

Master Recherche de Mathématiques
Probabilités

Jean-Yves DAUXOIS

CTU

Université de Franche-Comté
Année scolaire 2008-2009

Ce polycopié est encore récent. Veuillez m'excuser pour toutes les coquilles qu'il doit contenir. Merci d'avance à ceux qui voudront bien me les signaler.

Table des matières

Chapitre 1. Rappels ou compléments de Probabilités	7
1. Variables gaussiennes	8
2. Théorèmes des classes monotones et théorème de Carathéodory	14
Chapitre 2. Généralités sur les Processus Stochastiques	17
1. Notion de processus stochastique	18
2. Exemples classiques et fondamentaux de processus stochastiques	23
3. Bibliographie	26
Chapitre 3. Processus de Poisson	27
1. Définition et propriétés élémentaires	28
2. Caractérisation d'un processus de Poisson	32
3. Résultats asymptotiques	34
4. Bibliographie	37
Chapitre 4. Chaînes de Markov	39
1. Définition et matrice de transition	40
2. Exemples classiques de Chaînes de Markov	41
3. Équations de Chapman-Kolmogorov	48
4. Quelques formules de conditionnement	50
5. Classification des états	51
6. Périodicité	54
7. Temps d'atteinte, états récurrents et transients	56
8. Propriété de Markov forte	66
9. Récurrence positive et récurrence nulle	67
10. Loi stationnaire et théorèmes limites	70
11. Bibliographie	81
Chapitre 5. Espérances conditionnelles	83
1. Introduction	84
2. Espérance conditionnelle	85
Chapitre 6. Martingales	97
1. Martingales	98
2. Martingales bornées dans L^1	111
3. Inégalité de Doob et conséquences	115
4. Martingales bornées dans L^p , pour $p > 1$	117

CHAPITRE 1

Rappels ou compléments de Probabilités

1. Variables gaussiennes

1.1. Exemple fondamental.

EXERCICE 1.1. *Considérons n variables aléatoires X_1, \dots, X_n indépendantes et de loi respectivement $N(m_1, \sigma_1^2), \dots, N(m_n, \sigma_n^2)$.*

Pour $i = 1, \dots, n$, la variable aléatoire X_i est donc de densité

$$f_{X_i}(x) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp \left\{ -\frac{1}{2} \left(\frac{x - m_i}{\sigma_i} \right)^2 \right\},$$

par rapport à la mesure de Lebesgue λ sur \mathbb{R} .

1) *Montrer que l'on peut écrire la densité du vecteur de \mathbb{R}^n*

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

de la forme :

$$f_X(x_1, \dots, x_n) = \frac{1}{(\sqrt{2\pi})^n} \times \frac{1}{\sqrt{\det(\Gamma_X)}} \exp \left\{ -\frac{1}{2} (x - m)' \Gamma_X^{-1} (x - m) \right\},$$

où u' est la transposée du vecteur u et où m et Γ_X sont respectivement le vecteur des espérances et la matrice de covariance du vecteur aléatoire X .

2) *On rappelle que la fonction caractéristique (f.c.) d'une v.a.r. X_j de loi $N(m_j, \sigma_j^2)$ est :*

$$\varphi_{X_j}(\lambda_j) = \mathbb{E}(e^{i\lambda_j X_j}) = \exp \left\{ i\lambda_j m_j - \frac{1}{2} \lambda_j^2 \sigma_j^2 \right\},$$

pour tout λ_j de \mathbb{R} . En déduire que celle du vecteur $X = (X_1, \dots, X_n)$ est :

$$\varphi_X(\lambda) = \mathbb{E}(e^{i\lambda' X}) = \exp \left\{ i\lambda' m - \frac{1}{2} \lambda' \Gamma_X \lambda \right\},$$

pour tout λ vecteur de \mathbb{R}^n .

3) *En utilisant les f.c., montrer que toute combinaison linéaire des composantes de X est de loi normale sur \mathbb{R} dont on déterminera l'espérance et la variance.*

Solution de l'exercice.

1) En raison de l'indépendance des variables aléatoires X_i , la densité conjointe du vecteur X_1, \dots, X_n est :

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \frac{1}{(\sqrt{2\pi})^n} \frac{1}{\prod_{i=1}^n \sigma_i} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - m_i}{\sigma_i} \right)^2 \right\}.$$

Notons que la matrice Γ_X est diagonale en raison de l'indépendance des v.a.r. $(X_i)_{i=1,\dots,n}$. Comme toutes les variances σ_i sont strictement positives, on obtient aisément la matrice inverse

$$\Gamma_X^{-1} = \begin{pmatrix} 1/\sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & 1/\sigma_n^2 \end{pmatrix}.$$

On peut alors réécrire la densité conjointe du vecteur $X = (X_1, \dots, X_n)$ sous la forme

$$f_X(x_1, \dots, x_n) = \frac{1}{(\sqrt{2\pi})^n} \times \frac{1}{\sqrt{\det(\Gamma_X)}} \exp \left\{ -\frac{1}{2}(x - m)' \Gamma_X^{-1} (x - m) \right\},$$

puisque

$$\begin{aligned} & (x - m)' \Gamma_X^{-1} (x - m) \\ = & (x_1 - m_1, \dots, x_n - m_n) \begin{pmatrix} 1/\sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & 1/\sigma_n^2 \end{pmatrix} \begin{pmatrix} x_1 - m_1 \\ \vdots \\ x_n - m_n \end{pmatrix} \\ = & \sum_{i=1}^n \frac{(x_i - m_i)^2}{\sigma_i^2}. \end{aligned}$$

2) En raison de l'indépendance entre les composantes du vecteur X , on a, pour tout $\lambda = (\lambda_1, \dots, \lambda_n)$ de \mathbb{R}^n :

$$\varphi_X(\lambda) = \varphi_{X_1, \dots, X_n}(\lambda_1, \dots, \lambda_n) = \prod_{j=1}^n \varphi_{X_j}(\lambda_j),$$

d'où on tire :

$$\begin{aligned} \varphi_{X_1, \dots, X_n}(\lambda) &= \exp \left\{ i \sum_{j=1}^n \lambda_j m_j - \frac{1}{2} \sum_{j=1}^n \lambda_j^2 \sigma_j^2 \right\} \\ &= \exp \left\{ i \lambda' m - \frac{1}{2} \lambda' \Gamma_X \lambda \right\}. \end{aligned}$$

3) Une combinaison linéaire des composantes du vecteur X s'écrit de manière générale sous la forme :

$$\langle \lambda, X \rangle = \lambda' X$$

où $\lambda = (\lambda_1, \dots, \lambda_n)$ est un vecteur de \mathbb{R}^n . Il vient alors, pour tout u dans \mathbb{R} :

$$\begin{aligned} \varphi_{\langle \lambda, X \rangle}(u) &= \mathbb{E} (e^{iu \langle \lambda, X \rangle}) = \mathbb{E} (e^{i \langle u \lambda, X \rangle}) \\ &= \varphi_X(u \lambda) = \varphi_X(u \lambda_1, \dots, u \lambda_n) \\ &= \exp \left\{ i u \lambda' m - \frac{1}{2} u^2 \lambda' \Gamma_X \lambda \right\}. \end{aligned}$$

La fonction caractéristique de la v.a.r. $\langle \lambda, X \rangle$ est donc de la forme :

$$\varphi_{\langle \lambda, X \rangle}(u) = \exp \left\{ i u a - \frac{1}{2} u^2 b \right\},$$

avec $a = \lambda' m$ et $b = \lambda' \Gamma_X \lambda$. Par caractérisation, la v.a.r. $\langle \lambda, X \rangle$ est donc de loi $N(\lambda' m, \lambda' \Gamma_X \lambda)$. \diamond

1.2. Définition.

DÉFINITION 1.1. Un vecteur aléatoire $X = (X_1, \dots, X_n)$ de \mathbb{R}^n est dit vecteur gaussien si, pour tout $\lambda = (\lambda_1, \dots, \lambda_n)$ de \mathbb{R}^n , la v.a.r.

$$\lambda'X = \sum_{i=1}^n \lambda_i X_i$$

est une v.a.r. de loi normale. Autrement dit, si toute combinaison linéaire des composantes de (X_1, \dots, X_n) est de loi normale.

Si son vecteur des espérances est m et sa matrice de covariance est Γ_X , on note $X \sim N_n(m, \Gamma_X)$.

Remarquons que l'on peut en particulier en déduire que toutes les composantes du vecteur X sont des v.a.r. de loi normale (Exercice trivial !). En revanche, la réciproque est fautive. Un vecteur dont toutes les composantes sont de loi normale, n'est pas nécessairement un vecteur gaussien. On en verra un contre-exemple sous forme d'exercice à la fin de cette section.

La définition précédente implique également que tout sous vecteur d'un vecteur gaussien est encore un vecteur gaussien.

PROPOSITION 1.2. Si X est un vecteur gaussien de vecteur des espérances $m = (m_1, \dots, m_n)$ et de matrice de covariance Γ_X , alors, pour tout λ dans \mathbb{R}^n , la v.a.r. $\lambda'X = \langle \lambda, X \rangle$ est de loi $N(\lambda'm, \lambda'\Gamma_X\lambda)$.

EXERCICE 1.2. Démontrer cette proposition.

Preuve/solution de l'exercice. On utilise d'abord le fait que, par définition d'un vecteur gaussien, la v.a.r. $\lambda'X$ est de loi normale. Il ne reste plus qu'à calculer son espérance et sa variance. On sait que l'on peut écrire :

$$\begin{aligned} \mathbb{E}(\lambda'X) &= \lambda'\mathbb{E}X = \lambda'm \\ \text{et} \quad \Gamma_{\lambda'X} &= \lambda'\Gamma_X\lambda. \end{aligned} \quad \square$$

On peut aussi caractériser un vecteur gaussien par sa fonction caractéristique, grâce à la proposition suivante.

PROPOSITION 1.3. Pour qu'un vecteur X de \mathbb{R}^n soit un vecteur gaussien, il faut et il suffit qu'il existe un vecteur m de \mathbb{R}^n et une matrice Γ symétrique et positive de dimension $n \times n$ tels que, pour tout vecteur λ de \mathbb{R}^n , on ait :

$$\varphi_X(\lambda_1, \dots, \lambda_n) = \exp \left\{ i\lambda'm - \frac{1}{2}\lambda'\Gamma\lambda \right\}.$$

Dans ce cas, on a : $\mathbb{E}X = m$ et $\Gamma_X = \Gamma$.

EXERCICE 1.3. Démontrer cette proposition (Ind. Dans un sens on utilisera la proposition précédente et dans l'autre la définition précédente d'un vecteur gaussien).

Preuve/solution de l'exercice. Supposons que X soit un vecteur gaussien. Toute v.a.r. de la forme $\lambda'X$, pour λ dans \mathbb{R}^n , est donc de loi $N(\lambda'm, \lambda'\Gamma_X\lambda)$. Ainsi sa fonction caractéristique est :

$$\varphi_{\lambda'X}(u) = \mathbb{E}(e^{iu\lambda'X}) = \exp \left\{ iu\lambda'm - \frac{1}{2}u^2\lambda'\Gamma_X\lambda \right\}$$

En posant $u = 1$ dans cette équation, on obtient :

$$\mathbb{E}(e^{i\lambda'X}) = \exp \left\{ i\lambda'm - \frac{1}{2}\lambda'\Gamma_X\lambda \right\},$$

Ce qui est bien l'expression annoncée pour la fonction caractéristique.

Réciproquement, soit X un vecteur aléatoire dans \mathbb{R}^n ayant pour fonction caractéristique

$$\varphi_X(\lambda) = \exp \left\{ i\lambda'm - \frac{1}{2}\lambda'\Gamma_X\lambda \right\} = \mathbb{E} \left(e^{i\langle \lambda, X \rangle} \right),$$

pour tout λ dans \mathbb{R}^n . Notons maintenant $Y = \langle \lambda, X \rangle$ la variable aléatoire réelle dont la fonction caractéristique est, pour tout u dans \mathbb{R} :

$$\begin{aligned} \varphi_Y(u) &= \mathbb{E} \left(e^{iuY} \right) = \mathbb{E} \left(e^{iu\lambda'X} \right) = \mathbb{E} \left(e^{i\langle \lambda u, X \rangle} \right) \\ &= \exp \left\{ iu\lambda'm - \frac{1}{2}u^2\lambda'\Gamma_X\lambda \right\} \\ &= \exp \left\{ iua - \frac{1}{2}u^2b \right\} \end{aligned}$$

où $a = \lambda'm$ et $b = \lambda'\Gamma_X\lambda$. Par caractérisation, la v.a.r. Y est donc de loi $N(\lambda'm, \lambda'\Gamma_X\lambda)$. On a donc démontré que toute combinaison linéaire des composantes du vecteur X est de loi normale, et par définition il s'agit bien d'un vecteur gaussien. \square

Notons que, dans tout ce qui précède, la matrice Γ_X n'est pas supposée inversible. En revanche, la définition d'un vecteur gaussien par sa densité, par rapport à la mesure de Lebesgue dans \mathbb{R}^n , n'est possible que si cette matrice est inversible, comme l'affirme la proposition suivante.

PROPOSITION 1.4. *Soit X un vecteur gaussien dans \mathbb{R}^n d'espérance m et de matrice des covariances Γ_X . Lorsque Γ_X est inversible, le vecteur aléatoire X est dit vecteur aléatoire gaussien non dégénéré et sa loi est absolument continue par rapport à la mesure de Lebesgue dans \mathbb{R}^n et admet pour densité*

$$f_X(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\det(\Gamma_X)}} \exp \left\{ -\frac{1}{2}(x - m)'\Gamma_X^{-1}(x - m) \right\}.$$

Un vecteur gaussien de matrice de covariance Γ_X telle que $\det(\Gamma_X) = 0$ (i.e. Γ_X non inversible) est dit dégénéré et n'admet pas de densité par rapport à la mesure de Lebesgue dans \mathbb{R}^n .

1.3. Propriétés des vecteurs aléatoires gaussiens.

1.3.1. Transformation linéaire d'un vecteur gaussien.

PROPOSITION 1.5. *La transformée d'un vecteur gaussien de \mathbb{R}^n par une application linéaire de \mathbb{R}^n vers \mathbb{R}^p est encore un vecteur gaussien.*

EXERCICE 1.4. *Démontrer cette proposition (Ind. Calculer la fonction caractéristique du vecteur $Y = AX$ où A est la matrice de l'application linéaire considérée et X est le vecteur gaussien initial).*

Preuve/solution de l'exercice. Soit X un vecteur gaussien de \mathbb{R}^n , de vecteur des espérances m et de matrice de covariance Γ_X . Soit A la matrice associée à une transformation linéaire quelconque de \mathbb{R}^n vers \mathbb{R}^p . La matrice A est donc de dimension $p \times n$. Calculons la fonction caractéristique du vecteur aléatoire $Y = AX$.

D'après ce que l'on a vu au chapitre précédent, pour tout λ de \mathbb{R}^p , on a :

$$\begin{aligned}\varphi_Y(\lambda) &= \varphi_{AX}(\lambda) = \mathbb{E}\left(e^{i\langle \lambda, AX \rangle}\right) = \mathbb{E}\left(e^{i\langle A'\lambda, X \rangle}\right) \\ &= \varphi_X(A'\lambda) = \exp\left\{i\lambda'Am - \frac{1}{2}\lambda' A\Gamma_X A'\lambda\right\}.\end{aligned}$$

Par caractérisation, le vecteur Y est donc un vecteur gaussien dans \mathbb{R}^p de vecteur des espérances Am et de matrice de covariance $A\Gamma_X A'$, i.e.

$$Y \sim N_p(Am, A\Gamma_X A'). \quad \square$$

1.3.2. Vecteur gaussien et indépendance.

On sait que, d'une manière générale, l'indépendance entraîne la non corrélation (i.e. covariance nulle) mais que la réciproque est fautive. Dans le cas d'un vecteur gaussien il y a équivalence, comme le montre la proposition suivante.

