal{E}$

$$\mathbb{E}[X\mathbf{1}_{\{T \in C\}}] = \mathbb{E}[f(T)\mathbf{1}_{\{T \in C\}}] ; \forall C \in \mathcal{E}. \quad (2.24)$$

Examinons un exemple, celui où T prend un nombre fini de valeurs : t_1, t_2, \dots, t_n . On supposera que les t_i sont 2 à 2 distincts et $\mathbb{P}(T = t_i) > 0$. Alors $\mathcal{B} = \sigma(T) = \sigma(A_1, A_2, \dots, A_n)$ où l'on a posé $A_i = \{T = t_i\}$. On a vu :

$$\mathbb{E}(X|\mathcal{B}) = \sum_{i=1}^n \mathbb{E}(X|T = t_i)\mathbf{1}_{\{T=t_i\}}.$$

Posons

$$f(t) = \mathbb{E}(X|T = t), \quad t \in \{t_1, t_2, \dots, t_n\}.$$

Alors $\mathbb{E}(X|\mathcal{B}) = f(T)$.

On remarque sur cet exemple que f n'est pas unique, f est uniquement déterminée sur $\{t_1, t_2, \dots, t_n\}$. Rappelons que

$$\mathbb{E}(X|T = t) = \frac{1}{\mathbb{P}(T = t)} \mathbb{E}(X\mathbf{1}_{\{T=t\}}).$$

Cette quantité étant bien définie si $\mathbb{P}(T = t) > 0$, condition satisfaite si $t \in \{t_1, t_2, \dots, t_n\}$. Lorsque T est une v.a. quelconque, on a : $\mathbb{E}(X|T) = f(T)$. On peut noter $f(t) = \mathbb{E}(X|T = t)$, mais il s'agit d'une convention d'écriture, et non d'un conditionnement par l'événement $\{T = t\}$, qui peut être de probabilité nulle. Il est conseillé de se tenir à la notation (2.22).

Exemples :

Exemple 3) Soient X et Y deux v.a., X à valeurs réelles, X intégrable ou positive. On suppose que X et Y sont deux v.a. indépendantes. Alors $\mathbb{E}(X|Y) = \mathbb{E}(X)$.

Démonstration : Il s'agit de montrer :

$$\mathbb{E}(Xg(Y)) = \mathbb{E}(X)\mathbb{E}[g(Y)],$$

pour toute fonction g borélienne, positive et bornée. Mais X et Y sont deux v.a. indépendantes, cette égalité est réalisée. \square

Exemple 4) Soit (X, Y) un vecteur gaussien centré à valeurs dans \mathbb{R}^2 . On suppose que Y est non dégénérée, i.e. Y est presque sûrement non nulle. Alors $\mathbb{E}(X|Y) = aY$.

Démonstration : On procède comme dans le cadre des espaces euclidiens, en remplaçant l'orthogonalité par l'indépendance. On cherche un réel a tel que $X' = X - aY$ soit une v.a. indépendante de Y . Puisque (X, Y) est un vecteur gaussien, $(X', Y) = (X - aY, Y)$ l'est aussi. Par conséquent les v.a. X' et Y sont indépendantes si et seulement si :

$$\text{Cov}(X', Y) = \mathbb{E}(X'Y) = 0.$$

Mais

$$\mathbb{E}(X'Y) = \mathbb{E}(XY) - a\mathbb{E}(Y^2) = \mathbb{E}(XY) - a \text{Var}Y.$$

Sachant que $\text{Var}Y > 0$, a est unique et

$$a = \frac{\mathbb{E}(XY)}{\text{Var}Y},$$

a étant ainsi fixé, d'après la linéarité de l'espérance conditionnelle et le résultat de l'exemple 3, on a,

$$\mathbb{E}(X'|Y) = \mathbb{E}(X|Y) - a\mathbb{E}(Y|Y) = \mathbb{E}(X|Y) - aY = 0.$$

D'où

$$\mathbb{E}(X|Y) = aY = \frac{\mathbb{E}(XY)}{\text{Var}Y} Y. \quad (2.25)$$

□

L'application $X \rightarrow \mathbb{E}(X|Y)$ est une application de projection orthogonale sur le sous-espace vectoriel de dimension 1 engendrée par Y . Cet exercice sera repris dans un cadre plus général.

Proposition 2.3 Soit \mathcal{B} une sous-tribu de \mathcal{A} .

1) Soient X et Y deux variables aléatoires intégrables ou positives, alors

$$\mathbb{E}(X + Y|\mathcal{B}) = \mathbb{E}(X|\mathcal{B}) + \mathbb{E}(Y|\mathcal{B}).$$

2) Pour tout $a \in \mathbb{R}$ si X est intégrable et pour $a \geq 0$ si X est positif, $\mathbb{E}(aX|\mathcal{B}) = a\mathbb{E}(X|\mathcal{B})$.

3) Si X est \mathcal{B} -mesurable, alors $\mathbb{E}(XY|\mathcal{B}) = X\mathbb{E}(Y|\mathcal{B})$ lorsque $X \geq 0$ et $Y \geq 0$ ou Y et XY intégrables.

4) Si $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ alors $\mathbb{E}(X|\mathcal{B}) \in L^2(\Omega, \mathcal{B}, \mathbb{P})$.

5) Si \mathcal{B}_1 et \mathcal{B}_2 sont deux sous-tribus de \mathcal{A} , $\mathcal{B}_1 \subset \mathcal{B}_2$ alors

$$\mathbb{E}(\mathbb{E}(X|\mathcal{B}_2)|\mathcal{B}_1) = \mathbb{E}(X|\mathcal{B}_1), \quad (2.26)$$

pour $X \geq 0$ ou intégrable.

6) Si $X \leq Y$ p.s. alors $\mathbb{E}(X|\mathcal{B}) \leq \mathbb{E}(Y|\mathcal{B})$, lorsque $X \geq 0$ ou X et Y intégrables.

7) $|\mathbb{E}(X|\mathcal{B})| \leq \mathbb{E}(|X||\mathcal{B})$, X positive ou intégrable.

Démonstration : a) Montrons 1) lorsque X et Y sont intégrables. Soit $U = \mathbb{E}(X|\mathcal{B}) + \mathbb{E}(Y|\mathcal{B})$. U est une v.a. intégrable et \mathcal{B} -mesurable. Il s'agit de montrer,

$$\mathbb{E}[UZ] = \mathbb{E}[(X + Y)Z],$$

pour toute v.a. Z , \mathcal{B} -mesurable, bornée. Mais

$$\mathbb{E}((X + Y)Z) = \mathbb{E}(XZ) + \mathbb{E}(YZ) = \mathbb{E}[\mathbb{E}(X|\mathcal{B})Z] + \mathbb{E}[\mathbb{E}(Y|\mathcal{B})Z] = \mathbb{E}[UZ].$$

Le 2) se montre d'une manière analogue.

b) Montrons 3), lorsque cette fois $X \geq 0$ et $Y \geq 0$. On pose $V = X\mathbb{E}(Y|\mathcal{B})$. V est une v.a. positive, \mathcal{B} -mesurable. On considère une v.a. Z , \mathcal{B} -mesurable, positive et bornée. XZ étant \mathcal{B} -mesurable, on a :

$$\mathbb{E}[Z(XY)] = \mathbb{E}[(ZX)Y] = \mathbb{E}[ZX\mathbb{E}(Y|\mathcal{B})] = \mathbb{E}[ZV].$$

On en déduit 3).

c) La propriété 4) est une conséquence immédiate de la proposition 2.1.

d) La démonstration de 5) est aisée, plaçons-nous dans le cas positif. Comme dans b), Z désigne une v.a. test positive, \mathcal{B}_1 -mesurable. Posons $W = \mathbb{E}(X|\mathcal{B}_1)$. Par définition W est \mathcal{B}_1 -mesurable. De plus, Z étant \mathcal{B}_1 -mesurable est \mathcal{B}_2 -mesurable, donc,

$$\mathbb{E}[Z\mathbb{E}(X|\mathcal{B}_2)] = \mathbb{E}(ZX) = \mathbb{E}(Z\mathbb{E}(Y|\mathcal{B}_1)) = \mathbb{E}(ZW).$$

e) On peut écrire $Y = X + T$ avec T v.a. positive. D'après la propriété 1), $\mathbb{E}(Y|\mathcal{B}) = \mathbb{E}(X|\mathcal{B}) + \mathbb{E}(T|\mathcal{B})$, mais $T \geq 0$ implique $\mathbb{E}(T|\mathcal{B}) \geq 0$.

f) Si $X \geq 0$, $\mathbb{E}(X|\mathcal{B}) \geq 0$ et $\mathbb{E}(|X||\mathcal{B}) = \mathbb{E}(X|\mathcal{B})$, l'inégalité est une égalité. Supposons X élément de $L^1(\Omega, \mathcal{A}, \mathbb{P})$. En utilisant successivement l'inégalité : $-|X| \leq X \leq |X|$ et 6) on a : $-\mathbb{E}(|X||\mathcal{B}) \leq \mathbb{E}(X|\mathcal{B}) \leq \mathbb{E}(|X||\mathcal{B})$. \square

Proposition 2.4 Soient \mathcal{B}_1 et \mathcal{B}_2 deux sous-tribus de \mathcal{A} . Alors \mathcal{B}_1 et \mathcal{B}_2 sont indépendantes si et seulement si pour toute v.a. X , \mathcal{B}_2 -mesurable et bornée, $\mathbb{E}(X|\mathcal{B}_1) = \mathbb{E}(X)$.

Démonstration : On rappelle que deux tribus \mathcal{B}_1 et \mathcal{B}_2 sont indépendantes si et seulement si deux événements quelconques B_1 de \mathcal{B}_1 et B_2 de \mathcal{B}_2 sont indépendants, i.e. :

$$\mathbb{P}(B_1 \cap B_2) = \mathbb{P}(B_1)\mathbb{P}(B_2).$$

1) Supposons les deux tribus \mathcal{B}_1 et \mathcal{B}_2 indépendantes. Soit Z une v.a. bornée, \mathcal{B}_1 -mesurable. On a : $\mathbb{E}(XZ) = \mathbb{E}(X)\mathbb{E}(Z)$, d'où $\mathbb{E}(X|\mathcal{B}_1) = \mathbb{E}(X)$.

2) Réciproquement, on choisit $Z = \mathbf{1}_{B_1}$, $X = \mathbf{1}_{B_2}$ avec $B_1 \in \mathcal{B}_1$ et $B_2 \in \mathcal{B}_2$. Alors

$$\mathbb{E}(XZ) = \mathbb{P}(B_1 \cap B_2) = \mathbb{E}(X)\mathbb{E}(Z) = \mathbb{P}(B_2)\mathbb{P}(B_1).$$

Les deux tribus \mathcal{B}_1 et \mathcal{B}_2 sont indépendantes. \square

Remarques.

a) Si $\mathcal{B}_2 = \sigma(Y)$ et $\mathcal{B}_1 = \sigma(X)$, l'exemple apparaît comme un cas particulier de la proposition 2.4.

b) Soit Y une v.a. fixée. On sait que les deux v.a. X et Y sont indépendantes si et seulement si les tribus $\sigma(X)$ et $\sigma(Y)$ sont indépendantes. La proposition 2.4 affirme que X et Y sont indépendantes si et seulement si :

$$\mathbb{E}[f(X)|Y] = \mathbb{E}[f(X)], \quad (2.27)$$

pour toute fonction f borélienne, bornée et positive. En particulier $\mathbb{E}(X|Y) = \mathbb{E}(X)$. Mais cette relation n'implique pas, en général (2.27), sauf dans le cas gaussien !

2.2 Loi conditionnelle

Afin d'introduire cette nouvelle notion nous commençons par deux exemples :

Exemples :

Exemple 5) Ce premier exemple est relatif aux lois discrètes, il ne nécessite aucune théorie, tous les calculs peuvent être faits "à la main". Soient X et Y deux v.a. indépendantes, de loi Poisson de paramètre respectivement λ et μ . On pose $S = X + Y$. On veut calculer la loi de X sachant S . Il s'agit d'évaluer :

$$\mathbb{P}(X = n|S = m) = \frac{\mathbb{P}(X = n, X + Y = m)}{\mathbb{P}(X + Y = m)} = \frac{\mathbb{P}(X = n, Y = m - n)}{\mathbb{P}(X + Y = m)},$$

où $m \geq n \geq 0$. Mais $X + Y$ suit une loi de Poisson de paramètre $\lambda + \mu$, de plus X et Y sont indépendantes, d'où :

$$\begin{aligned} \mathbb{P}(X = n|S = m) &= \frac{\mathbb{P}(X = n)\mathbb{P}(Y = m - n)}{\mathbb{P}(X + Y = m)} = \frac{\lambda^n e^{-\lambda} \mu^{m-n} e^{-\mu}}{(\lambda + \mu)^m e^{-(\lambda + \mu)}} \\ &= C_m^n \rho^n (1 - \rho)^{m-n}, \end{aligned}$$

avec $\rho = \frac{\lambda}{\lambda + \mu}$. Par conséquent, conditionnellement à $\{S = m\}$, X suit une loi binomiale $\mathcal{B}(m, \rho)$.

Exemple 6) Le second exemple a trait aux lois continues. On se donne deux v.a. indépendantes X et Y , de loi exponentielle de paramètre 1. On pose $S = X + Y$. On va calculer $H(S) = \mathbb{E}(h(X)|S)$ pour toute fonction h borélienne bornée. Soient g une fonction borélienne bornée et $\Delta = \mathbb{E}(h(X)g(S))$. La fonction H est caractérisée par : $\Delta = \mathbb{E}[H(S)g(S)]$. Calculons Δ . On a :

$$\Delta = \mathbb{E}[h(X)g(X + Y)] = \iint_{x>0, y>0} h(x)g(x + y)e^{-(x+y)} dx dy.$$

x étant fixé, on fait le changement de variable $u = x + y$, il vient :

$$\Delta = \iint_{x>0, u \geq x} h(x)g(u)e^{-u} dx du = \int_0^\infty g(u)e^{-u} \left(\int_0^u h(x) dx \right) du.$$

On choisit $h = 1$. Alors $\Delta = \int_0^\infty g(u)ue^{-u}du$. Par conséquent la v.a. S admet pour densité $\mathbb{1}_{\{u>0\}}ue^{-u}$. On pose : $H(u) = \frac{1}{u} \int_0^u h(x)dx$, alors

$$\Delta = \int_0^\infty g(u)H(u)ue^{-u}du = \mathbb{E}[H(S)g(S)].$$

Mais $U_u(dx) = \frac{1}{u}\mathbb{1}_{[0,u]}(x)dx$ est la loi uniforme sur $[0, u]$. Nous dirons que conditionnellement à $S = u$, X suit la loi uniforme sur $[0, u]$,

$$\mathbb{E}[h(X)g(S)] = \int_0^\infty g(u) \left(\int_0^u h(x)U_u(dx) \right) \mu(du), \quad \forall f \geq 0, \quad \forall g \geq 0, \quad (2.28)$$

où $\mu(du) = \mathbb{1}_{\{u>0\}}ue^{-u}du$ est la loi de S .

Définition 2.1 Soient (E, \mathcal{E}) et (F, \mathcal{F}) deux espaces mesurables. On appelle noyau (positif) une application N définie sur $E \times \mathcal{F}$ à valeurs dans \mathbb{R}_+ telle que :

(i) pour tout $x \in E$, $A \in \mathcal{F} \rightarrow N(x, A)$ est une mesure positive sur (F, \mathcal{F}) ,

(ii) pour tout $A \in \mathcal{F}$, $x \rightarrow N(x, A)$ est mesurable de (E, \mathcal{E}) sur $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$.

On dit que le noyau N est une probabilité de transition de E vers F si pour tout x , $N(x, \cdot)$ est une probabilité. Lorsque $E = F$, N est appelé probabilité de transition sur E .

Remarques.

1) Une probabilité de transition est une famille "mesurable" de probabilités sur (F, \mathcal{F}) indexée par $E : (N(x, \cdot) ; x \in E)$.

2) Si g est une application mesurable définie sur F , positive alors

$$x \in E \rightarrow N(x, g) = \int f(y)N(x, dy) \text{ est } \mathcal{E}\text{-mesurable.} \quad (2.29)$$

3) Pour l'exemple 5, on a $E = F = \mathbb{N}$, $N(k, \cdot)$ est la loi binomiale $\mathcal{B}(k, \rho)$:

$$N(k, \cdot) = \sum_{i=0}^k C_k^i \rho^i (1 - \rho)^{k-i} \delta_i,$$

δ_i désignant la mesure de Dirac en i .

Quant à l'exemple 6, $E = F = \mathbb{R}_+$ et

$$N(u, dx) = \frac{1}{u}\mathbb{1}_{[0,u]}dx.$$

Exemple 7) :

Soient $f : (E \times F, \mathcal{E} \times \mathcal{F}) \rightarrow \mathbb{R}_+$ mesurable, μ une probabilité sur (F, \mathcal{F}) . On pose

$$N(x, A) = \int_A f(x, y)\mu(dy), \quad x \in E, \quad A \in \mathcal{F}.$$

N est un noyau. Si de plus $\int_F f(x, y)\mu(dy) = 1$, pour tout $x \in E$, N est une probabilité de transition.

Définition 2.2 Soient X et Y deux v.a. à valeurs dans (E, \mathcal{E}) , respectivement (F, \mathcal{F}) . On appelle loi conditionnelle de Y sachant X , une probabilité de transition N , de E vers F telle que :

$$\mathbb{E}(g(Y)|X) = \int g(y)N(X, dy) = N(X, g), \quad (2.30)$$

pour toute fonction borélienne positive g .

Remarques :

1) Si ν est une mesure positive sur (E, \mathcal{E}) , on note indifféremment :

$$\int_E f d\nu = \int_E f(x)\nu(dx) = \int_E f(x)d\nu(x) = \nu(f). \quad (2.31)$$

2) On a vu que $x \rightarrow \int g(y)N(x, dy)$ est mesurable, donc $\int g(y)N(X, dy)$ est une v.a. $\sigma(X)$ -mesurable. Ainsi pour démontrer que N est la loi conditionnelle de Y sachant X , il est nécessaire et suffisant de montrer :

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E} \left[f(X) \int_F g(y)N(x, dy) \right] = \int_E f(x) \left(\int_F g(y)N(x, dy) \right) \mu(dx)$$

où μ désigne la loi de X . C'est exactement la démarche que nous avons adoptée dans l'exemple 6.

3) La loi conditionnelle de Y sachant X n'est pas unique. Soient N_1 et N_2 deux lois conditionnelles de Y sachant X . On a seulement :

$$N_1(x, A) = N_2(x, A) \quad \forall A \in \mathcal{F},$$

pour μ -presque tout x , μ désignant la loi de X .

Dans le cas discret de l'exemple 5, N_1 et N_2 sont déterminés uniquement sur \mathbb{N} .

4) Formellement $N(x, dy)$ est la loi de Y sachant $X = x$. Si X suit une loi discrète, et x appartient au support de la loi de X , alors,

$$N(x, A) = \mathbb{P}(Y \in A | X = x) = \frac{\mathbb{P}(Y \in A, X = x)}{\mathbb{P}(X = x)}.$$

5) On montre que si E et F sont deux espaces métriques séparables et complets, munis de leur tribu borélienne, il existe alors une loi conditionnelle de Y sachant X . Rentrant dans ce cadre les espaces \mathbb{R}^n .

Soient X et Y deux v.a. à valeurs dans E et F respectivement. Si on connaît la loi μ de X et si N décrit la loi conditionnelle de Y sachant X , alors la loi du couple (X, Y) est déterminée. En effet

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)\mathbb{E}[g(Y)|X]] = \mathbb{E} \left[f(X) \int_F g(y)N(X, dy) \right] \quad (2.32)$$

pour toute $f : E \rightarrow \mathbb{R}_+$, $g : F \rightarrow \mathbb{R}_+$, mesurables, bornées.

On va s'intéresser au problème réciproque : calculer la loi conditionnelle de Y sachant X à l'aide de la loi de (X, Y) . Nous allons nous restreindre à $E = \mathbb{R}^n$ et $F = \mathbb{R}^m$, $n \geq 1$ et $m \geq 1$. En pratique, dans de nombreux cas, on rencontre des v.a. à valeurs dans \mathbb{R}^n , et dont la loi soit admet une densité (par rapport à la mesure de Lebesgue), soit est discrète. Lorsque X est discrète on supposera, pour simplifier que X est à valeurs dans \mathbb{Z}^n . Soit λ_d^n la mesure de comptage des points de \mathbb{Z}^n :

$$\lambda_d^n = \sum_{k \in \mathbb{Z}^n} \delta_k. \quad (2.33)$$

λ_d^n est une mesure positive de masse infinie ($\lambda_d^n(\mathbb{Z}^n) = +\infty$) mais σ -finie :

$$\lambda_d^n([-a, a] \times \dots \times [-a, a]) < \infty,$$

pour tout $a \in \mathbb{Z}$. Si X est à valeurs dans \mathbb{Z}^n , la loi de X admet une densité f par rapport à λ_d^n et

$$f(m) = \mathbb{P}(X = m), \quad m \in \mathbb{Z}^n.$$

On note λ_l^n la mesure de Lebesgue sur \mathbb{R}^n .

Proposition 2.5 *Soient X et Y deux v.a. à valeurs dans \mathbb{R}^n , respectivement \mathbb{R}^m . On suppose que la loi de (X, Y) admet une densité $\varphi(x, y)$ par rapport à la mesure $\nu_1 \otimes \nu_2$, où ν_1 (resp. ν_2) vaut λ_d^n ou λ_l^n . (resp. λ_d^m ou λ_l^m). On note :*

$$\alpha(x) = \int_{\mathbb{R}^m} \varphi(x, y) \nu_2(dy) ; \quad x \in \mathbb{R}^n, \quad (2.34)$$

$$N(x, dy) = \frac{\varphi(x, y)}{\alpha(x)} \mathbf{1}_{\{\alpha(x) > 0\}} \nu_2(dy) ; \quad x \in \mathbb{R}^n. \quad (2.35)$$

Alors,

- 1) La loi de X admet α comme densité par rapport à ν_1 .
- 2) La loi conditionnelle de Y sachant X est donnée par la probabilité de transition N .

Démonstration : a) Par définition, si $h : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ est mesurable et bornée, on a,

$$\mathbb{E}(h(X, Y)) = \int_{\mathbb{R}^n \times \mathbb{R}^m} h(x, y) \varphi(x, y) \nu_1(dx) \nu_2(dy). \quad (2.36)$$

En particulier si $h(x, y) = \mathbf{1}_{\{\alpha(x)=0\}}$,

$$\mathbb{P}(\alpha(X) = 0) = \int_{\mathbb{R}^n \times \mathbb{R}^m} \mathbf{1}_{\{\alpha(x)=0\}} \varphi(x, y) \nu_1(dx) \nu_2(dy).$$

On applique le théorème de Fubini :

$$\begin{aligned}\mathbb{P}(\alpha(X) = 0) &= \int_{\mathbb{R}^n} \mathbf{1}_{\{\alpha(x)=0\}} \left(\int_{\mathbb{R}^m} \varphi(x, y) \nu_2(dy) \right) \nu_1(dx) \\ &= \int_{\mathbb{R}^n} \mathbf{1}_{\{\alpha(x)=0\}} \alpha(x) \nu_1(dx) = 0\end{aligned}$$

Donc $\mathbb{P}(\alpha(X) = 0) = 0$. Ce qui signifie que presque sûrement, $\alpha(X) \neq 0$. Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g : \mathbb{R}^m \rightarrow \mathbb{R}$, deux fonctions boréliennes et bornées. On déduit de (2.36) :

$$\begin{aligned}\mathbb{E}[f(X)g(Y)] &= \mathbb{E} [f(X)g(Y)\mathbf{1}_{\{\alpha(X)>0\}}] \\ &= \int_{\mathbb{R}^n \times \mathbb{R}^m} f(x)g(y)\varphi(x, y)\mathbf{1}_{\{\alpha(x)>0\}}\nu_1(dx)\nu_2(dy).\end{aligned}\tag{2.37}$$

En particulier si $g = 1$,

$$\begin{aligned}\mathbb{E}[f(X)] &= \int_{\mathbb{R}^n} f(x) \left(\int_{\mathbb{R}^m} \varphi(x, y) \nu_2(dy) \right) \mathbf{1}_{\{\alpha(x)>0\}} \nu_1(dx) \\ &= \int_{\mathbb{R}^n} f(x) \alpha(x) \mathbf{1}_{\{\alpha(x)>0\}} \nu_1(dx) = \int_{\mathbb{R}^n} f(x) \alpha(x) \nu_1(dx).\end{aligned}$$

Ce qui signifie que la loi de X est $\alpha(x)\nu_1(dx)$. On revient à (2.37) et on applique à nouveau le théorème de Fubini :

$$\begin{aligned}\mathbb{E}[f(X)g(Y)] &= \int_{\mathbb{R}^n} f(x) \mathbf{1}_{\{\alpha(x)>0\}} \alpha(x) \left(\int_{\mathbb{R}^m} g(y) \frac{\varphi(x, y)}{\alpha(x)} \nu_2(dy) \right) \nu_1(dx) \\ &= \int_{\mathbb{R}^n} f(x) \mathbf{1}_{\{\alpha(x)>0\}} \alpha(x) \left(\int_{\mathbb{R}^m} g(y) N(x, dy) \right) \nu_1(dx).\end{aligned}$$

On a montré :

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E} \left(f(X) \int_{\mathbb{R}^m} g(y) N(X, dy) \right).\tag{2.38}$$

b) D'après l'exemple 7, N est bien un noyau. Il reste à montrer que N est une probabilité de transition :

$$\begin{aligned}N(x, \mathbb{R}^m) &= \int_{\mathbb{R}^m} \frac{\varphi(x, y)}{\alpha(x)} \mathbf{1}_{\{\alpha(x)>0\}} \nu_2(dy) = \frac{1}{\alpha(x)} \mathbf{1}_{\{\alpha(x)>0\}} \int_{\mathbb{R}^m} \varphi(x, y) \nu_2(dy) \\ &= \frac{1}{\alpha(x)} \mathbf{1}_{\{\alpha(x)>0\}} \alpha(x) = \mathbf{1}_{\{\alpha(x)>0\}}.\end{aligned}$$

On remarque qu'il faut modifier légèrement la définition de N , en définissant N par la formule (2.35) lorsque $\alpha(x) > 0$ et par une probabilité quelconque sur \mathbb{R}^m , quand $\alpha(x) = 0$, par exemple $N(x, dy) = \gamma(y)\nu_2(dy)$ convient, où $\gamma : \mathbb{R}^m \rightarrow \mathbb{R}_+$ est mesurable et $\int_{\mathbb{R}^m} \gamma(y)\nu_2(dy) = 1$. Alors

$$N(x, dy) = \left\{ \frac{\varphi(x, y)}{\alpha(x)} \mathbf{1}_{\{\alpha(x)>0\}} + \gamma(y) \mathbf{1}_{\{\alpha(x)=0\}} \right\} \nu_2(dy).\tag{2.39}$$

□

Nous allons à présent décrire aussi explicitement que possible le noyau N lorsque X et Y sont discrètes et/ou à densité. Lorsque l'une des v.a. est discrète on supposera pour simplifier qu'elle est à valeurs dans \mathbb{N} .

a) X est discrète

(i) Y est discrète

On note

$$p_{i,j} = \mathbb{P}(X = i, Y = j) ; i \in \mathbb{N}, j \in \mathbb{N}.$$

Alors,

$$\begin{aligned} \nu_1 = \nu_2 = \lambda_d^1 &= \sum_{k \in \mathbb{Z}} \delta_k \quad \text{et} \quad \varphi(i, j) = p_{i,j} \mathbf{1}_{\{i \geq 0, j \geq 0\}}, \\ \alpha(i) &= \int_{\mathbb{N}} \varphi(i, j) \nu_2(dj) = \sum_{j \geq 0} \varphi(i, j) = \sum_{j \geq 0} p_{i,j} = p_{i,\cdot}, \\ N(i, \{j\}) &= \frac{p_{i,j}}{p_{i,\cdot}}. \end{aligned}$$

On retrouve ainsi le résultat classique :

$$\mathbb{P}(Y = j | X = i) = \frac{\mathbb{P}(Y = j, X = i)}{\mathbb{P}(X = i)} = \frac{p_{i,j}}{p_{i,\cdot}}.$$

(ii) Y admet une densité

On suppose que Y est à valeurs dans \mathbb{R}^m . On a :

$$\begin{aligned} \nu_1 = \lambda_d^1 &= \sum_{k \in \mathbb{Z}} \delta_k \quad \text{et} \quad \nu_2(dy) = \lambda_l^m(dy) = dy_1 dy_2 \dots dy_m, \\ \alpha(i) &= \int_{\mathbb{R}^m} \varphi(i, y) dy = \mathbb{P}(X = i) ; i \in \mathbb{N}, \\ N(i, dy) &= \frac{\varphi(i, y)}{\alpha(i)} dy = \mathbb{P}(Y \in dy | X = i) ; i \in \mathbb{N}. \end{aligned}$$

En d'autres termes la probabilité $N(i, \cdot)$ admet $\frac{\varphi(i, y)}{\alpha(i)}$ comme densité (par rapport à la mesure de Lebesgue sur \mathbb{R}^m). De plus

$$N(i, f) = \int_{\mathbb{R}^m} f(y) N(i, dy) = \mathbb{E}[f(Y) | X = i], \quad f \geq 0.$$

Il s'agit d'un conditionnement usuel par l'événement $\{X = i\}$, lorsque celui-ci est de probabilité non nulle.

b) X admet une densité

On suppose que X est à valeurs dans \mathbb{R}^n .

(i) Y est discrète

$$\nu_1(dx) = \lambda_1^n(dx) = dx_1 dx_2 \dots dx_n \quad \text{et} \quad \nu_2 = \lambda_d^1 = \sum_{k \in \mathbb{Z}} \delta_k,$$

$$\alpha(x) = \int_{\mathbb{N}} \varphi(x, j) \nu_2(dj) = \sum_{k \in \mathbb{Z}} \varphi(x, k); \quad x \in \mathbb{R}^n,$$

$$N(x, \{j\}) = \frac{\varphi(x, j)}{\alpha(x)}, \quad j \in \mathbb{N}, x \in \mathbb{R}^n.$$

(ii) Y admet une densité

$$\nu_1(dx) = \lambda_1^n(dx) = dx_1 dx_2 \dots dx_n \quad \text{et} \quad \nu_2(dy) \lambda_1^m(dy) = dy_1 dy_2 \dots dy_m,$$

$$\alpha(x) = \int_{\mathbb{R}^m} \varphi(x, y) dy; \quad x \in \mathbb{R}^n,$$

$$N(x, dy) = \frac{\varphi(x, y)}{\alpha(x)} dy \quad x \in \mathbb{R}^n. \quad (2.40)$$

$N(x, \cdot)$ admet $\frac{\varphi(x, y)}{\alpha(x)}$ comme densité.

2.3 Le cas gaussien

Soit Z un vecteur gaussien à valeurs dans \mathbb{R}^{n+m} . On note X les n premières coordonnées de Z , et Y les m suivantes : $Z = (X, Y)$. On suppose que X admet une densité. Le but de ce paragraphe est de déterminer la loi de Y sachant X .

Rappelons que si $n = m = 1$, nous avons calculé $\mathbb{E}(Y|X)$ (voir exemple 4). Nous allons adapter cette approche au cas multidimensionnel.

Nous commençons par fixer quelques notations.

– K désigne la matrice de covariance de (X, Y) , on écrira :

$$K = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix},$$

- K_{11} est la matrice de covariance de X , carrée et d'ordre n ,
- $K_{12} = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))^*)$ est une matrice $n \times m$ (n lignes, m colonnes),
- $K_{21} = K_{12}^*$ est d'ordre $m \times n$,
- K_{22} est la matrice de covariance de Y , carrée et d'ordre m .

On a supposé que X admet une densité, condition équivalente à :

$$K_{11} \text{ est inversible.} \quad (2.41)$$

Proposition 2.6 1) Il existe une matrice A , d'ordre $m \times n$, et un vecteur gaussien X' , indépendant de X tel que

$$Y = AX + X' \quad (2.42)$$

2) De plus $A = K_{21}K_{11}^{-1}$ et le vecteur gaussien X' est caractérisé par,

$$\mathbb{E}(X') = \mathbb{E}(Y) - A\mathbb{E}(X), \quad K_{X'} = K_{22} - K_{21}K_{11}^{-1}K_{21}^*, \quad (2.43)$$

où $K_{X'}$ désigne la matrice de covariance de X' .

Remarque.

L'affirmation essentielle de la proposition 2.6 est le 1) et en particulier (2.42), on en déduit alors facilement le 2). La méthode développée est une extension de celle que nous avons utilisée en dimension 1.

Démonstration : (Proposition 2.6). **a)** La première étape consiste à montrer que si Z est un vecteur gaussien, $Z_1 = B_1Z$, $Z_2 = B_2Z$ où B_1 et B_2 sont deux matrices, alors les deux v.a. Z_1 et Z_2 sont indépendantes si et seulement si :

$$K_{Z_1, Z_2} = \mathbb{E}[(Z_1 - \mathbb{E}(Z_1))(Z_2 - \mathbb{E}(Z_2))^*] = 0. \quad (2.44)$$

On adopte les conventions du premier chapitre : les v.a. à valeurs multidimensionnelles sont représentées par une matrice unicolonne. Si Z_i est à valeurs dans \mathbb{R}^{n_i} , les deux v.a. Z_1 et Z_2 sont indépendantes si et seulement si :

$$\mathbb{E}[\exp\{i(u_1^*Z_1 + u_2^*Z_2)\}] = \mathbb{E}[\exp(iu_1^*Z_1)]\mathbb{E}[\exp(iu_2^*Z_2)], \quad (2.45)$$

pour tout $u_i \in \mathbb{R}^{n_i}$. Mais (2.45) est équivalente à :

$$u_1^*Z_1 \text{ et } u_2^*Z_2 \text{ sont indépendantes.} \quad (2.46)$$

Il est clair que (2.46) entraîne (2.45), en changeant dans (2.45), u_i en $\lambda_i u_i$, on montre que (2.45) entraîne (2.46). Mais

$$\tilde{Z} = \begin{pmatrix} u_1^*Z_1 \\ u_2^*Z_2 \end{pmatrix} = BZ.$$

Donc \tilde{Z} est un vecteur gaussien, à valeurs dans \mathbb{R}^2 ; on sait que les deux composantes de \tilde{Z} sont indépendantes si et seulement si,

$$\rho = \text{Cov}(u_1^*Z_1, u_2^*Z_2) = 0; \quad \forall u_1, \forall u_2. \quad (2.47)$$

Mais

$$u_1^*Z_1 - \mathbb{E}(u_1^*Z_1) = u_1^*Z_1 - u_1^*\mathbb{E}(Z_1) = u_1^*(Z_1 - \mathbb{E}(Z_1)),$$

$$Z_2^* u_2 - \mathbb{E}(Z_2^* u_2) = Z_2^* u_2 - \mathbb{E}(Z_2^*) u_2 = (Z_2^* - \mathbb{E}(Z_2^*)) u_2.$$

D'où

$$\rho \mathbb{E}[u_1^* (Z_1 - \mathbb{E}(Z_1)) (Z_2^* - \mathbb{E}(Z_2^*)) u_2] = u_1^* \mathbb{E}[(Z_1 - \mathbb{E}(Z_1)) (Z_2^* - \mathbb{E}(Z_2^*))] u_2.$$

Par conséquent,

$$\rho = u_1^* K_{Z_1, Z_2} u_2.$$

Il est à présent évident que : (2.47) est équivalent à (2.44).

b) Montrons (2.42). On cherche une matrice A telle que $X' = Y - AX$ soit indépendant de Y . Mais $X' = B_1 Z$ et $X = B_2 Z$ (rappelons que Z a pour coordonnées X et Y). On applique le résultat du a) : X' et X sont deux v.a. indépendantes si et seulement si,

$$K_{X', X} = \mathbb{E}[(Y - AX - \mathbb{E}(Y - AX))(X - \mathbb{E}(X))^*] = 0. \quad (2.48)$$

En utilisant la linéarité, il vient,

$$K_{X', X} = \mathbb{E}[(Y - \mathbb{E}(Y))(X - \mathbb{E}(X))^*] - \mathbb{E}[A(X - \mathbb{E}(X))(X - \mathbb{E}(X))^*],$$

$$K_{X', X} = K_{21} - A \mathbb{E}[(X - \mathbb{E}(X))(X - \mathbb{E}(X))^*] = K_{21} - A K_{11}.$$

Il est à présent aisé de montrer que (2.48) est équivalent à $K_{21} - A K_{11} = 0$. Mais K_{11} est inversible, donc A est unique et,

$$A = K_{21} K_{11}^{-1}.$$

c) Puisque $X' = Y - AX$, par linéarité, on a $\mathbb{E}(X') = \mathbb{E}(Y) - A \mathbb{E}(X)$ et

$$K_{X'} = \mathbb{E}[(Y_0 - AX_0)(Y_0 - AX_0)^*] = \mathbb{E}[(Y_0 - AX_0)(Y_0^* - X_0^* A^*)],$$

où $Y_0 = Y - \mathbb{E}(Y)$ et $X_0 = X - \mathbb{E}(X)$. On développe, il vient,

$$K_{X'} = \mathbb{E}[Y_0 Y_0^*] - A \mathbb{E}[X_0 Y_0^*] - \mathbb{E}[Y_0 X_0^*] A^* + A \mathbb{E}[X_0 X_0^*] A^*,$$

$$K_{X'} = K_{22} - A K_{12} - K_{21} A^* + A K_{11} A^*.$$

Mais

$$K_{21} = A K_{11} \implies -K_{21} A^* + A K_{11} A^* = -K_{21} A^* + K_{21} A^* = 0,$$

d'où

$$K_{X'} = K_{22} - A K_{12} = K_{22} - K_{21} K_{11}^{-1} K_{12} = K_{22} - K_{21} K_{11}^{-1} K_{21}^*.$$

□

Proposition 2.7 Soit N la loi conditionnelle de Y sachant X . Alors $N(x, \cdot)$ est la loi gaussienne $\mathcal{N}_n(Ax + m', K)$ avec $m' = \mathbb{E}(X') = \mathbb{E}(Y) - A \mathbb{E}(X)$, $K = K_{X'} = K_{22} - K_{21} K_{11}^{-1} K_{21}^*$.

Démonstration : a) Si U et V sont deux v.a. indépendantes, et f une fonction positive et mesurable, alors $\mathbb{E}[f(U, V)] = \mathbb{E}[F(U)]$, avec $F(u) = \mathbb{E}[f(u, V)]$.

En effet, si ν (resp. μ) désigne la loi de U (resp. V), $\nu \otimes \mu$ est la loi de (U, V) :

$$\mathbb{E}[f(U, V)] = \iint f(u, v)\nu(du)\mu(dv).$$

On applique le théorème de Fubini,

$$\mathbb{E}[f(U, V)] = \int \left(\int f(u, v)\mu(dv) \right) \nu(du) = \int F(u)\nu(du).$$

b) Soient f et g deux fonctions boréliennes, positives et

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

On applique la proposition 2.6, puis a) :

$$\Delta = \mathbb{E}[f(X)g(AX + X')] = \mathbb{E}[F(X)]$$

avec $F(x) = \mathbb{E}[f(x)g(AX + X')] = f(x)\mathbb{E}[g(AX + X')]$. Mais $AX + X'$ suit une loi $\mathcal{N}_n(AX + m', K) = N(x, \cdot)$, donc,

$$F(x) = f(x) \int g(y)N(x, dy),$$

$$\Delta = \mathbb{E} \left[f(X) \int g(y)N(X, dy) \right]$$

□

2.4 Introduction aux martingales

Définitions.

1) $(\Omega, \mathcal{A}, \mathbb{P})$ désigne un espace de probabilité usuel.

2) Une filtration indexée par I ($I = \{1, 2, \dots, n\}$ ou \mathbb{N}) est une famille $\mathcal{F} = (\mathcal{F}_i ; i \in I)$ croissante de sous-tribus de \mathcal{A} , ce qui signifie que \mathcal{F}_i est pour tout i une sous-tribu de \mathcal{A} et :

$$\mathcal{F}_i \subset \mathcal{F}_{i+1}, \quad (2.49)$$

pour tout $i \in I$, tel que $i + 1 \in I$.

3) Une famille de v.a. $\{X_i ; i \in I\}$ est dite une \mathcal{F} -martingale si

$$X_i \text{ est intégrable pour tout } i \in I, \quad (2.50)$$

$$X_i \text{ est } \mathcal{F}_i\text{-mesurable, } \forall i \in I, \quad (2.51)$$

$$X_i = \mathbb{E}(X_{i+1} | \mathcal{F}_i) \quad \forall i \in I, (i + 1) \in I. \quad (2.52)$$

4) On dit que $\{X_i ; i \in I\}$ est une martingale si ce processus est une \mathcal{F} -martingale avec

$$\mathcal{F}_i = \sigma\{X_j ; j \leq i, j \in I\}, \quad \mathcal{F} = (\mathcal{F}_i ; i \in I). \quad (2.53)$$

Exemples.

Exemple 8) Soit \mathcal{B} une sous-tribu de \mathcal{A} . On considère la filtration la plus simple, la filtration constante : $\mathcal{F}_i = \mathcal{B}, \forall i \in I$. On note $\mathcal{F} = (\mathcal{F}_i ; i \in I)$.

Soit $(X_i ; i \in I)$ une \mathcal{F} -martingale. D'après (2.51), X_i est \mathcal{B} -mesurable, donc $X_i = \mathbb{E}(X_{i+1}|\mathcal{B}) = X_{i+1}$. On en déduit que $(X_i ; i \in I)$ est une \mathcal{F} -martingale si et seulement si $i \rightarrow X_i$ est constante, égale à X , et de plus X est intégrable et \mathcal{B} -mesurable.

Exemple 9) Donnons à présent un exemple important de martingales. Soit $(\varepsilon_n ; n \geq 1)$ une suite de v.a. indépendantes. On suppose de plus $\mathbb{E}(|\varepsilon_n|) < +\infty$ et $\mathbb{E}(\varepsilon_n) = 0$, pour tout $n \geq 1$. On pose

$$X_n = \sum_{k=1}^n \varepsilon_k, \quad (2.54)$$

et $\mathcal{F}_n = \sigma(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$. Puisque $X_n - X_{n-1} = \varepsilon_n$, on a $\mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n)$. Il est clair que X_n est intégrable et \mathcal{F}_n -mesurable. Vérifions la propriété (2.52) :

$$\mathbb{E}(X_n \mathbf{1}_A) = \mathbb{E}(X_{n+1} \mathbf{1}_A), \quad \forall A \in \mathcal{F}_n. \quad (2.55)$$

Mais $X_{n+1} = X_n + \varepsilon_{n+1}$, par conséquent

$$\mathbb{E}(X_{n+1} \mathbf{1}_A) = \mathbb{E}(X_n \mathbf{1}_A) + \mathbb{E}(\varepsilon_{n+1} \mathbf{1}_A).$$

ε_{n+1} est une v.a. indépendante de \mathcal{F}_n , donc

$$\mathbb{E}(\varepsilon_{n+1} \mathbf{1}_A) = \mathbb{E}(\varepsilon_{n+1})\mathbb{P}(A) = 0.$$

L'égalité (2.55) en résulte immédiatement.

Cet exemple constitue en quelque sorte le prototype de martingale. Supposons que ε_n représente le gain (algébrique) d'un joueur, à la n -ième partie. X_n représente la fortune du joueur après n parties. On a fait l'hypothèse que les résultats des différentes parties sont indépendants et qu'en moyenne le jeu est équilibré : à chaque fois le joueur a autant de chance de gagner que de perdre, ce qui se traduit par $\mathbb{E}(\varepsilon_n) = 0, \forall n \geq 1$. Nous verrons (voir théorème 2.2) qu'il n'existe pas de stratégie optimale permettant d'avoir une espérance de gain positive.

Exemple 10) Soient $\mathcal{F} = (\mathcal{F}_n ; n \geq 0)$ une filtration, U une v.a. intégrable, \mathcal{A} -mesurable. On pose :

$$X_n = \mathbb{E}(U|\mathcal{F}_n). \quad (2.56)$$

Alors $(X_n ; n \geq 0)$ est une \mathcal{F} -martingale. Il est clair que (2.50) et (2.51) sont vérifiées. Par ailleurs (2.49) implique :

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) = \mathbb{E}[\mathbb{E}(U|\mathcal{F}_{n+1})|\mathcal{F}_n] = \mathbb{E}(U|\mathcal{F}_n) = X_n.$$

Donc (2.52) est satisfaite.

Pour simplifier nous supposerons dans la suite : $I = \mathbb{N}$. Les résultats et définitions s'adaptent aisément au cas où $I = \{1, 2, \dots, n\}$. Sauf mention contraire, $\mathcal{F} = (\mathcal{F}_n ; n \geq 0)$ désigne une filtration donnée et les martingales considérées dans la suite seront des \mathcal{F} -martingales.

Proposition 2.8 *Soient a et b deux réels, $(X_n)_{n \geq 0}$ et $(Y_n)_{n \geq 0}$ deux martingales. Alors*

1) $(aX_n + bY_n)_{n \geq 0}$ est une martingale.

2) On a

$$\mathbb{E}(X_m | \mathcal{F}_n) = X_n ; \quad \forall m \geq n \geq 0. \quad (2.57)$$

$$\mathbb{E}(X_n) = \mathbb{E}(X_0), \quad \forall n \geq 0. \quad (2.58)$$

Démonstration : 1) Posons $Z_n = aX_n + bY_n$. Il est évident que $(Z_n)_{n \geq 0}$ vérifie les deux propriétés (2.50) et (2.51). De plus

$$\begin{aligned} \mathbb{E}(Z_{n+1} | \mathcal{F}_n) &= \mathbb{E}[aX_{n+1} + bY_{n+1} | \mathcal{F}_n] = a\mathbb{E}(X_{n+1} | \mathcal{F}_n) + b\mathbb{E}(Y_{n+1} | \mathcal{F}_n) \\ &= aX_n + bY_n = Z_n. \end{aligned}$$

2) On suppose n fixé et on raisonne par récurrence sur $m \geq n$. Il est clair que la propriété est vraie pour $m = n$ et $m = n + 1$. On la suppose réalisée pour le rang $m \geq n$, montrons qu'elle a encore lieu pour $m + 1$:

$$\mathbb{E}(X_{m+1} | \mathcal{F}_n) = \mathbb{E}(\mathbb{E}(X_{m+1} | \mathcal{F}_m) | \mathcal{F}_n) = \mathbb{E}(X_m | \mathcal{F}_n) = X_n.$$

3) On prend l'espérance de part et d'autre de (2.57) : $\mathbb{E}(X_n) = \mathbb{E}(X_m)$, pour tout $m \geq n$, ce qui signifie que $n \rightarrow \mathbb{E}(X_n)$ est constante. \square

Définition 2.3 *Une application $T : \Omega \rightarrow \mathbb{N} \cup \{+\infty\}$ est un temps d'arrêt si :*

$$\{T \leq n\} \in \mathcal{F}_n, \quad \text{pour tout } n \geq 0. \quad (2.59)$$

On notera qu'un temps d'arrêt peut prendre la valeur $+\infty$.

Proposition 2.9 *1) Les temps constants sont des temps d'arrêt.*

2) *Une application $T : \Omega \rightarrow \mathbb{N} \cup \{+\infty\}$ est un temps d'arrêt si et seulement si :*

$$\{T = n\} \in \mathcal{F}_n, \quad \text{pour tout } n \geq 0. \quad (2.60)$$

En particulier les temps d'arrêt sont des v.a. \mathcal{A} -mesurables.

3) *Si T_1 et T_2 sont deux temps d'arrêt, alors $\sup(T_1, T_2)$, $\inf\{T_1, T_2\}$ et $T_1 + T_2$ sont des temps d'arrêt.*

4) Soit $(X_n; n \geq 0)$ une suite de v.a. à valeurs dans (U, \mathcal{U}) , telle que $X_n : (\Omega, \mathcal{F}_n) \rightarrow (U, \mathcal{U})$ soit mesurable pour tout $n \geq 0$. Pour tout F de \mathcal{U} , on note

$$D_F = \inf\{n \geq 0, X_n \in F\} \quad (2.61)$$

avec la convention $\inf \emptyset = +\infty$. Alors D_F est un temps d'arrêt.

Démonstration : a) L'assertion 1) est évidente.

b) Soit $n \geq 1$, on a

$$\{T = n\} = \{T \leq n\} \cap \{T > n - 1\} = \{T \leq n\} \cap \{T \leq n - 1\}^c$$

et $\{T \leq n\} = \bigcup_{k=0}^n \{T = k\}$. On en déduit l'équivalence entre (2.59) et (2.60). De plus $\{T = n\} \in \mathcal{F}_n \subset \mathcal{A}$, T est une v.a. \mathcal{A} -mesurable.

c) L'assertion 3 résulte des égalités :

$$\{\sup(T_1, T_2) \leq n\} = \{T_1 \leq n\} \cap \{T_2 \leq n\},$$

$$\{\inf(T_1, T_2) \leq n\} = \{T_1 \leq n\} \cup \{T_2 \leq n\},$$

$$\{T_1 + T_2 = n\} = \bigcup_{k=0}^n \{T_1 = k\} \cap \{T_2 = n - k\}.$$

d) On a,

$$\begin{aligned} \{D_F = n\} &= \{X_0 \notin F, \dots, X_{n-1} \notin F, X_n \in F\} \\ &= \{X_0 \in F^c\} \cap \dots \cap \{X_{n-1} \in F^c\} \cap \{X_n \in F\}. \end{aligned}$$

Donc D_F vérifie (2.60), D_F est bien un temps d'arrêt. Remarquons que :

$$\{D_F = +\infty\} = \{X_n \notin F, \forall n \geq 0\}. \quad (2.62)$$

□

Plaçons-nous dans le cadre de l'exemple . Le joueur peut décider de s'arrêter à un instant fixé à l'avance (il s'agit d'un temps d'arrêt constant). Il peut aussi se retirer du jeu dès que sa fortune est en dessous d'un seuil s , qu'il s'est fixé à l'avance ; il s'arrête en T :

$$T = \inf\{n \geq 0, X_n \leq s\}. \quad (2.63)$$

On a $T = D_F$ avec $F =]-\infty, s]$. T est un temps d'arrêt : le choix de s'arrêter en $T = n$ ne dépend que de X_0, X_1, \dots, X_n .

Proposition 2.10 Soit T un temps d'arrêt fini et $(X_n)_{n \geq 0}$ un processus adapté :

$$X_n \text{ est } \mathcal{F}_n\text{-mesurable pour tout } n \geq 0. \quad (2.64)$$

On pose $X_T(\omega) = X_{T(\omega)}(\omega)$. Alors X_T est une v.a.

Démonstration : On suppose que $X_n : (\Omega, \mathcal{F}_n) \rightarrow (U, \mathcal{U})$ est mesurable pour tout $n \geq 0$. Soit $A \in \mathcal{U}$, on a

$$\{X_T \in A\} = \bigcup_{n \geq 0} \{T = n, X_T \in A\} = \bigcup_{n \geq 0} \{T = n, X_n \in A\}.$$

Mais $\{T = n\}$ et $\{X_n \in A\}$ sont deux événements de \mathcal{F}_n donc $\{T = n, X_n \in A\} \in \mathcal{A}$. Par conséquent $\{X_T \in A\} \in \mathcal{A}$. \square

Théorème 2.2 (théorème d'arrêt) *Soit T un temps fini et $(X_n)_{n \geq 0}$ une martingale. On suppose que*

$$T \text{ est un temps d'arrêt borné} \tag{2.65}$$

ou

$$(X_{T \wedge n}; n \geq 0) \text{ est un processus borné,} \tag{2.66}$$

Alors

$$\mathbb{E}(X_T) = \mathbb{E}(X_0). \tag{2.67}$$

Démonstration : a) On commence par étudier le cas où T est borné. Il existe un entier k tel que

$$T \leq k. \tag{2.68}$$

D'après (2.57) et (2.60), on a :

$$\mathbb{E}(X_k | \mathcal{F}_n) = X_n \quad n \leq k; \quad \{T = n\} \in \mathcal{F}_n.$$

On en déduit :

$$\begin{aligned} \mathbb{E}(X_T) &= \sum_{n=0}^k \mathbb{E}(X_n \mathbf{1}_{\{T=n\}}) = \sum_{n=0}^k \mathbb{E}[\mathbb{E}(X_k | \mathcal{F}_n) \mathbf{1}_{\{T=n\}}] \\ &= \sum_{n=0}^k \mathbb{E}(X_k \mathbf{1}_{\{T=n\}}) = \mathbb{E} \left[X_k \left(\sum_{n=0}^k \mathbf{1}_{\{T=n\}} \right) \right] = \mathbb{E}(X_k). \end{aligned}$$

Mais d'après (2.58), $\mathbb{E}(X_k) = \mathbb{E}(X_0)$.

b) On suppose que (2.66) a lieu : il existe K tel que

$$|X_{T \wedge n}| \leq K, \quad \forall n \geq 0. \tag{2.69}$$

Nous allons utiliser le résultat de l'étape précédente. Posons

$$T_n = T \wedge n = \inf\{n, T\}.$$

T_n est un temps d'arrêt borné, donc $\mathbb{E}(X_{T_n}) = \mathbb{E}(X_0)$.

Mais T est un temps d'arrêt fini, la suite $n \rightarrow T_n$ est croissante et vaut T à partir d'un certain rang, donc X_{T_n} converge p.s. vers X_T . De plus d'après (2.69) la suite de v.a. X_{T_n} est bornée, une application du théorème de convergence dominée, conduit à : $\mathbb{E}(X_T) = \mathbb{E}(X_0)$. \square

Remarques.

- 1) En reprenant le contexte de l'exemple, le théorème 2.2 affirme que l'espérance du gain reste constante, pour toute règle d'arrêt T .
- 2) On peut montrer plus généralement, lorsque (2.66) a lieu, que $(X_{n \wedge T}; n \geq 0)$ est une martingale.
- 3) La propriété d'arrêt (2.67) est vraie avec des hypothèses plus faibles que (2.65) ou (2.66). Toutefois dans les exemples ces hypothèses sont souvent suffisantes.

Les martingales qui ne sont pas "trop grandes" convergent à l'infini. Il s'agit en règle générale de résultats difficiles à établir. Toutefois dans le cadre L^2 , il est facile d'obtenir un résultat de convergence.

Proposition 2.11 Soit $(X_n)_{n \geq 0}$ une martingale, de carré intégrable :

$$\sup_{n \geq 0} \mathbb{E}(X_n^2) < \infty \quad (2.70)$$

Alors X_n converge dans $L^2(\Omega)$ vers une v.a. X_∞ , de plus

$$X_n = \mathbb{E}(X_\infty | \mathcal{F}_n), \quad \forall n \geq 0. \quad (2.71)$$

Démonstration : 1) Soit $n \geq m$. On a :

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

Mais $\mathbb{E}(X_n X_m) = \mathbb{E}(\mathbb{E}(X_n X_m | \mathcal{F}_m)) = \mathbb{E}(X_m \mathbb{E}(X_n | \mathcal{F}_m)) = \mathbb{E}(X_m^2)$.

Par conséquent

$$\mathbb{E}((X_n - X_m)^2) = \mathbb{E}(X_n^2) - \mathbb{E}(X_m^2) = a_n - a_m; \quad n \geq m, \quad (2.72)$$

où l'on a posé : $a_n = \mathbb{E}(X_n^2)$. Si $n = m + 1$, on a $a_{m+1} - a_m = \mathbb{E}((X_{m+1} - X_m)^2) \geq 0$.

Ce qui signifie que $n \rightarrow a_n$ est croissante. La propriété (2.70) signifie que (a_n) est majorée donc (a_n) est une suite convergente. Puisque la suite (a_n) est de Cauchy, la relation (2.72) implique que $(X_n)_{n \geq 0}$ est une suite de Cauchy dans $L^2(\Omega)$. Cette suite converge dans $L^2(\Omega)$ vers une v.a. X_∞ , de carré intégrable.

2) Montrons (2.71). On utilise à nouveau (2.57) :

$$\begin{aligned} |\mathbb{E}(X_\infty | \mathcal{F}_n) - X_n| &= |\mathbb{E}(X_\infty | \mathcal{F}_n) - \mathbb{E}(X_{n+k} | \mathcal{F}_n)| = |\mathbb{E}[(X_\infty - X_{n+k}) | \mathcal{F}_n]| \\ &\leq \mathbb{E}(|X_\infty - X_{n+k}| | \mathcal{F}_n), \end{aligned}$$

avec $n \geq 0$ et $k \geq 0$.

On prend l'espérance de part et d'autre de l'inégalité précédente :

$$\mathbb{E}[|\mathbb{E}(X_\infty|\mathcal{F}_n) - X_n|] \leq \mathbb{E}[|X_\infty - X_{n+k}|]$$

Par ailleurs

$$\mathbb{E}[|X_\infty - X_{n+k}|] \leq \{\mathbb{E}[(X_\infty - X_{n+k})^2]\}^{1/2}.$$

Pour tout $n > 0$, la suite $(X_{n+k})_{k \geq 0}$ converge dans $L^2(\Omega)$ vers X_∞ . On fait tendre k vers $+\infty$, on obtient :

$$\mathbb{E}[|\mathbb{E}(X_\infty|\mathcal{F}_n) - X_n|] = 0.$$

D'où $X_n = \mathbb{E}(X_\infty|\mathcal{F}_n)$. □

Remarque.

Réciproquement soit X_∞ une v.a. appartenant à $L^2(\Omega)$, on définit une suite de v.a. $(X_n)_{n \geq 0}$ par (2.71). On a vu dans l'exemple que $(X_n)_{n \geq 0}$ est une martingale. De plus

$$\mathbb{E}[(X_\infty - X_n)^2] = \mathbb{E}(X_\infty^2) - 2\mathbb{E}(X_n X_\infty) + \mathbb{E}(X_n^2),$$

$$\mathbb{E}(X_n X_\infty) = \mathbb{E}(X_n \mathbb{E}(X_\infty|\mathcal{F}_n)) = \mathbb{E}(X_n X_n) = \mathbb{E}(X_n^2).$$

Donc

$$\mathbb{E}((X_\infty - X_n)^2) = \mathbb{E}(X_\infty^2) - \mathbb{E}(X_n^2) \geq 0.$$

En particulier $\mathbb{E}(X_n^2) \leq \mathbb{E}(X_\infty^2)$. Ce qui signifie que $(X_n)_{n \geq 0}$ est une martingale de carré intégrable. D'après la proposition 2.11, $(X_n)_{n \geq 0}$ converge dans $L^2(\Omega)$ vers \tilde{X} . X_∞ et \tilde{X} sont reliées par la relation

$$\mathbb{E}(\tilde{X}|\mathcal{F}_n) = \mathbb{E}(X_\infty|\mathcal{F}_n), \quad \forall n \geq 0.$$

Si l'on suppose de plus que $\mathcal{A} = \bigvee_{n \geq 0} \mathcal{F}_n$, en utilisant le théorème de classe monotone on peut montrer que $\tilde{X} = X_\infty$.

Les martingales de carré intégrable $(X_n)_{n \geq 0}$, s'identifient aux v.a. X_∞ de $L^2(\Omega)$ à travers (2.71), de plus X_∞ est la limite dans $L^2(\Omega)$ de $(X_n)_{n \geq 0}$.

2.5 Exercices

2.0. Soit N une v.a. à valeurs dans $\{0, \dots, n\}$; on note $\alpha_k = \mathbb{P}(N = k)$. On considère $(\varepsilon_n; n \geq 0)$ une suite de v.a. indépendantes, de même loi : $\mathbb{P}(\varepsilon_0 = 1) = p$, $\mathbb{P}(\varepsilon_0 = 0) = q$, avec $p + q = 1$, $p > 0$ et $q > 0$. On suppose que N est indépendante de la famille $(\varepsilon_n; n \geq 0)$. On définit alors la v.a. X par la relation : $X = \sum_{k=0}^N \varepsilon_k$.

a) Calculer la loi de X . Exprimer $\mathbb{E}(X)$, $\mathbb{E}(X^2)$ à l'aide de $\mathbb{E}(N)$ et $\mathbb{E}(N^2)$.

b) Soit $p' \in]0, 1[$. Déterminer la loi de N pour que la loi conditionnelle de N sachant $X = 0$ soit la loi binomiale $\mathcal{B}(n, p')$.

2.1. Soit (X, Y) un couple gaussien centré et on note par $\sigma(X)$ la tribu engendrée par X .

i) Prouver que :

$$\mathbb{E}[Y \mid \sigma(X)] = cX, \quad \text{où } c := \frac{\mathbb{E}(XY)}{\mathbb{E}(X^2)}.$$

ii) Soit (X_1, \dots, X_n) un vecteur aléatoire gaussien centré et $(a_1, \dots, a_n), (b_1, \dots, b_n) \in \mathbb{R}^n$. On définit $Y := \sum_{j=1}^n a_j X_j$ et $Z := \sum_{j=1}^n b_j X_j$. Que vaut $\mathbb{E}[Z \mid \sigma(Y)]$?

2.2. i) Soient Y, Y' deux variables aléatoires réelles gaussiennes, centrées, réduites, indépendantes. Montrer que, pour tous $t, s \geq 0$, les variables aléatoires :

$$e^{-t}\sqrt{1-e^{-2s}}Y + \sqrt{1-e^{-2t}}Y' \quad \text{et} \quad \sqrt{1-e^{-2(t+s)}}Z$$

ont la même loi. Ici Z est une variable gaussienne centrée réduite.

ii) Soit, pour tout $t \geq 0$, l'opérateur T_t défini par :

$$T_t f(x) := \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(e^{-t}x + \sqrt{1-e^{-2t}}y) e^{-y^2/2} dy.$$

Ici $f : \mathbb{R} \rightarrow \mathbb{R}$. Soit Y une variable gaussienne centrée réduite. Prouver que :

$$T_t f(x) = \mathbb{E} \left[f(e^{-t}x + \sqrt{1-e^{-2t}}Y) \right].$$

iii) On note par ν la mesure gaussienne centrée réduite sur \mathbb{R} : $\nu(dx) := (2\pi)^{-1/2} e^{-x^2/2} dx$ et par $L^p(\nu)$ l'espace $L^p(\mathbb{R}, \mathcal{B}(\mathbb{R}), \nu)$, $1 \leq p \leq \infty$. Prouver que, pour $p = 1$ et $p = \infty$, T_t est, pour tout $t \geq 0$, un opérateur borné de $L^p(\nu)$ dans $L^p(\nu)$.

iv) Prouver que, pour tous $t, s \geq 0$, $T_t(T_s f) = T_{t+s}f$. (On dit que $\{T_t : t \geq 0\}$ est un semi-groupe d'opérateurs, le semi-groupe d'Ornstein-Uhlenbeck.)

v) Prouver que, pour tout $t \geq 0$, et toutes fonctions $f, g : \mathbb{R} \rightarrow \mathbb{R}$ bornées,

$$\langle T_t f, g \rangle_{\nu} = \langle f, T_t g \rangle_{\nu}, \quad \text{où } \langle f, g \rangle_{\nu} := \int_{\mathbb{R}} f(x)g(x)\nu(dx).$$

vi) Soit (X, Y) un couple gaussien centré tel que $\mathbb{E}(X^2) = \mathbb{E}(Y^2) = 1$ et $\mathbb{E}(XY) = \rho \geq 0$. Montrer que, pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$, bornée :

$$\mathbb{E}[f(Y) \mid \sigma(X)] = T_t f(X), \quad \text{avec } \rho = e^{-t}.$$

Que se passe-t-il quand $\rho = 0$ ou $\rho = 1$? Qu'obtient-on dans le cas particulier où $f(x) = x$?

vii) Soient X_1, X_2, X_3 trois variables aléatoires réelles gaussiennes, centrées, réduites, indépendantes. On pose $Y_1 = X_1$ et

$$Y_2 = Y_1 e^{-t} + \sqrt{1-e^{-2t}} X_2, \quad Y_3 = Y_2 e^{-s} + \sqrt{1-e^{-2s}} X_3.$$

Montrer que, pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$, bornée, on a :

$$\mathbb{E}[f(Y_3) \mid \sigma(Y_1)] = T_{t+s} f(Y_1).$$

Calculer

$$\mathbb{E}[f(Y_3) \mid \sigma(Y_1, Y_2)].$$

En déduire le résultat du point **iv**), pour tous $t, s \geq 0$, $T_t(T_s f) = T_{t+s} f$.

2.3. Soit (X, Y, Z) un triplet gaussien centré.

i) Si on suppose seulement que X et Y sont indépendantes, montrer que :

$$\mathbb{E}[Z \mid \sigma(X, Y)] = \mathbb{E}[Z \mid \sigma(X)] + \mathbb{E}[Z \mid \sigma(Y)].$$

ii) On suppose que la matrice de covariance vaut

$$K = \begin{pmatrix} 1 & 0 & \rho_1 \\ 0 & 1 & \rho_2 \\ \rho_1 & \rho_2 & 1 \end{pmatrix}.$$

a) Montrer que l'on a $\rho_1^2 + \rho_2^2 \leq 1$.

b) Soit (X', Y', Z') un triplet gaussien centré de matrice de covariance égale à l'identité. Définissons :

$$\bar{X} := X', \bar{Y} := Y', \bar{Z} := \rho_1 X' + \rho_2 Y' + \sqrt{1 - \rho_1^2 - \rho_2^2} Z'.$$

Prouver que $(\bar{X}, \bar{Y}, \bar{Z})$ un triplet gaussien de même loi que (X, Y, Z) .

iii) Si la matrice de covariance est l'identité, quelle est la loi de

$$\frac{(X + YZ)}{\sqrt{1 + Z^2}}?$$

2.4. Soient X, Y deux variables aléatoires réelles définies sur $(\Omega, \mathcal{A}, \mathbb{P})$ et $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ une application mesurable. Soient, pour tout $x, y \in \mathbb{R}$, $g(x) := \mathbb{E}[f(x, Y)]$ et $h(y) := \mathbb{E}[f(X, y)]$.

i) On suppose que X, Y sont indépendantes.

a) Lorsque f est positive, montrer que

$$\mathbb{E}[f(X, Y)] = \mathbb{E}[g(X)] = \mathbb{E}[h(Y)].$$

b) Lorsque X suit la loi uniforme sur $[0, 1]$, montrer que

$$F_{X+Y}(t) = \int_0^1 F_Y(t-u) du, t \in \mathbb{R}.$$

ii) On suppose toujours que X, Y sont indépendantes. Montrer que

$$\mathbb{E}[f(X, Y) \mid \sigma(X)] = g(X) \text{ et } \mathbb{E}[f(X, Y) \mid \sigma(Y)] = h(Y).$$

iii) Soit \mathcal{B} une sous-tribu de \mathcal{A} . On ne suppose plus l'indépendance de X et Y , mais que X est \mathcal{B} -mesurable et que Y est indépendante de \mathcal{B} . Montrer que

$$\mathbb{E}[f(X, Y) \mid \mathcal{B}] = g(X).$$

Énoncer les hypothèses et un résultat similaire pour $h(Y)$.

2.5. Soient X, Y deux variables aléatoires positives, indépendantes, de même loi. On pose $U := \min\{X, Y\}$ et $V := \max\{X, Y\}$. Calculer $\mathbb{E}[U \mid \sigma(V)]$ lorsque :

- i) la loi commune est la loi exponentielle de paramètre λ ;
- ii) la loi commune est la loi uniforme sur $[0, 1]$.

2.6. Soit ε une variable aléatoire prenant seulement les valeurs ± 1 avec la probabilité $1/2$. Soit \mathcal{B} une sous-tribu de \mathcal{A} . Montrer que ε est indépendante de \mathcal{B} si et seulement si $\mathbb{E}[\varepsilon \mid \mathcal{B}] = 0$.

2.7. i) Soit Y une variable aléatoire ayant la densité $f_Y(y) := y^{-2}\mathbf{1}_{[1, \infty[}(y)$ et soit $X = Y$. Montrer que $\mathbb{E}(Y) = \infty$, mais que $\mathbb{E}[Y \mid \sigma(X)] < \infty$ p.s.

ii) Même question pour le couple aléatoire (X, Y) ayant la loi jointe donnée par $p_{ij} := [i(i+1)]^{-1}$, pour $i = j = 1, 2, \dots$

2.8. Soient $n \geq 1$ un entier fixé et p_1, p_2, p_3 trois réels positifs tels que $p_1 + p_2 + p_3 = 1$. On note :

$$p_{ij} = \begin{cases} n! \frac{p_1^i p_2^j p_3^{n-i-j}}{i! j! (n-i-j)!}, & \text{si } i + j \leq n \\ 0, & \text{sinon.} \end{cases}$$

- i) Montrer qu'il existe un couple (X, Y) tel que $\mathbb{P}(X = i, Y = j) = p_{ij}$.
- ii) Trouver les lois de X et de Y ainsi que celle de Y sachant X .
- iii) Calculer $\mathbb{E}(XY)$.

2.9. Soient X, Y deux variables aléatoires réelles. On suppose que Y suit la loi exponentielle de paramètre 1. On suppose que la loi conditionnelle de X sachant $Y = y$ est une loi de Poisson de paramètre y .

- i) Calculer les lois du couple (X, Y) , de X , ainsi que la loi conditionnelle de Y sachant X .
- ii) Montrer que $\mathbb{E}[(X - Y)^2] = 1$, en conditionnant d'abord par rapport à Y et ensuite par rapport à X .

2.10. i) Montrer que la fonction $f(x, y) = 2\mathbf{1}_{\{x, y \geq 0, x+y \leq 1\}}$ est une densité de probabilité d'un couple (X, Y) .

- ii) Trouver les lois marginales, ainsi que la loi conditionnelle de X sachant Y .
- iii) En déduire que la variable aléatoire $X/(1 - Y)$ est indépendante de Y et suit la loi uniforme sur $[0, 1]$.

2.11. Soient X, Y deux variables aléatoires réelles. On suppose que $X \sim \mathcal{N}(0, 1)$ et que la loi conditionnelle de Y sachant $X = x$, a la densité $(\alpha/2)e^{-\alpha|y|}$, où $\alpha := (1 + x^2)/2$.

- i) Trouver la densité de (X, Y) et celle de Y .
- ii) Montrer que Y admet des moments de tous ordres et que ceux impairs sont nuls.
- iii) Vérifier que la loi de Y est symétrique et trouver la fonction de répartition de $|Y|$.

2.12. Trouver la loi conditionnelle et l'espérance conditionnelle de Y sachant X , lorsque la densité du couple (X, Y) est donnée par :

- i) $f(x, y) = \lambda^2 e^{-\lambda y} \mathbf{1}_{\{0 \leq x \leq y\}}$.
- ii) $f(x, y) = x e^{-x(y+1)} \mathbf{1}_{\{x, y \geq 0\}}$.
- iii) $f(x, y) = (x/\sqrt{2\pi}) \exp(-x(y-x)/2) \mathbf{1}_{\{x, y > 0\}}$.
- iv) $f(x, y) = 4x(y-x) \exp(-(x+y)) \mathbf{1}_{\{0 \leq x \leq y\}}$.