PROPOSITION 1.6. *Soit X un vecteur gaussien dans \mathbb{R}^n . Pour que ses composantes X_1, \dots, X_n soient indépendantes, il faut et il suffit que la matrice de covariance soit diagonale.*

EXERCICE 1.5. *Démontrer cette proposition (Ind. Un seul sens nécessite une preuve. Pour ce faire on pourra utiliser les f.c. et la caractérisation qu'elles procurent pour l'indépendance entre v.a.r.).*

Preuve/solution de l'exercice. Il suffit, bien sûr, de montrer la réciproque. Supposons donc que Γ_X soit diagonale, i.e.

$$\Gamma_X = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{pmatrix}.$$

Comme X est un vecteur gaussien de loi $N_n(m, \Gamma_X)$, chacune de ses composantes X_j , pour $j = 1, \dots, n$, est de loi normale $N(m_j, \sigma_j^2)$ et de fonction caractéristique :

$$\varphi_{X_j}(\lambda_j) = \exp \left\{ i\lambda_j m_j - \frac{1}{2} \sigma_j^2 \lambda_j^2 \right\},$$

pour tout λ_j dans \mathbb{R} .

Par ailleurs, la fonction caractéristique du vecteur X est, pour tout λ dans \mathbb{R}^n :

$$\begin{aligned} \varphi_X(\lambda) &= \exp \left\{ i\lambda' m - \frac{1}{2} \lambda' \Gamma_X \lambda \right\} \\ &= \exp \left\{ i \sum_{j=1}^n \lambda_j m_j - \frac{1}{2} \sum_{j=1}^n \lambda_j^2 \sigma_j^2 \right\} \\ &= \exp \left\{ \sum_{j=1}^n \left\{ i\lambda_j m_j - \frac{1}{2} \lambda_j^2 \sigma_j^2 \right\} \right\} \\ &= \prod_{j=1}^n \exp \left\{ i\lambda_j m_j - \frac{1}{2} \lambda_j^2 \sigma_j^2 \right\} = \prod_{j=1}^n \varphi_{X_j}(\lambda_j). \end{aligned}$$

La caractérisation bien connue de l'indépendance au moyen des f.c. permet de conclure. \square

COROLLAIRE 1.7. *Si le couple (X, Y) est un vecteur gaussien, on a*

$$X \perp\!\!\!\perp Y \Leftrightarrow \text{Cov}(X, Y) = 0.$$

Preuve. Immédiate. \square

Nous attirons l'attention du lecteur sur le fait que deux variables aléatoires réelles gaussiennes et non corrélées ne sont pas nécessairement indépendantes. Pour s'assurer qu'elles le soient il faut pour cela qu'elles constituent un couple gaussien.

L'exercice suivant établit un contre-exemple.

EXERCICE 1.6. *Considérons une v.a.r. X de loi $N(0, 1)$ et ε une v.a.r. discrète de loi définie par :*

$$p(\varepsilon = 1) = \frac{1}{2} \quad \text{et} \quad p(\varepsilon = -1) = \frac{1}{2}$$

et telle que les v.a.r. ε et X soient indépendantes. On pose $Y = \varepsilon X$.

- 1) Calculer la loi de Y .*
 - 2) Montrer que X et Y ne sont pas corrélées (i.e. sont de covariance nulle).*
 - 3) Montrer qu'elles ne sont cependant pas indépendantes. (Ind. on pourra raisonner par l'absurde et considérer la v.a. $X + Y$)*
-

Solution de l'exercice.

1) On a :

$$\begin{aligned}
 F_Y(y) &= P(\varepsilon X \leq y) \\
 &= P(\{\varepsilon X \leq y\} \cap \{\varepsilon = 1\}) + P(\{\varepsilon X \leq y\} \cap \{\varepsilon = -1\}) \\
 &= P(\{X \leq y\} \cap \{\varepsilon = 1\}) + P(\{-X \leq y\} \cap \{\varepsilon = -1\}) \\
 &= P(X \leq y)P(\varepsilon = 1) + P(X \geq -y)P(\varepsilon = -1) \\
 &= \frac{1}{2}P(X \leq y) + \frac{1}{2}P(X \leq y) = F_X(y).
 \end{aligned}$$

Ainsi la v.a.r. Y est de loi $N(0, 1)$.

2) Puisque X et Y sont centrées et que ε et X sont indépendantes, on a :

$$\text{Cov}(X, Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y = \mathbb{E}(\varepsilon X^2) = \mathbb{E}\varepsilon\mathbb{E}X^2 = 0$$

3) En raisonnant par l'absurde, supposons qu'elles soient indépendantes. D'après ce que l'on a vu au début de ce chapitre, le couple (X, Y) serait gaussien et $X + Y$ serait alors de loi normale et donc absolument continue.

Or, en notant que $X + Y = (1 + \varepsilon)X$, on a :

$$P(X + Y = 0) = P(1 + \varepsilon = 0) = P(\varepsilon = -1) = \frac{1}{2},$$

ce qui contredit le fait que la v.a.r. $X + Y$ soit absolument continue. Les v.a.r. X et Y ne sont donc pas indépendantes. \diamond

2. Théorèmes des classes monotones et théorème de Carathéodory

Dans cette section nous ne donnerons pas les preuves des théorèmes car elles sont très techniques et longues. L'important est surtout de savoir utiliser ces résultats fondamentaux pour la suite de ce cours.

DÉFINITION 1.8. Une famille \mathcal{M} de parties d'un ensemble Ω est appelée **classe monotone** si

- i) $\Omega \in \mathcal{M}$,
- ii) si (A_n) est une suite croissante d'éléments de \mathcal{M} tels que $A_n \nearrow A$, alors $A \in \mathcal{M}$,
- iii) Si A et B sont des éléments de \mathcal{M} et si A est inclus dans B , alors $B \setminus A$ est dans \mathcal{M} .

Une tribu est bien sûr une classe monotone. On peut de même montrer facilement qu'une classe monotone fermée pour l'intersection finie est alors une tribu (Exercice à faire !).

THÉORÈME 1.9. (Théorème des classes monotones, version ensembliste) Si \mathcal{S} est une famille de parties de Ω fermée pour l'intersection finie, alors la classe monotone engendrée par \mathcal{S} coïncide avec la tribu engendrée par \mathcal{S} , i.e.

$$m(\mathcal{S}) = \sigma(\mathcal{S}).$$

THÉORÈME 1.10. (Théorème des classes monotones, version fonctionnelle) Soit \mathcal{H} un espace vectoriel de fonctions bornées définies sur Ω à valeurs dans \mathbb{R} contenant les constantes et stable par limites monotones bornées, i.e. si (f_n) est une suite croissante bornée de fonctions positives de \mathcal{H} , alors $f := \lim f_n$ est dans \mathcal{H} . Soit \mathcal{C} un sous-ensemble de \mathcal{H} stable par multiplication.

Alors \mathcal{H} contient toutes les fonctions bornées $\sigma(\mathcal{C})$ -mesurables.

Le théorème des classes monotones permet en particulier de montrer l'unicité dans le théorème fondamental suivant.

THÉORÈME 1.11. (**Théorème de Carathéodory**) Soit μ une mesure σ -additive sur l'algèbre de Boole \mathcal{A} de parties de Ω . Alors il existe une mesure $\tilde{\mu}$ sur la tribu $\sigma(\mathcal{A})$ telle que $\tilde{\mu}(A) = \mu(A)$, pour tout A de \mathcal{A} . De plus $\tilde{\mu}$ est unique et σ -finie si μ est σ -finie.

CHAPITRE 2

Généralités sur les Processus Stochastiques

Ce chapitre a pour but d'introduire formellement la notion de processus stochastique, d'en donner une définition relativement générale avant d'étudier plus précisément quelques cas particuliers dans les chapitres suivants. Il est certainement un peu abstrait et peu digeste (malgré mes efforts...) à la première lecture du polycopié. Il est d'ailleurs à noter qu'il se trouve à la marge du programme de cette unité. C'est pourquoi, il peut être conseillé de le sauter en première lecture. Ensuite, pour répondre à quelques unes des questions que vous pourriez vous poser sur les processus stochastiques étudiés ultérieurement (processus de Poisson, Chaînes de Markov et Martingales), il trouvera peut être son utilité, en particulier pour ceux qui désireraient en savoir un peu plus...

1. Notion de processus stochastique

1.1. Introduction. On s'intéresse à l'évolution d'un phénomène, d'une variable au cours du temps. On utilise le terme de processus quand on suppose l'intervention d'aléa ou d'un trop grand nombre de facteurs explicatifs qu'on ne peut prendre en compte et que l'on rassemble du coup dans l'aléa. Quatre grands types de processus se dégagent rapidement dont voici des exemples :

- →Évolution de la température relevée tous les matins à 7 heures.
 - Nombre mensuel d'immatriculations nouvelles relevées à Besançon.
 - Nombre de personnes "placées" dans le monde du travail par l'ANPE par mois.
- →Relevé par seconde de la dilatation d'un matériau sous l'effet de la chaleur.
 - Évaluation sensorielle d'un produit à des instants différents.
 - Relevé par heure de la concentration sanguine d'un certain type d'anticorps sur un patient.
- →Nombre d'appels reçus par un standard téléphonique depuis le début de la journée, i.e. dans l'intervalle $[8h, t]$ pour $t \in [8h, 17h]$.
 - Relevé en continu du nombre de voitures arrêtées à un feu rouge, nombre de clients à une caisse de supermarché ou à un guichet de banque.
- →Électrocardiogramme d'un patient
 - Variation d'un indice boursier.
 - Évolution de la proportion d'ozone dans l'atmosphère au cours de la journée dans le centre de Paris.

Ces processus peuvent être modélisés par la donnée d'un espace probabilisé (Ω, \mathcal{A}, P) , d'un espace probabilisable (E, \mathcal{E}) , d'un ensemble T et d'une famille de v.a. $(X_t)_{t \in T}$ de (Ω, \mathcal{A}, P) à valeurs dans (E, \mathcal{E}) . Les quatre grands types que nous avons vu précédemment sont différenciés par leur ensemble T , appelé **espace des temps**, et leur ensemble E , appelé **espace d'états**.

Si T est un espace dénombrable, \mathbb{N} par exemple, on parle de processus à **temps discret**. C'est le cas pour les deux premiers grands types introduits plus haut. Si l'espace T est de type continu, par exemple \mathbb{R} ou \mathbb{R}^+ , on parle de processus à **temps continu**. C'est le cas des deux derniers types. Notons que l'on peut aussi considérer des processus à espace des temps multidimensionnel, comme par exemple $T \subset \mathbb{R}^k$. Il en est ainsi pour la modélisation du mouvement de la vague sur l'océan. Pour un point t repéré par sa latitude et sa longitude, on relève la hauteur X_t de la vague. Il en est ainsi également en traitement de l'image où pour tout t coordonnées du pixel (dans \mathbb{N}^2) on associe un niveau de gris X_t pour des images en noir et blanc, ou un triplet de valeurs donnant les niveaux de rouge, bleu et vert. On parle alors de processus spatiaux mais nous n'aborderons pas ces derniers dans ce cours.

On peut également distinguer deux types d'espace d'états, selon qu'il est discret ou continu comme respectivement dans les types 1 et 3 et les types 2 et 4. Notons que nous n'avons considéré que des espaces d'états unidimensionnels et que bien sûr, sans difficulté majeure, on pourra également considérer des espaces d'états multimensionnels, par exemple $E \subset \mathbb{R}^k$. Il en est ainsi par exemple si on considère le déplacement au cours du temps d'une particule dans un liquide, repérée par ses coordonnées dans l'espace.

1.2. Définition-Mesurabilité. A la suite des exemples précédents, on souhaite définir une manière de représenter l'évolution d'un système au cours du temps.

Une représentation déterministe consiste à spécifier un espace T , dit espace des temps, un espace E , dit espace des états et une fonction

$$\begin{aligned} x : T &\rightarrow E \\ t &\mapsto x_t \end{aligned}$$

Ainsi on pourra représenter l'évolutions de systèmes déterministes par des équations par exemple du type : $x_t = t^2$ ou $x_t = A \cos at$, etc...

Mais, comme nous l'avons vu dans les exemples de la section précédente, il est parfois utile ou nécessaire de faire intervenir l'aléatoire. Pour cela on se donne un espace probabilisé (Ω, \mathcal{A}, P) et une famille $(X_t)_{t \in T}$ de v.a. de (Ω, \mathcal{A}, P) vers (E, \mathcal{E}) , i.e. pour tout $t \in T$ la fonction X_t de (Ω, \mathcal{A}) vers (E, \mathcal{E}) est mesurable, i.e.

$$\forall B \in \mathcal{E} : X_t^{-1}(B) = \{X_t \in B\} = \{\omega : X_t(\omega) \in B\} \in \mathcal{A}.$$

En résumé, on a donc :

$$\forall t \in T, \forall B \in \mathcal{E}, X_t^{-1}(B) \in \mathcal{A}.$$

Remarquons que l'on peut aussi considérer un processus comme une fonction de $T \times \Omega$ vers E ,

$$\begin{aligned} T \times \Omega &\rightarrow E \\ (t, \omega) &\mapsto X(t, \omega) \end{aligned}$$

et également comme une fonction de Ω vers l'ensemble des fonctions de T vers E , ensemble noté E^T ,

$$\begin{aligned} X : \Omega &\rightarrow E^T \\ \omega &\mapsto (t \mapsto X_t(\omega)) \end{aligned}$$

Dans la dernière formulation, pour une réalisation ω du hasard on alors une fonction de E^T . Chaque fonction obtenue pour un ω particulier est alors appelée **trajectoire**.

Revenons sur la notion de mesurabilité. Celle-ci, déjà introduite plus haut, est équivalente aux suivantes.

- $\forall n \in \mathbb{N}, \forall (t_1, \dots, t_n) \in T^n, \forall (B_1, \dots, B_n) \in \mathcal{E}^n$, on a :
 $\{(X_{t_1}, \dots, X_{t_n}) \in B_1 \times \dots \times B_n\} \in \mathcal{A}$
- $\forall n \in \mathbb{N}, \forall (t_1, \dots, t_n) \in T^n, \forall B \in \mathcal{E}^{\otimes n}$ tribu produit sur E^n
 $\{(X_{t_1}, \dots, X_{t_n}) \in B\} \in \mathcal{A}$

On a vu que l'on pouvait voir le processus X comme fonction de Ω vers E^T . Si l'on muni ce dernier espace d'une tribu, que l'on notera $\mathcal{E}^{\otimes T}$, on pourra alors également voir le processus X comme une fonction mesurable de (Ω, \mathcal{A}) vers $(E^T, \mathcal{E}^{\otimes T})$. Cette notion de mesurabilité sera alors équivalente à celle introduite précédemment.

DÉFINITION 2.1. Soit $(E_i, \mathcal{E}_i)_{i \in I}$ une famille d'espace mesurables. On note $\otimes_{i \in I} \mathcal{E}_i$ la tribu sur $\prod_{i \in I} E_i$ engendrée par les pavés $\prod_{i \in I} B_i$, où les B_i sont éléments de \mathcal{E}_i et sont tous égaux à E_i , sauf un nombre fini. Si T est un ensemble et (E, \mathcal{E}) un espace mesurable, on notera $\mathcal{E}^{\otimes T}$ la tribu engendrée par les pavés sur E^T .