2.13. Soit (X, Y) un couple gaussien tel que $\mathbb{E}(X) = m_x$, $\mathbb{E}(Y) = m_y$, $\text{Var}(X) = \sigma_x^2$, $\text{Var}(Y) = \sigma_y^2$ et $\rho(X, Y) = \rho$.

- i) Montrer que : $\mathbb{E}[X | \sigma(Y)] = m_x + \rho \sigma_x (Y - m_y) / \sigma_y$.
- ii) On définit la variance conditionnelle :

$$\text{Var}[X | \sigma(Y)] := \mathbb{E}[\{X - \mathbb{E}[X | \sigma(Y)]\}^2 | \sigma(Y)].$$

Alors $\text{Var}[X | \sigma(Y)] = \sigma_x^2(1 - \rho^2)$.

- iii) On suppose que $m_x = m_y = 1$, $\sigma_x^2 = 2$, $\sigma_y^2 = 1$ et que $\rho = -1/\sqrt{2}$. Faire le calcul de la loi conditionnelle de Y sachant X .
- iv) On suppose que $m_x = m_y = 0$ et que $\sigma_x^2 = \sigma_y^2 = 1$. Montrer que :

$$\mathbb{E}[X^2 | Y^2 = y^2] = \frac{1}{2} \{ \mathbb{E}[X^2 | Y = y] + \mathbb{E}[X^2 | Y = -y] \} = 1 + \rho^2(y^2 - 1).$$

2.14. Une variable aléatoire X_1 suit la loi uniforme sur $[0, 1]$. Si $X_1 = x_1$, alors X_2 est une variable aléatoire qui suit la loi uniforme sur $[x_1, x_1 + 1]$. Si $X_2 = x_2$, alors X_3 est une variable aléatoire qui suit la loi uniforme sur $[x_2, x_2 + 1]$. Les variables aléatoires X_4, \dots, X_n sont définies de la même manière pour tous $n \geq 4$. Calculer $\mathbb{E}(X_n)$.

2.15. Soit $\Omega = [0, 1]$, $\mathcal{A} = \mathcal{B}([0, 1])$ et \mathbb{P} la mesure de Lebesgue sur $[0, 1]$. Définissons les variables aléatoires :

$$X(\omega) := \begin{cases} 1, & \text{si } \omega \in [0, \frac{1}{2}] \\ 0, & \text{si } \omega \in]\frac{1}{2}, 1] \end{cases}, \quad Y(\omega) := \begin{cases} 1, & \text{si } \omega \in [0, \frac{3}{4}[\\ 0, & \text{si } \omega \in]\frac{3}{4}, 1] \end{cases}$$

$$Z(\omega) := \begin{cases} 1, & \text{si } \omega \in [\frac{1}{4}, \frac{3}{4}] \\ 0, & \text{si } \omega \notin [\frac{1}{4}, \frac{3}{4}]. \end{cases}$$

Montrer que $\mathbb{E}[X | \sigma(Z)] = \mathbb{E}(X)$ p.s., mais que $\mathbb{E}[X | \sigma(Y, Z)] \neq \mathbb{E}[X | \sigma(Y)]$.

2.16. Soit Z une variable aléatoire dont la loi est symétrique par rapport à zéro et soit X une variable aléatoire indépendante de Z et telle que $X \geq 1$ p.s. On note $Y := Z/X$. Calculer $\mathbb{E}(Z)$, $\mathbb{E}(Y)$ et $\mathbb{E}[Y | \sigma(X)]$. Montrer que $\mathbb{E}[Y | \sigma(X)] = \mathbb{E}(Y)$ p.s., mais que X et Y sont dépendantes.

* * *

2.17. Soient X, Y deux variables aléatoires de carré intégrable et \mathcal{B} une sous-tribu de l'espace des événements \mathcal{A} . Montrer que : $\mathbb{E}\{X \mathbb{E}[Y | \mathcal{B}]\} = \mathbb{E}\{Y \mathbb{E}[X | \mathcal{B}]\}$.

2.18. Soient $\mathcal{B}_1 \subset \mathcal{B}_2$ deux sous-tribus de \mathcal{A} , A un événement et X une variable aléatoire intégrable.

i) Montrer que :

$$\mathbb{E}[\mathbb{E}[X | \mathcal{B}_2) | \mathcal{B}_1] = \mathbb{E}[X | \mathcal{B}_1] \text{ et } \mathbb{E}[\mathbb{E}[X | \mathcal{B}_1) | \mathcal{B}_2] = \mathbb{E}[X | \mathcal{B}_1].$$

ii) Soit $\Omega = a, b, c$. Donner un exemple pour lequel

$$\mathbb{E}[\mathbb{E}[X | \mathcal{B}_2) | \mathcal{B}_1] \neq \mathbb{E}[\mathbb{E}[X | \mathcal{B}_1) | \mathcal{B}_2].$$

iii) Montrer que : $\mathbb{E}[\mathbb{P}(A | \mathcal{B}_2) | \mathcal{B}_1] = \mathbb{P}(A | \mathcal{B}_1)$ p.s.

2.19. Soit X une variable aléatoire réelle définie sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ de carré intégrable. Si \mathcal{B} est une sous-tribu de \mathcal{A} , on définit

$$\text{Var}[X | \mathcal{B}] := \mathbb{E}[\{X - \mathbb{E}[X | \mathcal{B}]\}^2 | \mathcal{B}].$$

Montrer que : $\text{Var}(X) = \mathbb{E}\{\text{Var}[X | \mathcal{B}]\} + \text{Var}\{\mathbb{E}[X | \mathcal{B}]\}$.

2.20. Soit $\{X_n : n \geq 1\}$ une suite de variables aléatoires sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. Soit \mathcal{B} une sous-tribu de \mathcal{A} .

i) (*convergence monotone*) On suppose que $X_1 \geq 0$ p.s., que pour tout $n \geq 1$, $X_{n+1} \geq X_n$ p.s. et que X_n est intégrable. Montrer que si $\lim_{n \uparrow \infty} X_n$ est intégrable, alors :

$$\lim_{n \uparrow \infty} \mathbb{E}[X_n | \mathcal{B}] = \mathbb{E}\left[\lim_{n \uparrow \infty} X_n | \mathcal{B}\right], \text{ p.s.}$$

ii) (*lemme de Fatou*) On suppose que pour tout $n \geq 1$, $X_n \geq 0$ p.s. et que X_n est intégrable. Montrer que :

$$\mathbb{E}\left[\liminf_{n \uparrow \infty} X_n | \mathcal{B}\right] \leq \liminf_{n \uparrow \infty} \mathbb{E}[X_n | \mathcal{B}], \text{ p.s.}$$

Lorsqu'on suppose que $X_n \leq Z$ p.s. pour une certaine variable aléatoire Z intégrable, montrer que :

$$\mathbb{E}\left[\limsup_{n \uparrow \infty} X_n | \mathcal{B}\right] \geq \limsup_{n \uparrow \infty} \mathbb{E}[X_n | \mathcal{B}], \text{ p.s.}$$

iii) (*convergence dominée*) On suppose que pour tout $n \geq 1$, X_n est intégrable et que $|X_n| \leq Z$ pour une certaine variable aléatoire Z intégrable. De plus on suppose qu'il existe une autre variable aléatoire X_∞ telle que $X_n \rightarrow X_\infty$ p.s. Montrer que :

$$\lim_{n \uparrow \infty} \mathbb{E}[X_n | \mathcal{B}] = \mathbb{E}[X_\infty | \mathcal{B}], \text{ p.s.}$$

iv) On suppose que pour tout $n \geq 1$, $X_n \geq 0$ p.s. et que X_n est intégrable. Montrer que :

$$\mathbb{E} \left[\sum_{n \geq 1} X_n \mid \mathcal{B} \right] = \sum_{n \geq 1} \mathbb{E} [X_n \mid \mathcal{B}], \text{ p.s.}$$

2.21. Inégalité de Jensen. Soit X une variable aléatoire réelle définie sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ intégrable et soit \mathcal{B} est une sous-tribu de \mathcal{A} . On considère $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction convexe telle que la variable aléatoire $\varphi(X)$ soit intégrable.

i) Dire pourquoi $\varphi(X)$ est une variable aléatoire.

ii) On suppose dans ce point seulement que X prend les valeurs réelles x et y avec les probabilités $\lambda \in]0, 1[$ et $1 - \lambda$. Majorer $\varphi(\mathbb{E}(X))$.

iii) On rappelle qu'une fonction réelle convexe admet des dérivées à gauche et à droite en chaque point. On note $\delta(x)$ la dérivée à droite de φ en $x \in \mathbb{R}$. Montrer que

$$\varphi(y) - \varphi(x) \geq \delta(x)(y - x), \forall x, y \in \mathbb{R}.$$

iv) Appliquer cette inégalité aux réels $y = X(\omega)$ et $x = \mathbb{E}[X \mid \mathcal{B}](\omega) =: Y(\omega)$.

v) Supposons d'abord que toutes les variables aléatoires de l'inégalité précédente sont bornées. Appliquer $\mathbb{E}(\cdot \mid \mathcal{B})$ et déduire l'inégalité de Jensen :

$$\varphi(\mathbb{E}[X \mid \mathcal{B}]) \leq \mathbb{E}[\varphi(X) \mid \mathcal{B}].$$

vi) On ne suppose plus que les variables sont bornées. Remplacer la variable X par $X_n = X \mathbf{1}_{\{|\mathbb{E}(X \mid \mathcal{B})| \leq n\}}$ et vérifier l'inégalité de Jensen pour X_n . Laisser ensuite n tendre vers l'infini et conclure.

vii) Que devient cette inégalité pour $\mathcal{B} = \{\emptyset, \Omega\}$?

viii) Soit $1 \leq p < \infty$. Montrer que $\mathbb{E}[\cdot \mid \mathcal{B}]$ est un opérateur de contraction sur $L^p(\Omega, \mathcal{A}, \mathbb{P})$. Précisément, si X est p -intégrable et si on note la norme $\|X\|_p := (\mathbb{E}|X|^p)^{1/p}$, alors

$$\|\mathbb{E}[X \mid \mathcal{B}]\|_p \leq \|X\|_p.$$

On pourra appliquer l'inégalité de Jensen avec la fonction convexe $\varphi(x) = |x|^p$.

ix) Soit $1 \leq p < \infty$. Montrer que si $\{X_n : n \geq 1\}$ est une suite de variables aléatoires telle que $X_n \rightarrow X_\infty$ dans L^p alors

$$\mathbb{E}[X_n \mid \mathcal{B}] \rightarrow \mathbb{E}[X_\infty \mid \mathcal{B}] \text{ dans } L^p.$$

2.22. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et \mathcal{B}, \mathcal{C} deux sous-tribus de \mathcal{A} , indépendantes. Soit X une variable aléatoire réelles définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ et on suppose que X est indépendante de \mathcal{C} . Soit \mathcal{D} la tribu engendrée par $\mathcal{B} \cup \mathcal{C}$. Montrer que $\mathbb{E}[X \mid \mathcal{D}] = \mathbb{E}[X \mid \mathcal{B}]$.

2.23. Soient $X, Y \in L^1(\Omega, \mathcal{A}, \mathbb{P})$. On suppose que

$$\mathbb{E}[X \mid \sigma(Y)] = Y, \text{ et } \mathbb{E}[Y \mid \sigma(X)] = X.$$

Montrer que $X = Y$ p.s. On pourra observer que, pour tout $x \in \mathbb{R}$ fixé :

$$0 \leq \mathbb{E} [(X - Y)\mathbf{1}_{\{Y \leq x < X\}}] = \mathbb{E} [(Y - X)\mathbf{1}_{\{x < X\} \cap \{x < Y\}}].$$

2.24. Soient \mathcal{F} et \mathcal{G} deux sous-tribus de \mathcal{A} et \mathcal{H} une sous-tribu de \mathcal{F} . On dit que \mathcal{F} et \mathcal{G} sont indépendantes conditionnellement à la tribu \mathcal{H} si

$$\mathbb{E}[XY | \mathcal{H}] = \mathbb{E}[X | \mathcal{H}] \mathbb{E}[Y | \mathcal{H}],$$

pour toutes les variables aléatoires X, Y bornées, avec $\sigma(X) \subset \mathcal{F}$ et $\sigma(Y) \subset \mathcal{G}$. On va montrer qu'une condition nécessaire et suffisante pour que \mathcal{F} et \mathcal{G} soient indépendantes conditionnellement à la tribu \mathcal{H} est

$$\mathcal{H} \supset \sigma\{\mathbb{E}[Y | \mathcal{F}] : Y \text{ bornée, } \mathcal{G}\text{-mesurable}\}.$$

i) Pour montrer que la condition est suffisante on pourra vérifier d'abord que

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

ii) Pour montrer que la condition est nécessaire il suffit de prouver que, pour toute variable aléatoire Y bornée, \mathcal{G} -mesurable,

$$\mathbb{E}[Y | \mathcal{H}] = \mathbb{E}[Y | \mathcal{F}].$$

Pour cela on pourra vérifier d'abord que

$$\mathbb{E}\{\mathbb{E}[Y | \mathcal{H}] \mathbb{E}[Y | \mathcal{F}]\} = \mathbb{E}\{\mathbb{E}[Y | \mathcal{H}]^2\} = \mathbb{E}\{\mathbb{E}[Y | \mathcal{F}]^2\}.$$

Conclure par le cas d'égalité dans l'inégalité de Cauchy-Schwarz.

2.25. *Indépendance et (in)dépendance conditionnellement à une tribu*

i) Soit (X, Y, Z) un triplet aléatoire discret dont la loi est donnée par :

$$\mathbb{P}(X = k, Y = m, Z = n) := p^3(1-p)^{m-3},$$

où $0 < p < 1$, $k = 1, \dots, m-1$, $m = 2, \dots, n-1$, $n = 3, 4, \dots$

a) Trouver lois des couples (X, Y) , (X, Z) et (Y, Z) et ensuite les marginales X, Y et Z . Montrer que X et Z sont dépendantes.

b) Calculer $\mathbb{P}(Z = n | X = k, Y = m)$, $k = 1, \dots, m-1$, $m = 2, \dots, n-1$. En déduire, $\mathbb{E}[Z | X = k, Y = m]$, pour $k = 1, \dots, m-1$, $m = 2, 3, \dots$ et ensuite $\mathbb{E}[Z | \sigma(X, Y)]$.

c) Par un calcul similaire à celui du point précédent exprimer, pour toute fonction mesurable et bornée g , $\mathbb{E}[g(Z) | \sigma(X, Y)]$. Montrer que cette quantité ne dépend pas de X .

d) En déduire que Z est indépendante de X , conditionnellement à la tribu $\sigma(Y)$, bien que Z et X sont dépendantes.

ii) Soient X, Y deux variables aléatoires indépendantes de loi de Bernoulli de paramètre p . On

note $Z := \mathbb{1}_{\{X+Y=0\}}$.

a) Calculer $\mathbb{E}[X \mid \sigma(Z)]$ et $\mathbb{E}[Y \mid \sigma(Z)]$.

b) Montrer que les variables aléatoires X et Y sont dépendantes conditionnellement à la tribu $\sigma(Z)$.

iii) Soient X_1, \dots, X_n des variables aléatoires indépendantes de loi exponentielle de paramètre λ . On note $S := X_1 + \dots + X_n$.

a) Trouver la loi conditionnelle de X_1 sachant S et calculer ensuite $\mathbb{E}[X_1 \mid \sigma(S)]$.

b) Montrer que les variables aléatoires X_1 et X_2 sont dépendantes conditionnellement à la tribu $\sigma(S)$.

c) Calculer $\mathbb{E}[X_1^2 \mid \sigma(S)]$ et $\mathbb{E}[X_1 X_2 \mid \sigma(S)]$.

iv) Donner un exemple de deux variables aléatoires X et Y à valeurs dans $\{-1, 0, 1\}$ telles que X et Y ne sont pas indépendantes, alors que $\mathbb{E}(X \mid Y) = \mathbb{E}(X)$.

2.26. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et X une variable aléatoire réelle positive ou nulle. Pour $A_1, A_2 \in \mathcal{A}$ on dit que $A_1 > A_2$ si $\mathbb{P}(A_1^c \cap A_2) = 0$. Soit \mathcal{B} une sous-tribu de \mathcal{A} et $A := \{X > 0\}$, $B := \{\mathbb{E}[X \mid \mathcal{B}] > 0\}$ et $\mathcal{C} := \{C \in \mathcal{B} : C > A\}$. Montrer que B est le plus petit élément de \mathcal{C} , c'est-à-dire : $B \in \mathcal{C}$ et $\forall C \in \mathcal{C}, C > B$.

2.27. Soient $p, q > 1$ deux réels conjugués, c'est-à-dire $1/p + 1/q = 1$.

i) Montrer que, pour tous $x, y \geq 0$, $xy \leq x^p/p + y^q/q$.

ii) Soit $X \in L^p(\Omega, \mathcal{A}, \mathbb{P})$ et $Y \in L^q(\Omega, \mathcal{A}, \mathbb{P})$. Si \mathcal{B} est une sous-tribu de \mathcal{A} , on note

$$B := \{\mathbb{E}[|X|^p \mid \mathcal{B}] > 0\} \cap \{\mathbb{E}[|Y|^q \mid \mathcal{B}] > 0\}.$$

Montrer que, sur B :

$$\frac{|XY|}{(\mathbb{E}[|X|^p \mid \mathcal{B}])^{1/p} (\mathbb{E}[|Y|^q \mid \mathcal{B}])^{1/q}} \leq \frac{|X|^p}{p \mathbb{E}[|X|^p \mid \mathcal{B}]} + \frac{|Y|^q}{q \mathbb{E}[|Y|^q \mid \mathcal{B}]}.$$

En déduire que, sur B :

$$\mathbb{E}[|XY| \mid \mathcal{B}] \leq (\mathbb{E}[|X|^p \mid \mathcal{B}])^{1/p} (\mathbb{E}[|Y|^q \mid \mathcal{B}])^{1/q}.$$

iii) Soit $B_X := \{\mathbb{E}[|X|^p \mid \mathcal{B}] = 0\}$. Montrer que, sur B_X , X est nulle.

iv) En déduire l'inégalité de Hölder :

$$\mathbb{E}[|XY| \mid \mathcal{B}] \leq (\mathbb{E}[|X|^p \mid \mathcal{B}])^{1/p} (\mathbb{E}[|Y|^q \mid \mathcal{B}])^{1/q}.$$

v) (inégalité de Markov) Montrer que si $a > 0$, alors

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[|X|^p \mid \mathcal{B}]}{a^p}.$$

2.28. Soient X, Y deux variables aléatoires. On suppose que $\mathbb{E}(|Y|) < \infty$. Montrer que $\mathbb{E}[Y \mid X]$ est presque sûrement constante si et seulement si, pour tout $t \in \mathbb{R}$, on a :

$$\mathbb{E}[Y \exp(itX)] = \mathbb{E}(Y) \mathbb{E}[\exp(itX)].$$

* * *

2.29. Soit $\{X_n : n \geq 1\}$ une suite de variables aléatoires indépendantes. On pose $Y_n = \sum_{j=1}^n X_j$, $n \geq 1$.

i) On suppose les variables X_n intégrables. Montrer que :

$$\left\{ Y_n - \sum_{j=1}^n \mathbb{E}(X_j) : n \geq 1 \right\}$$

est une martingale.

ii) On suppose que les variables X_n sont centrées et de carré intégrable. Montrer que :

$$\left\{ Y_n^2 - \sum_{j=1}^n \mathbb{E}(X_j^2) : n \geq 1 \right\}$$

est une martingale.

2.30. Soit $\{X_n : n \geq 1\}$ une suite de variables aléatoires indépendantes, de même loi. On note $\varphi(u) := \mathbb{E}[\exp iuX_1]$ la fonction caractéristique de la loi comune. Si $S_n = \sum_{j=1}^n X_j$, montrer que :

$$\left\{ \frac{\exp iuS_n}{\varphi(u)^n} : n \geq 1 \right\}$$

est une martingale (complexe).

2.31. Soit $\{X_n : n \geq 1\}$ une suite de variables aléatoires indépendantes telles que $\mathbb{E}(X_n) = m_n \neq 0$, $n \geq 1$ et $X_n \in L^p$, $\forall p$. Montrer que :

$$\left\{ Y_n = \prod_{j=1}^n \frac{X_j}{m_j} : n \geq 1 \right\}$$

est une martingale.

2.32. Soit $\{X_n : n \geq 1\}$ une suite de variables aléatoires indépendantes telles que $\mathbb{P}(X_n = 1) = p \in]0, 1[$, $\mathbb{P}(X_n = -1) = q = 1 - p$, $n \geq 1$. Pour tout $n \geq 1$, on introduit :

$$S_n := \sum_{j=1}^n X_j, Y_n := \left(\frac{q}{p}\right)^{S_n} \text{ et } Z_n := \frac{\cos\{\lambda[S_n - c]\}}{(\cos \lambda)^n}, c \in \mathbb{R}.$$

i) Montrer que Y est une martingale.

ii) Montrer que Z est une martingale si $\cos \lambda \neq 0$ et si $p = q$.

2.33. Soit ξ une variable aléatoire géométrique de paramètre $p \in]0, 1[$, c'est-à-dire $\mathbb{P}(\xi =$

$n) = p^n(1-p)$, $n \geq 0$. Soit, pour tout $n \geq 0$, \mathcal{F}_n la tribu engendrée par $\xi \wedge (n+1)$.

i) Montrer que :

$$\mathcal{F}_n = \sigma \{ \{ \xi = 0 \}, \{ \xi = 1 \}, \dots, \{ \xi = n \}, \{ \xi \geq n+1 \} \}.$$

ii) Montrer que :

$$X_n := \mathbf{1}_{\{\xi \leq n\}} - (1-p)(\xi \wedge n) \text{ et } Y_n := X_n^2 - p(1-p)(\xi \wedge n), \quad n \geq 0,$$

sont deux martingales.

2.34. i) Soit $\{\mathcal{F}_n : n \geq 0\}$ une filtration, T un temps d'arrêt et $A \in \mathcal{F}_T$. On pose :

$$T_A := \begin{cases} T & \text{sur } A \\ \infty & \text{sur } A^c. \end{cases}$$

Montrer que T_A est un temps d'arrêt.

ii) Si S est un temps d'arrêt tel que $T \leq S$, soit :

$$U := \begin{cases} T & \text{sur } A \\ S & \text{sur } A^c. \end{cases}$$

Montrer que U est un temps d'arrêt.

iii) Soit $\{M_n : n \geq 0\}$ une suite de variables aléatoires telles que, $\mathbb{E}[|M_n|] < \infty$, pour tout $n \geq 0$. Montrer que M est une martingale si et seulement si $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$, pour tout temps d'arrêt borné τ . Pour établir une partie, on pourra considérer des temps d'arrêt U avec $T \equiv n$, $A \in \mathcal{F}_n$ et $S \equiv n+1$.

2.35. Soit T un temps d'arrêt par rapport à une filtration $\{\mathcal{F}_n : n \geq 0\}$. Pour tout $n \in \mathbb{N}$, on pose :

$$\phi(n) := \inf\{k \geq 0 : \{T = n\} \in \mathcal{F}_k\}.$$

Prouver que $\phi(T)$ est un temps d'arrêt, majoré par T .

2.36. Soit $\{Y_n : n \geq 1\}$ une suite de variables aléatoires indépendantes telles que :

$$\mathbb{P}(Y_n = -1) = 1 - 2^{-n}, \quad \mathbb{P}(Y_n = 2^n - 1) = 2^{-n}.$$

Soit :

$$T := \inf\{n \geq 1 : Y_n \neq -1\}, \text{ et } X_n := \sum_{j=1}^n (-1)^j Y_j \mathbf{1}_{\{j \leq T\}}, \quad n \geq 1.$$

i) Montrer que X est une martingale.

ii) Prouver que $\mathbb{P}(T = \infty) > 0$.

iii) Soit $X^* := \sup_{n \geq 1} |X_n|$. Prouver que X^* est presque sûrement fini et que, sur $T = \infty$, $X_{2n-1} = 1$ et $X_{2n} = 0$, pour tout $n \geq 1$ et que p.s. X_n ne converge pas.

2.37. Soit $\{X_n : n \geq 0\}$ une suite de variables aléatoires. On note :

$$\bar{X}_n := \sup_{0 \leq j \leq n} X_j, \underline{X}_n := \inf_{0 \leq j \leq n} X_j, X_n^* := \sup_{0 \leq j \leq n} |X_j|.$$

i) On suppose que X est une sur-martingale, c'est-à-dire $X_n \in L^1$, pour tout n , et $\mathbb{E}[X_{n+1} | \mathcal{F}_n] \leq X_n$. Soit $x \geq 0$. Montrer que :

$$x\mathbb{P}(\bar{X}_n \geq x) \leq \mathbb{E}[X_0] - \mathbb{E}[X_n \mathbf{1}_{\{\bar{X}_n \leq x\}}],$$

$$x\mathbb{P}(\underline{X}_n \leq -x) \leq -\mathbb{E}[X_n \mathbf{1}_{\{\underline{X}_n \leq -x\}}] \leq \mathbb{E}(|X_n|).$$

On pourra introduire les temps d'arrêt $\tau := \inf\{k \geq 0 : X_k \geq x\} \wedge n$ et $\sigma := \inf\{k \geq 0 : X_k \leq -x\} \wedge n$.

ii) Montrer que si X est une sur-martingale positive, on a :

$$x\mathbb{P}(\bar{X}_\infty \geq x) \leq \mathbb{E}(X_0), \forall x \geq 0.$$

En particulier, \bar{X}_∞ est une variable aléatoire finie p.s.

iii) *Inégalité de Doob.* On suppose maintenant que X est une sous-martingale (c'est-à-dire que $-X$ est une sur-martingale) positive. Soit $p > 1$ et q son conjugué : $1/p + 1/q = 1$. Montrer que :

$$x\mathbb{P}(\bar{X}_n \geq x) \leq \mathbb{E}[X_n \mathbf{1}_{\{\bar{X}_n \geq x\}}], \quad x \geq 0, n \geq 0.$$

Établir

$$\mathbb{E}[\bar{X}_n^p] \leq \frac{p}{p-1} \mathbb{E}[X_n \bar{X}_n^{p-1}].$$

On pourra multiplier l'inégalité précédente par x^{p-2} , puis intégrer. Enfin, utiliser l'inégalité de Hölder pour obtenir :

$$\mathbb{E}[\bar{X}_n^p] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[X_n^p], \quad n \geq 0.$$

2.38. Soit $\{\varepsilon_n : n \geq 1\}$ une suite de variables aléatoires indépendantes, de même loi :

$$\mathbb{P}(\varepsilon_1 = 1) = p, \quad \mathbb{P}(\varepsilon_1 = -1) = q, \quad \text{avec } p, q > 0, p + q = 1.$$

On pose $S_0 = 0$, $S_n = \sum_{j=1}^n \varepsilon_j$, pour $n \in \mathbb{N}^*$, et, pour $a \in \mathbb{Z}$:

$$\tau_a := \inf\{n \in \mathbb{N} : S_n = a\}, \quad (\inf \emptyset = +\infty).$$

i) Montrer que τ_a est un temps d'arrêt pour la filtration $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_n = \sigma\{\varepsilon_1, \dots, \varepsilon_n\}$

ii) Soit $0 < u < 1$, $b > 0$ et $X_n := u^n b^{S_n}$. Montrer que $\{X_n : n \geq 0\}$ est une martingale, si $1 = u(pb + q/b)$.

iii) Prouver que :

$$\mathbb{E}[u^{\tau_a} \mathbf{1}_{\{\tau_a < \infty\}}] = \left(\frac{1 - \sqrt{1 - 4pqu^2}}{2qu}\right)^a$$

et que :

$$\mathbb{P}(\tau_a < \infty) = \inf\{1, (p/q)^a\}, \text{ pour } a \in \mathbb{N}.$$

Que se passe-t-il pour a négatif?

iv) On suppose $p < q$ et $a \in \mathbb{N}$. Vérifier que

$$\mathbb{P}(\tau_{-a} < \infty) = 1, \mathbb{P}(\tau_a < \infty) = (p/q)^a.$$

Retrouver ce résultat en utilisant la loi des grands nombres. Prouver que $1 + \bar{S}_\infty$ suit une loi géométrique, où $\bar{S}_\infty := \sup_{n \geq 0} S_n$.

v) On suppose maintenant $p = q = 1/2$. Montrer que, pour tout $a \in \mathbb{Z}$, $\tau_a < \infty$ p.s. Vérifier que

$$\{S_{n+\tau_a} - a : n \geq 0\} \text{ a la même loi que } \{S_n : n \geq 0\}.$$

Enfin, prouver que

$$\limsup_{n \uparrow \infty} S_n = +\infty, \liminf_{n \uparrow \infty} S_n = -\infty, \text{ p.s.}$$

2.39. Soit $\{X_n : n \geq 1\}$ une suite de variables aléatoires indépendantes telles que :

$$X_n := \begin{cases} 1 & \text{avec probabilité } (2n)^{-1} \\ 0 & \text{avec probabilité } 1 - n^{-1} \\ -1 & \text{avec probabilité } (2n)^{-1}. \end{cases}$$

Soit $Y_1 = X_1$ et pour $n \geq 2$:

$$Y_n := \begin{cases} X_n, & \text{si } Y_{n-1} = 0 \\ nY_{n-1}|X_n|, & \text{si } Y_{n-1} \neq 0. \end{cases}$$

i) Montrer que Y est une martingale par rapport à $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$.

ii) Montrer que $\{Y_n : n \geq 1\}$ ne converge pas presque sûrement. Cette suite converge-t-elle?

iii) Pourquoi le théorème de convergence des martingales ne s'applique pas?

2.40. Soit X une martingale telle que, pour tout $n \geq 1$, $\mathbb{E}(X_n) = 0$ et $\mathbb{E}(X_n^2) < \infty$. Montrer que :

$$\mathbb{P}(\max_{1 \leq j \leq n} X_j > x) \leq \frac{\mathbb{E}(X_n^2)}{\mathbb{E}(X_n^2) + x^2}, \quad x > 0.$$

2.41. Soit $\{Y_n : n \geq 1\}$ une suite de variables aléatoires indépendantes de même loi normale $\mathcal{N}(0, \sigma^2)$, où $\sigma > 0$. On pose $\mathcal{F}_n := \sigma(Y_1, \dots, Y_n)$ et $X_n := \sum_{i=1}^n Y_i$. On pose aussi, pour $u \in \mathbb{R}^*$,

$$Z_n^u = \exp\left(uX_n - \frac{1}{2}n u^2 \sigma^2\right).$$

i) Montrer que $\{Z_n^u; n \geq 1\}$ est une \mathcal{F}_n -martingale pour tout $u \in \mathbb{R}^*$.

ii) Montrer que, pour tout $u \in \mathbb{R}^*$, $\{Z_n^u; n \geq 1\}$ converge presque sûrement.

iii) Montrer que

$$K_n := \frac{1}{n} \left(uX_n - \frac{1}{2}n u^2 \sigma^2 \right), \quad n \geq 1$$

converge presque sûrement, et déterminer sa limite.

iv) Trouver la limite presque sûre de $\{Z_n^u : n \geq 1\}$ pour $u \in \mathbb{R}^*$.

v) Trouver $\lim_{n \rightarrow \infty} \mathbb{E}[Z_n^u]$. La martingale $\{Z_n^u : n \geq 1\}$ converge-t-elle dans L^1 ?

2.42. A l'instant 1, une urne contient une boule verte et une boule bleue. On tire une boule et on la remplace par deux boules de la même couleur que celle tirée, ce qui donne une nouvelle composition de l'urne à l'instant 2. On répète alors le procédé pour les instants successifs. On note Y_n le nombre de boules vertes dans l'urne à l'instant n , et $X_n = \frac{Y_n}{n+1}$ la proportion de boules vertes à cet instant. On pose $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$.

i) Montrer que $\mathbb{E}[Y_{n+1} | \mathcal{F}_n] = (Y_n + 1)X_n + Y_n(1 - X_n)$.

ii) Montrer que $\{X_n; n \geq 1\}$ est une \mathcal{F}_n -martingale convergeant presque sûrement vers une variable aléatoire U .

iii) En appliquant le théorème de la convergence dominée, montrer que pour tout $k \geq 1$, on a $\lim_{n \rightarrow \infty} \mathbb{E}[X_n^k] = \mathbb{E}[U^k]$.

On fixe $k \geq 1$. On pose alors, pour $n \geq 1$,

$$Z_n := \frac{Y_n(Y_n + 1) \dots (Y_n + k - 1)}{(n + 1)(n + 2) \dots (n + k)}.$$

En introduisant les variables aléatoires $\mathbf{1}_{\{Y_{n+1}=Y_n\}}$ et $\mathbf{1}_{\{Y_{n+1}=Y_n+1\}}$, montrer que $\{Z_n; n \geq 1\}$ est une \mathcal{F}_n -martingale.

iv) Exprimer la limite presque sûre de Z_n en fonction de la variable aléatoire U .

v) En déduire la valeur de $\mathbb{E}[U^k]$. Montrer que ces moments sont ceux d'une loi $\mathcal{U}_{[0,1]}$.

2.43. Soit une série d'expériences indépendantes, chacune d'entre elles produisant un succès avec probabilité p , un échec avec probabilité q , $p + q = 1$. Soit X_1 (resp. S) le numéro de la première (resp. de la deuxième) tentative qui se solde par un succès, et posons $X_2 = S - X_1$.

i) Indiquer la loi de X_1 , la loi du couple (X_1, X_2) et calculer $\mathbb{P}(X_1 = \ell | S = k)$.

ii) Calculer $\mathbb{E}[X_1 | S = k]$ à l'aide de la formule : $\mathbb{E}[X_1 | S = k] = \sum_{\ell} \ell \mathbb{P}(X_1 = \ell | S = k)$.

iii) Pour des variables aléatoires X_1 et X_2 indépendantes et de même loi *quelconque* (discrètes ou non), mais soit positives, soit intégrables, calculer $\mathbb{E}[X_1 | X_1 + X_2]$ et $\mathbb{E}[X_1 + X_2 | X_1]$.

2.44. On se donne une famille $(X_{k,\ell})_{k,\ell \geq 1}$ de variables de Bernoulli de paramètre p , indépendantes, et on considère une population d'individus qui évolue suivant la règle suivante : la première génération est constituée d'un seul individu, l'ancêtre. Le l -ème individu de la k -ème génération vit une unité de temps puis meurt en donnant naissance à $Y_{k,\ell} = 2X_{k,\ell}$ individus, qui, eux, appartiennent à la $(k + 1)$ -ème génération. L'effectif de la k -ème génération est noté Z_k (par conséquent, $Z_1 = 1$).

i) Calculer $\mathbb{E}[Y_{k,\ell}]$, ainsi que $\mathbb{P}(\frac{Z_k}{2} = j | Z_{k-1} = i)$.

ii) Calculer $\mathbb{E}[Z_k | Z_{k-1}]$ puis $\mathbb{E}[Z_k]$.

iii) Montrer que si $2p < 1$, avec probabilité 1 la population finit par s'éteindre en un temps fini.