Les notions équivalentes de la mesurabilité d'un processus stochastique, introduites précédemment, sont alors encore équivalentes à

- Le processus X est mesurable de (Ω, \mathcal{A}) vers $(E, \mathcal{E})^T$.

- Pour tout pavé $\prod_{t \in T} B_t$ de E^T on a :

$$X^{-1}\left(\prod_{t \in T} B_t\right) = \{\omega : X_t(\omega) \in B_t, \text{ pour tout } t\} \in \mathcal{A}$$

1.3. Loi d'un processus, Processus canoniques. La notion de loi d'un processus est calquée sur celle de loi d'une variable aléatoire.

DÉFINITION 2.2. On appelle **loi du processus** $(\Omega, \mathcal{A}, P, (X_t)_{t \in T})$ la loi image P_X de P par X .

La loi d'un processus est donc une probabilité sur l'ensemble des fonctions de T vers E .

THÉORÈME 2.3. La loi d'un processus $(\Omega, \mathcal{A}, P, (X_t)_{t \in T})$ est caractérisée par les valeurs de

$$P(\{(X_{t_1}, \dots, X_{t_n}) \in B_1 \times \dots \times B_n\})$$

pour tout n dans \mathbb{N} , tout (t_1, \dots, t_n) dans T^n et tout $B_1 \times \dots \times B_n$ dans \mathcal{E}^n . On dit encore que la loi P_X du processus X est caractérisée par l'ensemble des lois **marges de dimension finie**, i.e. par l'ensemble des lois $P_{(X_1, \dots, X_n)}$ de tout sous vecteur de dimension finie.

Preuve. Il suffit de remarquer que l'ensemble des pavés mesurables est stable par intersection finie. Le théorème de Carathéodory nous dit alors qu'une probabilité sur $(E, \mathcal{E})^T$ est caractérisée par ses valeurs sur les pavés $\prod_{t \in T} B_t$, i.e. par les $P_X(\prod_{t \in T} B_t)$. La définition d'un pavé implique alors clairement que la donnée de ces probabilités est équivalente à la donnée du système de probabilité du théorème. \square

DÉFINITION 2.4. Soient X et Y deux processus ayant même ensemble des temps T et même espace d'états (E, \mathcal{E}) . On dit que les processus X et Y sont **équivalents** s'ils ont même loi, i.e. si

$$\forall n \in \mathbb{N}, \forall (t_1, \dots, t_n) \in T^n, \forall (B_1, \dots, B_n) \in \mathcal{E}^n, \text{ on a :}$$

$$P(\{(X_{t_1}, \dots, X_{t_n}) \in B_1 \times \dots \times B_n\}) = P(\{(Y_{t_1}, \dots, Y_{t_n}) \in B_1 \times \dots \times B_n\})$$

Parmi tous les processus équivalents, i.e. ayant tous la même loi, on peut en distinguer un de manière unique. Il sera alors souvent pris comme représentant de sa classe d'équivalence.

Soit donc $(\Omega, \mathcal{A}, P, (X_t)_{t \in T})$ un processus défini sur (Ω, \mathcal{A}) et à valeurs dans $(E, \mathcal{E})^T$. Sa loi est P_X , image de P par la fonction mesurable X . Définissons l'ensemble $(\pi_t)_{t \in T}$ des applications coordonnées d'indice t , i.e. pour $t_0 \in T$:

$$\begin{aligned} \pi_{t_0} : E^T &\rightarrow E \\ (x_t)_{t \in T} &\mapsto x_{t_0} \end{aligned}$$

On peut, à la suite de ce que l'on a fait précédemment, montrer que la tribu $\mathcal{E}^{\otimes T}$ est la plus petite tribu rendant mesurable les applications coordonnées de E^T vers E , i.e. $\mathcal{E}^{\otimes T} = \sigma(\pi_{t_0}^{-1}(B) : \forall t_0 \in T, \forall B \in \mathcal{E}) = \sigma(\pi_t : t \in T)$, où si \mathcal{V} est une famille d'ensembles (resp.

famille de fonctions), la tribu $\sigma(\mathcal{V})$ est la tribu engendrée par (resp. rendant mesurable tous les éléments de) \mathcal{V} . Ainsi, sur $(E^T, \mathcal{E}^{\otimes T}, P_X)$ on peut définir une famille $(\pi_t)_{t \in T}$ de v.a. de $(E^T, \mathcal{E}^{\otimes T}, P_X)$ vers (E, \mathcal{E}) . Au sens de la première définition (vue en début de cette section), il s'agit donc d'un processus. Remarquons que ce processus est équivalent au processus $(X_t)_{t \in T}$. En effet, pour tout n dans \mathbb{N} , tout $(t_1, \dots, t_n) \in T^n$ et tout $(B_1, \dots, B_n) \in \mathcal{E}^n$, on a :

$$\begin{aligned} & P_X(\{(\pi_{t_1}, \dots, \pi_{t_n}) \in B_1 \times \dots \times B_n\}) \\ &= P(X^{-1}\{(\pi_{t_1}, \dots, \pi_{t_n}) \in B_1 \times \dots \times B_n\}) \\ &= P(\{\omega : (X_{t_1}(\omega), \dots, X_{t_n}(\omega)) \in B_1 \times \dots \times B_n\}) \\ &= P(\{(X_{t_1}, \dots, X_{t_n}) \in B_1 \times \dots \times B_n\}) \end{aligned}$$

On a donc pratiquement établi la proposition suivante, le reste étant aisé.

PROPOSITION 2.5. *L'ensemble des processus, ayant même espace des temps et d'états, qui sont équivalents entre eux forment une classe d'équivalence.*

*Parmi les éléments de cette classe d'équivalence, il en existe un unique défini sur $(E, \mathcal{E})^T$ par les applications coordonnées. Ce processus est appelé **processus canonique**.*

On remarque que travailler sur les processus canoniques revient à travailler au niveau des lois des processus, i.e. au niveau des classes d'équivalences. On appelle **version du processus** un élément quelconque de cette classe et on utilise le plus souvent la version canonique.

1.4. Processus canoniques ayant des répartitions finies données. Théorème de Kolmogorov. On vient de voir dans la section précédente qu'un processus stochastique est entièrement déterminé par ses marges de dimension finie. On peut se poser la question dans l'autre sens. Ayant un système de loi de dimension finie, existe-t-il un processus $(X_t)_{t \in T}$ ayant pour marges de dimension finie, ces lois ? De manière plus mathématique, on posera la question sous la forme suivante. Notons pour cela Σ l'ensemble des parties finies de T , i.e.

$$S \in \Sigma \text{ si } \exists n \in \mathbb{N} \text{ et } (t_1, \dots, t_n) \in T^n \text{ tel que } S = \{t_1, \dots, t_n\}.$$

Supposons que l'on ait un système $(\mu_s)_{s \in \Sigma}$ de lois de dimension finie. Existe-t-il une probabilité P sur (E^T, \mathcal{E}^T) , et si oui est-elle unique, telle que P ait pour système de marges de dimension finie la famille $(\mu_s)_{s \in \Sigma}$?

On remarque assez vite qu'une question de cohérence apparaît. Notons en effet π_S la projection de E^T sur E^S et $\pi_{S'}$, pour $S' \subset S$, la projection de E^S sur $E^{S'}$. On cherche donc une probabilité P telle que la probabilité $\pi_S * P$, image de P par π_S , soit égale à μ_S et ce pour tout S dans Σ . Or, clairement, pour $S' \subset S$, on doit avoir $\pi_{S'} * \pi_S * P$ à la fois égal à $\pi_{S'} * P = \mu_{S'}$ et égal à $\pi_{S'} * \mu_S$. Il faut donc nécessairement que le système de marges que l'on se donne $(\mu_s)_{s \in \Sigma}$ soit **cohérent**, i.e. :

$$\forall (S', S) \in \Sigma^2, S' \subset S : \pi_{S'} * \mu_S = \mu_{S'}.$$

L'intérêt du théorème suivant du à Kolmogorov est de montrer que cette condition, dont on vient de souligner la nécessité, est aussi suffisante. Mais donnons en premier lieu la définition d'un espace polonais.

DÉFINITION 2.6. *Un espace E est dit **polonais**, s'il est métrique, séparable (i.e. il existe une base dénombrable d'ouverts) et complet (toute suite de Cauchy est convergente).*

THÉORÈME 2.7. (Kolmogorov) *Soit E un espace d'états polonais muni de sa tribu borélienne (i.e. tribu engendrée par les ouverts pour la topologie) et T un espace des temps. Soit $(\mu_s)_{s \in \Sigma}$ un système de lois de dimension finie (μ_S loi sur E^S pour S partie finie de T). Si ce*

Le système de lois est cohérent alors il existe une unique probabilité P sur $(E, \mathcal{E})^T$ telle que ses marges de dimension finie coïncident avec le système de lois $(\mu_S)_{S \in \Sigma}$.

Remarque : On peut vérifier que les espaces mesurables suivants sont polonais :

- $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$;
- $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$;
- E est dénombrable muni de la tribu de toutes ses parties.

1.5. Modification d'un processus et processus indistinguables.

DÉFINITION 2.8. Soit $X = (\Omega, \mathcal{A}, P, (X_t)_{t \in T})$ et $X' = (\Omega, \mathcal{A}, P, (X'_t)_{t \in T})$ définis sur le même espace probabilisé (Ω, \mathcal{A}, P) , ayant même espace des temps T et même espace d'états E .

On dit que le processus X est une **modification** du processus X' si pour tout t dans T , les v.a. X_t et X'_t sont presque sûrement égales.

Ils sont dits **indistinguables** si l'on a P -presque-sûrement :

$$X_t = X'_t, \quad \forall t \in T.$$

Il est clair que deux processus modifications l'un de l'autre sont équivalents. Remarquons également que deux processus indistinguables sont modification l'un de l'autre mais, en revanche, la réciproque est fautive dès que l'espace des temps n'est pas dénombrable. L'indistinguabilité signifie qu'en dehors d'un ensemble négligeable sur Ω , toutes les trajectoires des deux processus coïncident ce que n'exigeait pas la notion de modification.

Exemples de processus qui sont modifications l'un de l'autre et qui ne sont pas indistinguables

- Soient $(X_t)_{t \in [0,1]}$ et $(X'_t)_{t \in [0,1]}$ deux processus définis sur le même espace probabilisé $([0, 1], \mathcal{B}_{[0,1]}, \lambda_{[0,1]})$, où $\lambda_{[0,1]}$ est la mesure de Lebesgue sur $[0, 1]$, et définis par :

$$\begin{aligned} X_t(\omega) &= \mathbb{1}_{]t,1]}(\omega) \\ X'_t(\omega) &= \mathbb{1}_{[t,1]}(\omega) \end{aligned}$$

On a pour tout t dans $[0, 1]$:

$$\lambda_{[0,1]}(X_t = X'_t) = \lambda_{[0,1]}(\{\omega : X_t(\omega) = X'_t(\omega)\}) = \lambda_{[0,1]}([0, 1] \setminus \{t\}) = 1$$

Les processus $(X_t)_{t \in [0,1]}$ et $(X'_t)_{t \in [0,1]}$ sont donc modifications l'un de l'autre mais ne sont pas indistinguables puisque les trajectoires, pour ω quelconque dans Ω sont :

$$\begin{aligned} t \mapsto X_t(\omega) &= \mathbb{1}_{]0,\omega[}(t) \\ t \mapsto X'_t(\omega) &= \mathbb{1}_{[0,\omega]}(t) \end{aligned}$$

et sont donc toujours différentes.

- Soit $(X_t)_{t \in [0,1]}$ un processus déterministe, tel que :

$$X_t(\omega) = 1, \quad \forall \omega \in \Omega \text{ et } \forall t \in [0, 1].$$

Soit maintenant U une v.a. de loi uniforme sur $[0, 1]$ et $(Y_t)_{t \in [0,1]}$ le processus défini, pour tout ω dans Ω par :

$$Y_t(\omega) = \begin{cases} 0 & \text{pour } t = U(\omega) \\ 1 & \text{ailleurs.} \end{cases}$$

On a alors clairement, pour tout t dans $[0, 1]$:

$$P(\{X_t = Y_t\}) = 1 - P(\{U = t\}) = 1.$$

Mais en revanche :

$$P(\{\forall t : X_t = Y_t\}) = 0.$$

Remarque : On verra ultérieurement qu'en revanche si deux processus sont modifications l'un de l'autre et s'ils sont tous les deux à trajectoires *p.s.* continues alors, nécessairement, ils sont également indistinguables.

2. Exemples classiques et fondamentaux de processus stochastiques

2.1. Suite de v.a. indépendantes. Supposons que les v.a. $(X_n)_{n \in \mathbb{N}}$ soient toutes indépendantes, à valeurs dans \mathbb{R} mais non nécessairement identiquement distribuées, de f.d.r. $(F_n)_{n \in \mathbb{N}}$. Les lois marginales du processus sont bien sûr caractérisées par les f.d.r. des vecteurs extraits. Or la f.d.r. multidimensionnelle d'un vecteur extrait X_{n_1}, \dots, X_{n_k} , où $\{n_1, \dots, n_k\}$ est une partie quelconque de \mathbb{N} , est :

$$F(x_{n_1}, \dots, x_{n_k}) = F_{n_1}(x_{n_1}) \cdots F_{n_k}(x_{n_k}).$$

On vérifie très facilement que ce système est cohérent et donc d'après le théorème de Kolmogorov, il existe un processus sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})^{\mathbb{N}}$ ayant les marges qui coïncident avec ce système de lois de dimension finie.

2.2. Processus stationnaires. Il a été vu en cours de Séries Temporelles qu'un processus avec pour espace des temps T et espace d'états E est dit **stationnaire (au sens fort)**, si pour tout n dans \mathbb{N} et tout $(t_1, \dots, t_n) \in T^n$, on a identité des lois des marges de dimension finie prises aux instants (t_1, \dots, t_n) et $(t_1 + h, \dots, t_n + h)$, pour $h > 0$, i.e.

$$\forall (B_1, B_2, \dots, B_n) \in \mathcal{E}^n, \text{ on a :}$$

$$P(\{(X_{t_1}, X_{t_2}, \dots, X_{t_n}) \in B_1 \times \cdots \times B_n\}) = P(\{(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_n+h}) \in B_1 \times \cdots \times B_n\})$$

On dit qu'il y a invariance des lois des marges de dimension finie par translation temporelle.

Si le processus est du second ordre (i.e. $X_t \in L^2(\Omega, \mathcal{A}, P)$, pour tout t), on dit qu'il est **stationnaire au sens faible** si sa moyenne est constante et la covariance entre X_t et X_{t+h} , pour t et $t+h$ dans T , ne dépend que de h .

2.3. Processus gaussiens.

DÉFINITION 2.9. *Un processus d'espace des temps T et d'espace d'états \mathbb{R} est dit **processus gaussien réel** si toutes ses marges de dimension finie sont des vecteurs gaussiens*

On sait qu'un vecteur gaussien est entièrement spécifié par la donnée de son espérance et sa matrice de covariance. On va donc pouvoir entièrement spécifier un processus gaussien en donnant sa fonction moyenne et sa fonction covariance.

En remarquant que tout vecteur extrait d'un vecteur gaussien est encore un vecteur gaussien, la cohérence d'un tel système de loi est vérifiée et on a donc la proposition suivante.

THÉORÈME 2.10. *Soit m une fonction de T vers \mathbb{R} et Γ une fonction symétrique de T^2 vers \mathbb{R} tel que pour toute partie finie $\{t_1, \dots, t_n\}$ de T , la matrice $(\Gamma(t_i, t_j))_{1 \leq i, j \leq n}$ soit définie positive (on dit que la fonction Γ est de type positif).*

Il existe alors un unique processus gaussien, à une équivalence près, dont les marges finies, pour $n \in \mathbb{N}$ et $(t_1, \dots, t_n) \in T^n$, sont un vecteur gaussien d'espérance $(m(t_1), \dots, m(t_n))^T$ et de matrice de covariance $(\Gamma(t_i, t_j))_{1 \leq i, j \leq n}$. Les fonctions m et Γ caractérisent entièrement la loi du processus gaussien.

2.4. Processus à accroissements indépendants.

DÉFINITION 2.11. Soit T un espace des temps inclus dans \mathbb{R} , E un espace polonais et X un processus de (Ω, \mathcal{A}, P) vers $(E, \mathcal{E})^T$.

On dit que le processus X est à **accroissements indépendants** (on écrit souvent **PAI**) si pour tout $t_1 < t_2 < \dots < t_n$ de T , les v.a.

$$X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, \dots, X_{t_n} - X_{t_{n-1}}$$

sont mutuellement indépendants. Si l'espace des temps admet un plus petit indice t_0 , on suppose que la famille précédente enrichie de la v.a. X_{t_0} est encore une famille de v.a. indépendantes.

Le processus est dit à **accroissements stationnaires** si la loi de $X_{t+h} - X_t$ dépend uniquement de h et non de t . Un processus à accroissements indépendants et stationnaires est noté **PAIS**.

Si l'espace des temps est \mathbb{N} (ou éventuellement \mathbb{Z}) on parle plutôt de **marche aléatoire**, appellation plus claire en considérant les v.a. (Z_n) définies, pour $n > 0$ par :

$$Z_0 = X_0, Z_1 = X_1 - X_0, Z_2 = X_2 - X_1, \dots, Z_n = X_n - X_{n-1},$$

On a en effet

$$X_n = Z_0 + Z_1 + \dots + Z_n$$

Le pas effectué au temps n par le processus, i.e. $Z_n = X_n - X_{n-1}$, est bien indépendants du passé.

Exemple : La marche aléatoire simple.

Soit X un processus stochastique à espace des temps \mathbb{N} , espace d'états \mathbb{Z} et défini de la manière suivante : $X_0 = 0$ et $Z_n = X_n - X_{n-1}$ est de loi $p\delta_1 + (1-p)\delta_{-1}$ et la famille des v.a. $(Z_n)_{n \in \mathbb{N}}$ est une famille indépendante. \diamond

Nous étudierons plus tard le mouvement brownien et le processus de Poisson qui sont des PAIS.

2.5. Martingales. Soit X un processus défini sur (Ω, \mathcal{A}, P) et à valeurs sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})^T$, où T est soit \mathbb{R}^+ , soit \mathbb{N} pour simplifier les notations. Définissons pour tout t dans T la tribu engendrée par les X_s pour $s \leq t$,

$$\mathcal{F}_t = \sigma(X_s; s \leq t \text{ et } s \in T),$$

dite histoire du processus au temps t . Il est évident que l'on a $\mathcal{F}_t \subset \mathcal{F}_{t'}$, pour tout $t \leq t'$. La famille de tribu (\mathcal{F}_t) est dite **filtration naturelle** associée au processus X .

DÉFINITION 2.12. On dit qu'un processus $(X_t)_{t \in T}$ à valeurs dans $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ est une **martingale** par rapport à sa filtration naturelle si l'on a :

$$\begin{aligned} \mathbb{E}|X_t| < +\infty, \quad \forall t \in T \\ \mathbb{E}(X_t / \mathcal{F}_s) = X_s, \quad \text{pour tout } s \leq t \text{ et } (s, t) \in T^2 \end{aligned}$$

Le signe $\mathbb{E}(X_t / \mathcal{F}_s)$ signifie l'espérance conditionnelle de X_t relativement à la tribu \mathcal{F}_s . Cette notion d'espérance conditionnelle sera définie précisément dans un chapitre ultérieur.

On peut utiliser les martingales pour modéliser des jeux équitables. En effet, si $X_t - X_s$ est le gain du joueur entre les instants s et t , on a dans un jeu équilibré : $\mathbb{E}(X_t - X_s / \mathcal{F}_s) = 0$.

Remarquons enfin, que la marche aléatoire simple que nous avons considéré précédemment est une martingale à temps discret si elle est symétrique, i.e. si $p = 1/2$, puisque :

$$\mathbb{E}|X_n| \leq \sum_{i=1}^n \mathbb{E}|Z_i| < +\infty$$

et $\mathbb{E}(X_{n+1} - X_n / \mathcal{F}_n) = \mathbb{E}(Z_{n+1} / \mathcal{F}_n) = \mathbb{E}Z_{n+1} = 0$

Nous reviendrons sur les martingales à temps discret, qui sont au programme de cette unité, dans un prochain chapitre.

2.6. Processus de Markov. Nous reprenons les mêmes notations qu'à la sous section précédente.

DÉFINITION 2.13. *Un processus $(\Omega, \mathcal{A}, P, (X_t)_{t \in T})$ à valeurs dans un espace (E, \mathcal{E}) est dit **processus de Markov** si pour tout $s \leq t$, $(s, t) \in T^2$ et tout $A \in \mathcal{E}$, on a :*

$$P(X_t \in A / \mathcal{F}_s) = P(X_t \in A / X_s).$$

On dit que le futur du processus ne dépend du passé qu'à travers le présent.

Ce processus est dit homogène si la loi de X_t sachant X_s ne dépend que de $t - s$ pour $s < t$.

Terminologie : Si l'espace des temps est discret, on parle de **chaîne de Markov**, s'il est continu on parle de processus de Markov (à temps continu). Les chaînes de markov sont au programme de cette unité et seront étudiées dans un chapitre ultérieur.

2.7. Processus de renouvellement. Processus de Poisson.

DÉFINITION 2.14. *Soient $(T_n)_{n \in \mathbb{N}}$ une suite de v.a. sur (Ω, \mathcal{A}, P) , positives, indépendantes et identiquement distribuées de f.d.r. F . Le processus $(\Omega, \mathcal{A}, P, (S_t)_{t \in \mathbb{R}^+})$ à valeurs dans $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+})$ défini par*

$$S_n = T_1 + \dots + T_n$$

*est dit **processus de renouvellement**.*

On remarque qu'un processus de renouvellement est un PAIS. Ces processus sont par exemple utilisés en fiabilité. Supposons que l'on dispose de matériels identiques (i.e. dont la loi d'attente de la panne est identique) et au comportement indépendant. Plaçons une première unité en marche à l'instant $t = 0$. Dès que celle-ci tombe en panne, on la remplace instantanément par une seconde et ainsi de suite. Le temps où aura lieu le $n^{\text{ième}}$ renouvellement est donc $S_n = T_1 + \dots + T_n$, où $(T_i)_{i=1, \dots, n}$ sont les temps d'attente de la panne pour les différents matériels.

A ce processus de renouvellement on peut associer un processus de comptage du nombre de renouvellement.

DÉFINITION 2.15. *Soit $(S_n)_{n \in \mathbb{N}}$ un processus de renouvellement défini par une suite de v.a. i.i.d. $(T_n)_{n \in \mathbb{N}}$ de f.d.r. F sur (Ω, \mathcal{A}, P) et à valeurs dans $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+})$*

*On appelle **processus de comptage des renouvellements** le processus $(N_t)_{t \in \mathbb{R}^+}$ défini sur (Ω, \mathcal{A}, P) et à valeurs dans \mathbb{N} par :*

$$\begin{aligned} N_t &= \sum_{n=1}^{+\infty} \mathbb{1}_{]0, t]}(S_n) \\ &= n \text{ si } S_n \leq t < S_{n+1} \\ &= \text{nombre de renouvellement survenus jusqu'au temps (inclus) } t. \end{aligned}$$

DÉFINITION 2.16. On appelle **processus de Poisson** de paramètre λ un processus de comptage de renouvellements associé à un processus de renouvellement où la loi des v.a. $(T_n)_{n \in \mathbb{N}}$ est exponentielle de paramètre λ .

On reviendra sur ces processus dans un autre chapitre.

3. Bibliographie

- Harald Cramer et M.R. Leadbetter. *Stationary and Related Stochastic Processes*. Wiley 1967.
- Didier Dacunha-Castelle et Marie Duflo. *Probabilités et Statistiques 2 : Problèmes à temps mobile*. Masson 1983.
- Claude Dellacherie et Paul-André Meyer. *Probabilités et potentiel, chapitres I à IV*. Hermann 1975.
- I.N. Kovalenko, N.Yu. Kuznetsov et V.M. Shurenkov. *Models of Random Processes*. CRC Press 1996.
- Michel Loève. *Probability Theory II, 4th Edition*. Springer 1978.
- M. Métivier. *Notions fondamentales de la théorie des probabilités*. Dunod 1972.

CHAPITRE 3

Processus de Poisson

On s'intéresse ici à modéliser un phénomène dont les instants d'occurrence forment une suite croissante de variables positives notées $(T_n)_{n \geq 0}$. Il en est ainsi par exemple pour modéliser les instants d'appels reçus par un standard téléphonique, les instants d'arrivée des patients aux urgences, de voitures à un feu rouge, de clients à une caisse de supermarché, des pannes sur un système réparable.

Mathématiquement il s'agit donc d'une suite croissante de v.a. $(T_n)_{n \geq 0}$:

$$0 = T_0 \leq T_1 \leq T_2 \leq \dots \leq T_n \leq \dots$$

On parle de processus ponctuels. Bien sûr, en posant

$$X_1 = T_1, X_2 = T_2 - T_1, \dots, X_n = T_n - T_{n-1},$$

il est équivalent de définir les instants $(T_n)_{n \geq 0}$ ou la suite $(X_n)_{n \geq 1}$ de v.a.r. positives, puisque l'on a :

$$T_n = X_1 + X_2 + \dots + X_n.$$

Les v.a. $(X_n)_{n \geq 1}$ sont souvent appelées temps inter-arrivées.

Notons N le processus défini sur \mathbb{R}^+ et à valeurs dans \mathbb{N} par

$$N_t = \sum_{n=1}^{+\infty} \mathbb{1}_{T_n \leq t}$$

appelé processus de comptage associé au processus ponctuel $(T_n)_{n \geq 0}$. Il est à nouveau clair que connaître la loi de $(T_n)_{n \geq 0}$ est équivalent à connaître celle du processus $N = (N_t)_{t \geq 0}$. On a en effet, pour tout n dans \mathbb{N} et tout $(t_1, \dots, t_n) \in \mathbb{R}^+ \times \dots \times \mathbb{R}^+$,

$$P(T_1 \leq t_1, T_2 \leq t_2, \dots, T_n \leq t_n) = P(N_{t_1} \geq 1, N_{t_2} \geq 2, \dots, N_{t_n} \geq n).$$

En posant l'hypothèse que les v.a. $(X_n)_{n \geq 1}$ sont i.i.d. de même f.d.r. F , on a vu au chapitre 2 que le processus $(T_n)_{n \geq 0}$ est alors un processus de renouvellement et $N = (N_t)_{t \geq 0}$ le processus de comptage des renouvellements. Nous allons étudier dans ce chapitre un cas particulier que constitue les processus de Poisson.

1. Définition et propriétés élémentaires

Il y a de très nombreuses manières de définir un processus de Poisson homogène sur \mathbb{R}^+ . Rappelons celle donnée au chapitre 2 et qui le présente comme un processus de renouvellement particulier.

DÉFINITION 3.1. *Un processus ponctuel $(T_n)_{n \geq 0}$ est dit processus de Poisson homogène sur \mathbb{R}^+ d'intensité λ si les temps d'inter-arrivées $(X_n)_{n \geq 1}$ sont i.i.d. (donc s'il est un processus de renouvellement) de loi exponentielle¹ $\mathcal{E}(\lambda)$. On dit de manière équivalente que $N = (N_t)_{t \geq 0}$ est un processus de Poisson.*

Remarquons que l'on a *p.s.* :

$$0 < T_1 < T_2 < \dots < T_n < \dots,$$

puisque les v.a. $(X_n)_{n \geq 1}$, de lois exponentielles, sont *p.s.* strictement positives. De plus le processus $N = (N_t)_{t \geq 0}$ est à trajectoires croissantes, continues à droite, à valeurs entières et ayant des sauts de taille 1.

Il y a un lien très étroit entre le processus de Poisson et la loi de Poisson comme le montre la proposition suivante.

¹Rappelons qu'une variable aléatoire réelle est dite de loi exponentielle de paramètre λ si elle est de loi absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R} et de densité $f(x) = \lambda \exp(-\lambda x) \mathbf{1}_{\mathbb{R}^+}(x)$.

PROPOSITION 3.2. Soit $N = (N_t)_{t \geq 0}$ un processus de Poisson d'intensité λ . Pour tout $t > 0$, la v.a.r. N_t est de loi de Poisson de paramètre λt , i.e.

$$P(N_t = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}.$$

EXERCICE 3.1. Démonstration de cette proposition.

1) En utilisant directement la définition d'un processus de Poisson, montrer que la formule est déjà vraie pour $n = 0$.

2) Montrer que loi de la somme de n v.a. i.i.d. de loi exponentielle $\mathcal{E}(\lambda)$ est une loi Gamma² $\gamma(n, \lambda)$. (Ind. on pourra utiliser la f.c. ou la transformée de Laplace)

3) Montrer que l'on a :

$$P(N_t = n) = P(T_n \leq t, T_n + X_{n+1} > t).$$

4) En calculant cette dernière probabilité, montrer que l'on a bien :

$$P(N_t = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}.$$

Preuve/solution de l'exercice.

1) Pour $n = 0$, il n'y a presque rien à montrer puisque

$$P(N_t = 0) = P(T_1 > t) = P(X_1 > t) = 1 - P(X_1 \leq t) = 1 - (1 - e^{-\lambda t}) = e^{-\lambda t}.$$

2) La transformée de Laplace d'une loi $\gamma(\alpha, \beta)$ est, pour $s < \beta$:

$$\begin{aligned} \mathcal{L}_{\gamma(\alpha, \beta)}(s) &= \int_{\mathbb{R}^+} e^{sx} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^{+\infty} x^{\alpha-1} e^{-(\beta-s)x} dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^{+\infty} \frac{u^{\alpha-1}}{(\beta-s)^\alpha} e^{-u} du = \frac{1}{(1 - \frac{s}{\beta})^\alpha}. \end{aligned}$$

D'où comme X_1, \dots, X_n sont i.i.d. de loi $\mathcal{E}(\lambda) = \gamma(1, \lambda)$, on a :

$$\begin{aligned} \mathcal{L}_{\sum_{i=1}^n X_i}(s) &= \mathbb{E}(e^{s \sum_{i=1}^n X_i}) = \prod_{i=1}^n \mathbb{E}(e^{s X_i}) \\ &= \prod_{i=1}^n \frac{1}{(1 - \frac{s}{\beta})} = \frac{1}{(1 - \frac{s}{\beta})^n} = \mathcal{L}_{\gamma(n, \lambda)}(s), \end{aligned}$$

ce qui, par caractérisation de la transformée de Laplace, permet de conclure.