2.45. i) On se livre à une suite d'expériences indépendantes, dont les résultats $(E_k)_{k \geq 1}$ sont

de deux types, par exemple "échec" ou "succès", avec probabilité p (resp. q), $p + q = 1$. On note N l'indice de la première expérience conduisant à un succès. Rappeler la loi de N et son espérance.

ii) On se donne une suite de variables aléatoires réelles indépendantes $(X_k)_{k \geq 0}$, de même densité f , densité supposée strictement positive sur \mathbb{R} . On note F la fonction de répartition associée à f . On note T le premier indice $k \geq 1$ tel que $X_k > X_0$. Quelle est la loi conditionnelle de T sachant que $X_0 = x$? Calculer $\mathbb{E}[T \mid X_0]$.

iii) Calculer, directement et également à l'aide de la question précédente, la loi de T et son espérance.

2.46. On suppose que le nombre $N_{a,b}$ d'instructions arrivant au microprocesseur d'un ordinateur pendant l'intervalle de temps $[a, b]$ est une variable aléatoire qui suit la loi de Poisson de paramètre $\lambda(b - a)$.

i) Calculer $\mathbb{E}(N_{a,b})$.

ii) On note X_k (respectivement D_k) le temps utilisé par le microprocesseur pour exécuter la k -ème instruction (respectivement le moment où le microprocesseur commence à exécuter cette k -ème instruction). On note A_k le nombre d'instruction arrivées dans la file d'attente pendant l'exécution de la k -ème instruction. Quelle est la loi conditionnelle de A_k sachant $(X_k, D_k) = (x, d)$? sachant $X_k = x$? En supposant que X_k a pour densité f , calculer $\mathbb{P}(A_k = \ell)$.

iii) On suppose que X_k suit la loi exponentielle de paramètre μ . Calculer $\mathbb{E}(X_k)$, $\mathbb{P}(A_k = \ell)$, $\mathbb{E}(A_k)$. Quelle condition faut-il imposer aux paramètres pour que l'ordinateur fonctionne correctement?

Chapitre 3

Chaînes de Markov

3.1 Définitions et exemples

Soit (E, \mathcal{E}) un espace mesurable; E est appelé l'espace d'état.

En pratique $E = \mathbb{N}, \mathbb{Z}, \mathbb{R}, \mathbb{Z}^d$ ou \mathbb{R}^d .

π désigne une probabilité de transition de E vers \mathcal{E} , rappelons :

- $x \rightarrow \pi(x, A)$ est mesurable pour tout $A \in \mathcal{E}$,
- $A \rightarrow \pi(x, A)$ est une probabilité sur (E, \mathcal{E}) pour tout x de E .

μ est une probabilité sur (E, \mathcal{E}) .

Définition 3.1 Une suite de v.a. $(X_n ; n \geq 0)$ à valeurs dans (E, \mathcal{E}) est dite une chaîne de Markov de loi initiale μ et de probabilité de transition π si

(1) $\mathbb{P}(X_0 \in \cdot) = \mu$

(2) La loi conditionnelle de X_{n+1} sachant X_0, X_1, \dots, X_n est $\pi(X_n, \cdot)$, $n \geq 0$.

Remarques :

1) La famille de v.a. $(X_n; n \geq 0)$ est indexée par \mathbb{N} . \mathbb{N} représente le temps.

2) Une chaîne de Markov modélise les changements "aléatoires" d'une particule à valeurs dans E , avec les règles suivantes :

(i) à l'instant 0, la position de la particule est aléatoire de loi μ .

(ii) à l'instant 1, sachant que $X_0 = x_0$, la particule se déplace aléatoirement suivant la probabilité $\pi(x_0, \cdot)$.

(iii) à l'instant 2, sachant que $X_1 = x_1$, elle saute aléatoirement avec la probabilité $\pi(x_1, \cdot)$. Cette probabilité ne dépend que de la position à l'état précédent.

3) La propriété (2) est équivalente à :

$$\mathbb{E}[f(X_{n+1})g(X_0, \dots, X_n)] = \mathbb{E}[\pi(X_n, f)g(X_0, \dots, X_n)],$$

pour toutes fonctions f et g mesurables et positives. Rappelons :

$$\pi(x, f) = \int_E f(y)\pi(x, dy), \quad f \geq 0; \quad \pi(x, \mathbf{1}_A) = \pi(x, A).$$

Lemme 3.1 1) Soient π une probabilité de transition sur (E, \mathcal{E}) et Φ une fonction $\mathcal{E} \times \mathcal{E}$ -mesurable et positive. Alors

$$x \rightarrow \int \Phi(x, y)\pi(x, dy) \text{ est } \mathcal{E}\text{-mesurable.}$$

2) Soient N_1 et N_2 deux probabilités de transition sur (E, \mathcal{E}) , et

$$N(x, A) = \int N_1(x, dy)N_2(y, A), \quad x \in E, A \in \mathcal{E}.$$

Alors N est une probabilité de transition sur (E, \mathcal{E}) . On note $N = N_1N_2$, le produit de N_1 avec N_2 . En particulier si $n \geq 1$, N^n désigne le produit de N_1 n fois par lui-même.

Proposition 3.1 Une suite de v.a. $(X_n; n \geq 0)$ à valeurs dans (E, \mathcal{E}) , définies sur $(\Omega, \mathcal{A}, \mathbb{P})$ est une chaîne de Markov de loi initiale μ , de probabilité de transition π ssi les deux conditions suivantes ont lieu :

$$\mathbb{E}[f_0(X_0)] = \int f(x_0)\mu(dx_0), \quad (3.1)$$

$$\begin{aligned} \mathbb{E}[f_0(X_0)\dots f_n(X_n)] &= \int f(x_0)\mu(dx_0) \int \pi(x_0, dx_1)f_1(x_1)\dots \\ &\dots \int \pi(x_{n-2}, dx_{n-1})f_{n-1}(x_{n-1}) \int \pi(x_{n-1}, dx_n)f_n(x_n), \quad n \geq 1, \end{aligned} \quad (3.2)$$

pour toutes fonctions f_0, f_1, \dots, f_n positives et mesurables.

Démonstration : a) Supposons que $(X_n; n \geq 0)$ soit une chaîne de Markov. On montre (3.2) par récurrence sur $n \geq 1$. En conditionnement pour X_0, X_1, \dots, X_{n-1} on obtient :

$$\begin{aligned} \mathbb{E}[f_0(X_0)\dots f_n(X_n)] &= \mathbb{E}[f_0(X_0)\dots f_{n-1}(X_{n-1})\mathbb{E}(f_n(X_n)|X_0, X_1, \dots, X_{n-1})] \\ &= \mathbb{E}[f_0(X_0)\dots f_{n-1}(X_{n-1})\pi(X_{n-1}, f_n)] \end{aligned}$$

Il suffit alors d'appliquer l'hypothèse de récurrence au rang $n - 1$.

b) Réciproquement, on peut écrire (3.2) sous la forme :

$$\mathbb{E}[f_0(X_0)\dots f_n(X_n)] = \mathbb{E}[f_0(X_0)\dots f_{n-1}(X_{n-1})\pi(X_{n-1}, f_n)].$$

On en déduit alors que $(X_n; n \geq 0)$ est une chaîne de Markov de probabilité de transition π et de loi initiale μ . \square

Nota : La proposition 3.1 signifie que la loi de (X_0, X_1, \dots, X_n) est entièrement déterminée par π et μ .

Intéressons-nous à présent au cas où E est **fini ou dénombrable**.

Une probabilité ν sur E s'identifie à une fonction $\nu : E \rightarrow [0, 1]$ telle que $\sum_{x \in E} \nu(x) = 1$, où l'on a posé $\nu(\{y\}) = \nu(y)$. Une probabilité de transition π sur E est identifiée à une fonction $\pi : E \times E \rightarrow [0, 1]$ telle que

$$\forall x \in E, \sum_{y \in E} \pi(x, y) = 1.$$

Dans ces conditions une chaîne de Markov $(X_n; n \geq 0)$ à valeurs dans E , de loi initiale μ et de probabilité de transition π est une suite de v.a. telle que :

$$(1)' \mathbb{P}(X_0 = x) = \mu(x)$$

$$(2)' \mathbb{P}(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) = \pi(x_n, x_{n+1}).$$

La proposition 3.1 devient :

$$\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = \mu(x_0)\pi(x_0, x_1)\dots\pi(x_{n-1}, x_n). \quad (3.3)$$

Plus particulièrement lorsque E est un ensemble fini à r éléments, π s'appelle une matrice de transition : $\pi = (\pi(x, y))_{x \in E, y \in E}$. Notons qu'une matrice de transition est alors une matrice carré d'ordre r , dont tous les éléments sont positifs ou nuls, et la somme des éléments situés sur une même ligne vaut 1, et ce quelle que soit la ligne considérée.

Si π_1 et π_2 sont deux matrices de transition, $\pi_1\pi_2$ désigne le produit usuel des deux matrices π_1 et π_2 :

$$\pi_1\pi_2(x, y) = \sum_{z \in E} \pi_1(x, z)\pi_2(z, y).$$

De plus $\pi_1\pi_2$ correspond exactement au produit de deux probabilités de transition. Remarquons que si E est dénombrable, la formule précédente garde un sens puisque $\pi_1(x, z) \geq 0$ et $\pi_2(z, y) \geq 0$.

Exemples.

Exemple 1)

$$E = \{1, 2\}, \quad \pi = (\pi(i, j))_{1 \leq i \leq 2, 1 \leq j \leq 2}, \quad \pi = \begin{pmatrix} \theta_1 & 1 - \theta_1 \\ 1 - \theta_2 & \theta_2 \end{pmatrix}_{0 \leq \theta_1 \leq 1, 0 \leq \theta_2 \leq 1}$$

Lorsque la particule est en 1 (resp. 2) elle reste en 1 (resp. 2) avec probabilité θ_1 (resp. θ_2) et va en 2 (resp. 1) avec probabilité $1 - \theta_1$ (resp. $1 - \theta_2$).

Les cas "extrêmes" correspondent à $\theta_1 = 0$ ou 1, $\theta_2 = 0$ ou 1.

– $\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$ partant de 1, le processus va ensuite en 2 et y reste, 2 est un état "absorbant". Le

cas $\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ est analogue.

– $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ partant de 1, la chaîne va ensuite en 2, puis revient au temps 2, sur le site 1, etc...

– $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ partant de i on reste en i !

On peut aussi étudier :

$$\begin{pmatrix} \theta_1 & 1 - \theta_1 \\ 0 & 1 \end{pmatrix} \text{ et } \begin{pmatrix} \theta_1 & 1 - \theta_1 \\ 1 & 0 \end{pmatrix}$$

Remarquons que lorsque $\theta_1 < 1$ et $\theta_2 < 1$, quel que soit le point de départ, le processus passe infiniment souvent en 1 et 2.

Exemple 2) Soient $(X_n; n \geq 0)$ une suite de v.a. à valeurs dans E , indépendantes et de même loi μ . Alors $(X_n; n \geq 0)$ est une chaîne de Markov de probabilité de transition $\pi(x, dy) = \mu(dy)$. π ne dépend pas de x . Cet exemple n'est pas très intéressant.

Exemple 3) Promenades (ou marches) aléatoires. Soient $(Y_n; n \geq 0)$ une suite de v.a. à valeurs dans \mathbb{R}^d , indépendantes. On note μ la loi de Y_0 , on suppose que les v.a. $Y_1, Y_2, \dots, Y_n \dots$ sont équidistribuées de loi ν . On pose :

$$X_n = Y_0 + Y_1 + \dots + Y_n, \quad n \geq 0.$$

On vérifie que $(X_n; n \geq 0)$ est une chaîne de Markov de loi initiale μ et de probabilité de transition $\pi : \pi(x, \cdot) = \delta_x * \nu$, où δ_x désigne la mesure de Dirac en x , et $*$ la convolution. Donc

$$\pi(x, f) = \int f(x + y)\nu(dy) = \mathbb{E}[f(x + Y_1)], \quad f \geq 0.$$

En particulier si ν est une probabilité sur $\{-1, 0, 1\}$, $(X_n; n \geq 0)$ est un **processus de naissance et de mort**.

Lorsque $\mu = \delta_0$ et $\nu = p\delta_1 + (1 - p)\delta_{-1}$, $(X_n; n \geq 0)$ est la **marche aléatoire de Bernouilli** sur Z issue de 0 de probabilité de transition π , avec $\pi(n, n + 1) = p$ et $\pi(n, n - 1) = q$ pour tout $n \in \mathbb{Z}$.

Proposition 3.2 Soit $(X_n; n \geq 0)$ une chaîne de Markov de probabilité de transition π . Pour tout $n \geq 0$ et $p \geq 1$, $\pi^p(x_n, \cdot)$ est la loi conditionnelle de X_{n+p} sachant $X_0 = x_0, X_1 = x_1, \dots, X_n = x_n$.

Démonstration : Lorsque $p = 1$, c'est exactement la définition d'une chaîne de Markov de probabilité de transition π . On montre la proposition 3.2 par récurrence sur $p \geq 1$. Soient $f : E^{n+1} \rightarrow \mathbb{R}$, $g : E \rightarrow \mathbb{R}$ mesurables et bornées. On note

$$\begin{aligned} \Delta &= \mathbb{E}[g(X_{n+p+1})f(X_0, X_1, \dots, X_n)] \\ &= \mathbb{E}[\mathbb{E}[g(X_{n+p+1})|(X_0, \dots, X_{n+p})]f(X_0, \dots, X_n)] \\ &= \mathbb{E}[\pi(X_{n+p}, g)f(X_0, \dots, X_n)]. \end{aligned}$$

On applique à présent l'hypothèse de récurrence : $\Delta = \mathbb{E}[g_1(X_n)f(X_0, \dots, X_n)]$ avec $g_1(x) = \int \pi(y, g)\pi^p(x, dy) = (\pi^p \cdot \pi)(x, g) = \pi^{p+1}(x, g)$. \square

Remarque.

Lorsque E est un ensemble fini ou dénombrable, l'analogie de la proposition 3.2 est :

$$\mathbb{P}(X_{n+p} = x_{n+p} | X_n = x_n, \dots, X_0 = x_0) = \pi^p(x_n, x_{n+p}). \quad (3.4)$$

3.2 Chaîne canonique. Temps d'arrêt

On définit l'espace canonique $\Omega_0 = E^{\mathbb{N}}$: l'ensemble des suites : $n \rightarrow \omega_n$, à valeurs dans E . On note $(X_n; n \geq 0)$ les applications coordonnées :

$$X_n(\omega) = \omega_n, \quad n \geq 0 \text{ où } \omega = (\omega_n; n \geq 0). \quad (3.5)$$

Pour tout $n \geq 0$, on note \mathcal{A}_n la σ -algèbre engendrée par X_0, X_1, \dots, X_n . \mathcal{A}_n est constituée par les événements : $B = \{(X_0, X_1, \dots, X_n) \in A\}$ où A est un élément de $\mathcal{E}^{\otimes(n+1)}$.

Notons que $\mathcal{A}_n \subset \mathcal{A}_{n+1}$.

\mathcal{A} désigne la tribu engendrée par les applications coordonnées $(X_n; n \geq 0)$. \mathcal{A} est également la tribu engendrée par $\bigcup_{n \geq 0} \mathcal{A}_n$.

π une probabilité de transition sur (E, \mathcal{E}) et μ une probabilité sur (E, \mathcal{E}) .

Le résultat essentiel de ce paragraphe est :

Théorème 3.1 (Ionescu-Tulcea) *Il existe une unique probabilité \mathbb{P}_μ sur Ω_0 telle que les applications coordonnées soient une chaîne de Markov de loi initiale μ et de probabilité de transition π .*

Ainsi, une chaîne de Markov de loi initiale μ et de probabilité de transition π détermine de manière unique une probabilité sur l'espace des suites. Pour simplifier les notations on notera par la suite $\mathbb{P}_x = \mathbb{P}_{\delta_x}$.

Proposition 3.3 1) $\mathbb{P}_x(X_0 = x) = 1$.
 2) $\mathbb{P}_x(f(X_n)) = \pi^n(x, f)$ où $f : (E, \mathcal{E}) \rightarrow (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$
 Lorsque E est dénombrable, $\mathbb{P}_x(X_n = y) = \pi^n(x, y)$.
 3) Pour tout A de \mathcal{A} , $x \rightarrow \mathbb{P}_x(A)$ est \mathcal{E} -mesurable et

$$\mathbb{P}_\mu(A) = \int_E \mathbb{P}_x(A) \mu(dx). \quad (3.6)$$

Démonstration : Le 1) résulte de la définition de \mathbb{P}_x .

2) est une réécriture de la proposition 3.2.

Démontrons 3). Si $A \in \mathcal{A}_n$, on peut écrire $\mathbf{1}_A$ sous la forme : $\mathbf{1}_A = f(X_0, X_1, \dots, X_n)$, d'où

$$\mathbb{P}_x(A) = \mathbb{E}_x(f(x, X_1, \dots, X_n)),$$

$$\mathbb{P}_x(A) = \int \pi(x, dx_1) \int \pi(x_1, dx_2) \dots \int f(x, x_1, \dots, x_n) \pi(x_{n-1}, dx_n).$$

$$\mathbb{P}_\mu(A) = \int \pi(dx_0) \int \pi(x_0, dx_1) \int \pi(x_1, dx_2) \dots \int f(x_0, x_1, \dots, x_n) \pi(x_{n-1}, dx_n).$$

On en déduit la mesurabilité de $x \rightarrow \mathbb{P}_x(A)$, $A \in \bigcup_n \mathcal{A}_n$. On montre que $x \rightarrow \mathbb{P}_x(A)$ est mesurable pour tout A de \mathcal{A} en raisonnant par classe monotone.

D'une manière analogue on montre : $\mathbb{P}_\mu(A) = \int_E \mathbb{P}_x(A)\mu(dx)$ pour tout A de $\bigcup_n \mathcal{A}_n$. On pose $\mathbb{Q}(A) = \int_E \mathbb{P}_x(A)\mu(dx)$, $A \in \mathcal{A}$. Sachant que \mathbb{P}_x est une probabilité sur \mathcal{A} , il est facile de vérifier que \mathbb{Q} est une probabilité sur (Ω_0, \mathcal{A}) . De plus $\mathbb{Q} = \mathbb{P}_\mu$ sur \mathcal{A}_n , ce qui signifie que sous \mathbb{Q} , (X_n) est une chaîne de Markov de loi initiale μ et de probabilité de transition π . D'après le théorème 3.1, $\mathbb{Q} = \mathbb{P}_\mu$ sur \mathcal{A} . \square

Notations.

a) L'opérateur de translation $\theta : E^{\mathbb{N}} \rightarrow E^{\mathbb{N}}$ est défini par

$$\theta((x_n; n \geq 0)) = (x_{n+1}; n \geq 0). \quad (3.7)$$

θ est une application mesurable.

b) On note $\theta_n = \theta^n = \theta \circ \dots \circ \theta$ (n fois). On a :

$$X_n \circ \theta(\omega) = X_{n+1}(\omega), \quad X_n \circ \theta_p(\omega) = X_{n+p}(\omega). \quad (3.8)$$

En effet soit $\omega = (\omega_n; n \geq 0)$. On a successivement,

$$\theta_p(\omega) = (\omega_{n+p}; n \geq 0), \quad X_n(\theta_p(\omega)) = \omega_{n+p} = X_{n+p}(\omega).$$

Théorème 3.2 (*Propriété de Markov*) Soit $(\Omega_0, \mathcal{A}, (X_n; n \geq 0), (\mathbb{P}_x)_{x \in E})$ la chaîne canonique à valeurs dans E . Alors

$$\mathbb{E}_\mu[\Phi \circ \theta_n | \mathcal{A}_n] = \mathbb{E}_{X_n}(\Phi) \quad p.s., \quad (3.9)$$

pour toute v.a.r. Φ , positive ou bornée, \mathcal{A} -mesurable.

Remarques.

1) (3.9) s'écrit sous la forme : $\mathbb{E}_\mu[\Phi \circ \theta_n | \mathcal{A}_n] = h(X_n)$ avec $h(x) = \mathbb{E}_x(\Phi)$. Puisque $x \rightarrow \mathbb{E}_x(\Phi)$ est mesurable, le membre de droite de (3.9) est une v.a., $\sigma(X_n)$ -mesurable.

2) Si $\Phi = f(X_k)$, avec f mesurable positive ou bornée. Alors $\Phi \circ \theta_n = f(X_k \circ \theta_n) = f(X_{n+k})$ et $\mathbb{E}_x(\Phi) = \pi^k(x, f)$. Le théorème 3.2 est une extension de la proposition 3.2.

Démonstration : (théorème 3.2)

Il suffit de prendre $\Phi = \psi(X_0, \dots, X_p)$. Il s'agit par conséquent de montrer :

$$\mathbb{E}_\mu[\psi(X_0, \dots, X_p) \circ \theta_n | \mathcal{A}_n] = \mathbb{E}_{X_n}(\psi(X_0, \dots, X_p)). \quad (3.10)$$

Mais $\psi(X_0, \dots, X_p) \circ \theta_n = \psi(X_0 \circ \theta_n, \dots, X_p \circ \theta_n) = \psi(X_n, \dots, X_{n+p})$.

(3.10) est alors équivalente à :

$$\mathbb{E}_\mu[\psi(X_n, X_{n+1}, \dots, X_{n+p}) | \mathcal{A}_n] = \mathbb{E}_{X_n}(\psi(X_0, \dots, X_p)). \quad (3.11)$$

On montre cette égalité par récurrence sur $p \geq 0$. Si $p = 0$, on a

$$\mathbb{E}_\mu[\psi(X_n) | \mathcal{A}_n] = \mathbb{E}_\mu[\psi(X_n) | X_0, \dots, X_n] = \psi(X_n).$$

Mais $\mathbb{E}_x(\psi(X_0)) = \psi(x)$. L'égalité (3.11) est bien réalisée si $p = 0$. Supposons à présent que (3.11) est réalisée pour un certain $p \geq 0$, montrons que (3.11) a encore lieu lorsque p est remplacé par $p + 1$. Sachant que $\mathcal{A}_n \subset \mathcal{A}_{n+p}$, on écrit :

$$\xi = \mathbb{E}_\mu[\psi(X_n, X_{n+1}, \dots, X_{n+p+1}) | \mathcal{A}_n] = \mathbb{E}_\mu[\mathbb{E}_\mu[\psi(X_n, \dots, X_{n+p+1}) | \mathcal{A}_{n+p}] | \mathcal{A}_n].$$

On applique la propriété de Markov, il vient

$$\xi = \mathbb{E}_\mu \left[\int_E \psi(X_n, \dots, X_{n+p}, y) \pi(X_{n+p}, dy) | \mathcal{A}_n \right]$$

On utilise l'hypothèse de récurrence :

$$\xi = \mathbb{E}_{X_n} \left(\int_E \psi(X_0, \dots, X_p, y) \pi(X_p, dy) \right)$$

Mais $\mathbb{E}_x \left[\int_E \psi(X_0, \dots, X_p, y) \pi(X_p, dy) \right] = \mathbb{E}_x[\psi(X_0, \dots, X_p, X_{p+1})]$. Donc

$$\xi = \mathbb{E}_{X_n}(\psi(X_0, \dots, X_{p+1})).$$

□

Nous allons à présent définir la notion de temps d'arrêt. Ces variables aléatoires vont jouer un rôle essentiel dans le paragraphe suivant.

Définitions.

- 1) Une filtration $\mathcal{F} = (\mathcal{F}_n; n \geq 0)$ est une suite croissante de tribus incluses dans une tribu donnée $\mathcal{F}' : \mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}'$ pour tout $n \geq 0$.
- 2) Un processus $(X_n; n \geq 1)$ est une suite de v.a. On dit que ce processus est \mathcal{F} -adapté si pour tout $n \geq 0$, X_n est \mathcal{F}_n -mesurable.
- 3) Un temps d'arrêt est une v.a. $T : (\Omega, \mathcal{A}) \rightarrow (\mathbb{N} \cup \{+\infty\}), \mathcal{P}(\mathbb{N} \cup \{+\infty\})$ telle que, pour tout $n \geq 0$, $\{T \leq n\} \in \mathcal{F}_n$.

Un temps d'arrêt est relatif à une filtration. Il peut être infini. Une v.a. constante est un temps d'arrêt.

Rappelons les résultats suivants :

- Proposition 3.4** 1) Soit T une v.a. de $(\Omega, \mathcal{F}') \rightarrow \mathbb{N} \cup \{+\infty\}$. T est un temps d'arrêt ssi $\{T = n\} \in \mathcal{F}_n$ pour tout $n \geq 0$.
- 2) Si T et T' sont deux temps d'arrêt, alors $T \wedge T'$, $T \vee T'$ et $T + T'$ sont des temps d'arrêt.
- 3) Si $(X_n; n \geq 0)$ est un processus adapté à valeurs dans (U, \mathcal{U}) , la v.a. $D_F = \inf\{n \geq 0; X_n \in F\}$ où $F \in \mathcal{U}$ est un temps d'arrêt. (on fait la convention : $\inf \emptyset = +\infty$).

Proposition 3.5 Soit T un temps d'arrêt.

- 1) On définit $\mathcal{F}_T = \{A \in \mathcal{F}'; A \cap \{T \leq n\} \in \mathcal{F}_n, \forall n\}$. Alors \mathcal{F}_T est une sous-tribu de \mathcal{F}' .
- 2) T une v.a. \mathcal{F}_T -mesurable.
- 3) Si $(X_n; n \geq 0)$ est un processus adapté à valeurs dans (U, \mathcal{U}) et $T < +\infty$, alors X_T est \mathcal{F}_T -mesurable, la v.a. X_T étant définie par $X_T(\omega) = X_{T(\omega)}(\omega)$.
- 4) Si T et T' sont deux temps d'arrêt $\mathcal{F}_{T \wedge T'} = \mathcal{F}_T \cap \mathcal{F}_{T'}$; si de plus $T \leq T'$ alors $\mathcal{F}_T \subset \mathcal{F}_{T'}$.

Démonstration : 1) $\Omega \cap \{T \leq n\} = \{T \leq n\} \in \mathcal{F}_n$, Ω est élément de \mathcal{F}_T . Si $A \in \mathcal{F}_T$,

$$B = A^c \cap \{T \leq n\} = (A^c \cup \{T > n\}) \cap \{T \leq n\} = (A \cap \{T \leq n\})^c \cap \{T \leq n\}.$$

Donc B appartient à \mathcal{F}_n . Par conséquent $A^c \in \mathcal{F}_T$.

Soit (A_k) une suite d'éléments de \mathcal{F}_T . On note $A = \bigcup_k A_k$ et,

$$C = A \cap \{T \leq n\} = \bigcup_k (A_k \cap \{T \leq n\}).$$

Puisque $A_k \in \mathcal{F}_T$, \mathcal{F}_n est une tribu, $C \in \mathcal{F}_n$. Ce qui signifie que $A \in \mathcal{F}_T$. \mathcal{F}_T est une sous-tribu de \mathcal{F}' .

2) $\{T = n\} \in \mathcal{F}_T$ car $\{T = n\} \cap \{T \leq m\}$ est égal à \emptyset si $n > m$, et égal à $\{T = n\}$ sinon. Donc cet événement appartient à \mathcal{F}_m .

3) Soit $A \in \mathcal{U}$. Alors $\{X_T \in A\} \cap \{T \leq n\} = \bigcup_{k=0}^n \{T = k, X_k \in A\} \in \mathcal{F}_n$. Par conséquent $X_T : (\Omega, \mathcal{F}_T) \rightarrow (U, \mathcal{U})$ est une application mesurable.

4) a) Si $T \leq T'$. Soit $A \in \mathcal{F}_T$. $A \in \{T' \leq n\} = (A \in \{T \leq n\}) \cap \{T' \leq n\}$ donc $A \in \mathcal{F}_{T'}$. Ce qui prouve $\mathcal{F}_T \subset \mathcal{F}_{T'}$.

b) Revenons au cas général : de $T \wedge T' \leq T$ et $T \wedge T' \leq T'$ on tire $\mathcal{F}_{T \wedge T'} \subset \mathcal{F}_T$ et $\mathcal{F}_{T \wedge T'} \subset \mathcal{F}_{T'}$ d'où $\mathcal{F}_{T \wedge T'} \subset \mathcal{F}_T \cap \mathcal{F}_{T'}$. Montrons l'inclusion réciproque, soit $A \in \mathcal{F}_T \cap \mathcal{F}_{T'}$.

$$A \cap \{T \wedge T' \leq n\} = A \cap (\{T \leq n\} \cup \{T' \leq n\}) = (A \cap \{T \leq n\}) \cup (A \cap \{T' \leq n\}) \in \mathcal{F}_n.$$

Par conséquent $\mathcal{F}_T \cap \mathcal{F}_{T'} \subset \mathcal{F}_{T \wedge T'}$. □

Remarque.

Soient T le temps d'arrêt constant égal à k et $\mathcal{F}_T = \{A \in \mathcal{F}'; A \cap \{T \leq n\} \in \mathcal{F}_n; \forall n \geq 0\}$. Mais $\{T \leq n\} = \Omega$ si $k \leq n$, et \emptyset si $k > n$. Par conséquent $\mathcal{F}_T = \{A \in \mathcal{F}'; A \in \mathcal{F}_n, \forall n \geq k\} = \mathcal{F}_k$. Il n'y a donc aucune ambiguïté, lorsque $T = k$, \mathcal{F}_T coïncide avec \mathcal{F}_k .

Revenons aux chaînes de Markov sur l'espace canonique Ω_0 . \mathcal{A}_n désigne la tribu engendrée par les v.a. X_0, X_1, \dots, X_n et $\mathcal{A}_n \subset \mathcal{A}$. Il est clair que $(\mathcal{A}_n; n \geq 0)$ est une filtration. Un temps d'arrêt sera toujours relatif à $(\mathcal{A}_n; n \geq 0)$. Si T est un temps d'arrêt, on définit : $\theta_T : E^{\mathbb{N}} \rightarrow E^{\mathbb{N}}$ par

$$\theta_T(\omega) = \theta_{T(\omega)}(\omega) \text{ si } \{T(\omega) < \infty\}. \quad (3.12)$$

En particulier si $T(\omega) = k$, $X_n \circ \theta_T(\omega) = X_n \circ \theta_k(\omega) = X_{n+k}(\omega)$. Donc

$$X_n \circ \theta_T = X_{n+T} \text{ sur } \{T < \infty\}. \quad (3.13)$$

Cette propriété se généralise, si T et T' sont deux temps d'arrêt,

$$X_{T'} \circ \theta_T = X_{T'+T} \text{ sur } \{T < \infty, T' < \infty\}. \quad (3.14)$$

Théorème 3.3 (*Propriété de Markov forte*) Soient $(\Omega_0, \mathcal{A}, (X_n; n \geq 0), (\mathbb{P}_x)_{x \in E})$ la chaîne canonique à valeurs dans E et T un temps d'arrêt. Alors

$$\mathbb{E}_\mu[\Phi \circ \theta_T \mathbf{1}_{\{T < \infty\}} | \mathcal{A}_T] = \mathbb{E}_{X_T}(\Phi) \mathbf{1}_{\{T < \infty\}} \quad p.s.$$

pour toute v.a.r. Φ , positive ou bornée, \mathcal{A} -mesurable.

Démonstration : Puisque $x \rightarrow \mathbb{E}_x(\Phi)$ est \mathcal{E} -mesurable et $X_T \mathbf{1}_{\{T < \infty\}}$ est une v.a. \mathcal{A}_T -mesurable, la v.a. $\mathbb{E}_{X_T}(\Phi) \mathbf{1}_{\{T < \infty\}}$ est \mathcal{A}_T -mesurable. Soit A un élément de \mathcal{A}_T . Posons

$$a = \mathbb{E}_\mu[\mathbf{1}_A \Phi \circ \theta_T \mathbf{1}_{\{T < \infty\}}].$$

On a :

$$a = \sum_{n \geq 0} \mathbb{E}_\mu[\mathbf{1}_A \mathbf{1}_{\{T=n\}} \Phi \circ \theta_T] = \sum_{n \geq 0} \mathbb{E}_\mu[\mathbf{1}_{A \cap \{T=n\}} \Phi \circ \theta_n]$$

Mais $A \cap \{T = n\} \in \mathcal{A}_n$, on conditionne par \mathcal{A}_n , il vient :

$$a = \sum_{n \geq 0} \mathbb{E}_\mu[\mathbf{1}_{A \cap \{T=n\}} \mathbb{E}_\mu[\Phi \circ \theta_n | \mathcal{A}_n]]$$

On applique le théorème 3.2,

$$\begin{aligned} a &= \sum_{n \geq 0} \mathbb{E}_\mu[\mathbf{1}_{A \cap \{T=n\}} \mathbb{E}_{X_n}(\Phi)] = \sum_{n \geq 0} \mathbb{E}_\mu[\mathbf{1}_A \mathbf{1}_{\{T=n\}} \mathbb{E}_{X_n}(\Phi)] \\ &= \mathbb{E}_\mu[\mathbb{E}_{X_T}(\Phi) \mathbf{1}_A \mathbf{1}_{\{T < \infty\}}]. \end{aligned}$$

□

Exemple.

Soit $\Phi = h(X_1)$. Alors $\Phi \circ \theta_T \mathbf{1}_{\{T < \infty\}} = h(X_1 \circ \theta_T) \mathbf{1}_{\{T < \infty\}} = h(X_{T+1}) \mathbf{1}_{\{T < \infty\}}$. Pour ce choix de Φ , on a,

$$\mathbb{E}[\Phi \circ \theta_T | \mathcal{A}_T] = \mathbb{E}[h(X_{T+1}) | \mathcal{A}_T] = \mathbb{E}_{X_T}(h(X_1)) = \pi(X_T, h).$$

Corollaire 3.1 Soit T un temps d'arrêt fini. Le processus $(X_{T+n}; n \geq 0)$ est une chaîne de Markov de loi initiale ν , de probabilité de transition π , ν désignant la loi de X_T sous \mathbb{P}_μ .

Démonstration : Posons $Y_n = X_{T+n}$. Il est clair que,

$$\mathbb{P}_\mu(Y_0 \in \cdot) = \mathbb{P}_\mu(X_T \in \cdot) = \nu.$$

Il s'agit de montrer : $\mathbb{E}_\mu[f(Y_{n+1}) | Y_0, \dots, Y_n] = \pi(Y_n, f)$, où f est une fonction \mathcal{E} -mesurable bornée. On introduit $\Phi = f(X_1)$. Alors $f(Y_{n+1}) = f(X_{T+n+1}) = \Phi \circ \theta_U$ où l'on a posé $U = T + n$. U est la somme de deux temps d'arrêt, U est donc un temps d'arrêt. Soit $0 \leq k \leq n$. Puisque $Y_k = X_{T+k}$, Y_k est une v.a. \mathcal{A}_{T+k} -mesurable; mais $T + k \leq T + n$, donc $\mathcal{A}_{T+k} \subset \mathcal{A}_{T+n}$.

Par conséquent pour tout $k \in \{0, 1, \dots, n\}$, Y_k est $\mathcal{A}_{T+n} = \mathcal{A}_U$ -mesurable. Ce qui signifie : $\sigma(Y_0, \dots, Y_n) \subset \mathcal{A}_U$. On en déduit,

$$a = \mathbb{E}_\mu[f(Y_{n+1})|Y_0, \dots, Y_n] = \mathbb{E}_\mu[\mathbb{E}_\mu[\Phi \circ \theta_U | \mathcal{A}_U] | Y_0, Y_1, \dots, Y_n].$$

Mais d'après le théorème 3.3, $\mathbb{E}_\mu[\Phi \circ \theta_U] = \mathbb{E}_{X_U}(\Phi)$. De plus $\sigma(Y_0, Y_1, \dots, Y_n) = \sigma(X_T, X_{T+1}, \dots, X_{T+n-1}, X_U)$. En particulier X_U est $\sigma(Y_0, Y_1, \dots, Y_n)$ -mesurable. D'où

$$a = \mathbb{E}_{X_U}(\Phi) = \mathbb{E}_{X_U}(f(X_1)) = \pi(X_U, f) = \pi(Y_n, f).$$

□

Remarque.

En particulier si $T = k$, le processus $(X_{k+n}; n \geq 0)$ est une chaîne de Markov de probabilité de transition π .

3.3 Potentiel. États récurrents, états transients

π est une probabilité de transition sur (E, \mathcal{E}) et $(\Omega_0, \mathcal{A}, (\mathbb{P}_x)_{x \in E})$ est sa réalisation canonique. On supposera que \mathcal{E} contient les points : $\{x\} \in \mathcal{E}$, pour tout x de E .

Définition 3.2 *Le noyau*

$$U = \sum_{n \geq 0} \pi^n = Id + \pi + \pi^2 + \dots + \pi^n + \dots$$

s'appelle le noyau potentiel de la chaîne.

On a donc

$$U(x, B) = \sum_{n \geq 0} \pi^n(x, B) = \sum_{n \geq 0} \mathbb{P}_x(X_n \in B) = \mathbb{E}_x \left(\sum_{n \geq 0} \mathbf{1}_{\{X_n \in B\}} \right); B \in \mathcal{E}. \quad (3.15)$$

Plus généralement si f est \mathcal{E} -mesurable et positive,

$$U(x, f) = \mathbb{E}_x \left(\sum_{n \geq 0} f(X_n) \right).$$

Remarquons :

$$U = Id + \pi(Id + \pi + \pi^2 + \dots + \pi^n + \dots) = Id + \pi U.$$

D'une manière analogue $U = Id + U\pi$. En particulier $Id = (Id - \pi)U = U(Id - \pi)$, U est "l'inverse" de $Id - \pi$.

Définitions.

1) Pour tout A de \mathcal{E} , N_A est le nombre de fois où la chaîne $(X_n; n \geq 0)$ visite A :

$$N_A = \sum_{n \geq 0} \mathbf{1}_{\{X_n \in A\}}. \quad (3.16)$$

Lorsque $A = \{x\}$, on note $N_x = N_{\{x\}}$.

Par conséquent, U admet l'interprétation probabiliste suivante :

$$U(x, A) = \mathbb{E}_x(N_A), \quad x \in E, \quad A \in \mathcal{E}. \quad (3.17)$$

2) Soient $x \in E$ et σ_x le temps d'arrêt :

$$\sigma_x = \inf\{n \geq 1, X_n = x\}$$

avec la convention $\inf \emptyset = +\infty$. σ_x est le premier instant (plus grand que 1) de retour en x . Remarquons :

$$(i) \quad 1 \leq \sigma_x \leq +\infty; \quad (ii) \quad \text{sur } \{\sigma_x < \infty\}, \quad X_{\sigma_x} = x. \quad (3.18)$$

3) Un état $x \in E$ est dit récurrent si $\mathbb{P}_x(\sigma_x < \infty) = 1$. Lorsque $\mathbb{P}_x(\sigma_x < \infty) < 1$, x est dit transient ou transitoire.

Théorème 3.4 *Soit $x \in E$. Il n'y a que deux cas possibles :*

1) x est récurrent et alors $\mathbb{P}_x(N_x = \infty) = 1$.

2) x est transient et alors $\mathbb{P}_x(N_x < \infty) = 1$.

Remarque.

On a x est récurrent si et seulement si $\mathbb{P}_x(N_x = \infty) = 1$, x est transient si et seulement si $\mathbb{P}_x(N_x < \infty) = 1$.

Démonstration : (théorème 3.4)

1) On définit par récurrence la suite $(\sigma^{(n)}; n \geq 1)$ de temps d'arrêt :

$$\sigma^{(1)} = \sigma_x; \quad \sigma^{(n+1)} = \inf\{k > \sigma^{(n)}; X_k = x\}. \quad (3.19)$$

$\{\sigma^{(n)}; n \geq 1\}$ est l'ensemble des temps de passages successifs en x , plus précisément :

$$\{\sigma^{(n)}; \sigma^{(n)} < \infty \text{ et } n \geq 1\} = \{n \geq 1; X_n = x\}. \quad (3.20)$$

Puisque $\sigma^{(n+1)} \geq \sigma^{(n)} + 1$, la suite $n \rightarrow \sigma^{(n)}$ est strictement croissante, de plus,

$$\sigma^{(n+1)} = \sigma^{(n)} + \inf\{k > 0, X_{k+\sigma^{(n)}} = x\}, \quad \text{sur } \{\sigma^{(n)} < \infty\}.$$

Par conséquent,

$$\sigma^{(n+1)} = \sigma^{(n)} + \sigma^{(1)} \circ \theta_{\sigma^{(n)}}, \quad \text{sur } \{\sigma^{(n)} < \infty\}. \quad (3.21)$$

2) Soit $x \in E$ fixé, on pose $a_n = \mathbb{P}_x(\sigma^{(n)} < \infty)$. Montrons

$$a_n = a_1^n; \quad n \geq 1. \quad (3.22)$$

On déduit de (3.21),

$$a_{n+1} = \mathbb{P}_x(\sigma^{(n+1)} < \infty) = \mathbb{P}_x(\sigma^{(n)} < \infty, \sigma^{(1)} \circ \theta_{\sigma^{(n)}} < \infty).$$

Posons $\Phi = \mathbf{1}_{\{\sigma^{(1)} < \infty\}} = \mathbf{1}_{\{\sigma_x < \infty\}}$. Alors

$$a_{n+1} = \mathbb{E}_x \left[\mathbf{1}_{\{\sigma^{(n)} < \infty\}} \Phi \circ \theta_{\sigma^{(n)}} \right].$$

L'événement $\{\sigma^{(n)} < \infty\}$ appartient à $\mathcal{A}_{\sigma^{(n)}}$, donc

$$a_{n+1} = \mathbb{E}_x \left[\mathbf{1}_{\{\sigma^{(n)} < \infty\}} \mathbb{E}_x \left[\mathbf{1}_{\{\sigma^{(n)} < \infty\}} \Phi \circ \theta_{\sigma^{(n)}} \mid \mathcal{A}_{\sigma^{(n)}} \right] \right].$$

On applique le théorème 3.3, il vient,

$$a_{n+1} = \mathbb{E}_x \left[\mathbf{1}_{\{\sigma^{(n)} < \infty\}} \mathbb{E}_{X_{\sigma^{(n)}}}(\Phi) \right]$$

Mais d'après (3.18) (ii), sur $\{\sigma^{(n)} < \infty\}$, $X_{\sigma^{(n)}} = x$. D'où,

$$a_{n+1} = \mathbb{P}_x(\sigma^{(n)} < \infty) \mathbb{E}_x(\Phi) = a_n \mathbb{P}_x(\sigma^{(1)} < \infty) = a_n a_1; \quad n \geq 1.$$

La suite $(a_n)_{n \geq 1}$, est géométrique de raison a_1 , (3.22) s'en déduit immédiatement.

3) On remarque que $\{\sigma^{(n+1)} < \infty\} \subset \{\sigma^{(n)} < \infty\}$, donc

$$\mathbb{P}_x \left(\bigcap_{n \geq 1} \{\sigma^{(n)} < \infty\} \right) = \lim_{n \rightarrow \infty} \mathbb{P}_x(\sigma^{(n)} < \infty).$$

D'après la relation (3.20), on a : $\bigcap_{n \geq 1} \{\sigma^{(n)} < \infty\} = \{N_x = \infty\}$. Il est alors aisé d'en déduire que si $a_1 < 1$ (resp. $a_1 = 1$), $\mathbb{P}_x(N_x = \infty) = 0$ (resp. $\mathbb{P}_x(N_x = \infty) = 1$). \square

Corollaire 3.2 *Si x est transient, sous \mathbb{P}_x , N_x est une v.a. géométrique de paramètre $p = 1 - \mathbb{P}_x(\sigma_x < \infty) = \mathbb{P}_x(\sigma_x = \infty)$:*

$$\mathbb{P}_x(N_x = k) = (1 - p)^{k-1} p; \quad k \geq 1.$$

Démonstration : Soit $k \geq 2$. On utilise la suite $(\sigma^{(n)}; n \geq 1)$ définie par (3.19). On a :

$$\{N_x \geq k\} = \{\sigma^{(k-1)} < \infty\}, \quad \mathbb{P}_x \text{ presque sûrement.} \quad (3.23)$$

Donc $\mathbb{P}_x(N_x \geq k) = \mathbb{P}_x(\sigma^{(k-1)} < \infty) = a_{k-1} = a_1^{k-1} = (1 - p)^{k-1}$. \square

Théorème 3.5 Soit x un point de E . On a les équivalences :

(i) x est récurrent $\iff U(x, \{x\}) = \infty$.

(ii) x est transient $\iff U(x, \{x\}) < \infty$.

Dans le second cas $U(x, \{x\}) = \frac{1}{1 - \mathbb{P}_x(\sigma_x < \infty)} = \frac{1}{\mathbb{P}_x(\sigma_x = \infty)}$.

Démonstration : a) Si x est récurrent, $\mathbb{P}_x(N_x = \infty) = 1$. Mais $U(x, \{x\}) = \mathbb{E}_x(N_x)$. Donc $U(x, \{x\}) = \infty$.

b) Supposons x transient. Rappelons que si ξ est une v.a. de loi géométrique de paramètre p , i.e., $\mathbb{P}(\xi = n) = p(1-p)^{n-1}$ pour tout $n \geq 1$, alors $\mathbb{E}(\xi) = 1/p$. On déduit alors du corollaire 3.2,

$$U(x, \{x\}) = \mathbb{E}_x(N_x) = 1/p.$$

□

Nous allons à présent étudier quelques exemples.

Exemple 1 : On reprend l'exemple de la chaîne de Markov à valeurs dans $E = \{1, 2\}$ (exemple 1, paragraphe 1). Etudions le cas où $\theta_1 < 1$ et $\theta_2 < 1$. On a :

$$\mathbb{P}_1(\sigma_1 \geq k+1) = \mathbb{P}_1(X_1 = 2, X_2 = 2, \dots, X_k = 2) = \pi(1, 2)\pi(2, 2)\dots\pi(2, 2).$$

Par conséquent,

$$\mathbb{P}_1(\sigma_1 \geq k+1) = (1 - \theta_1)\theta_2^{k-1}; \quad k \geq 2.$$

Mais $\theta_2 < 1$, si on fait tendre k vers l'infini, on a $\mathbb{P}_1(\sigma_1 = +\infty) = 0$. Donc $\mathbb{P}_1(\sigma_1 < \infty) = 1$, l'état 1 est récurrent. Une approche analogue montrerait que 2 est également récurrent.

Exemple 2 : Marches aléatoires de Bernouilli à valeurs dans \mathbb{Z} .

On conserve les notations de l'exemple 3 du paragraphe 1. On suppose que $(Y_n; n \geq 0)$ est une suite de v.a. à valeurs dans \mathbb{Z} , indépendantes, et que les v.a. Y_n , pour $n \geq 1$, ont la même loi ν . La chaîne associée est $X_n = \sum_{k=0}^n Y_k$.

$$\pi(x, x+1) = \nu(1) = p, \quad \pi(x, x-1) = \nu(-1) = q,$$

où $p > 0$, $q > 0$, et $p + q = 1$. Montrons que tous les états sont récurrents (resp. transients) si $p = 1/2$ (resp. $p \neq 1/2$). Pour ce faire on va calculer $\pi^n(x, x)$ puis étudier la finitude de $U(x, x)$. Soit $x \in \mathbb{Z}$. On a,

$$\pi^n(x, x) = \mathbb{P}\left(x + \sum_{k=1}^n Y_k = x\right) = \mathbb{P}\left(\sum_{k=1}^n Y_k = 0\right). \quad (3.24)$$

Posons $\bar{Y}_n = \left(\frac{1 + Y_n}{2}\right)$. Les v.a. $(\bar{Y}_n; n \geq 0)$ sont indépendantes, ont même loi et

$$\mathbb{P}(\bar{Y}_1 = 1) = p, \quad \mathbb{P}(\bar{Y}_1 = 0) = 1 - p.$$

Alors

$$\pi^n(x, x) = \mathbb{P} \left(\sum_{k=1}^n (2\bar{Y}_k - 1) = 0 \right) = \mathbb{P} \left(2 \left(\sum_{k=1}^n \bar{Y}_k \right) = n \right). \quad (3.25)$$

La v.a. $\left(\sum_{k=1}^n \bar{Y}_k \right)$ suit une loi binomiale $\mathcal{B}(n, p)$, elle est en particulier à valeurs dans \mathbb{N} , donc

$$\pi^n(x, x) = 0 \text{ si } n \text{ est impair,} \quad (3.26)$$

et

$$\pi^{2n}(x, x) = \mathbb{P} \left(\sum_{k=1}^n \bar{Y}_k = n \right) = C_{2n}^n (pq)^n = \frac{(2n)!}{(n!)^2} (pq)^n; \quad n \geq 0. \quad (3.27)$$

Rappelons la formule de Stirling :

$$n! \sim \sqrt{2\pi n} n^{n+1/2} e^{-n} \text{ lorsque } n \rightarrow \infty. \quad (3.28)$$

– si $p = 1/2$

$$\pi^{2n}(x, x) = \frac{(2n)!}{(n!)^2} \left(\frac{1}{2} \right)^{2n} \sim \frac{1}{\sqrt{\pi n}}. \quad (3.29)$$

Puisque $U(x, x) = \sum_{n \geq 0} \pi^{2n}(x, x)$, $U(x, x) = \infty$; par conséquent tous les états sont récurrents.

– si $p \neq 1/2$

On sait que la fonction $x \in [0, 1] \rightarrow x(1-x)$ admet un unique maximum pour $x = 1/2$, et ce maximum vaut $1/4$. Par conséquent $p(1-p) < 1/4$ et $a = 4p(1-p) < 1$. On en déduit,

$$\pi^{2n}(x, x) = a^n \left(\frac{(2n)!}{(n!)^2} \left(\frac{1}{2} \right)^{2n} \right) \sim \frac{1}{\sqrt{\pi}} \frac{a^n}{\sqrt{n}}. \quad (3.30)$$

La série de terme général $\pi^{2n}(x, x)$ est convergente, donc $U(x, x) < \infty$, x est transient.

Proposition 3.6 Soient $x \neq y$. Alors

$$U(x, \{y\}) = \mathbb{P}_x(\sigma_y < \infty) U(y, \{y\}),$$

avec la convention $0 \times (+\infty) = 0$.

Démonstration : Rappelons que $U(x, \{y\}) = \mathbb{E}(N_y)$. On a,

$$\mathbb{E}_x(N_y) = \mathbb{E}_x \left[\sum_{n \geq 0} \mathbf{1}_{\{X_n=y\}} \right] = \mathbb{E}_x \left[\left(\sum_{n \geq \sigma_y} \mathbf{1}_{\{X_n=y\}} \right) \mathbf{1}_{\{\sigma_y < \infty\}} \right].$$

(la deuxième égalité a lieu car $x \neq y$).

$$\begin{aligned} U(x, \{y\}) &= \mathbb{E}_x \left(\left(\sum_{m \geq 0} \mathbf{1}_{\{X_{m+\sigma_y}=y\}} \right) \mathbf{1}_{\{\sigma_y < \infty\}} \right) \\ &= \mathbb{E}_x \left(\left(\sum_{m \geq 0} \mathbf{1}_{\{X_m=y\}} \right) \circ \theta_{\sigma_y} \mathbf{1}_{\{\sigma_y < \infty\}} \right). \end{aligned}$$

En introduisant la v.a. N_y , on réécrit la formule précédente sous la forme suivante :

$$U(x, \{y\}) = \mathbb{E}_x(N_y \circ \theta_{\sigma_y} \mathbf{1}_{\{\sigma_y < \infty\}}) = \mathbb{E}_x(\mathbb{E}_x(N_y \circ \theta_{\sigma_y} \mathbf{1}_{\{\sigma_y < \infty\}} | \mathcal{A}_{\sigma_y})).$$

On applique la propriété de Markov forte :

$$U(x, \{y\}) = \mathbb{E}_x \left(\mathbf{1}_{\{\sigma_y < \infty\}} \mathbb{E}_{X_{\sigma_y}}(N_y) \right) = \mathbb{E}_x \left(\mathbf{1}_{\{\sigma_y < \infty\}} \mathbb{E}_y(N_y) \right).$$

D'où $U(x, \{y\}) = \mathbb{P}_x(\sigma_y < \infty) \mathbb{E}_y(N_y) = \mathbb{P}_x(\sigma_y < \infty) U(y, \{y\})$. □

Remarque.

Soit y fixé. Puisque $\mathbb{P}_x(\sigma_y < \infty) \leq 1$, la fonction $x \in E \rightarrow U(x, \{y\})$ atteint son maximum pour $y = x$.

Donnons une première application de la proposition 3.6.

Proposition 3.7 *Si E est fini, il existe au moins un état récurrent.*

Démonstration : En effet

$$\sum_{y \in E} N_y = N_E = +\infty.$$

On prend l'espérance de part et d'autre de cette égalité, il vient,

$$+\infty = \sum_{y \in E} \mathbb{E}_x(N_y) = \sum_{y \in E} U(x, \{y\}).$$

Puisque la somme comporte un nombre fini de termes, il existe $y \in E$ tel que $U(x, \{y\}) = +\infty$. Si $y = x$, alors $U(x, \{x\}) = +\infty$, x est alors récurrent. Etudions à présent le cas où $y \neq x$. On applique la proposition 3.6 :

$$U(x, \{y\}) = \mathbb{P}_x(\sigma_y < \infty) U(y, \{y\}) \leq U(y, \{y\}).$$

Par conséquent $U(y, \{y\}) = +\infty$, y est récurrent. □

On suppose désormais que E est un ensemble fini ou dénombrable.

Définitions.

- 1) Un point x de E conduit à y de E si $x = y$ ou $x \neq y$ et $\mathbb{P}_x(\sigma_y < \infty) > 0$. On note $x \rightarrow y$.
- 2) Deux points x et y de E communiquent si $x \rightarrow y$ et $y \rightarrow x$. On note $x \longleftrightarrow y$.

- 3) Une partie A de \mathcal{E} est dite fermée (ou absorbante) si A est non vide et si pour tout x de A , $\mathbb{P}_x(X_n \in A, \forall n \geq 0) = 1$.
 x appartenant à E , est dit absorbant si $\{x\}$ est fermé. Par conséquent x est absorbant si $\mathbb{P}_x(X_n = x; \forall n \geq 0) = 1$.
- 4) La chaîne est dite irréductible si la seule partie fermée non vide est E .
- 5) \mathcal{R} désigne l'ensemble des états récurrents.
- 6) On note $U(x, y) = U(x, \{y\})$.

Lemme 3.2 Soient x et y deux éléments distincts de E . On a alors la série d'équivalences :

$$\begin{aligned} x \longrightarrow y &\iff \mathbb{P}_x(\sigma_y < \infty) > 0 \iff \mathbb{E}_x(N_y) = U(x, y) > 0 \\ &\iff \mathbb{P}_x(\exists n \geq 0, X_n = y) > 0 \iff \exists n; \pi^n(x, y) = \mathbb{P}_x(X_n = y) > 0. \end{aligned}$$

En particulier, x ne conduit pas à y si et seulement si

$$\mathbb{P}_x(\sigma_y < \infty) = 0 \iff U(x, y) = 0 \iff \mathbb{P}_x(X_n \neq y; \forall n \geq 0) = 0.$$

Démonstration : Par définition $x \rightarrow y \iff \mathbb{P}_x(\sigma_y < \infty) > 0$.

N_y est une v.a. positive. Donc $\mathbb{E}_x(N_y) = 0 \iff \mathbb{P}_x$ presque sûrement, $N_y = 0$. Mais $N_y = \sum_{n \geq 0} \mathbf{1}_{\{X_n = y\}}$. Donc

$$\mathbb{E}_x(N_y) = 0 \iff \forall n \geq 0, \mathbf{1}_{\{X_n = y\}} = 0, \mathbb{P}_x \text{ p.s.} \iff \mathbb{P}_x(X_n \neq y; \forall n \geq 0) = 1.$$

Il est clair que cette dernière condition est équivalente à $\mathbb{P}_x(\sigma_y < \infty) = 0$. Posons $A_n = \{X_n \neq y\}$, $B_n = A_n^c = \{X_n = y\}$.

$$\mathbb{P}_x \left(\bigcap_{n \geq 0} A_n \right) = 1 \iff \mathbb{P}_x \left(\left(\bigcap_{n \geq 0} A_n \right)^c \right) = \mathbb{P}_x \left(\bigcup_{n \geq 0} B_n \right) = 0.$$

Puisque $B_m \subset \left(\bigcup_{n \geq 0} B_n \right)$, la relation précédente implique $\mathbb{P}_x(B_m) = 0$, pour tout $m \geq 0$. Réciproquement si $\mathbb{P}_x(B_m) = 0$, pour tout $m \geq 0$, alors

$$\mathbb{P}_x \left(\bigcup_{n \geq 0} B_n \right) \leq \sum_{n \geq 0} \mathbb{P}_x(B_n) = 0.$$

Par conséquent

$$\begin{aligned} U(x, y) = 0 &\iff \mathbb{P}_x(X_n \neq y; \forall n \geq 0) = 1 \iff \mathbb{P}_x(\exists n \geq 0; X_n = y) = 0 \\ &\iff \forall n \geq 0, \mathbb{P}_x(X_n = y) = \pi^n(x, y) = 0. \end{aligned}$$

Par ailleurs

$$\{\sigma_y < \infty\} = \{\forall n \geq 0, X_n = y\} = \bigcup_{n \geq 0} \{X_n = y\}.$$

Donc $\mathbb{P}_x(\sigma_y < \infty) = \mathbb{P}(\forall n \geq 0, X_n = y)$. □

Proposition 3.8 1) La relation "conduit à" est transitive : si $x \rightarrow y$ et $y \rightarrow z$ alors $x \rightarrow z$.
2) La relation "communiquer" est une relation d'équivalence.

Démonstration : Par construction la relation \longleftrightarrow est réflexive et symétrique. Montrons que \rightarrow est transitive. Il est clair que si $x = y$ ou $y = z$, on a $x \rightarrow z$. Supposons $x \neq y$ et $y \neq z$. Soit $T = \inf\{n > \sigma_y; X_n = z\}$. Si $T < \infty$, on a $X_T = z$, mais σ_z est le premier instant plus grand que 1, où la chaîne visite z , donc $\sigma_z \leq T$. Si $T = +\infty$, la relation précédente est encore réalisée. Par ailleurs, $T = \sigma_y + \sigma_z \circ \theta_{\sigma_y}$, d'où $\sigma_z \leq \sigma_y + \sigma_z \circ \theta_{\sigma_y}$. On en déduit,

$$\mathbb{P}_x(T < \infty) = \mathbb{P}_x(\sigma_y + \sigma_z \circ \theta_{\sigma_y} < \infty) \leq \mathbb{P}_x(\sigma_z < \infty).$$

Mais

$$\mathbb{P}_x(\sigma_y + \sigma_z \circ \theta_{\sigma_y} < \infty) = \mathbb{P}(\sigma_y < \infty, \sigma_z \circ \theta_{\sigma_y} < \infty) = \mathbb{E}_x(\mathbf{1}_{\{\sigma_y < \infty\}} \mathbf{1}_{\{\sigma_z \circ \theta_{\sigma_y} < \infty\}}).$$

On applique la propriété de Markov forte avec $\Phi = \mathbf{1}_{\{\sigma_z < \infty\}}$ et le temps d'arrêt σ_y , il vient,

$$\begin{aligned} \mathbb{P}_x(\sigma_y + \sigma_z \circ \theta_{\sigma_y} < \infty) &= \mathbb{E}_x(\mathbf{1}_{\{\sigma_y < \infty\}} \mathbb{E}_{X_{\sigma_y}}(\Phi)) = \mathbb{P}_x(\sigma_y < \infty) \mathbb{E}_y(\Phi) \\ &= \mathbb{P}_x(\sigma_y < \infty) \mathbb{P}_y(\sigma_z < \infty). \end{aligned}$$

Sachant que $x \rightarrow y$ et $y \rightarrow z$, alors $\mathbb{P}_x(\sigma_y < \infty) > 0$ et $\mathbb{P}_y(\sigma_z < \infty) > 0$. On a montré $\mathbb{P}_x(\sigma_z < \infty) \geq \mathbb{P}_x(\sigma_y < \infty) \mathbb{P}_y(\sigma_z < \infty) > 0$, ce qui signifie $x \rightarrow z$. Il reste à établir la transitivité de la relation \longleftrightarrow . Supposons que $x \longleftrightarrow y$ et $y \longleftrightarrow z$, alors $(x \rightarrow y$ et $y \rightarrow z)$ et $(z \rightarrow y$ et $y \rightarrow x)$. D'après la transitivité de \rightarrow , on a $x \rightarrow z$ et $z \rightarrow x$. D'où $x \longleftrightarrow z$. \square

La proposition qui suit et en particulier l'assertion 1, joue un rôle important dans les exercices.

Proposition 3.9 Soient x un état récurrent et $y \in E$, $y \neq x$. On suppose que $x \rightarrow y$.
1) Alors y est récurrent, $y \rightarrow x$. De plus

$$U(x, y) = U(y, x) = \infty \quad \text{et} \quad \mathbb{P}_x(\sigma_y < \infty) = \mathbb{P}_y(\sigma_x < \infty) = 1.$$

2) Soit $E = \bigcup_{i \in I} E_i$ la partition de E en classes d'équivalence pour la relation "communiquer". Alors chaque E_i ne contient que des états récurrents, ou que des états transients.

Démonstration : a) La première étape de la démonstration consiste à établir que si $x \rightarrow y$, x récurrent alors y est récurrent. Puisque $x \rightarrow y$, d'après le lemme 3.2, il existe n tel que $\pi^n(x, y) > 0$. Mais $U(x, y) = \sum_{m \geq 0} \pi^m(x, y)$, tous les termes étant positifs ou nuls,

$$U(x, y) \geq \sum_{m \geq n} \pi^m(x, y) = \sum_{m \geq 0} \pi^{n+m}(x, y).$$

Puisque E est dénombrable, π^{n+m} est le produit de π^m et de π^n :

$$\pi^{n+m}(x, y) = \sum_{z \in E} \pi^m(x, z) \pi^n(z, y) \geq \pi^m(x, x) \pi^n(x, y).$$

On a utilisé à nouveau le fait que tous les termes de la somme sont positifs ou nuls. Par conséquent,

$$U(x, y) \geq \pi^n(x, y) \left(\sum_{m \geq 0} \pi^m(x, x) \right) = \pi^n(x, y) U(x, x).$$

Mais $\pi^n(x, y) > 0$, x est récurrent ($U(x, x) = \infty$) alors $U(x, y) = \infty$. Sachant que $\mathbb{P}_x(\sigma_y < \infty) > 0$, une application de la proposition 3.6 conduit à : $U(y, y) = \infty$. Ce qui signifie que y est récurrent.

b) Montrons dans une seconde étape que :

$$(x \rightarrow y, x \text{ récurrent}) \implies \mathbb{P}_y(\sigma_x < \infty) = 1.$$

Sachant que x est récurrent, sous \mathbb{P}_x , $N_x = +\infty$, on a :

$$\mathbb{P}_x(\sigma_y < \infty) = \mathbb{P}_x(\sigma_y < \infty, \sigma_x \circ \theta_{\sigma_y} < \infty).$$

On applique la propriété de Markov forte avec $\Phi = \mathbf{1}_{\{\sigma_x < \infty\}}$ et σ_y comme temps d'arrêt :

$$\mathbb{P}_x(\sigma_y < \infty) = \mathbb{E}_x(\mathbf{1}_{\{\sigma_y < \infty\}} \mathbb{E}_{X_{\sigma_y}}(\Phi)) = \mathbb{P}_x(\sigma_y < \infty) \mathbb{P}_y(\sigma_x < \infty).$$

D'où : $\mathbb{P}_x(\sigma_y < \infty)(1 - \mathbb{P}_y(\sigma_x < \infty)) = 0$. Mais $x \rightarrow y$ donc $\mathbb{P}_x(\sigma_y < \infty) > 0$. La relation précédente implique que $\mathbb{P}_y(\sigma_x < \infty) = 1$. Cette identité signifie en particulier que $y \rightarrow x$.

c) Résumons les deux étapes précédentes, on a montré : $(x \text{ récurrent}, x \rightarrow y) \implies y \text{ récurrent}, y \rightarrow x$ et $\mathbb{P}_y(\sigma_x < \infty) = 1$. On choisit $x' = y$, $y' = x$. D'après ce qui précède, x' est récurrent et $x' \rightarrow y'$; donc $y' = x \rightarrow x' = y$ et $\mathbb{P}_{y'}(\sigma_{x'} < \infty) = \mathbb{P}_x(\sigma_y < \infty) = 1$. On applique la proposition 3.6 :

$$U(x, y) = \mathbb{P}_x(\sigma_y < \infty) U(y, y) = U(y, y).$$

y est récurrent, donc $U(x, y) = U(y, y) = \infty$. Par symétrie on montre $U(y, x) = \infty$.

d) Il reste à montrer 2). Si tous les éléments de E_i sont transients, le résultat est établi. Sinon il existe $x \in E_i$, x récurrent. Si $y \in E_i$, alors $x \rightarrow y$. D'après le 1), y est récurrent. \square

Théorème 3.6 *Supposons $\mathcal{R} \neq \emptyset$. Il existe alors une partition $(\mathcal{R}_i; i \in J)$ de \mathcal{R} telle que pour tout i de J , \mathcal{R}_i soit non vide et*

- 1) pour tout x et y de \mathcal{R}_i , $U(x, y) = \infty$ et $\mathbb{P}_x(N_y = \infty) = 1$
- 2) si $x \in \mathcal{R}_i$ et $y \in \mathcal{R}_j$ avec $i \neq j$ alors $U(x, y) = 0$ et $\mathbb{P}_x(\sigma_y < \infty) = 0$.

Démonstration : On note $(E_i; i \in I)$ les classes d'équivalence pour la relation d'équivalence \longleftrightarrow . On pose $J = \{i \in I; \mathcal{R} \cap E_i \neq \emptyset\}$ et $\mathcal{R}_i = \mathcal{R} \cap E_i$, $i \in J$. Par construction $(\mathcal{R}_i, i \in J)$ est une partition de \mathcal{R} . Il existe de plus un état récurrent x_i dans chaque \mathcal{R}_i .

1) Soient x et y deux éléments de \mathcal{R}_i . Puisque $x_i \longleftrightarrow y$ et $x_i \longleftrightarrow x$, d'après la proposition 3.8, x et y communiquent. Mais x est récurrent et $x \longleftrightarrow y$, d'après la proposition 3.9, $U(x, y) = U(y, x) = \infty$ et $\mathbb{P}_x(\sigma_y < \infty) = \mathbb{P}_y(\sigma_x < \infty) = 1$. Si $x = y$, il est clair que $\mathbb{P}_x(N_x = \infty) = 1$.

Supposons $y \neq x$. On a alors,

$$\mathbb{P}_x(N_y = \infty) = \mathbb{P}_x(\sigma_y < \infty, N_y \circ \sigma_y < \infty).$$

On applique la propriété de Markov forte,

$$\mathbb{P}_x(N_y = \infty) = \mathbb{E}_x(\mathbf{1}_{\{\sigma_y < \infty\}} \mathbb{P}_y(N_y = \infty)) = \mathbb{P}_x(\sigma_y < \infty) \mathbb{P}_y(N_y = \infty) = 1.$$

2) Etudions à présent le cas où $x \in \mathcal{R}_i$, $y \in \mathcal{R}_j$ et $i \neq j$, x et y sont donc dans deux classes d'équivalence disjointes. Si $x \rightarrow y$, x étant récurrent, d'après la proposition 3.9, y conduit à x . Donc $x \longleftrightarrow y$, ce qui est impossible. Par conséquent x ne conduit pas à y , d'après le lemme 3.8, $\mathbb{P}_x(\sigma_y < \infty) = 0$ et $U(x, y) = 0$. \square

Nous allons nous intéresser à la caractérisation des parties fermées.

Proposition 3.10 *Soit F une partie non vide de E . On a les équivalences :*

(i) F partie fermée \iff (ii) $\forall x \in F, \pi(x, F) = 1 \iff$ (iii) $\forall x \in F, \forall y \notin F, U(x, y) = 0$.

Démonstration : a) Il est clair que l'on a les quatre propriétés suivantes :

$$1 - \mathbb{P}_x(X_n \in F; \forall n \geq 0) = \mathbb{P}_x(\exists n \geq 0, X_n \notin F) = \mathbb{P}_x\left(\bigcup_{n \geq 0} \{X_n \notin F\}\right), \quad (3.31)$$

$$\mathbb{P}_x\left(\bigcup_{n \geq 0} B_n\right) = 0 \text{ si et seulement si } \mathbb{P}_x(B_n) = 0 \quad \forall n \geq 0, \quad (3.32)$$

(Utiliser $B_m \subset \bigcup_{n \geq 0} B_n$ et $\mathbb{P}_x(\bigcup_{n \geq 0} B_n) \leq \sum_{n \geq 0} \mathbb{P}_x(B_n)$).

$$U(x, A) = \sum_{n \geq 0} \pi^n(x, A) = 0 \iff \forall n \geq 0, \pi^n(x, A) = 0, \quad (3.33)$$

$$U(x, A) = \sum_{y \in A} U(x, y) = 0 \iff U(x, y) = 0, \quad \forall y \in A. \quad (3.34)$$

Soit F un sous-ensemble non vide de E . F est une partie fermée si et seulement si $\mathbb{P}_x(X_n \in F; \forall n \geq 0) = 1$ pour tout $x \in F$. En utilisant à la fois (3.31), (3.32), (3.33) et (3.34), on montre la série d'équivalence : F est un ensemble fermé $\iff \pi^n(x, F^c) = 0 \quad \forall n \geq 0; \forall x \in F \iff U(x, F^c) = 0 \iff U(x, y) = 0 \quad \forall x \in F, \forall y \in F^c$.

b) On a établi : (i) \iff (iii) et (i) $\implies \pi(x, F) = 1, \forall x \in F$. En particulier (i) \implies (ii). Montrons la réciproque : $\pi^n(x, F) = 1; \forall n \geq 0, \forall x \in F$. On raisonne par récurrence sur n . Pour $n = 1$, la propriété est vérifiée. Supposons $\pi^n(x, F) = 1, \forall x \in F$. On a

$$\pi^{n+1}(x, F) = \mathbb{P}_x(X_{n+1} \in F) = \mathbb{E}_x(\pi(X_n, F)).$$

Mais $\mathbb{P}_x(X_n \in F) = \pi^n(x, F) = 1$, donc \mathbb{P}_x p.s. $X_n \in F$ et $\pi(X_n, F) = 1$. Ce qui prouve $\pi^{n+1}(x, F) = 1$. \square

Remarques.

- 1) Soit F une partie fermée de E . On peut restreindre π à F .
 2) Soit $\mathcal{R} = \bigcup_{i \in J} \mathcal{R}_i$ la partition de l'ensemble des points récurrents définie par le théorème 3.6. Alors chaque \mathcal{R}_i est une partie fermée. Montrons que si $x \in \mathcal{R}_i$ et $y \in \mathcal{R}_i$ alors $U(x, y) = 0$. Si y est récurrent, $y \in \mathcal{R}_j$, pour un certain $j \neq i$. D'après le théorème 3.6, $U(x, y) = 0$. Examinons le cas où y est transient. Si $x \rightarrow y$, d'après la proposition 3.9, y est récurrent. Ce qui est impossible, donc x ne conduit pas à y , d'après le lemme 3.2, $U(x, y) = 0$.