²Rappelons qu'une variable aléatoire réelle est dite de loi gamma de paramètres (α, β) si elle est de loi absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R} et de densité

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbf{1}_{\mathbb{R}^+}(x).$$

3) On a :

$$P(N_t = n) = P(T_n \leq t, T_{n+1} > t) = P(T_n \leq t, T_n + X_{n+1} > t).$$

4) Par indépendance entre T_n et X_{n+1} on peut écrire :

$$\begin{aligned} P(T_n \leq t, T_n + X_{n+1} > t) &= \int_{u \leq t, u+y > t} f_{T_n, X_{n+1}}(u, y) du dy \\ &= \int_0^t \frac{\lambda^n}{\Gamma(n)} u^{n-1} e^{-\lambda u} \left(\int_{t-u}^{+\infty} \lambda e^{-\lambda y} dy \right) du \\ &= \int_0^t \frac{\lambda^n}{\Gamma(n)} u^{n-1} e^{-\lambda u} e^{-\lambda(t-u)} du \\ &= \frac{\lambda^n}{\Gamma(n)} e^{-\lambda t} \int_0^t u^{n-1} du = \frac{\lambda^n}{\Gamma(n)} e^{-\lambda t} \left[\frac{u^n}{n} \right]_0^t \\ &= \frac{\lambda^n t^n e^{-\lambda t}}{(n-1)!n} = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \end{aligned}$$

ce qui est bien la formule annoncée. \square

Avant de présenter d'autres propriétés élémentaires du processus de Poisson, introduisons un lemme technique.

LEMME 3.3. Soient X_1, \dots, X_n des v.a. sur \mathbb{R} i.i.d. absolument continues (i.e. de loi absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R}) de densité f . Notons $X_{(1)} < \dots < X_{(n)}$ les n statistiques d'ordre associées. Le vecteur $(X_{(1)}, \dots, X_{(n)})$ est alors absolument continu de densité sur \mathbb{R}^n

$$f_{(X_{(1)}, \dots, X_{(n)})}(u_1, \dots, u_n) = n! \prod_{i=1}^n f(u_i) \mathbb{1}_{u_1 < u_2 < \dots < u_n}$$

Preuve. Soit B un borélien de \mathbb{R}^n et Σ_n l'ensemble des permutations de l'ensemble $\{1, 2, \dots, n\}$. On a :

$$\begin{aligned} P((X_{(1)}, \dots, X_{(n)}) \in B) &= \sum_{\sigma \in \Sigma_n} P(\{(X_{\sigma(1)}, \dots, X_{\sigma(n)}) \in B\} \cap \{X_{\sigma(1)} < \dots < X_{\sigma(n)}\}) \\ &= \sum_{\sigma \in \Sigma_n} \int_B \mathbb{1}_{u_1 < u_2 < \dots < u_n} \prod_{i=1}^n f(u_i) du_1 \dots du_n \\ &= \int_B n! \prod_{i=1}^n f(u_i) \mathbb{1}_{u_1 < u_2 < \dots < u_n} du_1 \dots du_n, \end{aligned}$$

ce qui prouve bien le lemme. \square

PROPOSITION 3.4. Soit $(T_n)_{n \geq 0}$ un processus de Poisson homogène sur \mathbb{R}^+ , de processus de comptage associé $(N_t)_{t \geq 0}$. On a alors :

i) Pour tout $n \geq 1$, la loi du vecteur (T_1, \dots, T_n) a pour densité

$$\lambda^n e^{-\lambda t_n} \mathbb{1}_{t_1 < t_2 < \dots < t_n}$$

par rapport à la mesure de Lebesgue sur \mathbb{R}^n .

- ii) Pour tout $n \geq 1$, la loi conditionnelle de (T_1, \dots, T_n) sachant l'événement $\{N_t = n\}$ est celle des statistiques d'ordre de n v.a.r. i.i.d. de loi uniforme sur $[0, t]$, i.e.

$$f_{(T_1, \dots, T_n)}^{N_t=n}(t_1, \dots, t_n) = \frac{n!}{t^n} \mathbb{1}_{t_1 < t_2 < \dots < t_n \leq t}$$

EXERCICE 3.2. *Démonstration de cette proposition.*

1) Pour montrer le i), utiliser le théorème de changement de variables multidimensionnel et la définition d'un processus de Poisson.

2) Pour montrer le i), montrer en premier lieu que pour tout borélien B de $\mathbb{R}^+ \times \dots \times \mathbb{R}^+$, on a :

$$P((T_1, \dots, T_n) \in B, N_t = n) = \int \mathbb{1}_B(t_1, \dots, t_n) \mathbb{1}_{t_1 < t_2 < \dots < t_n \leq t} \lambda^n e^{-\lambda t} dt_1 \dots dt_n.$$

3) En déduire que l'on a :

$$P((T_1, \dots, T_n) \in B \mid N_t = n) = \int_B \frac{n!}{t^n} \mathbb{1}_{t_1 < t_2 < \dots < t_n \leq t} dt_1 \dots dt_n$$

et conclure.

Preuve/solution de l'exercice.

- 1) C'est une application aisée du théorème de changement de variables multidimensionnel. En effet, on a :

$$\begin{cases} T_1 &= X_1 \\ T_2 &= X_1 + X_2 \\ &\vdots \\ T_n &= X_1 + \dots + X_n. \end{cases}$$

Ainsi, le vecteur (T_1, \dots, T_n) est la transformation linéaire du vecteur (X_1, \dots, X_n) par φ définie sur $\mathbb{R}^+ \times \dots \times \mathbb{R}^+$ et à valeurs dans $\Delta_n = \{(t_1, \dots, t_n) \in \mathbb{R}^n; t_1 < t_2 < \dots < t_n\}$, qui est donc un \mathcal{C}^1 -difféomorphisme de Jacobien égal à 1. La formule du changement de variable multidimensionnel nous donne donc :

$$\begin{aligned} f_{(T_1, \dots, T_n)}(t_1, \dots, t_n) &= f_{(X_1, \dots, X_n)}(\varphi^{-1}(t_1, \dots, t_n)) |J_{\varphi^{-1}}| \mathbb{1}_{\Delta}(t_1, \dots, t_n) \\ &= \lambda e^{-\lambda t_1} \lambda e^{-\lambda(t_2 - t_1)} \lambda e^{-\lambda(t_3 - t_2)} \dots \lambda e^{-\lambda(t_n - t_{n-1})} \mathbb{1}_{t_1 < t_2 < \dots < t_n} \\ &= \lambda^n e^{-\lambda t_n} \mathbb{1}_{t_1 < t_2 < \dots < t_n} \end{aligned}$$

- 2) Soit B un borélien de $\mathbb{R}^+ \times \dots \times \mathbb{R}^+$. On a :

$$\begin{aligned} P((T_1, \dots, T_n) \in B, N_t = n) &= P((T_1, \dots, T_n) \in B, T_n \leq t < T_{n+1}) \\ &= \int \mathbb{1}_B(t_1, \dots, t_n) \mathbb{1}_{t_n \leq t < t_{n+1}} \lambda^{n+1} e^{-\lambda t_{n+1}} \mathbb{1}_{t_1 < t_2 < \dots < t_{n+1}} dt_1 \dots dt_{n+1} \\ &= \int \mathbb{1}_B(t_1, \dots, t_n) \mathbb{1}_{t_1 < t_2 < \dots < t_n \leq t} \lambda^n \left(\int_t^{+\infty} \lambda e^{-\lambda t_{n+1}} dt_{n+1} \right) dt_1 \dots dt_n \\ &= \int \mathbb{1}_B(t_1, \dots, t_n) \mathbb{1}_{t_1 < t_2 < \dots < t_n \leq t} \lambda^n e^{-\lambda t} dt_1 \dots dt_n. \end{aligned}$$

3) De ce qui précède on tire

$$\begin{aligned} P((T_1, \dots, T_n) \in B \mid N_t = n) &= \frac{P((T_1, \dots, T_n) \in B, N_t = n)}{P(N_t = n)} \\ &= \int_B \frac{n!}{t^n} \mathbb{1}_{t_1 < t_2 < \dots < t_n \leq t} dt_1 \cdots dt_n, \end{aligned}$$

ce qui prouve bien que la densité de (T_1, \dots, T_n) sachant $\{N_t = n\}$ est de densité

$$\frac{n!}{t^n} \mathbb{1}_{t_1 < t_2 < \dots < t_n \leq t}$$

qui correspond à la densité des n statistiques d'ordre dans un échantillon de lois uniformes sur $[0, t]$. \square

2. Caractérisation d'un processus de Poisson

Nous allons voir sous forme de théorème deux séries de conditions nécessaires et suffisantes pour avoir un processus de Poisson. Chacune d'elle aurait pu être utilisée comme définition d'un processus de Poisson.

THÉORÈME 3.5. *Soit $N = (N_t)_{t \geq 0}$ le processus de comptage d'un processus ponctuel sur \mathbb{R}^+ . C'est un processus de Poisson d'intensité λ si, et seulement si,*

- i) $N_0 = 0$;
- ii) le processus N est à accroissements indépendants ;
- iii) pour tout $0 \leq s < t$, la v.a.r. $N_t - N_s$ est de loi de Poisson de paramètre $\lambda(t - s)$.

Conséquence : La v.a. $N_t - N_s$ compte le nombre d'arrivées dans l'intervalle $]s, t]$ et les nombres d'arrivées observées sur deux intervalles disjoints sont indépendants en probabilité. \diamond

EXERCICE 3.3. *Démonstration de la condition nécessaire de ce théorème. La condition suffisante est directe.*

Soit $n \in \mathbb{N}$, un ensemble de n réels ordonnés $0 \leq t_1 < t_2 < \dots < t_n$, et n entiers strictement positifs k_1, k_2, \dots, k_n . Notons $k = k_1 + k_2 + \dots + k_n$ et $(U_{(1)}, \dots, U_{(n)})$ les n statistiques d'ordre d'un échantillon de v.a. de loi uniforme sur $[0, t_n]$.

1) En utilisant en particulier la proposition précédente, montrer que l'on peut écrire :

$$\begin{aligned} P(N_{t_1} = k_1, N_{t_2} - N_{t_1} = k_2, \dots, N_{t_n} - N_{t_{n-1}} = k_n) &= \\ P\left(\sum_{i=1}^k \mathbb{1}_{\{U_{(i)} \leq t_1\}} = k_1, \sum_{i=1}^k \mathbb{1}_{\{t_1 < U_{(i)} \leq t_2\}} = k_2, \dots, \sum_{i=1}^k \mathbb{1}_{\{t_{n-1} < U_{(i)} \leq t_n\}} = k_n\right) & \end{aligned}$$

2) Se convaincre que dans la dernière expression on peut remplacer les statistiques d'ordre par les variables initiales et donc non ordonnées. Montrer alors, en utilisant la loi multinomiale, que l'on a :

$$\begin{aligned} P(N_{t_1} = k_1, N_{t_2} - N_{t_1} = k_2, \dots, N_{t_n} - N_{t_{n-1}} = k_n) &= \\ \frac{(\lambda t_1)^{k_1}}{k_1!} e^{-\lambda t_1} \frac{(\lambda(t_2 - t_1))^{k_2}}{k_2!} e^{-\lambda(t_2 - t_1)} \dots \frac{(\lambda(t_n - t_{n-1}))^{k_n}}{k_n!} e^{-\lambda(t_n - t_{n-1})}. & \end{aligned}$$

3) *Conclure.*

Preuve/solution de l'exercice.

1) On a :

$$\begin{aligned}
& P(N_{t_1} = k_1, N_{t_2} - N_{t_1} = k_2, \dots, N_{t_n} - N_{t_{n-1}} = k_n) \\
= & P(N_{t_1} = k_1, N_{t_2} - N_{t_1} = k_2, \dots, N_{t_n} - N_{t_{n-1}} = k_n | N_{t_n} = k) P(N_{t_n} = k) \\
= & P\left(\sum_{i=1}^k \mathbb{1}_{\{T_i \leq t_1\}} = k_1, \sum_{i=1}^k \mathbb{1}_{\{t_1 < T_i \leq t_2\}} = k_2, \dots, \sum_{i=1}^k \mathbb{1}_{\{t_{n-1} < T_i \leq t_n\}} = k_n | N_{t_n} = k\right) \times P(N_{t_n} = k) \\
= & P\left(\sum_{i=1}^k \mathbb{1}_{\{U_{(i)} \leq t_1\}} = k_1, \sum_{i=1}^k \mathbb{1}_{\{t_1 < U_{(i)} \leq t_2\}} = k_2, \dots, \sum_{i=1}^k \mathbb{1}_{\{t_{n-1} < U_{(i)} \leq t_n\}} = k_n\right) \times P(N_{t_n} = k),
\end{aligned}$$

où la dernière égalité est justifiée par le *ii*) de la proposition précédente.

2) Compter le nombre de statistiques d'ordre plus petites que t_1 (par exemple) revient à compter le nombre de variables initiales plus petites que t_1 . De même pour celles comprises entre t_1 et t_2 .

Par ailleurs le vecteur aléatoire

$$\left(\sum_{i=1}^k \mathbb{1}_{\{U_i \leq t_1\}}, \sum_{i=1}^k \mathbb{1}_{\{t_1 < U_i \leq t_2\}}, \dots, \sum_{i=1}^k \mathbb{1}_{\{t_{n-1} < U_i \leq t_n\}} \right)$$

est de loi multinomiale de paramètres

$$\left(k, \frac{t_1}{t_n}, \frac{t_2 - t_1}{t_n}, \dots, \frac{t_n - t_{n-1}}{t_n} \right).$$

On a donc :

$$\begin{aligned}
& P(N_{t_1} = k_1, N_{t_2} - N_{t_1} = k_2, \dots, N_{t_n} - N_{t_{n-1}} = k_n) \\
= & P\left(\sum_{i=1}^k \mathbb{1}_{\{U_i \leq t_1\}} = k_1, \sum_{i=1}^k \mathbb{1}_{\{t_1 < U_i \leq t_2\}} = k_2, \dots, \sum_{i=1}^k \mathbb{1}_{\{t_{n-1} < U_i \leq t_n\}} = k_n\right) \times P(N_{t_n} = k) \\
= & \frac{k!}{k_1! k_2! \dots k_n!} \left(\frac{t_1}{t_n}\right)^{k_1} \left(\frac{t_2 - t_1}{t_n}\right)^{k_2} \dots \left(\frac{t_n - t_{n-1}}{t_n}\right)^{k_n} \frac{(\lambda t_n)^k}{k!} e^{-\lambda t_n} \\
= & \frac{(\lambda t_1)^{k_1}}{k_1!} e^{-\lambda t_1} \frac{(\lambda(t_2 - t_1))^{k_2}}{k_2!} e^{-\lambda(t_2 - t_1)} \dots \frac{(\lambda(t_n - t_{n-1}))^{k_n}}{k_n!} e^{-\lambda(t_n - t_{n-1})}.
\end{aligned}$$

3) On constate alors aisément que les v.a. $N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}}$ sont indépendantes et de loi de Poisson de paramètre respectivement $\lambda t_1, \lambda(t_2 - t_1), \dots, \lambda(t_n - t_{n-1})$.

Pour la condition suffisante, il n'y a rien à faire ! La condition nécessaire a entièrement spécifié la loi du processus N (i.e. pour tout n dans \mathbb{N} , tout (t_1, \dots, t_n) on peut donner la loi du vecteur $(N_{t_1}, N_{t_2}, \dots, N_{t_n})$) et on a vu qu'il était équivalent de se donner la loi du processus $(T_n)_{n \geq 1}$ que celle du processus $(N_t)_{t > 0}$. \square

DÉFINITION 3.6. *Un processus de comptage N est dit stationnaire et d'évènements rares de taux λ si pour tout $t > 0$ et $\Delta t > 0$ on a :*

$$\begin{aligned} P(N_{t+\Delta t} - N_t = 1) &= \lambda\Delta t + o(\Delta t) \\ \text{et } P(N_{t+\Delta t} - N_t > 1) &= o(\Delta t). \end{aligned}$$

Rappel : $o(\cdot)$ est une fonction telle que $\lim_{h \rightarrow 0} o(h)/h = 0$. ◇

THÉORÈME 3.7. *Soit N le processus de comptage d'un processus ponctuel sur \mathbb{R}^+ . C'est un processus de Poisson d'intensité λ si, et seulement si, le processus N est un processus stationnaire d'évènements rares à accroissements indépendants.*

Cette propriété d'évènements rares du processus de Poisson donne une autre interprétation du taux ou intensité associée à un processus de Poisson.