Théorème 3.7 1) La chaîne est irréductible ssi $U(x, y) > 0$ pour tout x et y de E .

2) Lorsque la chaîne est irréductible,

(i) Soit $U(x, y) < \infty$ pour tout x et y de E , tous les états sont transients, la chaîne est dite transiente.

(ii) Soit $U(x, y) = \infty$ pour tout x et y de E , tous les états sont récurrents, la chaîne est dite récurrente.

Démonstration : 1) Montrons l'équivalence concernant l'irréductibilité de la chaîne. Supposons que $U(x, y) > 0$, pour tout x et y de E . Soient A une partie fermée et $x \in A$. Raisonnons par l'absurde, supposons qu'il existe $y \notin A$. D'après la proposition 3.10, $U(x, y) = 0$. Cette conclusion est à rejeter, donc $A^c = \emptyset$, ce qui signifie que $E = A$, la chaîne est irréductible.

Etablissons à présent la réciproque : on suppose que E est la seule partie fermée. Soit x un élément de E fixé. On note $F = \{y \in E; U(x, y) > 0\}$. Puisque $U(x, x) \geq \pi^0(x, x) = 1$, x est élément de F . Montrons que F est une partie fermée. D'après la proposition 3.10, il s'agit de montrer que si $y \in F$, et $z \notin F$ alors $U(y, z) = 0$. Par conséquent $U(x, y) > 0$ et $U(x, z) = 0$. D'après le lemme 3.2, ces deux identités impliquent : $x \rightarrow y$ et $x \rightarrow z$. Si $y \rightarrow z$ par transitivité de la relation "conduit à", $x \rightarrow z$ ce qui est exclu, donc $y \rightarrow z$, en utilisant à nouveau le lemme 3.2 on a $U(y, z) = 0$.

2) On suppose que la chaîne est irréductible. Supposons qu'il existe un état récurrent x_0 . D'après le théorème 3.6, $x_0 \in \mathcal{R}_{i_0}$, et \mathcal{R}_{i_0} est une partie fermée (voir la remarque 2) qui suit la proposition 3.10). La chaîne étant irréductible, $\mathcal{R}_{i_0} = E$, les états sont tous récurrents et d'après le théorème 3.6, pour tous x et y de E , $U(x, y) = \infty$. Examinons à présent le cas où tous les états sont transients. On sait que $U(y, y) < \infty, \forall y \in E$. On déduit directement de la proposition 3.6, la finitude de $U(x, y)$. \square

3.4 Mesure invariante et convergence

Définition 3.3 Soit ν une mesure sur E , positive, non identiquement nulle. On dit que ν est une mesure stationnaire (ou invariante) associée à la probabilité de transition π si

$$\nu \cdot \pi = \nu. \quad (3.35)$$

Remarques

- 1) $\nu = (\nu(i) : i \in E)$ est une mesure invariante si $\nu(i) \geq 0, \forall i \in E, \exists i \in E$ tel que $\nu(i) > 0$ et

$$\nu(j) = \sum_{i \in E} \nu(i) \pi(i, j); \text{ pour tout } j \in E. \quad (3.36)$$

2) Si $\lambda > 0$ et ν est une mesure stationnaire alors $\lambda\nu$ est une mesure stationnaire. Si ν est de masse totale finie, i.e. $\sum_{i \in E} \nu(i) < \infty$, $\frac{1}{\lambda\nu}$ est une probabilité invariante, où $\lambda\nu = \sum_{i \in E} \nu(i)$.

3) En raisonnant par récurrence sur $n \geq 1$, il est facile de montrer que si ν est stationnaire, alors

$$\nu\pi^n = \nu; \quad n \geq 1. \quad (3.37)$$

L'interprétation probabiliste de cette identité est la suivante : si ν est une probabilité invariante, et si X_0 a pour loi ν alors X_n a pour loi ν , pour tout $n \geq 0$, (X_n) désignant une chaîne de Markov de probabilité de transition π . Ceci justifie la terminologie stationnaire ou invariante.

4) Il n'y a pas a priori unicité de la mesure invariante.

a) Considérons une chaîne de Markov avec deux états absorbants a et b . Alors δ_a et δ_b sont deux probabilités invariantes.

b) Soit (X_n) une marche aléatoire à valeurs dans \mathbb{Z}^d , alors la mesure uniforme ρ_u (i.e. $\rho_u(i) = 1; \forall i \in \mathbb{Z}^d$) est une mesure invariante. En effet, en conservant les notations de l'exemple 3, paragraphe 1, on a : $\pi(x, y) = \nu(y - x)$.

$$\rho_u\pi(j) = \sum_{i \in \mathbb{Z}^d} \rho_u(i)\pi(i, j) = \sum_{i \in \mathbb{Z}^d} \nu(j - i) = \sum_{i \in \mathbb{Z}^d} \nu(i) = 1 = \rho_u(j).$$

ρ_u est bien une mesure invariante. Il est évident que $\rho_u(\mathbb{Z}^d) = +\infty$.

Examinons à présent le cas particulier de la marche de Bernoulli : $d = 1$, $\nu = p\delta_1 + q\delta_{-1}$ avec $p > 0$, $q > 0$, $p + q = 1$. On pose $\rho(i) = (p/q)^i$. Montrons que ρ est invariante :

$$\begin{aligned} \rho\pi(j) &= \sum_i \rho(i)\pi(i, j) = \rho(j-1)\pi(j-1, j) + \rho(j+1)\pi(j+1, j) \\ &= \left(\frac{p}{q}\right)^{j-1} p + \left(\frac{p}{q}\right)^{j+1} q = \left(\frac{p}{q}\right)^j \left(\frac{q}{p}p + \frac{p}{q}q\right) = \left(\frac{p}{q}\right)^j = \rho(j). \end{aligned}$$

Si $p \neq q$, la marche aléatoire possède deux mesures invariantes. De plus ces deux mesures invariantes sont de masse totale infinie. On peut remarquer de plus que la chaîne est transiente.

5) Supposons que E soit fini et

$$\lim_{n \rightarrow \infty} \pi^n(i, j) = \nu(j), \quad \text{pour tout } i \in E \text{ et } j \in E. \quad (3.38)$$

Il est évident que $\nu(j) \geq 0$, pour tout j de E . $\pi^n(i, \cdot)$ étant une probabilité,

$$\sum_{j \in E} \pi^n(i, j) = 1. \quad (3.39)$$

La somme comportant un nombre fini de termes, en passant à la limite, $n \rightarrow \infty$, on obtient $\sum_{j \in E} \nu(j) = 1$. ν est une probabilité sur E . Montrons que ν est invariante. On a :

$$\pi^{n+1}(i, j) = (\pi^n \cdot \pi)(i, j) = \sum_{k \in E} \pi^n(i, k)\pi(k, j).$$

On passe à la limite, $n \rightarrow \infty$, on obtient,

$$\nu(j) = \sum_{k \in E} \nu(k) \pi(k, j) = (\nu \pi)(j).$$

L'interprétation probabiliste de (3.38) est la suivante : (3.38) a lieu si et seulement si pour tout $i \in E$, X_n converge en loi sous \mathbb{P}_i lorsque $n \rightarrow \infty$, vers une distribution ν indépendante de i .

Théorème 3.8 Soient x un état récurrent et $T = \inf\{n \geq 1; X_n = x\}$.

Alors $\mu_x(y) = \mathbb{E}_x \left(\sum_{n=0}^{T-1} \mathbf{1}_{\{X_n=y\}} \right)$ définit une mesure stationnaire. De plus $\mu_x(x) = 1$.

Démonstration : Le point x étant fixé, nous allons noter $\mu = \mu_x$ pour alléger les notations. Il est clair que $\mu_x(y) \geq 0$ pour tout y de E .

1)

$$\mu \pi(z) = \sum_{y \in E} \mu(y) \pi(y, z) = \sum_{y \in E} \mathbb{E}_x \left(\sum_{n=0}^{\infty} \mathbf{1}_{\{n \leq T-1, X_n=y\}} \pi(y, z) \right).$$

Puisque les quantités intervenant sont positives, on peut permuter \mathbb{E}_x et $\sum_{n \geq 0}$. Mais $\{n \leq T-1\} = \{n < T\}$ et $\{T \leq n\} = \{n < T\}^c \in \mathcal{A}_n$. On en déduit que l'événement $\{n \leq T-1, X_n = y\}$ appartient à \mathcal{A}_n . De plus sur $\{X_n = y\}$, on a $\pi(y, z) = \mathbb{P}_y(X_1 = z) = \mathbb{P}_{X_n}(X_1 = z)$. Une application directe de la propriété de Markov conduit à :

$$\mathbb{E}_x(\mathbf{1}_{\{X_{n+1}=z\}} | \mathcal{A}_n) = \mathbb{P}_{X_n}(X_1 = z),$$

d'où

$$\begin{aligned} \mu \pi(z) &= \sum_{y \in E} \sum_{n=0}^{\infty} \mathbb{E}_x (\mathbf{1}_{\{n \leq T-1, X_n=y, X_{n+1}=z\}}) \\ &= \sum_{n=0}^{\infty} \mathbb{E}_x \left(\sum_{y \in E} \mathbf{1}_{\{n \leq T-1, X_n=y, X_{n+1}=z\}} \right) = \sum_{n \geq 0} \mathbb{E}_x (\mathbf{1}_{\{n \leq T-1, X_{n+1}=z\}}). \end{aligned}$$

On pose $m = n + 1$ et on permute \mathbb{E}_x avec $\sum_{n \geq 0}$, il vient,

$$\mu \pi(z) = \mathbb{E}_x \left(\sum_{m=1}^T \mathbf{1}_{\{X_m=z\}} \right) = \mathbb{E}_x \left(\sum_{m=0}^{T-1} \mathbf{1}_{\{X_m=z\}} + \xi \right) = \mu(z) + \mathbb{E}_x(\xi),$$

avec $\xi = \mathbf{1}_{\{X_T=z\}} - \mathbf{1}_{\{X_0=z\}}$.

Mais x est récurrent, $\mathbb{P}_x(T < \infty) = 1$, et $\mathbb{P}_x(X_0 = x) = 1$, donc $\xi = \mathbf{1}_{\{x=z\}} - \mathbf{1}_{\{x=z\}} = 0$. Nous avons montré que $\mu \pi = \mu$.

2) Par définition pour tout n tel que $1 \leq n < T$, on a : $X_n \neq x$. Mais $\mathbb{P}_0(X_0 = x) = 1$, donc $\mu(x) = 1$. μ est donc non identiquement nulle. μ est une mesure invariante, si $\mu(y) < \infty$, pour tout $y \in E$. Puisque $\mu \pi = \mu$, d'après (3.37) on a $\mu \pi^n = \mu$, pour tout $n \geq 1$.

En particulier,

$$\mu(x) = 1 = \mu\pi^n(x) = \sum_z \mu(z)\pi^n(z, x) \geq \mu(y)\pi^n(y, x),$$

car la somme ne comporte que des termes positifs. S'il existe $n \geq 0$, tel que $\pi^n(y, x) > 0$, alors $\mu(y) \leq 1/\pi^n(y, x) < +\infty$. Sinon, d'après le lemme 3.2, y ne conduit pas à x . x étant récurrent, si $x \rightarrow y$, d'après la proposition 3.9, $y \rightarrow x$, ce qui est impossible. Par conséquent x ne conduit pas à y , une nouvelle application du lemme 3.2, permet d'affirmer que $\mathbb{P}_x(\sigma_y < \infty) = 0$, donc

sous \mathbb{P}_x , $\sum_{n=0}^{T-1} \mathbf{1}_{\{X_n=y\}} = 0$. Ce qui signifie que $\mu(y) = 0$. □

Remarques

1) On peut préciser le support de la mesure μ_x , c'est-à-dire l'ensemble des points y tels que $\mu_x(y) > 0$. D'après la définition de μ_x , si x ne conduit pas à y , alors le lemme 3.2 implique que $\mu_x(y) = 0$. Réciproquement si $\mu_x(y) = 0$ alors $\mathbb{P}_x(\exists n < T, X_n = y) = \mathbb{P}_x(\exists n \leq T, X_n = y) = 0$. Soit $(\sigma^{(k)}; k \geq 1)$ la suite des temps de passages successifs en x , voir (3.19). En utilisant la propriété de Markov forte, il n'est pas difficile de montrer que

$$\mathbb{P}_x(\exists n \leq \sigma^{(1)}; X_n = y) = 0 \iff \mathbb{P}_x(\exists n \leq \sigma^{(k)}; X_n = y) = 0, \forall k \in \mathbb{N}. \quad (3.40)$$

On fait tendre k vers $+\infty$, on obtient $\mathbb{P}_x(\exists n, X_n = y) = \mathbb{P}_x(\sigma_y < \infty) = 0$. Donc x ne conduit pas à y . On a montré,

$$\{y \in E, \mu_x(y) > 0\} = \{y \in E; x \rightarrow y\}. \quad (3.41)$$

Puisque x est récurrent, d'après la proposition 3.9,

$$\{y \in E; x \rightarrow y\} = \{y \in E; x \longleftrightarrow y\}.$$

Soit $\mathcal{R} = \bigcup_i \mathcal{R}_i$ la partition de l'ensemble des points récurrents (voir théorème 3.6), $x \in \mathcal{R}_{i_0}$, alors le support de μ_x est \mathcal{R}_{i_0} . On a alors l'implication $\mu_x(y) > 0 \implies y$ est récurrent. On verra plus loin (proposition 3.11) une réciproque partielle de cette propriété.

2) Si l'ensemble des points récurrents \mathcal{R} possède au moins deux classes \mathcal{R}_1 et \mathcal{R}_2 , il existe au moins deux mesures invariantes. En effet on choisit x_1 dans \mathcal{R}_1 et x_2 dans \mathcal{R}_2 . μ_{x_1} et μ_{x_2} sont des mesures invariantes, d'après ce qui précède le support de μ_{x_i} est \mathcal{R}_i . Donc μ_{x_1} n'est pas proportionnelle à μ_{x_2} . On peut remarquer que la chaîne n'est pas irréductible.

Proposition 3.11 *Soit μ une probabilité invariante. Si $\mu(x) > 0$ alors x est récurrent.*

Démonstration : Soit x tel que $\mu(x) > 0$. D'après la propriété (3.37), $\mu\pi^n = \mu$. On en déduit :

$$\begin{aligned} \sum_{y \in E} \mu(y) \left(\sum_{n=0}^{\infty} \pi^n(y, x) \right) &= \sum_{n=0}^{\infty} \left(\sum_{y \in E} \mu(y) \pi^n(x, y) \right) \\ &= \sum_{n=0}^{\infty} \mu\pi^n(x) = \sum_{n=0}^{\infty} \mu(x) = +\infty \end{aligned}$$

Mais $U(y, x) = \sum_{n \geq 0} \pi^n(y, x)$, donc $\sum_{y \in E} \mu(y)U(y, x) = +\infty$. Rappelons que d'après la proposition 3.6, $U(y, x) \leq U(x, x)$. Par conséquent,

$$+\infty \leq \sum_y \mu(y)U(x, x) = \left(\sum_y \mu(y) \right) U(x, x) = U(x, x).$$

Donc $U(x, x) = +\infty$, ce qui signifie que x est récurrent. \square

Remarque

Si μ est une mesure invariante de masse totale infinie, alors $\mu(x) > 0$ n'implique pas nécessairement que x est récurrent. Il suffit de considérer la marche aléatoire Bernoulli sur \mathbb{Z} ($\nu = p\delta_1 + q\delta_{-1}$, $p \neq q$). On a montré que ρ_u la mesure uniforme sur \mathbb{Z} est une mesure invariante, à l'évidence pour tout x de \mathbb{Z} , $\rho_u(x) = 1 > 0$ et tout état x est transient.

Théorème 3.9 *Si la chaîne de Markov est récurrente et irréductible alors la mesure stationnaire est unique à une constante multiplicative près.*

Démonstration : Soit a un élément fixé de E . On pose $q_n(y) = \mathbb{P}_a(n < T, X_n = y)$; $y \in E, n \geq 0$, où $T = \inf\{n \geq 1, X_n = a\}$.

1) Montrons par récurrence sur $m \geq 0$:

$$\lambda \geq \lambda(a) \left(\sum_{0 \leq n \leq m} q_n \right) \text{ sur } E \setminus \{a\} \quad (3.42)$$

où λ désigne une mesure positive telle que $\lambda\pi = \lambda$ (λ est une mesure invariante pour la chaîne de Markov si λ n'est pas identiquement nulle). Si $m = 0$, on a $\sum_{0 \leq n \leq m} q_n = q_0$ et

$$q_0(y) = \mathbb{P}_a(X_0 = y) = \delta_a(y), \quad (3.43)$$

car $T \geq 1$. En particulier si $y \neq a$, $\lambda(a)q_0(y) = 0 \leq \lambda(y)$. Ce qui prouve (3.42) lorsque $n = 0$. Supposons à présent (3.42), et montrons que cette identité est encore réalisée si m est remplacé

par $m + 1$. Sachant que $\lambda = \lambda.\pi$ et $\lambda \geq \lambda(a) \left(\sum_{0 \leq n \leq m} q_n \right)$ on en déduit

$$\lambda = \lambda.\pi = \lambda(a) \left(\sum_{0 \leq n \leq m} q_n \right) = \lambda(a) \left(\sum_{0 \leq n \leq m} q_n \pi \right).$$

En particulier si $y \in E$,

$$\lambda(y) \geq \lambda(a) \left(\sum_{0 \leq n \leq m} (q_n \pi)(y) \right) \quad (3.44)$$

Mais $(q_n \pi)(y) = \sum_{z \in E} q_n(z) \pi(z, y) = \sum_{z \in E} \mathbb{P}_a(n < T, X_n = z) \pi(z, y)$.

Puisque $\{n < T\} = \{T \leq n\}^c$, cet événement appartient à \mathcal{A}_n , une application de la propriété de Markov conduit à : $\mathbb{P}_a(n < T, X_n = z, X_{n+1} = y) = \mathbb{P}_a(n < T, X_n = z)\pi(z, y)$. Par conséquent

$$(q_n\pi)(y) = \sum_{z \in E} \mathbb{P}_a(n < T, X_n = z, X_{n+1} = y) = \mathbb{P}_a(n < T, X_{n+1} = y).$$

Si $T = n + 1$, alors $X_T = a$, donc si y est différent de a , les deux conditions $n < T$ et $X_{n+1} = y$ impliquent $T > n + 1$, d'où $(q_n\pi)(y) = q_{n+1}(y)$. On revient à (3.44),

$$\lambda(y) \geq \lambda(a) \left(\sum_{0 \leq n \leq m} q_{n+1}(y) \right) = \lambda(a) \left(\sum_{1 \leq n \leq m+1} q_n(y) \right) = \lambda(a) \left(\sum_{0 \leq n \leq m+1} q_n(y) \right).$$

La dernière égalité résulte de (3.43) : $q_0(y) = 0$ si $y \neq a$.

2) Soit μ_a la mesure invariante définie par le théorème 3.8. Vérifions

$$\lambda \geq \lambda(a)\mu_a, \text{ sur } E, \quad (3.45)$$

où λ désigne une mesure positive telle que $\lambda\pi = \lambda$. Soit $y \in E$, $y \neq a$. D'après (3.42),

$$\lambda(y) \geq \lambda(a) \left(\sum_{0 \leq n \leq m} q_n(y) \right).$$

On fait tendre m vers $+\infty$, à la limite, on obtient

$$\lambda(y) \geq \lambda(a) \left(\sum_{n \geq 0} q_n(y) \right). \quad (3.46)$$

$$\text{Mais } \mu_a(y) = \mathbb{E} \left[\sum_{n=0}^{T-1} \mathbf{1}_{\{X_n=y\}} \right] = \mathbb{E} \left[\sum_{n \geq 0} \mathbf{1}_{\{n < T, X_n=y\}} \right].$$

On peut permuter \mathbb{E} et \sum , on obtient : $\mu_a(y) = \sum_{n \geq 0} q_n(y)$. L'inégalité (3.45) est alors une conséquence immédiate de (3.46). Si $y = a$, on a vu que $\mu_a(a) = 1$, il est clair que $\lambda(a) \geq \lambda(a)\mu_a(a)$. Ce qui achève la démonstration de (3.45).

La chaîne est récurrente irréductible, $\{y \in E, a \rightarrow y\} = E$. On a montré (voir (3.41)) que $\{y \in E; \mu_a(y) > 0\} = \{y \in E; a \rightarrow y\} = E$. On déduit de (3.45) :

$$\text{si } \lambda \text{ est une mesure invariante alors } \lambda(y) > 0, \text{ pour tout } y \text{ de } E. \quad (3.47)$$

(Il suffit en effet de choisir a tel que $\lambda(a) > 0$).

3) On est à présent en mesure de vérifier l'unicité, à une constante multiplicative près. Soit λ une mesure invariante pour π . On choisit a tel que $\lambda(a) > 0$. Posons $\lambda' = \lambda - \lambda(a)\mu_a$. D'après (3.45), λ' est une mesure positive. De plus elle vérifie $\lambda'\pi = \lambda'$, en effet :

$$\lambda'\pi = (\lambda - \lambda(a)\mu_a)\pi = \lambda\pi - \lambda(a)\mu_a\pi = \lambda - \lambda(a)\mu_a = \lambda.$$

Par ailleurs $\lambda'(a) = \lambda(a) - \lambda(a)\mu_a(a) = \lambda(a) - \lambda(a) = 0$. Appliquons (3.47) en changeant λ et λ' : λ' est identiquement nulle (sinon $\lambda'(y) > 0, \forall y \in E$). Il est maintenant clair que $\lambda = \lambda(a)\mu_a$, λ est unique à une constante multiplicative près. \square

Remarque.

Toute mesure invariante λ vérifie $\lambda(x) > 0$, pour tout x de E .

Théorème 3.10 (théorème ergodique) Soient X_n une chaîne de Markov récurrente et irréductible, λ une mesure invariante et x_0 un point quelconque de E .

1) Pour tout y de E ,

$$\frac{\sum_{0 \leq m < n} \mathbf{1}_{\{X_m=y\}}}{\sum_{0 \leq m < n} \mathbf{1}_{\{X_m=x_0\}}} \xrightarrow[n \rightarrow \infty]{\mathbb{P}_{x_0} \text{ p.s.}} \frac{\lambda(y)}{\lambda(x_0)}.$$

2) Plus généralement, pour toute fonction $f : E \rightarrow \mathbb{R}$ positive ou λ -intégrable, on a :

$$\frac{\sum_{0 \leq m < n} f(X_m)}{\sum_{0 \leq m < n} \mathbf{1}_{\{X_m=x_0\}}} \xrightarrow[n \rightarrow \infty]{\mathbb{P}_{x_0} \text{ p.s.}} \frac{\int_E f d\lambda}{\lambda(x_0)}.$$

où $\int_E f d\lambda = \sum_{y \in E} f(y)\lambda(y)$.

Remarques.

1) La chaîne étant récurrente et irréductible, λ est une unique à une constante multiplicative près, les deux rapports $\frac{\lambda(y)}{\lambda(x_0)}$ et $\frac{\int_E f d\lambda}{\lambda(x_0)}$ ne dépendent pas du choix de cette constante. De plus $\lambda(x_0) > 0$, les limites sont bien définies.

2) Puisque la chaîne est récurrente et irréductible, $\sum_{0 \leq m < n} \mathbf{1}_{\{X_m=y\}}$ converge, \mathbb{P}_{x_0} p.s., lorsque

n tend vers $+\infty$, vers N_y , et cette v.a. est infinie. Les deux suites $\left(\sum_{0 \leq m < n} \mathbf{1}_{\{X_m=y\}} \right)_{n \geq 0}$ et

$\left(\sum_{0 \leq m < n} \mathbf{1}_{\{X_m=x_0\}} \right)_{n \geq 0}$ sont croissantes et convergent p.s. vers $+\infty$.

Démonstration : (théorème 3.10)

Soit x_0 un point fixé de E . On note $(\sigma_k; k \geq 1)$ la suite des temps de passages successifs en x_0 . On pose $\sigma_0 = 0$ et

$$Y_k = \sum_{\sigma_k \leq n < \sigma_{k+1}} \mathbf{1}_{\{X_n=y\}}; k \geq 0.$$

On a : $\sigma_{k+1} - \sigma_k = \sigma_1 \circ \theta_{\sigma_k}$ et

$$Y_k = \sum_{0 \leq n < \sigma_{k+1} - \sigma_k} \mathbf{1}_{\{X_{n+\sigma_k}=y\}} = \left(\sum_{0 \leq n < \sigma_1} \mathbf{1}_{\{X_n=y\}} \right) \circ \theta_{\sigma_k} = Y_0 \circ \theta_{\sigma_k}.$$

Les v.a. (Y_k) sont indépendantes et ont même loi. Vérifions à cet effet :

$$\mathbb{E}_{x_0}[f_0(Y_0)\dots f_k(Y_k)] = \mathbb{E}_{x_0}[f_0(Y_0)]\dots\mathbb{E}_{x_0}[f_k(Y_0)]; k \geq 0$$

pour toutes fonctions boréliennes bornées f_0, \dots, f_k . On raisonne par récurrence sur k . Si $k = 0$, le résultat est évident. Explicitons le passage $k \implies k + 1$. On pose $Z = f_0(Y_0)\dots f_k(Y_k)$. Les v.a. Y_i sont \mathcal{A}_{σ_i} -mesurables, de plus $\sigma_1 \leq \dots \leq \sigma_k$, donc Z est une v.a. ma_{σ_k} -mesurable et bornée. Posons

$$\Delta = \mathbb{E}_{x_0}[f_0(Y_0)\dots f_k(Y_k)f_{k+1}(Y_{k+1})] = \mathbb{E}_{x_0}[Zf_{k+1}(Y_{k+1})].$$

Puisque $Y_{k+1} = Y_0 \circ \theta_{\sigma_{k+1}}$, une application de la propriété de Markov forte, conduit à :

$$\Delta = \mathbb{E}_{x_0}(Z\mathbb{E}_{X_{\sigma_k}}(f_{k+1}(Y_0))) = \mathbb{E}_{x_0}(Z)\mathbb{E}_{x_0}(f_{k+1}(Y_0)),$$

car σ_k est fini et $X_{\sigma_k} = x_0$.

Il suffit ensuite d'appliquer l'hypothèse de récurrence. La v.a. Y_0 est par définition positive et,

$$\mathbb{E}_{x_0}(Y_0) = \mathbb{E}_{x_0}\left(\sum_{0 \leq n < \sigma_1} \mathbf{1}_{\{X_n=y\}}\right) = \mu_{x_0}(y),$$

où μ_{x_0} est la mesure invariante introduite au théorème 3.8. Y_0 est donc intégrable, on peut appliquer la loi forte des grands nombres :

$$\frac{1}{n}(Y_0 + \dots + Y_{n-1}) \xrightarrow[n \rightarrow \infty]{} \mu_{x_0}(y) \quad \mathbb{P}_{x_0} \text{ p.s.}$$

Mais $\sum_{i=0}^{n-1} Y_i = \sum_{i=0}^{n-1} \sum_{\sigma_i \leq k \leq \sigma_{i+1}} \mathbf{1}_{\{X_k=y\}} = \sum_{0 \leq k \leq \sigma_n} \mathbf{1}_{\{X_k=y\}}$. Posons : $L_k = \sum_{n < k} \mathbf{1}_{\{X_n=x_0\}}$. Sur $\{\sigma_l < m \leq \sigma_{l+1}\}$ et sous \mathbb{P}_{x_0} , on a :

$$\sum_{k < l} Y_k = \sum_{n < \sigma_l} \mathbf{1}_{\{X_n=y\}} \leq \sum_{n < m} \mathbf{1}_{\{X_n=y\}} \leq \sum_{n < \sigma_{l+1}} \mathbf{1}_{\{X_n=y\}} = \sum_{k < l+1} Y_k,$$

et $L_m = l + 1$ (sous \mathbb{P}_x). Par conséquent :

$$\frac{1}{L_m} \sum_{k < L_m-1} Y_k \leq \frac{\sum_{n < m} \mathbf{1}_{\{X_n=y\}}}{\sum_{n < m} \mathbf{1}_{\{X_n=x_0\}}} \leq \frac{1}{L_m} \sum_{k < L_m} Y_k.$$

Par ailleurs $\lim_{m \rightarrow \infty} L_m = N_{x_0} = +\infty$. On en déduit immédiatement le résultat. Lorsque $\mathbf{1}_{\{y\}}$ est remplacé par f , f étant positive ou λ -intégrable, on pose :

$$Y_k = \sum_{\sigma_k \leq n < \sigma_{k+1}} f(X_n).$$

On montre comme précédemment que les v.a. $(Y_k)_{k \geq 0}$ sont indépendantes et ont même loi. Pour déterminer la limite, il faut calculer $\mathbb{E}_{x_0}(Y_0)$:

$$\mathbb{E}_{x_0}(Y_0) = E_{x_0}\left(\sum_{0 \leq n < \sigma_1} f(X_n)\right) = \sum_{y \in E} f(y)\mathbb{E}_{x_0}\left(\sum_{0 \leq n < \sigma_1} \mathbf{1}_{\{X_n=y\}}\right) = \int f d\mu_{x_0}.$$

Par conséquent,

$$\mathbb{E}_{x_0}(Y_0) = \frac{\int f d\lambda}{\lambda(x_0)}.$$

□

Définition 3.4 Une chaîne récurrente et irréductible est dite positive si "sa" mesure invariante est bornée (i.e. : $\sum_{x \in E} \lambda(x) < \infty$). Lorsque $\sum_{x \in E} \lambda(x) = +\infty$, on dit que la chaîne est récurrente nulle. Notons qu'une chaîne positive admet une unique probabilité invariante.

Théorème 3.11 Soient une chaîne irréductible et récurrente et $x_0 \in E$.

1) Si cette chaîne est de plus positive, alors pour tout y de E ,

$$(i) \quad \mathbb{E}_y(T_y) = \frac{1}{\lambda(y)},$$

où λ est la probabilité invariante et $T_x = \inf\{n \geq 1, X_n = x\}$.

$$(ii) \quad \frac{1}{n} \left(\sum_{k < n} \mathbf{1}_{\{X_k=y\}} \right) \xrightarrow{n \rightarrow \infty} \lambda(y), \quad \mathbb{P}_{x_0} \text{ p.s.}$$

(iii) Soit $f : E \rightarrow \mathbb{R}$ positive ou λ -intégrable, alors

$$\frac{1}{n} \left(\sum_{k < n} f(X_k) \right) \xrightarrow{n \rightarrow \infty} \sum_{z \in E} f(z) \lambda(z).$$

2) Si la chaîne est de plus nulle, $\mathbb{E}_x(T_x) = +\infty$, pour tout x de E , \mathbb{P}_{x_0} presque sûrement,

$$\frac{1}{n} \left(\sum_{k < n} \mathbf{1}_{\{X_k=y\}} \right) \rightarrow 0 \quad \text{et} \quad \frac{1}{n} \left(\sum_{k < n} f(X_k) \right) \rightarrow 0, \quad \text{lorsque } n \rightarrow \infty.$$

Démonstration : On a les deux identités :

$$T_x = \sum_{0 \leq n < T_x} 1 \quad \text{et} \quad 1 = \sum_{y \in E} \mathbf{1}_{\{X_n=y\}}; \quad n \geq 0, x \in E.$$

D'où

$$T_x = \sum_{0 \leq n < T_x} \left(\sum_{y \in E} \mathbf{1}_{\{X_n=y\}} \right) = \sum_{y \in E} \left(\sum_{0 \leq n < T_x} \mathbf{1}_{\{X_n=y\}} \right).$$

On prend l'espérance sous \mathbb{E}_x , il vient,

$$\mathbb{E}_x(T_x) = \sum_{y \in E} \mathbb{E}_x \left(\sum_{0 \leq n < T_x} \mathbf{1}_{\{X_n=y\}} \right) = \sum_{y \in E} \mu_x(y) = \mu_x(E),$$

où μ_x est la mesure invariante définie au théorème 3.8.

a) Supposons la chaîne récurrente et positive. La probabilité invariante vaut $\lambda = c\mu_x$,

mais $\mu_{x_0}(x_0) = 1$, d'où $c = \lambda(x_0)$ et $\mathbb{E}_{x_0}(T_{x_0}) = \mu_{x_0}(E) = \lambda(E)/c = 1/c = 1/\lambda(x_0)$. On applique le théorème ergodique à $f = 1$, on a,

$$\frac{n}{\sum_{0 \leq m < n} \mathbf{1}_{\{X_m = x_0\}}} \xrightarrow{n \rightarrow \infty} \frac{\lambda(E)}{\lambda(x_0)} = \frac{1}{\lambda(x_0)}, \quad \mathbb{P}_{x_0} \text{ p.s.}$$

On en déduit sans peine, que \mathbb{P}_{x_0} p.s., $\frac{1}{n} \left(\sum_{0 \leq m < n} \mathbf{1}_{\{X_m = x_0\}} \right)$ converge vers $\lambda(x_0)$, lorsque $n \rightarrow \infty$. Soit f une fonction positive ou intégrable par rapport à λ . On a l'identité :

$$\frac{1}{n} \left(\sum_{0 \leq m < n} f(X_m) \right) = \left(\frac{\sum_{0 \leq m < n} f(X_m)}{\sum_{0 \leq m < n} \mathbf{1}_{\{X_m = x_0\}}} \right) \left(\frac{1}{n} \left(\sum_{0 \leq m < n} \mathbf{1}_{\{X_m = x_0\}} \right) \right) \quad (3.48)$$

On applique le théorème ergodique, le membre de gauche de (3.48) converge, \mathbb{P}_{x_0} p.s. vers $\int_E f d\lambda$.

b) Lorsque la chaîne est récurrente irréductible, $\mu_x(E) = +\infty$, donc $\mathbb{E}_x(T_x) = +\infty$. Montrons que les deux limites sont nulles. Soit A une partie finie de E . La fonction $f = \mathbf{1}_A$ est donc λ -intégrable. On applique le théorème ergodique :

$$\lim_{n \rightarrow \infty} \left(\frac{\sum_{0 \leq m < n} \mathbf{1}_{\{X_m \in A\}}}{\sum_{0 \leq m < n} \mathbf{1}_{\{X_m = x_0\}}} \right) = \frac{\lambda(A)}{\lambda(x_0)}; \quad \mathbb{P}_{x_0} \text{ p.s.} \quad (3.49)$$

Mais $\sum_{0 \leq m < n} \mathbf{1}_{\{X_m \in A\}} \leq n$, donc

$$\frac{\sum_{0 \leq m < n} \mathbf{1}_{\{X_m \in A\}}}{\sum_{0 \leq m < n} \mathbf{1}_{\{X_m = x_0\}}} \leq \frac{n}{\sum_{0 \leq m < n} \mathbf{1}_{\{X_m = x_0\}}}; \quad \forall n \geq 0.$$

On prend la \liminf , $n \rightarrow \infty$, de part et d'autre de cette inégalité, et on utilise (3.49),

$$\liminf_{n \rightarrow \infty} \frac{n}{\sum_{0 \leq m < n} \mathbf{1}_{\{X_m = x_0\}}} \geq \frac{\lambda(A)}{\lambda(x_0)}.$$

On choisit A_k une suite d'ensembles finis tels que $\bigcup_{k \geq 0} A_k = E$. Alors $\lambda(A_k) \rightarrow +\infty$. Le membre de gauche ne dépend pas de A . On remplace A par A_k dans l'inégalité précédente puis en faisant tendre $k \rightarrow +\infty$, on a

$$\liminf_{n \rightarrow \infty} \frac{n}{\sum_{0 \leq m < n} \mathbf{1}_{\{X_m = x_0\}}} = +\infty.$$

Ce qui signifie que la limite du rapport est infinie, il est alors évident que

$$\frac{1}{n} \sum_{0 \leq m < n} \mathbf{1}_{\{X_m = x_0\}} \xrightarrow{n \rightarrow \infty} 0, \quad \mathbb{P}_{x_0} \text{ p.s.}$$

On montre d'une manière analogue, que la deuxième limite faisant intervenir f , est nulle. \square

Remarques.