Preuve. Nous ne prouvons ici que la condition nécessaire, même si la condition suffisante n'est pas d'une grande difficulté.

D'après la proposition précédente, on sait que N est un processus à accroissements indépendants et que pour tout $t > 0$ et tout $\Delta t > 0$ la v.a. $N_{t+\Delta t} - N_t$ est de loi de Poisson $\mathcal{P}(\lambda\Delta t)$. D'où :

$$P(N_{t+\Delta t} - N_t = 1) = \lambda\Delta t e^{-\lambda\Delta t} = \lambda\Delta t(1 - \lambda\Delta t + o(\Delta t)) = \lambda\Delta t + o(\Delta t).$$

De même, on a :

$$\begin{aligned} P(N_{t+\Delta t} - N_t > 1) &= 1 - e^{-\lambda\Delta t} - (\lambda\Delta t)e^{-\lambda\Delta t} \\ &= 1 - (1 - \lambda\Delta t + o(\Delta t)) - (\lambda\Delta t + o(\Delta t)) = o(\Delta t). \end{aligned}$$

ce qui prouve la condition nécessaire. □

PROPOSITION 3.8. *Un processus de Poisson est un processus de Markov.*

Preuve. Nous aurions pu ici utiliser la remarque du chapitre 2 disant qu'un PAI est un processus de Markov. Nous allons cependant retrouver le résultat directement. Soit $t > 0$, $n \in \mathbb{N}$, un ensemble de n réels ordonnés $0 = t_0 < t_1 < \dots < t_n < t$ et n entiers strictement positifs $k_1 \leq \dots \leq k_n$. Par indépendance des accroissements d'un processus de Poisson, on a pour tout $k \geq k_n$:

$$\begin{aligned} &P(N_t = k | N_{t_1} = k_1, \dots, N_{t_n} = k_n) \\ &= P(N_t - N_{t_n} = k - k_n | N_{t_1} = k_1, N_{t_2} - N_{t_1} = k_2 - k_1, \dots, N_{t_n} - N_{t_{n-1}} = k_n - k_{n-1}) \\ &= P(N_t - N_{t_n} = k - k_n) = P(N_t - N_{t_n} = k - k_n | N_{t_n} = k_n) \\ &= P(N_t = k | N_{t_n} = k_n), \end{aligned}$$

ce qu'il fallait démontrer. □

3. Résultats asymptotiques

THÉORÈME 3.9. *Soit N un processus de Poisson d'intensité λ . On a alors :*

$$\begin{aligned} i) \quad &p.s. \lim_{t \rightarrow +\infty} \frac{N_t}{t} = \lambda \\ ii) \quad &\sqrt{t} \left(\frac{N_t}{t} - \lambda \right) \xrightarrow{\mathcal{L}} N(0, \lambda), \text{ quand } t \rightarrow +\infty. \end{aligned}$$

EXERCICE 3.4. *Démonstration de ce théorème.*

1) *En utilisant l'inégalité de Tchebychev, montrer que l'on a :*

$$\forall \varepsilon > 0 : P\left(\left|\frac{N_{k^2}}{k^2} - \lambda\right| \geq \varepsilon\right) \leq \frac{\lambda}{\varepsilon^2 k^2}.$$

2) *Utiliser alors le théorème de Borel-Cantelli pour en déduire que l'on a p.s. :*

$$\lim_{k \rightarrow +\infty} \frac{N_{k^2}}{k^2} = \lambda.$$

3) *En notant K la fonction définie par :*

$$K(t) = [\sqrt{t}],$$

où $[\cdot]$ est la partie entière, montrer que l'on peut écrire :

$$\frac{K^2(t)}{(K(t)+1)^2} \frac{N_{K^2(t)}}{K^2(t)} \leq \frac{N_t}{t} \leq \frac{(K(t)+1)^2}{K^2(t)} \frac{N_{(K(t)+1)^2}}{(K(t)+1)^2}$$

4) *En déduire alors que l'on a bien le i) du théorème.*

5) *Pour le ii), calculer la transformée de Laplace d'une loi de Poisson $\mathcal{P}(\lambda)$*

6) *En déduire que l'on peut écrire :*

$$\mathbb{E}\left(e^{u\sqrt{t}\left(\frac{N_t}{t} - \lambda\right)}\right) = \exp\left\{\frac{\lambda u^2}{2} + o(1)\right\}.$$

7) *En conclure le résultat ii) du théorème.*

Preuve/solution de l'exercice.

1) D'après l'inégalité de Tchebychev, on a pour tout $k \in \mathbb{N}$:

$$\forall \varepsilon > 0 : P(|N_{k^2} - \mathbb{E}N_{k^2}| \geq \varepsilon) \leq \frac{\text{Var}(N_{k^2})}{\varepsilon^2}.$$

Or, on sait que, toujours pour tout k dans \mathbb{N} , la v.a. N_{k^2} est de loi de Poisson $\mathcal{P}(\lambda k^2)$. Ainsi, il vient :

$$\begin{aligned} \forall \varepsilon > 0 : P(|N_{k^2} - \lambda k^2| \geq \varepsilon k^2) &\leq \frac{\lambda k^2}{\varepsilon^2 k^4} = \frac{\lambda}{\varepsilon^2 k^2} \\ \iff \forall \varepsilon > 0 : P\left(\left|\frac{N_{k^2}}{k^2} - \lambda\right| \geq \varepsilon\right) &\leq \frac{\lambda}{\varepsilon^2 k^2}. \end{aligned}$$

2) Du 1) on tire :

$$\sum_{k=0}^{+\infty} P\left(\left|\frac{N_{k^2}}{k^2} - \lambda\right| \geq \varepsilon\right) \leq \sum_{k=0}^{+\infty} \frac{\lambda}{\varepsilon^2 k^2} = \frac{\lambda}{\varepsilon^2} \sum_{k=0}^{+\infty} \frac{1}{k^2} < +\infty.$$

Ainsi, en notant $A_n = \{|\frac{N_{n^2}}{n^2} - \lambda| \geq \varepsilon\}$, on a d'après le théorème de Borel-Cantelli :

$$P(\overline{\lim}_n A_n) = 0 \text{ ou encore } P(\underline{\lim}_n A_n^c) = 1.$$

On a donc :

$$\begin{aligned}
& p.s. \varliminf_n A_n^c \iff p.s. : \cup_n \cap_{k \geq n} A_k^c \\
\iff & p.s. \exists n \text{ tel que, pour tout } k \geq n, \text{ on ait } A_k^c \\
\iff & p.s. \exists n \text{ tel que, pour tout } k \geq n, \text{ on ait } \left| \frac{N_{k^2}}{k^2} - \lambda \right| < \varepsilon \\
\iff & p.s. \lim_{k \rightarrow +\infty} \frac{N_{k^2}}{k^2} = \lambda.
\end{aligned}$$

3) Notons maintenant K la fonction définie par :

$$K(t) = [\sqrt{t}],$$

où $[\cdot]$ est la partie entière. Par croissance du processus de comptage on a, puisque $K^2(t) \leq t < (K(t) + 1)^2$:

$$N_{K^2(t)} \leq N_t \leq N_{(K(t)+1)^2}$$

D'où :

$$\begin{aligned}
& \frac{N_{K^2(t)}}{(K(t) + 1)^2} \leq \frac{N_t}{t} \leq \frac{N_{(K(t)+1)^2}}{K^2(t)} \\
\iff & \frac{K^2(t)}{(K(t) + 1)^2} \frac{N_{K^2(t)}}{K^2(t)} \leq \frac{N_t}{t} \leq \frac{(K(t) + 1)^2}{K^2(t)} \frac{N_{(K(t)+1)^2}}{(K(t) + 1)^2}
\end{aligned}$$

4) Comme on a :

$$\frac{K^2(t)}{(K(t) + 1)^2} \rightarrow 1, \text{ lorsque } t \rightarrow +\infty$$

et

$$p.s. : \lim_{k \rightarrow +\infty} \frac{N_{k^2}}{k^2} = \lambda,$$

l'inégalité obtenue en 3) nous assure que :

$$p.s. \lim_{t \rightarrow +\infty} \frac{N_t}{t} = \lambda.$$

5) La transformée d'une loi de Poisson de paramètre λ est :

$$\mathcal{L}(u) = \exp\{\lambda(e^u - 1)\}.$$

6) On a :

$$\mathbb{E} \left(e^{u\sqrt{t}(\frac{N_t}{t} - \lambda)} \right) = e^{-u\lambda\sqrt{t}} \mathbb{E} \left(e^{\frac{u}{\sqrt{t}} N_t} \right) = \exp\{-u\lambda\sqrt{t} + \lambda t(e^{\frac{u}{\sqrt{t}}} - 1)\}.$$

Or, le développement limité de l'exponentielle nous donne :

$$e^{u/\sqrt{t}} - 1 = \frac{u}{\sqrt{t}} + \frac{u^2}{2t} + o\left(\frac{1}{t}\right).$$

D'où :

7) en faisant tendre $t \rightarrow +\infty$, on obtient :

$$\mathbb{E} \left(e^{u\sqrt{t}(\frac{N_t}{t} - \lambda)} \right) = \exp \left\{ \frac{\lambda u^2}{2} + o(1) \right\} \xrightarrow{t \rightarrow +\infty} e^{\frac{\lambda u^2}{2}}.$$

On reconnaît à la limite la transformée de Laplace d'une loi $N(0, \lambda)$, ce qui achève la démonstration du théorème. \square

4. Bibliographie

- Pierre Brémaud. *Introduction aux Probabilités : modélisation des phénomènes aléatoires*. Springer 1997.
- Christiane Coccozza-Thivent. *Processus stochastiques et fiabilité des systèmes*. Springer 1997.
- Paul S. toulouse. *Thèmes de Probabilités et Statistique*. Dunod 1999.

CHAPITRE 4

Chaînes de Markov

1. Définition et matrice de transition

Dans le Chapitre 2, nous avons donné la définition générale d'un processus de Markov. Nous allons ici nous intéresser de près au cas où l'espace des états et le temps sont discrets. On parle alors de chaînes de Markov à espace des états discret.

La définition vue au Chapitre 2 peut être alors réécrite de la manière suivante.

DÉFINITION 4.1. Soit (Ω, \mathcal{A}, P) un espace probabilisé. Soit $(X_n)_{n \in \mathbb{N}}$ une famille de v.a. de (Ω, \mathcal{A}, P) et à valeurs dans (E, \mathcal{E}) où E est un espace dénombrable et $\mathcal{E} = \mathcal{P}(E)$.

On dit que $(X_n)_{n \in \mathbb{N}}$ est une **chaîne de Markov** à valeurs dans l'espace d'état E si, pour tout $n \geq 1$ et toute suite $(i_0, i_1, \dots, i_{n-1}, i, j)$ d'éléments de E telle que $P(X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0)$ soit strictement positive, on ait :

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0) = P(X_{n+1} = j | X_n = i).$$

On dit que le futur du processus ne dépend du passé qu'à travers le présent.

L'espace E sera souvent \mathbb{Z} ou \mathbb{N} ou une partie de ceux-ci. Mais, afin de garder le maximum de généralité, nous ne spécifierons pas, sauf nécessaire, l'espace E . Notons

$$p_{i,j}^{n,n+1} = P(X_{n+1} = j | X_n = i).$$

la probabilité de passage de l'état i à l'état j entre les instants n et $n+1$ et $P^{n,n+1} = (p_{i,j}^{n,n+1})_{(i,j) \in E^2}$ la matrice finie ou dénombrable (suivant que E est fini ou dénombrable) de toutes les probabilités de transition en une étape à l'instant n . Cette matrice $P^{n,n+1}$ est appelée matrice de transition en une étape à l'instant n .

Une telle notation permet de prendre en compte les situations où les probabilités de passage d'un état i quelconque à un état j quelconque dépendent des instants (n et $n+1$) où on les considère. Même si de telles situations peuvent se rencontrer dans la pratique, nous nous concentrerons davantage dans la suite sur celles où ces probabilités ne dépendent pas de l'instant n . On parle alors de chaîne de Markov homogène.

DÉFINITION 4.2. Une chaîne de Markov est dite **homogène** si sa matrice de transition $P^{n,n+1}$ (en une étape à l'instant n) ne dépend pas de n , i.e. si, pour tout $n \geq 0$ et tout (i, j) dans E^2 , on a :

$$p_{i,j}^{n,n+1} = P(X_{n+1} = j | X_n = i) = p_{ij}.$$

La matrice des probabilités de transition en une étape est alors notée $P = (p_{i,j})_{(i,j) \in E^2}$ et est simplement appelée **matrice de passage ou de transition** de la chaîne.

PROPOSITION 4.3. Toute matrice de transition $P = (p_{i,j})_{(i,j) \in E^2}$ vérifie les propriétés suivantes :

$$\begin{aligned} p_{i,j} &\geq 0, \forall (i, j) \in E^2; \\ \sum_{j \in E} p_{i,j} &= 1, \forall i \in E. \end{aligned}$$

EXERCICE 4.1. Démontrer cette proposition.

Preuve/solution de l'exercice. Le premier point découle trivialement du fait que les p_{ij} sont des probabilités et le second de l'égalité

$$\sum_{j \in E} p_{i,j} = P(X_{n+1} \in E | X_n = i) = 1$$

pour tout i dans E . □

Terminologie : Une matrice vérifiant les assertions de la propriété 4.3 est appelée **matrice stochastique**.

PROPOSITION 4.4. *Une chaîne de Markov est entièrement déterminée par la donnée de sa matrice de transition P et sa loi initiale π_0 .*

EXERCICE 4.2. *Démontrer cette proposition. (Ind. on pourra utiliser les résultats de la section 1.3 du Chapitre 2)*

Preuve/solution de l'exercice. Pour tout $n \geq 1$ et tout (i_0, i_1, \dots, i_n) élément de E^n on a, grâce à la propriété de Markov :

$$\begin{aligned} & P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) \\ &= P(X_n = i_n | X_{n-1} = i_{n-1}) \cdots P(X_1 = i_1 | X_0 = i_0) P(X_0 = i_0) \\ &= p_{i_{n-1}, i_n} \cdots p_{i_0, i_1} \pi_0. \end{aligned}$$

Il est aisé de voir que ce résultat permet alors de déterminer toute probabilité faisant intervenir les états du processus X_{j_1}, \dots, X_{j_k} aux instants quelconques $j_1 < \dots < j_k$, pour tout $k \geq 1$. Le théorème de Kolmogorov vu au Chapitre 2 permet de conclure. □

Une chaîne de Markov est généralement représentée schématiquement par un **graphe de Markov**, dont les sommets sont les différents états possibles de la chaîne. Deux sommets i et j sont alors reliés par une flèche, avec l'étiquette $p_{i,j}$, allant de i à j quand la chaîne peut passer de l'état i à l'état j en une étape, i.e. quand $p_{i,j} > 0$.

2. Exemples classiques de Chaînes de Markov

Nous allons détailler quelques exemples de chaînes de Markov. Nous nous attarderons bien sûr pas sur la suite de v.a. indépendantes et à valeurs dans E qui constitue trivialement une chaîne de Markov.

2.1. La chaîne à deux états. La chaîne à deux états est sans aucun doute le cas le plus simple de chaîne de Markov. Notons, par exemple, par 1 et 2 ses états et sa matrice de transition est

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

avec p et q tous deux dans $[0, 1]$. Le graphe, très simple, d'une telle chaîne est celui de la Figure 1.

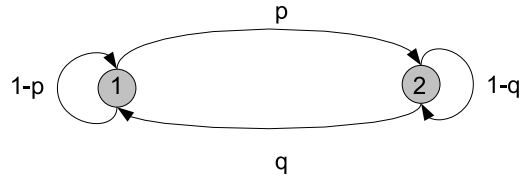


FIGURE 1. Chaîne de Markov à deux états

2.2. La marche aléatoire unidimensionnelle.

EXEMPLE 2.1. *La marche aléatoire simple.*

Rappelons qu'au Chapitre 2, nous avons défini la **marche aléatoire simple** comme un processus stochastique X à espace des temps \mathbb{N} , espace d'états \mathbb{Z} et défini de la manière suivante : $X_0 = 0$ et $Z_n = X_n - X_{n-1}$ est de loi $p\delta_1 + (1-p)\delta_{-1}$, où $p \in]0, 1[$ et la famille des v.a. $(Z_n)_{n \in \mathbb{N}}$ est une famille indépendante. La propriétés de Markov d'un tel processus est trivialement vérifiée puisque le processus est à accroissements indépendants. L'espace des états d'une telle chaîne est \mathbb{Z} et sa matrice de transition est telle que, pour tout $(i, j) \in \mathbb{Z}^2$, on ait :

$$p_{i,j} = \begin{cases} p & \text{si } j = i + 1; \\ 1 - p & \text{si } j = i - 1; \\ 0 & \text{sinon.} \end{cases}$$

Le graphe de la marche aléatoire simple est donné dans la figure 2.

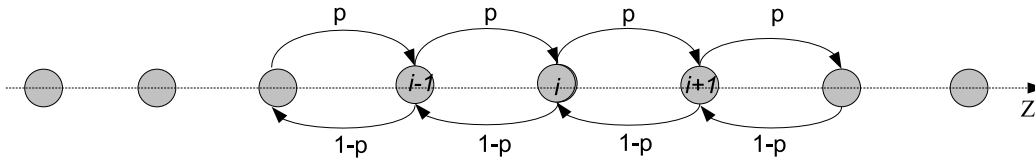


FIGURE 2. Marche aléatoire simple

Une telle chaîne de Markov peut, par exemple, modéliser le déplacement d'une puce sur une échelle horizontale.

Une marche aléatoire simple est dite *symétrique* si $p = 1/2$. ◇

EXEMPLE 2.2. *La marche aléatoire unidimensionnelle.*

Une généralisation naturelle de la marche aléatoire simple est la marche aléatoire **unidimensionnelle**, avec espace des états \mathbb{Z} , telle que si à l'instant n la chaîne est dans l'état i alors à l'instant $n + 1$ elle ne peut être que dans les états $i - 1, i$ et $i + 1$ avec les probabilités q_i, r_i et p_i respectivement, où les réels q_i, r_i et p_i sont positifs et tels que $q_i + r_i + p_i = 1$, pour tout i dans \mathbb{Z} . On a donc :

$$\begin{aligned} P(X_{n+1} = i - 1 | X_n = i) &= q_i; \\ P(X_{n+1} = i | X_n = i) &= r_i; \\ P(X_{n+1} = i + 1 | X_n = i) &= p_i. \end{aligned}$$

La matrice de transition est alors :

$$P = \begin{pmatrix} \cdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots \\ \cdots & 0 & q_i & r_i & p_i & 0 & \cdots & \cdots \\ \cdots & \cdots & 0 & q_{i+1} & r_{i+1} & p_{i+1} & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

Le graphe de la marche aléatoire unidimensionnelle est donné dans la Figure 3 :

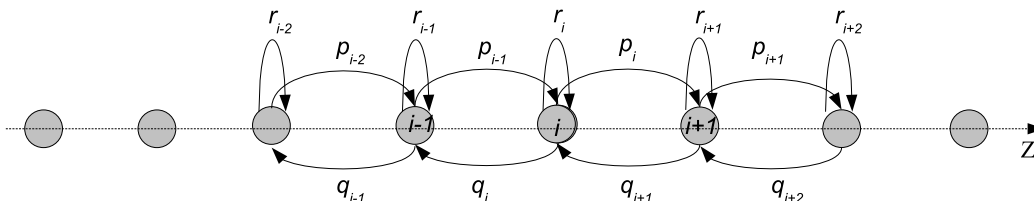


FIGURE 3. Marche aléatoire unidimensionnelle

Si l'on reprend l'exemple de la puce, cela revient à lui donner un droit de repos et également de lui permettre de changer de règle de conduite en fonction de sa position sur l'échelle.

EXEMPLE 2.3. *Fortune d'un joueur (seul contre la banque)*

Un cas particulier de la marche aléatoire que nous venons de présenter est donné par la modélisation des gains d'un joueur contre une banque que l'on suppose de capital infini.

EXERCICE 4.3. *Considérons la fortune d'un joueur à une partie de hasard. La règle du jeu est supposée telle que, si la fortune du joueur à un instant n est k alors à l'instant $n + 1$ elle est augmentée de 1 avec une probabilité p_k et diminuée de 1, avec une probabilité $q_k = 1 - p_k$. On suppose de plus qu'une fois le joueur ruiné la partie est terminée.*

1) *Donner l'espace des états de la chaîne de Markov modélisant cette partie.*

2) *Quelle est sa matrice de transition ?*

3) *Quel est son graphe ?*

4) *Que dire de l'état 0 ?*

Solution de l'exercice.

- 1) C'est une marche aléatoire unidimensionnelle sur \mathbb{N} avec de plus $q_0 = 0$.
 2) La matrice de passage est :

$$P = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ q_1 & 0 & p_1 & 0 & \cdots & \cdots & \cdots & \vdots \\ 0 & q_2 & 0 & p_2 & 0 & \cdots & \cdots & \vdots \\ 0 & 0 & q_3 & 0 & p_3 & 0 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & 0 & q_i & 0 & p_i & 0 & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

- 3) Le graphe d'une telle marche aléatoire est donné dans la Figure 4:

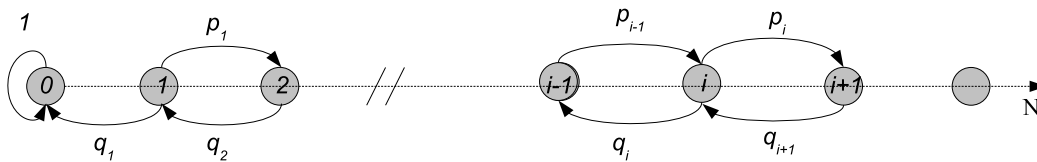


FIGURE 4. Diagramme de la chaîne de Markov modélisant les gains d'un joueur seul contre la banque

- 4) L'état 0 est alors dit **absorbant**, car une fois atteint cet état, on ne peut plus le quitter. \diamond

EXEMPLE 2.4. Fortune d'un joueur A contre un joueur B

EXERCICE 4.4. *Considérons maintenant la situation d'une partie entre deux joueurs A et B dont la somme de leurs fortunes est égale à a euros. À chaque partie, le joueur A gagne un euro de son adversaire avec une probabilité p et donne un euro à B avec une probabilité $q = 1 - p$. Le jeu s'arrête alors dès que l'un des deux joueurs est ruiné.*

On note X_n la fortune du joueur A à l'instant n (celle de B est bien sûr $a - X_n$). On considère la chaîne de Markov $(X_n)_{n \in \mathbb{N}}$.

1) *Se convaincre qu'il s'agit bien d'une chaîne de Markov. Donner l'espace de ses états.*

2) *Donner sa matrice de transition.*

3) *Donner le graphe de cette chaîne.*

4) *Que dire des états 0 et a ?*

Solution de l'exercice.

- 1) l'espace des états est $E = \{0, 1, \dots, a\}$.
- 2) La matrice de transition est :

$$P = \begin{pmatrix} 1 & 0 & \dots & \dots & \dots & \dots & 0 \\ q & 0 & p & 0 & \dots & \dots & \vdots \\ 0 & q & 0 & p & 0 & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & q & 0 & p & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & q & 0 & p \\ 0 & \vdots & \vdots & \vdots & \vdots & 0 & 1 \end{pmatrix}$$

- 3) Le graphe d'une telle marche aléatoire est donné dans la Figure 5 :

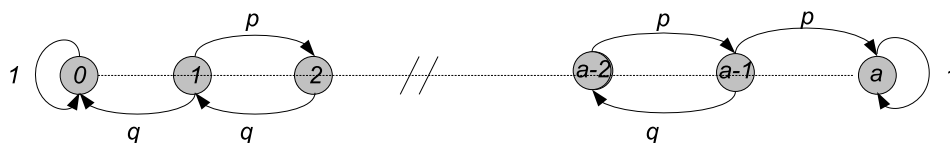


FIGURE 5. Diagramme de la chaîne de Markov modélisant les gains d'un joueur A contre un joueur B

- 4) Les extrémités de l'espace des états E , i.e. les états 0 et a sont appelés **barrières** et sont dans ce cas des barrières absorbantes. \diamond

EXEMPLE 2.5. Modèle de diffusion d'Ehrenfest

On peut bien sûr trouver des applications des marches aléatoires dans bien d'autres domaines que celui de la théorie des jeux. Ainsi en est-il du modèle de diffusion d'Ehrenfest, utilisé pour modéliser la diffusion d'un gaz à travers une membrane séparant deux récipients A et B . Ce modèle est basé sur le même principe que celui du tirage dans deux urnes A et B , contenant à elles deux a boules. Nous entamons son étude dans l'exercice suivant.

EXERCICE 4.5. *On considère donc deux urnes A et B , contenant à elles deux a boules. À chaque instant, le numéro d'une des boules est choisi au hasard et celle qui porte ce numéro est alors changée d'urne. On note par X_n le nombre de boules que contient l'urne A à l'instant n . Le processus $(X_n)_{n \in \mathbb{N}}$ est alors une marche aléatoire à espaces des états $E = \{0, 1, \dots, a\}$.*

- 1) Donner sa matrice de transition.
 - 2) Donner son graphe.
 - 3) Que dire des états 0 et a ?
-

Solution de l'exercice.

1) Sa matrice de transition est :

$$P = \begin{pmatrix} 0 & 1 & 0 & \dots & \dots & \dots & \dots & 0 \\ 1/a & 0 & (a-1)/a & 0 & \dots & \dots & \dots & \vdots \\ 0 & 2/a & 0 & (a-2)/a & 0 & \dots & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & (a-1)/a & 0 & 1/a \\ 0 & \dots & \dots & \dots & \dots & \dots & 1 & 0 \end{pmatrix}$$

2) Le graphe du modèle de diffusion d'Erhenfest est donné dans la Figure 6 :

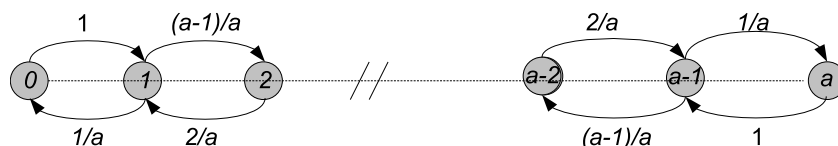


FIGURE 6. Diagramme du Modèle de diffusion d'Ehrenfest

3) Les barrières 0 et a ne sont plus ici absorbantes mais **réfléchissantes**, i.e. quand la chaîne atteint l'état 0, elle provient de l'état 1 et y retourne juste après. De même quand on arrive à l'état a , uniquement possible en passant juste avant par $a-1$, alors on retourne de manière sûre à l'état $a-1$. De plus on constate aisément que, dans un tel modèle, les transferts sont plus probables dans le sens du récipient le plus plein vers celui qui l'est moins. \diamond

2.3. La marche aléatoire multidimensionnelle. La marche aléatoire multidimensionnelle symétrique est une chaîne de Markov à espace d'états $E = F^n$, où F est tout ou partie de \mathbb{Z} , et telle que pour tout n -uplé $\mathbf{k} = (k_1, \dots, k_n)$ et $\mathbf{l} = (l_1, \dots, l_n)$ de E , on ait

$$P(X_{m+1} = \mathbf{l} | X_m = \mathbf{k}) = p_{\mathbf{k}, \mathbf{l}} = \begin{cases} \frac{1}{2n} & \text{si } \sum_{i=1}^n |l_i - k_i| = 1; \\ 0 & \text{sinon.} \end{cases}$$

Le graphe d'une telle marche aléatoire spatiale est alors celui de la Figure 7:

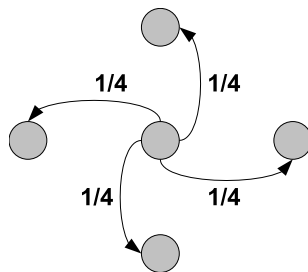


FIGURE 7. Diagramme de la marche aléatoire simple symétrique en dimension 2

2.4. La file d'attente à temps discret discret. Une file d'attente est utilisée pour modéliser le nombre de clients en attente d'un service à un guichet ou à une caisse de supermarché, ou bien le nombre de requêtes en attente pour un serveur informatique, etc...

Nous allons voir que, dans le cas où le temps est discrétisé, ou si l'on raisonne par unité de temps, alors on obtient une chaîne de Markov. Utilisons l'interprétation en termes de clients et guichet pour présenter davantage ce modèle.

On suppose qu'à chaque unité de temps des clients arrivent, en nombre aléatoire, à un guichet et que celui ne peut servir qu'un client à la fois avec un temps de service d'une unité de temps. Notons Y_n le nombre de clients arrivant au n ème instant. On suppose que la famille des (Y_n) est indépendante et identiquement distribuées de loi donnée, pour tout n dans \mathbb{N} , par

$$P(Y_n = k) = a_k \geq 0 \quad \forall k \in \mathbb{N}$$

$$\sum_{k=0}^{+\infty} a_k = 1.$$

Quand aucun client n'est en attente, aucun service n'est effectué. On note par X_n le nombre de clients en attente à chaque instant n et si celui-ci est égal à i alors X_{n+1} sera égal $i - 1 + Y_n$ si $i \geq 1$ et à Y_n si $i = 0$. On peut résumer ceci de la manière suivante :

$$X_{n+1} = (X_n - 1)^+ + Y_n.$$

Ainsi la matrice de transition d'une telle file d'attente à temps discret est :

$$P = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \cdots & \cdots \\ a_0 & a_1 & a_2 & a_3 & \cdots & \cdots \\ 0 & a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & 0 & a_0 & a_1 & a_2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

Il n'est bien sûr pas possible de faire un graphe pour une telle chaîne de Markov, sauf pour des cas particuliers très simples. Intuitivement il apparaît assez facilement que si l'espérance du nombre de clients arrivant par unité de temps est strictement supérieure à 1 (i.e. $\mathbb{E}Y_n > 1$) alors la taille de la file d'attente s'accroîtra de manière illimitée. En revanche, si cette espérance est strictement inférieure à 1, alors on verra que la taille de la file d'attente s'approchera d'un équilibre avec le temps. La situation où cette espérance est égale à 1 est, nous le verrons, source de grande instabilité du système.