1) Supposons la chaîne irréductible, récurrente et positive.

On sait que $\frac{1}{n} \left(\sum_{0 \leq m < n} \mathbf{1}_{\{X_m=y\}} \right)$ converge \mathbb{P}_{x_0} p.s. vers $\lambda(y)$, où x et y sont deux éléments de E .

De plus : $0 \leq \frac{1}{n} \left(\sum_{0 \leq m < n} \mathbf{1}_{\{X_m=y\}} \right) \leq \frac{n}{n} = 1$. On peut appliquer le théorème de convergence dominée :

$$\lim_{n \rightarrow \infty} \mathbb{E}_x \left(\frac{1}{n} \left(\sum_{0 \leq m < n} \mathbf{1}_{\{X_m=y\}} \right) \right) = \lambda(y).$$

Mais

$$\mathbb{E}_x \left(\frac{1}{n} \sum_{0 \leq m < n} \mathbf{1}_{\{X_m=y\}} \right) = \frac{1}{n} \sum_{0 \leq m < n} \mathbb{P}_x(X_m = y) = \frac{1}{n} \sum_{0 \leq m < n} \pi^m(x, y).$$

Par conséquent

$$\lim_{n \rightarrow \infty} \left(\frac{\pi(x, y) + \dots + \pi^n(x, y)}{n} \right) = \lambda(y); \quad \forall x \in E. \quad (3.50)$$

On a une réciproque partielle de cette propriété, supposons que pour tout x et y de E , la limite de la moyenne de Césaro de $\pi^n(x, y)$ converge : $\frac{1}{n}(\pi(x, y) + \dots + \pi^n(x, y))$ converge vers un réel, indépendant de x , que l'on note $\lambda(y)$. Puisque $0 \leq \pi^n(x, y) \leq 1$ et $\sum_{y \in E} \pi^n(x, y) = 1$, on a, $0 \leq \lambda(y) \leq 1$,

$$0 \leq \frac{1}{n}(\pi(x, y) + \dots + \pi^n(x, y)) \leq 1, \quad (3.51)$$

et

$$\sum_{y \in E} \lambda(y) = \lim_{n \rightarrow \infty} \frac{1}{n} \left(\sum_{y \in E} \pi(x, y) + \dots + \sum_{y \in E} \pi^n(x, y) \right) = 1.$$

Ce qui signifie que λ est une probabilité. Montrons qu'elle est invariante. En utilisant à nouveau (3.51), on a :

$$\lambda\pi(y) = \sum_{z \in E} \lambda(z)\pi(z, y) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{z \in E} \sum_{k=1}^n \pi^k(x, z)\pi(z, y)$$

où $x \in E$. Mais

$$\sum_{z \in E} \sum_{k=1}^n \pi^k(x, z)\pi(z, y) = \sum_{k=1}^n \left(\sum_{z \in E} \pi^k(x, z)\pi(z, y) \right) = \sum_{k=1}^n \pi^k\pi(x, y) \sum_{k=1}^n \pi^{k+1}(x, y).$$

De plus

$$\frac{1}{n} \sum_{k=1}^n \pi^{k+1}(x, y) = \frac{1}{n} \sum_{k=2}^{n+1} \pi^k(x, y) = \left(\frac{n+1}{n} \right) \left[\frac{1}{n+1} \sum_{k=1}^{n+1} \pi^k(x, y) \right] - \frac{\pi(x, y)}{n}.$$

On en déduit : $\lambda\pi(y) = \lambda(y)$ pour tout y de E , λ est invariante. On sait que si une suite a_n converge, sa moyenne de Cesaro (i.e. la suite $(\frac{1}{n}(a_1 + \dots + a_n))_{n \geq 1}$) converge vers la même limite. L'approche que nous venons de développer généralise celle basée sur la propriété (3.38).

2) Si $(X_n)_{n \geq 0}$ est une suite de v.a.r. indépendantes, de même loi. (X_n) est une chaîne de Markov de probabilité de transition π , où $\pi(x, y) = \nu(y)$, ν désignant la loi commune des X_i . Supposons que les v.a. X_n soient à valeurs dans E fini ou dénombrable. Quitte à restreindre E , on peut supposer que E est le support de ν c'est-à-dire que $\nu(y) > 0$ pour tout y de E . On en déduit que $\pi(x, y) = \nu(y) > 0$ et $\pi(y, x) = \nu(x) > 0$, pour tout x et y de E , donc $x \rightarrow y$ et $y \rightarrow x$. Il y a une seule classe d'équivalence, la chaîne est nécessairement irréductible. Par définition, $\pi^n(x, y) = \mathbb{P}(X_n = y | X_0 = x) = \mathbb{P}(X_n = y) = \nu(y) > 0$. Par conséquent,

$$U(x, x) = \sum_{n \geq 0} \pi^n(x, x) = \sum_{n \geq 0} \nu(x) = +\infty.$$

Les états sont tous récurrents. La chaîne est récurrente irréductible. Il est facile de vérifier que ν est une probabilité invariante : si X_0 a pour loi ν , X_1 a pour loi ν . (X_n) est alors récurrente positive. Supposons que $E = \mathbb{Z}^d$ et ν admette un moment d'ordre 1 : $\int_{\mathbb{Z}^d} |x| \nu(dx) < +\infty$. On choisit $f(x) = x$, f est intégrable par rapport à ν , d'après le théorème ergodique on a :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \left(\sum_{i=1}^n X_i \right) = \int_{\mathbb{Z}^d} x f d\nu = \int_{\mathbb{Z}^d} x d\nu(x). \quad (3.52)$$

On retrouve, dans le cas discret (on a supposé que X_n est à valeurs dans \mathbb{Z}^d) la loi forte des grands nombres. Dans le cas général, où X_n est une chaîne de Markov à valeurs dans \mathbb{Z}^d , récurrente irréductible et positive de probabilité invariante ν , admettant un moment d'ordre 1, le résultat (3.52) est encore vrai. Dans ce cas les v.a. (X_0, \dots, X_n) ne sont pas indépendantes, le théorème ergodique apparaît comme une généralisation de la loi forte des grands nombres. On peut remarquer que la démonstration du théorème ergodique repose de manière essentielle sur la loi forte des grands nombres !

3.5 Exercices

3.1. Soit $\{X_n : n \geq 0\}$ une chaîne de Markov à valeurs dans (E, \mathcal{E}) , de loi initiale μ et de probabilité de transition π . Étudier si les suites suivantes sont des chaînes de Markov et calculer, éventuellement, les probabilités de transition :

- i) $Y_n := X_{2n}$;
- ii) $Y'_n := X_{n+k}$ et $Y''_n := X_{ln}$, où k, l sont deux entiers $k \geq 1, l > 1$;
- iii) $\tilde{Y}_m := X_{n_m}$, où $\{n_m : m \geq 0\} \subset \mathbb{N}$ est une suite croissante non-bornée;
- iv) $\tilde{Y}_n := f(X_n)$, où $f : (E, \mathcal{E}) \rightarrow (F, \mathcal{F})$ est une bijection;
- v) $Z_n := (X_n, X_{n+1})$;
- vi) $\tilde{Z}_n := (X_n, f(X_n))$, où $f : (E, \mathcal{E}) \rightarrow (F, \mathcal{F})$ est quelconque.

3.2. Soit $\{\xi_n : n \geq 0\}$ une suite de variables aléatoires indépendantes telles que $\mathbb{P}(\xi_n = \pm 1) = 1/2$. Étudier si les suites suivantes sont des chaînes de Markov ou pas :

- i) $S_n := \sum_{i=1}^n \xi_i$;
- ii) $M_n := \max\{S_k : 0 \leq k \leq n\}$;
- iii) $X_n := M_n - S_n$;
- iv) $Y_n := \xi_n \xi_{n+1}$;
- v) $Z_n := (\xi_n + \xi_{n+1})/2$.

3.3. Soient $\{X_n : n \geq 0\}$ et $\{Y_n : n \geq 0\}$ deux chaînes de Markov à valeurs dans \mathbb{Z} . On pose $Z_n := X_n + Y_n$. La suite $\{Z_n : n \geq 0\}$ est-elle nécessairement une chaîne de Markov ?

3.4. On suppose que le temps qu'il fera demain dépend des deux jours précédents. On suppose que :

$$\mathbb{P}\{\text{il pleut demain} \mid \text{il a plu hier et aujourd'hui}\} = 0,7$$

$$\mathbb{P}\{\text{il pleut demain} \mid \text{il a plu aujourd'hui mais pas hier}\} = 0,5$$

$$\mathbb{P}\{\text{il pleut demain} \mid \text{il a plu hier mais pas aujourd'hui}\} = 0,4$$

$$\mathbb{P}\{\text{il pleut demain} \mid \text{il n'a pas plu ni hier, ni aujourd'hui}\} = 0,2.$$

Montrer qu'on peut modéliser ceci par une chaîne de Markov. Quelle est la probabilité, sachant qu'il a plu lundi et mardi qu'il pleuve jeudi ?

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

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

i) Montrer que pour $(\alpha, \beta) \neq (0,0)$:

$$\pi^n = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{(1 - \alpha - \beta)^n}{\alpha + \beta} \begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix}.$$

Que se passe-t-il lorsque $\alpha = 0$ ou $\beta = 0$ ou $\alpha = \beta = 0$? On supposera pour la suite de l'exercice que $(\alpha, \beta) \neq (0, 0)$.

ii) Vérifier (par récurrence) que, pour toute loi initiale μ :

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

iii) Si $(\alpha, \beta) \neq (1, 1)$, montrer que $\{X_n : n \geq 0\}$ converge en loi vers une variable aléatoire de loi ν . Que vaut ν ? On supposera pour la suite de l'exercice que $(\alpha, \beta) \neq (1, 1)$.

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

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

v) Calculer

$$\text{Cov}_\nu(X_n, X_{n+1}) := \mathbb{E}_\nu(X_n X_{n+1}) - \mathbb{E}_\nu(X_n) \mathbb{E}_\nu(X_{n+1}).$$

Les variables aléatoires $\{X_n : n \geq 0\}$ sont-elles indépendantes?

vi) Calculer le potentiel de la chaîne.

vii) On note $S_n := \sum_{i=1}^n X_i$. Montrer que :

$$\mathbb{E}_\nu(S_n) = \frac{n\alpha}{\alpha + \beta}, \quad \text{Var}_\nu(S_n) \leq C n,$$

où C est une constante.

viii) (*Loi faible des grands nombres*) En déduire que :

$$\lim_{n \uparrow \infty} \frac{S_n}{n} = \frac{\alpha}{\alpha + \beta}$$

en probabilité sous \mathbb{P}_ν , sous \mathbb{P}_0 et sous \mathbb{P}_1 .

3.6. Soit $\{X_n : n \geq 0\}$ une chaîne de Markov à valeurs dans $E = \{0, 1, \dots, N\}$ de probabilité de transition π donnée par :

$$\pi(x, y) := \begin{cases} p & \text{si } y = x + 1 \\ 1 - p & \text{si } y = 0 \\ 0 & \text{sinon,} \end{cases} \quad 0 \leq x \leq N - 1,$$

où $0 < p < 1$ et avec N un état absorbant.

i) Classifier les états de la chaîne.

ii) Soit $T := \inf\{n \geq 0 : X_n = N\}$. Calculer, pour $x \in E$, $\mathbb{E}_x(T)$.

iii) Une pièce donne pile avec une probabilité p et elle est lancée successivement plusieurs fois. On fixe $N \in \mathbb{N}^*$. Montrer que, avec probabilité 1, on obtient N fois consécutives pile avec un nombre fini de lancers. Combien de lancers sont nécessaires en moyenne?

3.7. Soit $\{X_n : n \geq 0\}$ une chaîne de Markov à valeurs dans (E, \mathcal{E}) , de loi initiale μ et de

probabilité de transition π . Soit $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$. Une fonction $f : E \rightarrow \mathbb{R}$ est dite *harmonique* ou *invariante* si

$$\int_E \pi(x, dy) |f(y)| < \infty \text{ et } \int_E \pi(x, dy) f(y) = f(x), \forall x \in E.$$

i) Soit f invariante telle que $\int_E |f(y)| \mu(dy) < \infty$. Montrer que $\{f(X_n) : n \geq 0\}$ est une \mathcal{F}_n -martingale.

ii) Supposons que E est fini et qu'il y a deux états absorbants a et b : $\pi(a, a) = \pi(b, b) = 1$. Soit $\sigma(x) := \inf\{n \geq 0 : X_n = x\}$ et $T := \sigma(a) \wedge \sigma(b)$.

a) Montrer que : $\mathbb{P}(\forall n \geq 0, X_n = c \mid X_0 = c) = 1$, si $c = a$ ou b .

b) Supposons que, pour tout $x \in E$, $\pi(x, \{a, b\}) > 0$. On note

$$\xi := \inf\{\pi(x, \{a, b\}) : x \in E \setminus \{a, b\}\} > 0$$

et

$$A_n := \{X_i \neq a, X_i \neq b, \text{ pour tout } 0 \leq i \leq n\}.$$

Montrer que :

$$\mathbb{P}(A_n \mid X_0 = x) \leq (1 - \xi)^n.$$

En déduire que, pour tout $x \in E$, $\mathbb{P}_x(T < \infty) = 1$.

iii) On suppose que, pour tout $x \in E$, $\mathbb{P}_x(T < \infty) = 1$ et soit f invariante, bornée et telle que $f(a) \neq f(b)$. Calculer

$$\rho_{x,a} = \mathbb{P}_x(X_T = a) \text{ et } \rho_{x,b} = \mathbb{P}_x(X_T = b).$$

iv) Étudier le cas où $E = \{1, \dots, M\}$, $M \geq 3$ entier, $\pi(1, 1) = \pi(M, M) = 1$ et $\pi(i, i+1) = \pi(i, i-1) = 1/2$, pour $2 \leq i \leq M-1$.

3.8. Soit $\{X_n : n \geq 0\}$ une chaîne de Markov sur $E = \{1, 2, 3\}$ ayant la probabilité de transition

$$\pi := \begin{pmatrix} 0 & 1 - \alpha & \alpha \\ 1 - \beta & 0 & \beta \\ 1/3 & 1/3 & 1/3 \end{pmatrix}.$$

On définit la suite $\{Y_n : n \geq 0\}$ par

$$Y_n := \begin{cases} 1, & \text{si } X_n = 1 \text{ ou } 2 \\ 2, & \text{si } X_n = 3. \end{cases}$$

Montrer que si $\alpha = \beta$, alors $\{Y_n : n \geq 0\}$ est une chaîne de Markov sur $F = \{1, 2\}$.

3.9. Soit $\{\xi_n : n \geq 1\}$ une suite de variables aléatoires indépendantes à valeurs dans $E = \{1, \dots, N\}$ et de loi uniforme sur E .

i) Soit $X_n := \text{card}\{\xi_1, \dots, \xi_n\}$. Montrer que $\{X_n : n \geq 0\}$ est une chaîne de Markov et calculer

sa probabilité de transition π .

ii) Soit $v = (x_1, \dots, x_N)^*$ un vecteur propre pour la valeur propre λ de π . Vérifier que :

$$\left(\lambda - \frac{j}{N}\right) x_j = \left(1 - \frac{j}{N}\right) x_{j+1}.$$

En déduire que π admet N valeurs propres distinctes, λ_i , $i = 1, \dots, N$, et que les vecteurs propres correspondants v^i peuvent être écrits :

$$v^i := \left(\frac{C_i^1}{C_N^1}, \frac{C_i^2}{C_N^2}, \dots, \frac{C_i^i}{C_N^i}, 0, \dots, 0\right)^*.$$

iii) Montrer que π^* admet λ_i , $i = 1, \dots, N$, comme valeurs propres et $w^i = (w_1^i, \dots, w_N^i)$ comme vecteur propre associé à λ_i , avec

$$w_j^i = (-1)^{j-i} C_{N-i}^{j-i}, \quad j = 1, \dots, N.$$

iv) Vérifier que v^i est orthogonal à w^j si $i \neq j$ (on pourra calculer $(\pi v^i)^* w^j$).

v) En déduire :

$$\pi^n = \sum_{i=1}^N \frac{v^i (w^i)^*}{(w^i)^* v^i} \lambda_i^n.$$

En particulier :

$$\pi_{jk}^n = \begin{cases} C_{N-j}^{N-k} \sum_{l=0}^{k-j} (-1)^{k-j-l} C_{k-j}^l \left(\frac{l+j}{N}\right)^N, & \text{si } k \geq j \\ 0, & \text{si } k < j. \end{cases}$$

3.10. On joue à pile ou face, avec la probabilité p d'obtenir face. Les lancers sont indépendants. On dit qu'on est dans l'état $E_1 := (F, F)$ à l'instant n si l'on a tiré face en n et $n-1$; de même on note $E_2 := (P, P)$, $E_3 := (P, F)$, $E_4 := (F, P)$.

i) Montrer que la suite des lancers constitue une chaîne de Markov et calculer sa probabilité de transition π .

ii) Trouver la probabilité de passer de E_i à E_j pendant le laps de temps k .

iii) Calculer la mesure stationnaire. Que peut-on dire quant à la récurrence de la chaîne?

3.11. Modèle de Wright. Une population cellulaire comprend $2N$ cellules qui peuvent être de deux types : A ou a . Si la population mère compte j cellules de type A et $2N - j$ cellules de type a ; à la génération suivante le nombre de cellules de type A vaut $\sum_{i=1}^{2N} Z_i$, où les variables Z_i sont indépendantes et de même loi $\mathbb{P}(Z_i = 1) = j/(2N)$ et $\mathbb{P}(Z_i = 0) = (2N - j)/(2N)$. On note X_n le nombre de cellules de type A de la n -ième génération.

i) Vérifier que $\{X_n : n \geq 0\}$ est une chaîne de Markov et donner sa probabilité de transition π .

ii) Que vaut $\pi(0, j)$ et $\pi(2N, j)$? La chaîne est-elle irréductible? Quels sont les états récurrents?

iii) Montrer que $\{X_n : n \geq 0\}$ est une martingale qui converge vers une variable X_∞ . Que valent les probabilités de "fixation" en A et a ?

3.12. Une population cellulaire comprend N cellules qui peuvent être de deux types : A ou a . On passe d'une génération à la suivante en dédoublant chaque cellule mère et en choisissant au hasard N cellules parmi les $2N$. Les choix successifs sont supposés indépendants. On note X_n le nombre de cellules de type A de la n -ième génération.

i) Montrer que $\{X_n : n \geq 0\}$ est une chaîne de Markov dont la probabilité de transition vaut :

$$\pi(x, y) := \frac{C_{2x}^y C_{2(N-x)}^{N-y}}{C_{2N}^N}.$$

ii) Montrer que 0 et N sont des états absorbants. Si on note $T_i := \inf\{n \geq 0 : X_n = i\}$, prouver que :

$$\mathbb{P}_x(T_N < T_0) = x/N \text{ et } \mathbb{P}_x(T_0 < T_N) = 1 - x/N.$$

3.13. Modèle de Ehrenfest. Étant donné deux enceintes séparées par une paroi poreuse et contenant ensemble N particules diffusant à travers cette paroi, on décrit le nombre aléatoire X_n de particules se trouvant dans la première enceinte ($0 \leq X_n \leq N$) aux instants successifs $n \in \mathbb{N}$ de transition des particules par une chaîne Markov de probabilité de transition :

$$\pi(x, x-1) := \frac{x}{N}, \quad \pi(x, x+1) := \frac{N-x}{N}, \quad 0 \leq x \leq N.$$

À chaque instant n , les probabilités que la transition se fasse de la première enceinte vers la seconde ou de la seconde enceinte vers la première sont donc proportionnelles aux nombres de particules en présence dans la première et la deuxième enceinte :

$$\frac{\pi(x, x-1)}{x} = \frac{\pi(x, x+1)}{N-x}.$$

Montrer que cette chaîne est récurrente irréductible et trouver sa mesure invariante.

3.14. Modèle de Laplace-Bernoulli. N boules noires et N boules blanches sont placées dans deux urnes de sorte que chacune contienne N boules. Après chaque unité de temps on choisit au hasard une boule de chaque urne ; les deux boules ainsi choisies changent d'urne. On note par X_n le nombre de boules noires dans la première urne. Montrer que $\{X_n : n \geq 0\}$ est une chaîne de Markov irréductible et trouver sa mesure stationnaire.

3.15. Soit X_n le nombre de particules présentes à chaque instant n dans un volume donné V . On fait l'hypothèse que, pendant l'intervalle de temps $[n, n+1[$, chacune des X_n particules a la probabilité p , $0 < p < 1$, de quitter le volume considéré V , et que, d'autre part, pendant cet intervalle de temps, un nombre aléatoire de particules suivant une loi de Poisson, de paramètre λ , entre dans V . On suppose, en outre, que les divers phénomènes aléatoires ainsi considérés sont indépendants les uns des autres.

i) Montrer que $\{X_n : n \geq 0\}$ est une chaîne de Markov de probabilité de transition :

$$\pi(x, y) := \sum_{k=(x-y)^+}^x C_x^k p^k (1-p)^{x-k} e^{-\lambda} \frac{\lambda^{y-x+k}}{(y-x+k)!}.$$

ii) Calculer $\mathbb{E}_x(e^{itX_1})$.

iii) En déduire la fonction caractéristique de X_1 , lorsque X_0 suit une loi de Poisson de paramètre θ . En déduire que la loi de Poisson de paramètre λ/p est une mesure stationnaire pour $\{X_n : n \geq 0\}$.

iv) Montrer que $\{X_n : n \geq 0\}$ est une chaîne irréductible, récurrente et positive.

v) Calculer

$$\lim_{n \uparrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} X_i.$$

3.16. Soit $\{X_n : n \geq 0\}$ une chaîne de Markov à valeurs dans \mathbb{N} de probabilité de transition $\pi : \pi(n, n+1) := \theta_n, \pi(n, 0) := 1 - \theta_n$, où $0 < \theta_n < 1$.

i) Soit $T := \inf\{n \geq 1 : X_n = 0\}$. Calculer la loi de T lorsque $X_0 = 0$. En déduire que 0 est un état transitoire si et seulement si $\prod_{n=0}^{\infty} \theta_n > 0$.

ii) Étudier la récurrence et la transience des autres états.

iii) Calculer

$$\nu(k) := \mathbb{E}_0 \left(\sum_{n=0}^{T-1} \mathbf{1}_{\{X_n=k\}} \right).$$

Vérifier que $(\nu(k) : k \geq 0)$ est une mesure invariante.

3.17. On considère une machine qui tombe en panne en une suite de temps aléatoires ξ_1, ξ_2, \dots et on suppose que $\xi_1, \xi_2 - \xi_1, \dots$ sont indépendantes, de même loi. On note $p_k := \mathbb{P}(\xi_1 = k)$, $k \geq 0$. On suppose que si la machine tombe en panne elle est réparée instantanément. Soit X_n la période de temps qui sépare n de la prochaine panne.

i) Montrer que $\{X_n : n \geq 0\}$ est une chaîne de Markov de probabilité de transition $\pi(0, k) = p_k, \pi(k+1, k) = 1$, pour $k \geq 0$.

ii) On suppose maintenant $p_k > 0$ pour tout $k > 0$. Montrer que $\{X_n : n \geq 0\}$ est récurrente et irréductible.

iii) Calculer la loi de X_1 en fonction de celle de X_0 . En déduire que si $\sum_{k \geq 1} kp_k < \infty$, il existe une seule mesure invariante.

iv) Soit $T := \inf\{n \geq 1 : X_n = 0\}$. Calculer

$$\nu(k) := \mathbb{E}_0 \left(\sum_{n=0}^{T-1} \mathbf{1}_{\{X_n=k\}} \right).$$

Que peut-on remarquer ?

3.18. Soit $\{X_n : n \geq 0\}$ une chaîne de Markov à valeurs dans \mathbb{N} , de probabilité de transition $\pi :$

$$\pi(0, 0) = 1, \pi(i, j) := e^{-i} \frac{i^j}{j!}, i \geq 1, j \geq 0.$$

i) Classifier les états de cette chaîne. Quels sont les états récurrents ? transients ?

ii) Montrer que $f(j) = j$ est harmonique et que, pour tout $x \in \mathbb{N}$, $\{X_n : n \geq 0\}$ est une

\mathbb{P}_x -martingale qui converge vers une variable aléatoire X_∞ .

iii) Quelle est la loi de X_∞ ?

3.19. Une souris se promène sur un domaine carré à 4×4 cases (les déplacements diagonaux sont interdits). À chaque étape, elle passe de la case qu'elle occupe à l'une des r cases voisines adjacentes avec la probabilité $1/r$ (si une case est centrale, $r = 4$; si une case est au bord, mais pas au coin, $r = 3$; si une case est au coin, $r = 2$). Soit X_n la position de la souris à l'étape n .

i) Prouver que la chaîne $\{X_n : n \geq 0\}$ est irréductible et récurrente.

ii) Calculer la probabilité invariante.

3.20. L'épreuve d'un livre est lue par une suite infinie d'éditeurs qui cherchent les fautes. Chaque faute est détectée avec la probabilité p à chaque lecture. Entre deux lectures, l'imprimeur corrige les fautes détectées, mais introduit un nombre aléatoire de nouvelles erreurs distribué suivant une loi de Poisson (des erreurs peuvent être introduites même s'il n'y a pas de faute détectée). On suppose que les phénomènes aléatoires sont indépendantes et que les nombres de nouvelles erreurs après chaque lecture sont identiquement distribués. On note X_n le nombre d'erreurs après le n -ième cycle éditeur-imprimeur. Trouver l'expression de la fonction génératrice $G_n(s) = \mathbb{E}(s^{X_n})$. Calculer la loi stationnaire de la chaîne $\{X_n : n \geq 0\}$.

3.21 Soit $\{X_n : n \geq 0\}$ une chaîne de Markov à valeurs dans (E, \mathcal{E}) , dénombrable, et de probabilité de transition π . On suppose qu'il existe un état $i_0 \in E$ tel que i_0 conduit à tout autre état $j \in E$ et que si on note $T_{i_0} := \inf\{n \geq 0 : X_n = i_0\}$, alors $\mathbb{P}_j(T_{i_0} < \infty) = 1$ pour tout $j \in E$. Montrer que la chaîne est récurrente.

3.22. Chaîne de naissance et de mort Soit $\{X_n : n \geq 0\}$ une chaîne de Markov à valeurs dans \mathbb{N} de matrice de transition π :

$$\pi(0,0) = r_0, \pi(0,1) = p_0, \pi(i,j) := \begin{cases} p_i & \text{si } j = i + 1 \\ r_i & \text{si } j = i \\ q_i & \text{si } j = i - 1, \end{cases} \quad \text{pour } i \geq 1,$$

où $0 < p_i, q_i \leq 1$, $p_0 + r_0 = 1$ et pour $i \geq 1$ $r_i \geq 0$, $p_i + q_i + r_i = 1$. On note

$$\gamma_0 = 1, \gamma_i := \frac{q_1 \cdots q_i}{p_1 \cdots p_i}.$$

i) Étudier les fonctions u vérifiant $\pi u(i) = u(i)$, $i \in \{a+1, \dots, b-1\}$, pour $a, b \in \mathbb{N}$, $a < b$.

ii) Soit $T_i := \inf\{n \geq 0 : X_n = i\}$ et $\tau := T_a \wedge T_b$. Pour $a \leq i \leq b$ montrer que $\mathbb{P}_i(\tau < \infty) = 1$. Exprimer $v(i) := \mathbb{P}_i(X_\tau = b)$ en fonction des γ_i .

iii) Exprimer $\{T_0 = \infty\}$ en fonction des événements $\{T_n < T_0\}$ et calculer $\mathbb{P}_1(T_n \leq T_0)$. En déduire la valeur de $\mathbb{P}_1(T_0 = \infty)$ et une condition de récurrence de la chaîne en fonction des γ_i .

iv) Montrer que toute mesure ν satisfaisant

$$\nu(i)\pi(i,j) = \nu(j)\pi(j,i), \quad i, j \in \mathbb{N}$$

est de la forme :

$$\nu(i) := \alpha \zeta_i,$$

où on a posé

$$\zeta_0 = 1, \zeta_i = \frac{p_0 \cdots p_{i-1}}{q_1 \cdots q_i}.$$

Montrer que la mesure définie par la formule précédente est stationnaire.

v) Sous quelles conditions existe-t-il une probabilité stationnaire ? L'exprimer alors en fonction des ζ_i .

vi) On suppose $p_i = p$ pour tout $i \in \mathbb{N}$ et $q_i = q$, pour tout $i \geq 1$. Étudier la récurrence de la chaîne suivant les différentes valeurs de p, q . Sous quelles hypothèses sur p et q existe-t-il une probabilité stationnaire ?

3.23. Soit $\{X_n : n \geq 0\}$ une chaîne de Markov sur $E = \{0, 1, \dots, N\}$ ayant la probabilité de transition

$$\pi := \begin{pmatrix} q_0 & p_0 & 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ q_1 & 0 & p_1 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & q_2 & 0 & p_2 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & q_{N-2} & 0 & p_{N-2} & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & q_{N-1} & 0 & p_{N-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 0 & q_N & p_N \end{pmatrix},$$

où $0 < p_i, q_i < 1$ et $p_i + q_i = 1$, $i = 0, \dots, N$.

i) Soit $A := \{0, N\}$. Déterminer les probabilités pour que la chaîne atteigne l'état N avant d'atteindre l'état 0 ,

$$p_i(N, A) := \mathbb{P}_i(T_A < \infty, X_{T_A} = N), T_A := \inf\{n \geq 0 : X_n \in A\}.$$

ii) Trouver $p_i(N, A)$ dans le cas où $p_i = p = 1 - q$, $i = 0, \dots, N$.

iv) Sous les hypothèses du point précédent, trouver la mesure stationnaire.

3.24. Chaîne autorégressive d'ordre 1 (AR1). Soit $\{\xi_n : n \geq 0\}$ une suite de variables aléatoires réelles indépendantes de même loi, centrées et de variance σ^2 . Soit X_0 une variable aléatoire réelle, indépendante de cette suite. Pour tout réel θ , $|\theta| < 1$ on définit par récurrence :

$$X_0^\theta := X_0, X_{n+1}^\theta := \theta X_n^\theta + \xi_{n+1}, n \geq 0.$$

i) Montrer que $\{X_n^\theta : n \geq 0\}$ est une chaîne de Markov.

ii) Soit

$$Y_n^\theta := \theta^n X_0 + \xi_1 + \theta \xi_2 + \dots + \theta^{n-1} \xi_n, n \geq 0.$$

Montrer que $\{Y_n^\theta : n \geq 0\}$ converge dans L^2 et presque sûrement vers une variable aléatoire Y^θ de loi μ^θ . Prouver que $\{X_n^\theta : n \geq 0\}$ converge en loi vers μ^θ .

iii) Montrer que :

$$\lim_{n \uparrow \infty} \mathbb{E}(X_n^\theta) = 0 \text{ et } \lim_{n \uparrow \infty} \mathbb{E}[(X_n^\theta)^2] = \frac{\sigma^2}{1 - \theta^2}.$$

3.25. Modèle Galton-Watson. Soit $\nu = \{p_k : k \geq 0\}$ une probabilité sur \mathbb{N} . Soit la chaîne $\{X_n : n \geq 0\}$ définie sur \mathbb{N} par :

$$\mathbb{P}(X_0 = 1) = 1, \mathbb{P}(X_{n+1} = j \mid X_n = i) = \nu^{*i}(j), \nu^{*0} = \delta_0.$$

- i) Décrire le modèle associé à cette chaîne.
 ii) Observer que 0 est un état absorbant. Classifier les autres états.
 iii) Soit $G_n(s) = \mathbb{E}(s^{X_n})$ la fonction génératrice de X_n . Calculer G_n en fonction de $G(s) := \sum_{k \geq 0} s^k \mathbb{P}_1(X_1 = k) = \mathbb{E}_1(s^{X_1})$.
 iv) Soit $A := \{\exists n : X_n = 0\}$ l'événement extinction de la chaîne. Prouver que $\mathbb{P}(A) = \rho$ où ρ est la plus petite racine de l'équation $G(s) = s$.
 v) Montrer que $\mathbb{P}(A) < 1 \Leftrightarrow \mathbb{E}_1(X_1) = \sum_{k \geq 0} k p_k > 1$.
 vi) Montrer que $\mathbb{E}[X_n \mid X_{n-1}] = m X_{n-1}$, où m est l'espérance de ν , $m = \sum_{k \geq 0} k p_k$. En déduire $\mathbb{E}(X_n)$. Retrouver le résultat en utilisant la fonction G_n .
 vii) Supposons $m > 1$ et soit ρ l'unique solution de $G(s) = s$ dans $[0, 1[$. Montrer que $M_n = \rho^{X_n}$ est une martingale et retrouver le fait que $\rho = \mathbb{P}(T_0 < \infty)$, où $T_0 = \inf\{n \geq 0 : X_n = 0\}$.
 viii) On note σ^2 la variance de ν . Montrer que $\mathbb{E}[X_n^2 \mid X_{n-1}] = \sigma^2 X_{n-1} + m^2 X_{n-1}^2$. En déduire $\text{Var}(X_n)$. Retrouver le résultat en utilisant la fonction G_n .
 ix) Supposons $m > 1$ et soit $W_n = X_n/m^n$. Montrer que W_n est une martingale bornée dans L^2 qui converge vers une variable aléatoire W_∞ . Vérifier que la fonction caractéristique φ de W_∞ satisfait $G(\varphi(t)) = \varphi(mt)$.
 x) Que vaut $\mathbb{P}_i(A)$?

3.26. i) Montrer que, pour toute variable aléatoire à valeurs dans \mathbb{R}_+ , $\mathbb{E}(X \mid X > 0) \leq \mathbb{E}(X^2)/\mathbb{E}(X)$.

ii) Soit $\{X_n : n \geq 0\}$ un processus de Galton-Watson avec $X_0 = 1$ et $\mathbb{P}(X_1 = k) = qp^k$, $k \geq 0$, avec $p > 1/2$ et $q = 1 - p$. Montrer que :

$$\mathbb{E} \left(\frac{X_n}{\mu^n} \mid X_n > 0 \right) \leq \frac{2p}{p-q}, \mu := p/q.$$

iii) Montrer que :

$$\lim_{n \uparrow \infty} \mathbb{E} \left(\frac{X_n}{\mu^n} \mid X_n > 0 \right) = \frac{p}{p-q}.$$

3.27. Soit $\{p(x) : x \in \mathbb{Z}\}$ une probabilité sur \mathbb{Z} , et soit, pour $s \in [0, 1]$,

$$\hat{p}(s) := \sum_{x \in \mathbb{Z}} p(x) e^{2i\pi s x}.$$

i) Montrer que $\int_0^1 \hat{p}(s) ds = p(0)$.

ii) Montrer que $\hat{p}(s) \neq 1$ pour tout $s \neq 0$ si et seulement si le p.g.c.d. des x tels que $p(x) \neq 0$ vaut 1. On pourra considérer

$$\text{Re}(1 - \hat{p}(s)) = \sum_{x \in \mathbb{Z}} p(x) (1 - \cos(2\pi s x)) \geq 0.$$

iii) Soit π la probabilité de transition sur \mathbb{Z} définie par $\pi(x, y) := p(x - y)$. Vérifier que le potentiel $U := \sum_{n \geq 0} \pi^n$ a la même forme, soit $U(x, y) = u(x - y)$.

iv) On suppose que p est symétrique, c'est-à-dire $p(-x) = p(x)$, pour tout $x \in \mathbb{Z}$, et que $\hat{p}(s) \neq 1$ pour tout $s \neq 0$. Montrer que :

a) \mathbb{Z} est une classe de récurrence si $\int_0^1 ds/(1 - \hat{p}(s)) = \infty$;

b) tous les états sont transitoires si $\int_0^1 ds/(1 - \hat{p}(s)) < \infty$.

v) Soit ν la loi de probabilité sur \mathbb{Z} définie par :

$$\nu(-2) = \nu(-1) = \nu(1) = \nu(2) = 1/4.$$

Soit ξ_1, ξ_2, \dots une suite de variables aléatoires indépendantes de loi ν . On pose $S_0 = 0$ et $S_n = \xi_1 + \dots + \xi_n$, pour $n \geq 1$. Montrer que $\{S_n : n \geq 0\}$ est une chaîne de Markov de probabilité de transition $\pi(x, y) = \nu(x - y)$, $x, y \in \mathbb{Z}$.

vi) Montrer que $\hat{\nu}(s) < 1$, pour tout $s \neq 0$ (modulo 1).

vii) Pour $y \in \mathbb{Z}$, soit $p_n(y) := \mathbb{P}(S_n = y)$ et $U(y) := \sum_{n \geq 0} p_n(y)$. Prouver que :

a) $\hat{p}_n(s) = [\hat{\nu}(s)]^n$;

b) $\hat{U}(s) = 1/(1 - \hat{\nu}(s))$.

viii) Montrer que $\int_0^1 ds/(1 - \hat{\nu}(s)) = \infty$.

ix) Montrer que la chaîne est irréductible.

x) Que vaut $\sum_{n \geq 0} \mathbf{1}_{\{S_n = y\}}$?

Chapitre 4

Simulation

4.1 Introduction

L'usage de méthodes de simulation pour résoudre des problèmes déterministes tels que calculs d'intégrales, de solutions d'équations intégrales ou aux dérivées partielles, présente rarement un intérêt clair en ce qui concerne la rapidité des algorithmes. L'avantage réel est qu'elles sont souvent faciles à programmer et font gagner du temps. Donnons un exemple : soit f une fonction continue et $I = \int_0^1 f(t)dt$. On peut approximer I par un schéma aléatoire, en considérant une suite $(U_n)_{n \geq 1}$ de v.a. indépendantes, chacune de loi uniforme sur $[0, 1]$ et

$$X_n = \frac{1}{n}(f(U_1) + \dots + f(U_n)); \quad n \geq 1.$$

Les v.a. $(f(U_n))_{n \geq 1}$ sont indépendantes, équidistribuées et possèdent un moment d'ordre 1. D'après la loi forte des grands nombres, X_n converge presque sûrement, vers $\mathbb{E}(f(U_1))$. Mais $\mathbb{E}(f(U_1)) = \int_0^1 f(t)dt = I$.

On utilise plutôt la simulation pour résoudre des problèmes de nature aléatoire. Considérons l'exemple suivant. Un système en réseau est formé de sept composants fonctionnant en série ou en parallèle selon le schéma suivant : L'ensemble est en état de marche tant qu'il existe un trajet allant de A à B ne rencontrant aucun appareil en panne. On suppose que les différents éléments constitutifs de ce montage ont des durées de vie aléatoire T_1, T_2, \dots, T_7 . La durée de vie de l'ensemble est :

$$T = T_7 \vee [(T_1 \vee T_2 \vee T_3) \wedge (T_6 \vee (T_4 \wedge T_5))]. \quad (4.1)$$

où $x \vee y = \sup(x, y)$ et $x \wedge y = \inf(x, y)$. On notera,

$$T = f(T_1, T_2, \dots, T_7). \quad (4.2)$$

Si deux appareils ont pour durée de vie ξ et ξ' , ces deux éléments, placés en série (resp. en parallèle) ont pour durée de vie $\xi \wedge \xi'$ (resp. $\xi \vee \xi'$). On en déduit aisément (4.1). On suppose connue la loi de (T_1, T_2, \dots, T_7) et on voudrait évaluer la durée de vie moyenne de ce système, c'est-à-dire $\mathbb{E}(T)$. Un utilisateur pourrait chercher à placer les différents composants

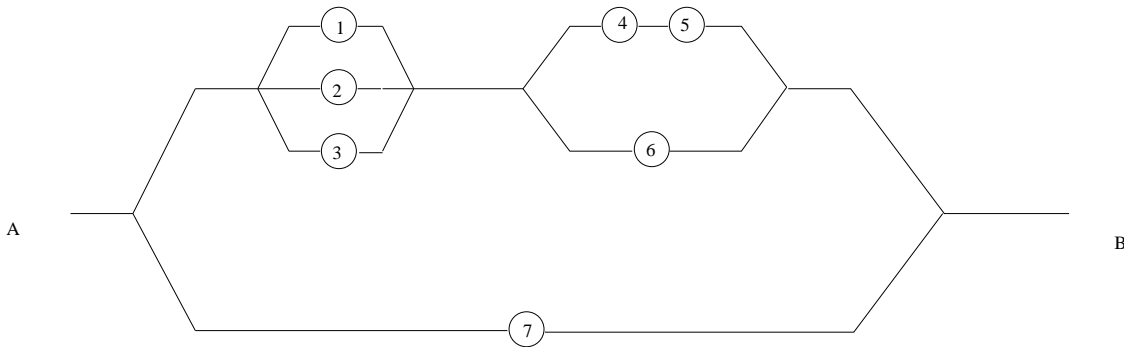


FIG. 4.1 – Système en réseau

en cherchant à optimiser la durée de vie, d'où l'intérêt de connaître cette quantité.

Connaissant la loi de (T_1, T_2, \dots, T_7) , la relation (4.1) permet d'exprimer $\mathbb{E}(T)$ comme une intégrale itérée d'ordre sept, mais le calcul peut s'avérer long et surtout impossible à mener jusqu'au bout. En pratique on ne cherche pas une valeur exacte de $\mathbb{E}(T)$, une valeur approchée suffit. L'idée consiste à considérer une suite de v.a. indépendantes,

$$((T_1^{(k)}, \dots, T_7^{(k)}); k \geq 1) \text{ de même loi que } (T_1, T_2, \dots, T_7).$$

Avec de bonnes hypothèses sur la loi de (T_1, T_2, \dots, T_7) , la loi forte des grands nombres nous dit que :

$$\frac{1}{n} \left(\sum_{k=1}^n T^{(k)} \right) \text{ converge p.s. vers } \mathbb{E}(T), \quad (4.3)$$

où $T^{(k)} = f(T_1^{(k)}, \dots, T_7^{(k)})$.

Cette approche repose sur le fait qu'il est possible de reproduire, de manière indépendante à chaque fois, le fonctionnement de l'appareil. Pour diverses raisons évidentes, coût, perte de temps, il est hors de question de réaliser physiquement ces différentes expériences.

La question est alors la suivante : est-il possible de simuler par une machine, des v.a. de lois données ? Afin de simplifier l'approche, nous allons nous restreindre au cas des variables aléatoires à valeurs réelles. La simulation de v.a.r. repose sur la simulation de v.a. de loi uniforme sur $[0, 1]$. Comment choisir un réel au hasard, appartenant à $[0, 1]$? La première idée consiste à choisir aléatoirement les chiffres de son développement décimal. Il s'agit de choisir uniformément un chiffre parmi $0, 1, \dots, 9$ et de recommencer la procédure ; ceci peut être réalisé en tirant d'une urne une boule parmi 10 boules indiscernables, numérotées de 0 à 9. On note $\alpha_1, \alpha_2, \dots, \alpha_n, \dots$ les résultats des tirages successifs. Pour tout n on associe $U_n = \frac{\alpha_1}{10} + \dots + \frac{\alpha_n}{10^n}$. U_n a même loi que la partie entière de $10^n U$, où U suit une loi uniforme sur $[0, 1]$. Ce procédé permet d'approximer une v.a. de loi uniforme d'aussi près que l'on veut, mais on a recours à une opération physique de tirages successifs. On peut y remédier en utilisant des tables de nombres au hasard, mais cette solution ne convient pas à un usage intensif de v.a. Les ordinateurs,

calculatrices, utilisent une méthode tout à fait différente, basée sur la création de nombres pseudo-aléatoires. Soient m un entier assez grand, et ξ une v.a. à valeurs dans $\{0, 1, \dots, m - 1\}$ et de loi uniforme. La v.a. $U' = \xi/m$ est une approximation d'ordre $\frac{1}{m}$ d'une v.a. U de loi uniforme sur $[0, 1]$: U' a même loi que la partie entière de mU . En effet soit $0 \leq k \leq m - 1$, on a

$$\mathbb{P}\left(U' = \frac{k}{m}\right) = \mathbb{P}(\xi = k) = \frac{1}{m},$$

$$\mathbb{P}(\lfloor mU \rfloor = k) = \mathbb{P}(k \leq mU < k + 1) = \mathbb{P}\left(\frac{k}{m} \leq U < \frac{k + 1}{m}\right) = \frac{1}{m}.$$

où $\lfloor a \rfloor$ désigne la partie entière de a . Un procédé très employé pour générer des suites (ξ_k) d'entiers à valeurs dans $\{0, 1, \dots, m - 1\}$ est la suivante :

$$\xi_{n+1} = a\xi_n \pmod{m} \tag{4.4}$$

où a est un entier fixé. La valeur initiale ξ_0 est appelée la graine ou racine. On ne peut échapper à l'arbitraire du choix de la racine, mais en contrepartie on a l'avantage de pouvoir reproduire à volonté la même suite "aléatoire" sans pour autant la garder en mémoire, il suffit en effet d'utiliser la même racine.

Le caractère aléatoire de telles suites soulève des doutes, on construit des suites périodiques. On teste statistiquement, via la loi faible des grands nombres, le théorème central limite, le test du chi-deux, les propriétés aléatoires d'un tel type de générateur. Les couples (a, m) qui ont passé avec succès les tests statistiques sont les suivants :

a	2100005341	69069	$7^5 = 16807$	3^{17}	$2^{22} - 2^{14} + 4$	$2^{27} - 2^{18} - 1$
m	$2^{31} - 1$	2^{32}	$2^{31} - 1$	2^{29}	$2^{31} - 1$	$2^{35} - 2^5 + 1$

Nota : $2^{32} = 4294967296$; $2^{29} = 536870912$; $2^{31} - 1 = 2147483647$.

Dans la suite on suppose que l'on dispose d'un générateur de nombres aléatoires parfait permettant de fournir des v.a., indépendantes, de loi uniforme, et en aussi grand nombre que l'on veut.

En matlab : `rand(n,m)` pour une matrice aléatoire $n \times m$.

4.2 Procédés de simulations de variables aléatoires à valeurs réelles

Simulations de variables aléatoires courantes

a) v.a. discrètes.

Soit X une v.a. discrète : $X(\Omega) = \{x_i; i \in I\}$, I peut être fini ou dénombrable. On note $p_i = \mathbb{P}(X = x_i)$. Si U désigne une v.a. de loi uniforme sur $[0, 1]$, alors

$$X' = x_0 \mathbb{1}_{\{U < p_0\}} + x_1 \mathbb{1}_{\{p_0 \leq U < p_0 + p_1\}} + \dots + x_i \mathbb{1}_{\{p_0 + \dots + p_{i-1} \leq U < p_0 + \dots + p_i\}} + \dots \tag{4.5}$$

a même loi que X . En effet,

$$\mathbb{P}(X' = x_0) = \mathbb{P}(U < p_0) = p_0,$$

$$\mathbb{P}(X' = x_i) = \mathbb{P}(p_0 + \dots + p_{i-1} \leq U < p_0 + \dots + p_i) = p_i, \text{ lorsque } i \geq 1.$$

Si X prend un nombre fini de valeurs x_0, x_1, \dots, x_n , alors

$$X' = x_0 \mathbb{1}_{\{U < p_0\}} + x_1 \mathbb{1}_{\{p_0 \leq U < p_0 + p_1\}} + \dots + x_n \mathbb{1}_{\{p_0 + \dots + p_{i-1} \leq U < 1\}}. \quad (4.6)$$

Pour la simulation :

```

coin=rand(1,m) ;
cumprob=[0 cumsum(p)] ;
for j=1 :n
ind=find((coin>cumprob(j))&(coin<=cumprob(j+1))) ;
end ;
Xprime=x(ind) ;

```

En particulier si X suit une loi de Bernoulli, $\mathbb{P}(X = 0) = q$ et $\mathbb{P}(X = 1) = p$, avec $p > 0$, $q > 0$, $p + q = 1$, alors

$$X' = \mathbb{1}_{\{U \leq p\}}. \quad (4.7)$$

```
rand<=p
```

b) Loi binomiale.

Si X suit la loi binomiale $\mathcal{B}(n, p)$, on peut utiliser le schéma général précédent ou

$$X' = \sum_{i=1}^n \varepsilon_i = \sum_{i=1}^n \mathbb{1}_{\{U_i \leq p\}} \quad (4.8)$$

où $(U_i; 1 \leq i \leq n)$ désignent n v.a. indépendantes et de loi uniforme sur $[0, 1]$.

```
x=sum(rand(1,n)<=p) ;
```

c) Loi exponentielle de paramètre λ .

On prend

$$X' = -\lambda \ln U \quad (4.9)$$

En effet

$$\mathbb{P}(X' \leq x) = \mathbb{P}(-\lambda \ln U \leq x) = \mathbb{P}(U \geq e^{-x/\lambda}) = 1 - e^{-x/\lambda}; \quad x \geq 0.$$

On en déduit que X' admet pour densité $\mathbb{1}_{\{x > 0\}} e^{-x/\lambda}$.

```
x=-log(rand)*lambda ;
```

d) Loi géométrique.

X a pour loi $\mathcal{G}(p)$, $0 < p < 1$, ce qui signifie que

$$\mathbb{P}(X = k) = p(1 - p)^{k-1}; \quad k \geq 1.$$

– Utilisons le schéma général de simulation des v.a. discrètes :

$$X' = \mathbb{1}_{\{U < p\}} + \sum_{i \geq 1} i \mathbb{1}_{\{(1-p)^i \leq 1-U < (1-p)^{i-1}\}}, \quad (4.10)$$

car

$$\sum_{i=0}^n p_i = \sum_{i=1}^n p(1-p)^{i-1} = 1 - (1-p)^n.$$

`x=floor(-log(rand)/(-log(1-p))) ;`

– On peut utiliser une procédure utilisant le fait que la loi de X est celle du nombre d'essais indépendants nécessaires, pour obtenir la réalisation d'un événement de probabilité p :

$$X' = \sum_{i \geq 1} i \mathbb{1}_{\{\varepsilon_1=0, \dots, \varepsilon_{i-1}=0, \varepsilon_i=1\}} \quad (4.11)$$

où $(\varepsilon_i)_{i \geq 1}$ est une suite de v.a. de Bernoulli, indépendantes et de même loi : $\mathbb{P}(\varepsilon_1 = 1) = p$, $\mathbb{P}(\varepsilon_1 = 0) = q$. Si de plus on simule ε_i par des v.a. U_i de loi uniforme ($\varepsilon_i = \mathbb{1}_{\{U_i \leq p\}}$), on a

$$X' = \sum_{i \geq 1} i \mathbb{1}_{\{U_1 > p, \dots, U_{i-1} > p, U_i \leq p\}}. \quad (4.12)$$

e) Loi de Poisson.

Soit X une v.a. de loi de Poisson $\mathcal{P}(\lambda)$: $\mathbb{P}(X = n) = \frac{\lambda^n}{n!} e^{-\lambda}$, pour tout $n \geq 0$, où $\lambda > 0$. Il n'est pas aisé de calculer $p_0 + \dots + p_n$, en fonction de n , c'est la raison pour laquelle on utilise un schéma de simulation reposant sur une suite $(Y_n)_{n \geq 1}$ de v.a., indépendantes, de loi exponentielle. On choisit :

$$X' = \sum_{i \geq 1} i \mathbb{1}_{\{Y_1 + \dots + Y_i \leq \lambda < Y_1 + \dots + Y_{i+1}\}}. \quad (4.13)$$

Montrons que X' suit la loi $\mathcal{P}(\lambda)$. En effet,

$$\mathbb{P}(X' = 0) = \mathbb{P}(\lambda < Y_1) = \int_{\lambda}^{\infty} e^{-t} dt = e^{-\lambda}.$$

Soit $i \geq 1$, on a

$$\mathbb{P}(X' = i) = \mathbb{P}(Y_1 + \dots + Y_i \leq \lambda < Y_1 + \dots + Y_i + Y_{i+1}).$$

Mais $Y_1 + \dots + Y_i$ suit la loi $\gamma(i)$, Y_{i+1} suit $\gamma(1)$ et ces deux v.a. sont indépendantes, d'où

$$\begin{aligned} \mathbb{P}(X' = i) &= \frac{1}{(i-1)!} \iint_{\mathbb{R}_+} \mathbb{1}_{\{u \leq \lambda < u+v\}} u^{i-1} e^{-(u+v)} du dv \\ &= \frac{1}{(i-1)!} \int_0^{\lambda} \left(\int_{\lambda-u}^{\infty} e^{-v} dv \right) u^{i-1} e^{-u} du = \frac{e^{-\lambda}}{(i-1)!} \int_0^{\lambda} u^{i-1} du \\ &= \frac{\lambda^i}{i!} e^{-\lambda}. \end{aligned}$$

Si on simule Y_i avec $-\ln(U_i)$, alors $\{Y_1 + \dots + Y_i \leq \lambda\} = \{U_1 U_2 \dots U_i \geq e^{-\lambda}\}$, d'où

$$X' = \sum_{i \geq 1} \mathbf{1}_{\{U_1 U_2 \dots U_i \geq e^{-\lambda} > U_1 U_2 \dots U_{i+1}\}}. \quad (4.14)$$

f) Loi normale.

randn

Le résultat clé est,

Proposition 4.1 *Soient U et V deux v.a. indépendantes, U de loi uniforme sur $[0, 1]$ et V une v.a. de loi exponentielle. On pose*

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

Alors X et Y sont indépendantes et ont pour loi commune la loi gaussienne réduite et centrée.

Démonstration : Posons $\Delta = \mathbb{E}[f(X, Y)]$, où f est une fonction borélienne et bornée. On a,

$$\begin{aligned} \Delta &= \mathbb{E}[f(\sqrt{2V} \cos(2\pi U), \sqrt{2V} \sin(2\pi U))] \\ &= \iint f(\sqrt{2v} \cos(2\pi u), \sqrt{2v} \sin(2\pi u)) e^{-v} \mathbf{1}_{\{v>0, 0<u<1\}} du dv \end{aligned}$$

On fait le changement de variable : $\rho = \sqrt{2v}$ et $\theta = 2\pi u$, il vient

$$\Delta = \frac{1}{2\pi} \iint f(\rho \cos \theta, \rho \sin \theta) e^{-\rho^2/2} \mathbf{1}_{\{\rho>0, 0<\theta<2\pi\}} \rho d\rho d\theta.$$

On pose cette fois $x = \rho \cos \theta$, $y = \rho \sin \theta$, on obtient,

$$\Delta = \frac{1}{2\pi} \iint f(x, y) e^{-(x^2+y^2)/2} dx dy.$$

□

Nous avons vu qu'il est possible de simuler une v.a. de loi exponentielle, par conséquent, on sait simuler un couple de v.a. indépendantes, chacune de loi $\mathcal{N}(0, 1)$. Il est évident qu'en itérant ce procédé, on sait simuler un vecteur gaussien de loi $\mathcal{N}_d(0, Id)$. Si K est une matrice carrée symétrique et positive, soit A une racine carrée de K : $AA^* = K$. Alors si X a pour loi $\mathcal{N}_d(0, Id)$, le vecteur $m + AX$ suit $\mathcal{N}_d(m, K)$.

Procédés généraux de simulation

Soit X une v.a.r. On désigne par F sa fonction de répartition : $F(x) = \mathbb{P}(X \leq x)$. F est une fonction croissante, et continue à droite. On note F^{-1} son inverse continue à droite :

$$F^{-1}(x) = \inf\{t \in \mathbb{R}; F(t) > x\}; \quad 0 < x < 1. \quad (4.15)$$

Si X admet une densité f strictement positive sur (a, b) , alors F est continue et strictement croissante sur $]a, b[$, F^{-1} coïncide avec la fonction réciproque de la restriction de F à $]a, b[$. Lorsque X est une v.a. discrète prenant un nombre fini de valeurs $x_0 < x_1 < \dots < x_n$, alors

$$F^{-1}(x) = x_0 \mathbf{1}_{\{0 < x < p_0\}} + \sum_{i=1}^n x_i \mathbf{1}_{\{p_0 + \dots + p_{i-1} \leq x < p_0 + \dots + p_i\}}. \quad (4.16)$$

On peut montrer que pour toute fonction de répartition F , et pour tout x réel, on a :

$$\{F^{-1}(y) < x\} = \{y < F(x)\}, \quad y \notin D \quad (4.17)$$

où D désigne un ensemble au plus dénombrable inclus dans $]0, 1[$. Pour les deux exemples considérés,

- $D = \emptyset$, si X admet une densité,
- $D = \{p_0, p_0 + p_1, \dots, p_0 + p_1 + \dots + p_{n-1}\}$, si X est discrète.

Dans le second cas une représentation graphique de F permet de vérifier aisément que (4.17) a lieu.

Proposition 4.2 (*méthode de la fonction réciproque*) Soit X une v.a.r. de fonction de répartition F , U une v.a. de loi uniforme sur $[0, 1]$. Alors $X' = F^{-1}(U)$ a même loi que X .

Démonstration : Calculons la fonction de répartition G de X' :

$$G(x) = \mathbb{P}(X' \leq x) = \mathbb{P}(F^{-1}(U) \leq x), \quad x \in \mathbb{R}.$$

Mais $\mathbb{P}(U \in D) = 0$, donc d'après (4.17),

$$G(x) = \mathbb{P}(U \notin D, F^{-1}(U) \leq x) = \mathbb{P}(U \notin D; U \leq F(x)) = \mathbb{P}(U \leq F(x)) = F(x).$$

On a montré que G coïncide avec F , X et X' ont même loi. □

Remarques.

1) Si X suit une loi exponentielle de paramètre λ , alors $F(x) = 1 - e^{-x/\lambda}$ pour $x \geq 0$. F^{-1} est la fonction réciproque de F , donc $F^{-1}(x) = -\lambda \ln(1 - x)$. Par conséquent $X' = -\lambda \ln(U')$ suit la loi exponentielle de paramètre λ , avec $U' = 1 - U$; U' suit la loi uniforme sur $[0, 1]$. On retrouve le schéma de simulation c) de la section 2.1.

2) Si X suit une loi discrète, $X(\Omega) = \{x_0, x_1, \dots, x_n\}$, avec $x_0 < x_1 < \dots < x_n$. On note $p_i = P(X = x_i)$, $0 \leq i \leq n$. En utilisant (4.16), on a :

$$X' = x_0 \mathbf{1}_{\{0 < U < p_0\}} + \sum_{i=1}^n x_i \mathbf{1}_{\{p_0 + \dots + p_{i-1} \leq U < p_0 + \dots + p_i\}}.$$

On retrouve le schéma défini par la formule (4.6).

4) En principe cette approche permet de simuler n'importe quelle v.a. à valeurs réelles. Il faut toutefois remarquer qu'il faut connaître la fonction de répartition; ce qui est le cas de la loi exponentielle, et des lois discrètes. Mais pour les lois gaussiennes ou gamma cet algorithme ne

s'applique pas.

Nous allons à présent définir **la méthode de rejet**. Soit Y une v.a. admettant une densité g , on suppose que Y est simulable. On considère une v.a. X de densité f , et on suppose qu'il existe une constante c telle que

$$\frac{f(x)}{g(x)} \leq c, \text{ pour tout } x \text{ réel.}$$

L'algorithme de simulation d'une v.a. de densité f est le suivant :

Étape 1 : simuler Y et U une v.a. de loi uniforme, U et Y indépendantes.

Étape 2 : si $U \leq f(Y)/cg(Y)$ poser $X' = Y$, sinon revenir à l'étape 1.

On peut décrire ce schéma de la manière suivante :

$$X' = Y_1 \mathbf{1}_{\{U_1 \leq h(Y_1)\}} + \sum_{i \geq 1} Y_{i+1} \mathbf{1}_{\{U_1 > h(Y_1), \dots, U_i > h(Y_i), U_{i+1} \leq h(Y_{i+1})\}}. \quad (4.18)$$

où l'on a posé $h(x) = f(x)/cg(x)$. $(Y_i)_{i \geq 1}$ et $(U_i)_{i \geq 1}$ désignent deux suites de v.a. indépendantes, Y_i a même loi que Y_1 , U_i suit la loi uniforme sur $[0, 1]$; on suppose de plus que ces deux familles de v.a. sont indépendantes.

Proposition 4.3 *La v.a. définie par la méthode de rejet ou par la formule (4.18) a pour densité f .*

Démonstration : On note N le nombre d'itérations et $p = \mathbb{P}(U_1 \leq h(Y_1))$. Soient φ une fonction borélienne positive et $I = \mathbb{E}[\varphi(X')]$. On a,

$$I = \sum_{i=1}^{\infty} \mathbb{E}(\varphi(X') \mathbf{1}_{\{N=i\}})$$

$$I = \mathbb{E}[\varphi(Y_1) \mathbf{1}_{\{U_1 \leq h(Y_1)\}}] + \sum_{i \geq 1} \mathbb{E}[\varphi(Y_{i+1}) \mathbf{1}_{\{U_1 > h(Y_1), \dots, U_i > h(Y_i), U_{i+1} \leq h(Y_{i+1})\}}].$$

Les v.a. $U_1, \dots, U_{i+1}, Y_1, \dots, Y_{i+1}$ étant indépendantes, on a

$$I = \mathbb{E}[\varphi(Y_1) \mathbf{1}_{\{U_1 \leq h(Y_1)\}}] + \sum_{i \geq 1} (\mathbb{P}(U_1 > h(Y_1)))^i \mathbb{E}[\varphi(Y_{i+1}) \mathbf{1}_{\{U_{i+1} \leq h(Y_{i+1})\}}].$$

De plus (Y_{i+1}, U_{i+1}) a même loi que (Y_1, U_1) , donc

$$I = \mathbb{E}[\varphi(Y_1) \mathbf{1}_{\{U_1 \leq h(Y_1)\}}] + \left(\sum_{i \geq 1} (1-p)^i \right) \mathbb{E}[\varphi(Y_1) \mathbf{1}_{\{U_1 \leq h(Y_1)\}}],$$

$$I = \frac{1}{p} \mathbb{E}[\varphi(Y_1) \mathbf{1}_{\{U_1 \leq h(Y_1)\}}].$$

Mais U_1 est indépendante de Y_1 et suit une loi uniforme sur $[0, 1]$, Y_1 a pour densité g d'où

$$\begin{aligned} I &= \frac{1}{p} \iint \varphi(y)g(y) \mathbf{1}_{\{0 \leq u \leq h(y)\}} dy du = \frac{1}{p} \int \varphi(y)g(y) \left(\int_0^{h(y)} du \right) dy \\ &= \frac{1}{p} \int \varphi(y)h(y)g(y) dy. \end{aligned}$$

Or $gh = f/c$, on en déduit,

$$I = \frac{1}{pc} \int \varphi(y)f(y) dy.$$

En particulier si $\varphi = 1$, alors $I = 1$ et $pc = \int f(y) dy = 1$. On a montré que X' a pour densité f . \square

Remarques.

1) La v.a. $1/g(Y)$ est bien définie, en effet

$$\mathbb{P}(g(Y) = 0) = \int \mathbf{1}_{\{g(y)=0\}} g(y) dy = 0.$$

Donc $g(Y) > 0$ presque sûrement.

2) Le nombre d'itérations nécessaires N est une v.a. de loi géométrique, de paramètre $p = \mathbb{P}(U_1 \leq h(Y_1)) = \frac{1}{c}$. En particulier $\mathbb{E}(N) = c$.

3) On peut modifier l'algorithme précédent en évitant d'avoir recours aux v.a. U_i . On commence avec U_1 et Y_1 et on étudie le cas où il y a rejet, c'est-à-dire lorsque $U_1 > h(Y_1)$. Posons $U'_2 = \frac{U_1 - h(Y_1)}{1 - h(Y_1)}$. Alors conditionnellement à $\{U_1 > h(Y_1)\}$, U'_2 suit une loi uniforme sur $[0, 1]$. Il s'agit de montrer :

$$\mathbb{P}(U'_2 < x | U_1 > h(Y_1)) = x, \quad \text{avec } x \in [0, 1]. \quad (4.19)$$

Cette propriété est équivalente à

$$\mathbb{P}(U'_2 < x, U_1 > h(Y_1)) = x \mathbb{P}(U_1 > h(Y_1)). \quad (4.20)$$

Mais $\mathbb{P}(U'_2 < x, U_1 > h(Y_1)) = \mathbb{P}(h(Y_1) < U_1 < h(Y_1) + x(1 - h(Y_1)))$.

Sachant que U_1 suit une loi uniforme et que U_1 et Y_1 sont deux v.a. indépendantes,

$$\mathbb{P}(h(Y_1) < U_1 < h(Y_1) + x(1 - h(Y_1)) | Y_1) = h(Y_1) + x(1 - h(Y_1)) - h(Y_1) = x(1 - h(Y_1))$$

On en déduit,

$$\mathbb{P}(U'_2 < x, U_1 > h(Y_1)) = x(1 - \mathbb{E}(h(Y_1))).$$

D'une manière analogue : $\mathbb{P}(U_1 > h(Y_1)) = \mathbb{E}[1 - h(Y_1)]$, (4.20) s'en déduit donc immédiatement.

On continue ensuite l'algorithme avec Y_2 et U'_2 . On a ainsi remplacé l'utilisation de U_2 par le calcul supplémentaire de U'_2 . Les appels au générateur de nombres aléatoires étant coûteux en

temps, on cherche à les utiliser avec parcimonie. Dans le même esprit on cherchera à prendre c le plus petit possible.

Nous allons donner une application de la méthode de rejet à la simulation des lois gamma. Sauf pour la loi exponentielle, on ne connaît pas explicitement la fonction de répartition des lois gamma. La méthode de la fonction réciproque ne s'applique pas. Rappelons que si $X_{a,b}$ suit la loi $\gamma(a, b)$, $X_{a,b}$ admet pour densité,

$$f_{a,b}(x) = \frac{x^{a-1}}{\Gamma(a)b^a} e^{-x/b} \mathbf{1}_{\{x>0\}}$$

où $a > 0$ et $b > 0$. On montre facilement que $\frac{1}{b}X_{a,b}$ suit $\gamma(a, 1)$. On se ramène donc à la simulation de $X_{a,1}$. Par ailleurs si ξ et ξ' sont deux v.a. indépendantes, ξ de loi $\gamma(a, 1)$ et ξ' de loi $\gamma(a', 1)$ alors $\xi + \xi'$ a pour loi $\gamma(a + a', 1)$. Soient $a > 0$, n la partie entière de a , ξ_1, \dots, ξ_n , n v.a. indépendantes et de loi exponentielle (donc de loi $\gamma(1, 1)$).

(i) si a est entier, i.e. $n = a$,

$$\xi_1 + \dots + \xi_n \text{ a même loi que } X_{a,1}. \quad (4.21)$$

On a vu que l'on sait simuler des v.a. de loi exponentielle, on est donc capable de simuler une v.a. de loi $\gamma(a, 1)$ avec a entier.

(ii) si a n'est pas entier, $a > n$, alors

$$X_{a,1} \text{ a même loi que } \xi_1 + \dots + \xi_n + \hat{X}_{a-n,1}, \quad (4.22)$$

où $\xi_1, \xi_2, \dots, \xi_n$ et $\hat{X}_{a-n,1}$ sont indépendantes et $\hat{X}_{a-n,1}$ suit la loi $\gamma(a - n, 1)$. On est donc conduit à simuler des v.a. $X_{a,1}$, avec $0 < a < 1$.

Proposition 4.4 Soit $0 < a < 1$. La méthode de rejet appliquée avec $f_{a,1}$ et g_a définie par

$$g_a(x) = \frac{ae}{a+e} (x^{a-1} \mathbf{1}_{\{0 \leq x < 1\}} + e^{-x} \mathbf{1}_{\{x \geq 1\}}),$$

permet de simuler une v.a. $X_{a,1}$ de loi $\gamma(a, 1)$.

Démonstration : 1) Puisque g_a est positive et

$$\int_0^1 x^{a-1} dx = \frac{1}{a}; \quad \int_1^\infty e^{-x} dx = \frac{1}{e} \quad \text{et} \quad \frac{1}{a} + \frac{1}{e} = \frac{a+e}{ae},$$

alors $\int_{\mathbb{R}} g_a(x) dx = \int_0^\infty g_a(x) dx = 1$; g_a est la densité d'une v.a. Y_a . Soit G_a la fonction de répartition de Y_a . Il est facile de vérifier :

$$G_a(x) = \begin{cases} 0; & x < 0, \\ \frac{e}{a+e} x^a; & 0 \leq x \leq 1, \\ \frac{e}{a+e} + \frac{ae}{a+e}(e^{-1} - e^{-x}); & x > 1. \end{cases}$$

On en déduit :

$$y = G_a(x); 0 \leq x \leq 1 \iff x = \left(\frac{a+e}{e} y \right)^{1/a}; 0 \leq y \leq \frac{e}{a+e},$$

et

$$y = G_a(x); x > 1 \iff x = -\ln \left(\frac{a+e}{ae} (1-y) \right); y > \frac{e}{a+e}.$$

Par conséquent,

$$G_a^{-1}(x) = \left(\frac{a+e}{e} x \right)^{1/a} \mathbf{1}_{\{0 < x < \frac{e}{a+e}\}} - \ln \left(\frac{a+e}{ae} (1-x) \right) \mathbf{1}_{\{x > \frac{e}{a+e}\}}; 0 < x < 1. \quad (4.23)$$

Puisque l'on a su calculer explicitement G , la v.a. Y_a est simulable par la méthode de la fonction réciproque.

2) Pour appliquer la méthode de rejet, il reste à montrer que $f_{a,1}(x) \leq c g_a(x)$, pour tout $x > 0$. Mais $e^{-y} \leq 1$ si $y \geq 0$ et $x^{a-1} \leq 1$ lorsque $x \geq 1$, donc $c = \frac{a+e}{ae\Gamma(a)}$ convient. Soit $h_a = f_{a,1}/c g_a$. On a :

$$h_a(y) = e^{-y} \mathbf{1}_{\{0 \leq y < 1\}} + y^{a-1} \mathbf{1}_{\{y \geq 1\}}.$$

On peut appliquer la proposition 4.3, dans ce cas particulier l'algorithme de rejet est le suivant :

Étape 1 : simulation d'une v.a. Y de densité g_a en posant $Y = G_a^{-1}(U)$, U v.a. uniforme sur $[0, 1]$, G_a^{-1} définie par (4.23).

Étape 2 : simulation d'une v.a. V de loi uniforme sur $[0, 1]$ et indépendante de U .

Si $V \leq e^{-Y} \mathbf{1}_{\{0 \leq Y < 1\}} + Y^{a-1} \mathbf{1}_{\{Y \geq 1\}}$, on pose $X = Y$, sinon on recommence l'étape 1. \square