2.5. Un modèle de gestion des stocks. Dans un service de gestion des stocks, il arrive à chaque unité de temps une demande d'une quantité Y_n du produit géré (ce peut être par exemple le cumul de toutes les demandes écoulées dans l'heure qui précède). On suppose que les Y_n , pour $n \in \mathbb{N}$, sont indépendants et identiquement distribués de loi générale

$$P(Y_n = k) = a_k, \quad \forall k \in \mathbb{N},$$

avec bien sûr les $a_k \geq 0$ et tels que $\sum_{k=0}^{+\infty} a_k = 1$. On note par X_n l'état du stock à l'instant n que l'on suppose toujours inférieur à un entier $S > 0$ fixé (ce peut être par exemple la capacité maximale de stockage). Les X_n sont alors à valeurs dans $S - \mathbb{N}$, où une valeur négative signifie que le service de gestion n'a pas pu honorer la demande et est donc en dette du montant $-X_n$. Après chaque réponse à une demande, est fait un inventaire dont la police, basée sur le choix d'un entier $s < S$ fixé, est la suivante. Si la quantité de stock est inférieure ou égale à s alors le stock est réapprovisionné de manière à annuler la dette (si elle a lieu) et remettre l'état du

stock à S . Si la quantité de stock est comprise entre s et S , aucun réapprovisionnement n'est effectué. On a alors le lien suivant entre les quantités demandées et l'état du stock :

$$X_{n+1} = \begin{cases} X_n - Y_{n+1} & \text{si } s < X_n \leq S; \\ S - Y_{n+1} & \text{si } X_n \leq s. \end{cases}$$

La matrice de transition est alors :

$$P = (p_{i,j})_{(i,j) \in S-\mathbb{N} \times S-\mathbb{N}}$$

$$= \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \cdots & \cdots & \cdots & \cdots \\ 0 & a_0 & a_1 & a_2 & a_3 & \cdots & \cdots & \cdots \\ 0 & 0 & a_0 & a_1 & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & a_0 & a_1 & a_2 & a_3 & \cdots \\ a_0 & a_1 & a_2 & a_3 & \cdots & \cdots & \cdots & \cdots \\ a_0 & a_1 & a_2 & a_3 & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

2.6. Un exemple de chaîne de Markov en génétique. On se propose de modéliser les variations dans la représentation d'un gène dans une population. Par souci de simplicité, on ne s'intéresse qu'à un modèle haploïde simple d'une population de taille fixée à $2N$ gènes et que chaque gène est de type a ou de type A . La répartition des gènes dans la population est déterminée par la règle suivante. Si dans la population parente, on dénombre i gènes de type a et $2N - i$ de type A , alors chaque gène dans la population suivante est de type a avec la probabilité $p_i = i/2N$ et de type A avec la probabilité $q_i = 1 - p_i$ et cela de manière indépendante pour chaque gène. Si on note X_n le nombre de gènes de type a dans la n ième population alors, le processus $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov à espace des états $E = \{0, 1, \dots, 2N\}$. Les probabilités de transition sont alors de la forme :

$$p_{i,j} = P(X_{n+1} = j | X_n = i) = C_{2N}^j p_i^j q_i^{2N-j}, \quad \forall (i, j) \in E^2.$$

On note que bien sûr les états 0 et $2N$ sont absorbants. D'autres modèles ont été proposés, particulièrement en proposant d'autres expressions pour les probabilités p_i et q_i . On pourra consulter, entre autres, le livre de Karlin et Taylor.

3. Équations de Chapman-Kolmogorov

Dans la première section nous avons uniquement parlé des probabilités de transition en une étape d'un état à un autre. Bien sûr, on peut s'intéresser aux probabilités de passer en deux, trois ou plus étapes d'un état à un autre. Notons ainsi P_n la **matrice des probabilités de transition en n étapes** dont le terme général est, pour tout $(i, j) \in E^2$:

$$p_{i,j}^{(n)} = P(X_{m+n} = j | X_m = i) = P(X_n = j | X_0 = i), \quad \forall m \in \mathbb{N}$$

où la dernière égalité est justifiée par l'homogénéité de la chaîne.

L'**équation de Chapman-Kolmogorov** permet d'exprimer la matrice P_n en fonction de la matrice de transition P .

THÉORÈME 4.5. (Chapman-Kolmogorov) On a, pour tout n dans \mathbb{N} :

$$P_n = P^n.$$

Le produit matriciel est un outil bien connu pour des matrices de taille finie. On étend sans difficulté sa définition à des matrices infinies.

EXERCICE 4.6. *Démontrer ce théorème. (Ind. On pourra raisonner par récurrence)*

Preuve/Solution de l'exercice. La formule est triviale pour $n = 0$ puisque $P^0 = I$. Elle est également vraie pour $n = 1$. Supposons maintenant qu'elle soit vraie jusqu'au rang n . L'homogénéité de la chaîne nous permet alors d'écrire, pour tout $(i, j) \in E^2$:

$$\begin{aligned} p_{i,j}^{(n+1)} &= P(X_{n+1} = j | X_0 = i) = \sum_{k \in E} P(X_{n+1} = j, X_n = k | X_0 = i) \\ &= \sum_{k \in E} P(X_{n+1} = j | X_n = k) P(X_n = k | X_0 = i) = \sum_{k \in E} p_{k,j} p_{i,k}^{(n)}. \end{aligned}$$

On a donc, en prenant l'écriture matricielle et en utilisant l'hypothèse de récurrence,

$$P_{n+1} = P_n P = P^n P = P^{n+1},$$

ce qui prouve bien le résultat. □

COROLLAIRE 4.6. *On a pour tout n et m dans \mathbb{N} ,*

$$P_{m+n} = P_m P_n = P_n P_m,$$

c'est à dire que pour tout (i, j) dans E^2 , on a :

$$p_{i,j}^{(m+n)} = \sum_{k \in E} p_{i,k}^{(m)} p_{k,j}^{(n)} = \sum_{k \in E} p_{i,k}^{(n)} p_{k,j}^{(m)}.$$

Preuve. Cela découle directement de l'associativité du produit matriciel. En effet, pour $(m, n) \in \mathbb{N}^2$, on a

$$P_{m+n} = P^{m+n} = P^m P^n = P_m P_n$$

et cette formule est bien évidemment symétrique en m et n . □

PROPOSITION 4.7. *Pour tout entier n positif, la matrice de transition en n étapes P_n est également une matrice stochastique.*

Preuve. Soit $n \in \mathbb{N}$. Les éléments de la matrice P_n sont bien sûr positifs puisqu'il s'agit de probabilités et on a de plus :

$$\sum_{j \in E} p_{i,j}^{(n)} = \sum_{j \in E} P(X_n = j | X_0 = i) = P(X_n \in E | X_0 = i) = 1,$$

ce qui prouve bien la proposition. □

Notons que la matrice stochastique P_n est la matrice de transition (en une étape) de la chaîne de Markov $(X_{nm})_{m \in \mathbb{N}}$.

4. Quelques formules de conditionnement

Notons (\mathcal{F}_n) la filtration naturelle associée au processus (X_n) (Cf. Chapitre 2), i.e., pour tout n :

$$\mathcal{F}_n = \sigma(X_k; k = 0, \dots, n).$$

On peut montrer que tout événement de la tribu \mathcal{F}_n est une réunion dénombrable d'événements disjoints de cette même tribu et de la forme $\{X_0 = i_0, \dots, X_n = i_n\}$. Ainsi, tout événement A de \mathcal{F}_n est tel que l'événement $A \cap \{X_n = i\}$ reste dans la tribu \mathcal{F}_n . Si ce dernier n'est pas l'ensemble vide alors on a forcément $A \subset \{X_n = i\}$, ce qui revient à dire que le présent est donc fixé.

PROPOSITION 4.8. *Pour tout entier n et tout événement A de la tribu \mathcal{F}_n , on a :*

$$P(X_{n+1} = j | A, X_n = i) = P(X_{n+1} = j | X_n = i) = p_{i,j},$$

dès que $P(A, X_n = i) > 0$.

EXERCICE 4.7. *Démonstration de cette proposition.*

1) *Démontrer d'abord le cas (très facile) $n = 1$.*

2) *On suppose maintenant $n > 0$. Démontrer (proprement !) que l'on a :*

$$P(X_{n+1} = j, \cup_{k \in K} B_k, X_n = i) = p_{i,j} P(\cup_{k \in K} B_k, X_n = i).$$

(Ind. On pourra utiliser le résultat rappelé précédemment à savoir que tout événement A de la tribu \mathcal{F}_n est une réunion dénombrable, sur un ensemble d'indices que l'on va noter K , d'événements disjoints de la forme $B_k = \{X_0 = i_0^k, \dots, X_n = i_n^k\}$, pour $k \in K$)

Preuve/Solution de l'exercice.

1) Il n'y a rien à montrer pour $n = 0$, puisque dans ce cas, l'événement $A \cap \{X_0 = i\}$ n'est pas vide que si $A = \{X_0 = i\}$ et le résultat de la proposition n'est que la définition de $p_{i,j}$.

2) Supposons maintenant $n > 0$. On a donc, pour tout A de la tribu \mathcal{F}_n , l'existence d'un ensemble dénombrable d'indices K et d'une famille d'événements B_k , pour $k \in K$, tels que :

$$A = \cup_{k \in K} B_k.$$

Or, pour tout k dans K , si l'on a

$$P(B_k, X_n = i) = P(X_0 = i_0^k, \dots, X_n = i_n^k, X_n = i) > 0$$

(ce qui implique que $i_n^k = i$!) alors l'égalité de la proposition avec $A = B_k$ n'est rien d'autre que la propriété de Markov.

On peut alors écrire, pour tout $(i, j) \in E^2$:

$$\begin{aligned}
 P(X_{n+1} = j, \cup_{k \in K} B_k, X_n = i) &= \sum_{k \in K} P(X_{n+1} = j, B_k, X_n = i) \\
 &= \sum_{k \in K} P(X_{n+1} = j | B_k, X_n = i) P(B_k, X_n = i) \\
 &= p_{i,j} \sum_{k \in K} P(B_k, X_n = i) \\
 &= p_{i,j} P(\cup_{k \in K} B_k, X_n = i).
 \end{aligned}$$

La dernière probabilité étant strictement positives puisqu'elle l'est pour chaque B_k , on a donc démontré que :

$$p_{i,j} = P(X_{n+1} = j | \cup_{k \in K} B_k, X_n = i) = P(X_{n+1} = j | A, X_n = i),$$

ce qui était le résultat cherché. \square

Il est important de noter que le résultat de la proposition n'est vrai qu'en prenant des événements du présent de la forme $X_n = i$ et non pour des événements de forme générale $X_n \in B$. On trouvera dans le livre de Foata et Fuchs, dont cette section est largement inspirée, un contre-exemple (*Processus Stochastiques*, Dunod 2002, p. 73). On notera également que l'on pourrait par le même genre de raisonnement montrer que ce type de résultat reste vrai pour des événements du futur prenant en compte d'autres temps que $n+1$. Ainsi par exemple, on a :

$$\begin{aligned}
 &P(X_{n+1} = j_{n+1}, \dots, X_{n+r} = j_{n+r} | A, X_n = i) \\
 &= P(X_{n+1} = j_{n+1}, \dots, X_{n+r} = j_{n+r} | X_n = i)
 \end{aligned}$$

ou encore

$$P(X_{n+r} = j | A, X_n = i) = P(X_{n+r} = j | X_n = i) = p_{i,j}^{(r)}.$$

5. Classification des états

DÉFINITION 4.9. On dit que l'état j est **accessible** à partir de i si il existe un entier n tel que $p_{i,j}^{(n)} > 0$ et l'on note $i \longrightarrow j$.

EXERCICE 4.8. On reprend l'exemple de modélisation de la fortune du joueur A contre le joueur B (Cf. Exemple 2.4). Étudier l'accessibilité entre les états en prouvant vos affirmations.

Solution de l'exercice. Tous les états sont accessibles depuis $\{1, \dots, a-1\}$ mais seulement 0 est accessible depuis 0 et seulement a est accessible depuis a . En effet, par exemple, $3 \longrightarrow 1$ puisque

$$P(X_2 = 1 | X_0 = 3) = P(X_2 = 1, X_1 = 2 | X_0 = 3) = p_{2,1} p_{3,2} = q^2 > 0.$$

En revanche, $0 \not\rightarrow 1$ car $P(X_n = 1 | X_0 = 0) = 0$, quelque soit n . \diamond

PROPOSITION 4.10. Une condition nécessaire et suffisante pour que $i \longrightarrow j$ est que $P(\exists n : X_n = j | X_0 = i) > 0$

EXERCICE 4.9. *Démontrer cette proposition.*

Preuve/Solution de l'exercice. Comme on a, pour tout n dans \mathbb{N} :

$$\begin{aligned} P(\exists m \geq 0 : X_m = j | X_0 = i) &= P(\cup_{m \in \mathbb{N}} \{X_m = j\} | X_0 = i) \\ &\geq P(\{X_n = j\} | X_0 = i), \end{aligned}$$

la condition est bien nécessaire.

Réciproquement, comme on a :

$$P(\exists n \geq 0 : X_n = j | X_0 = i) = P(\cup_{n \in \mathbb{N}} \{X_n = j\} | X_0 = i) \leq \sum_{n \in \mathbb{N}} p_{i,j}^{(n)},$$

elle est bien suffisante. □

PROPOSITION 4.11. *La relation d'accessibilité est réflexive et transitive.*

EXERCICE 4.10. *Démontrer cette proposition.*

Preuve/Solution de l'exercice. On a bien sûr $i \longrightarrow i$ puisque $p_{i,i}^{(0)} = P(X_0 = i | X_0 = i) = 1$. Par ailleurs, supposons que pour des états i, j et k on ait : $i \longrightarrow j$ et $j \longrightarrow k$. Alors, d'après la définition, il existe m et n , entiers positifs, tels que : $p_{i,j}^{(m)} > 0$ et $p_{j,k}^{(n)} > 0$. D'où, en utilisant le corollaire 4.6, il vient :

$$p_{i,k}^{(m+n)} = \sum_{l \in E} p_{i,l}^{(m)} p_{l,k}^{(n)} \geq p_{i,j}^{(m)} p_{j,k}^{(n)} > 0,$$

ce qui prouve bien que l'on a : $i \longrightarrow k$ et la relation est bien transitive. □

EXERCICE 4.11. *Montrer que la relation d'accessibilité n'est pas symétrique en trouvant un contre-exemple par exemple dans l'Exemple 2.4*

Solution de l'exercice. On a vu que $1 \longrightarrow 0$ mais que $0 \not\rightarrow 1$. ◇

DÉFINITION 4.12. *On dit que deux états i et j d'une chaîne de Markov **communiquent**, si $i \longrightarrow j$ et $j \longrightarrow i$. On note $i \longleftrightarrow j$.*

PROPOSITION 4.13. *La relation de communication entre états est une relation d'équivalence. On a donc :*

i) (Réflexive) Tout état i de la chaîne communique avec lui même, i.e. $i \longleftrightarrow i$;

- ii) (*Symétrique*) Si un état i communique avec un état j , alors la réciproque est vraie, i.e. $i \longleftrightarrow j \iff j \longleftrightarrow i$;
- iii) (*Transitive*) Si un état i communique avec un état j qui lui même communique avec un état k , alors l'état i communique avec l'état k , i.e. si $i \longleftrightarrow j$ et $j \longleftrightarrow k$ alors $i \longleftrightarrow k$.

Preuve. Les propriétés de réflexivité et de transitivité sont obtenues avec les résultats obtenus précédemment sur la relation d'accessibilité (Cf. proposition 4.11). La symétrie découle directement de la définition de la relation de communication. \square

On a vu que tout état d'une chaîne de Markov communique avec lui même, puisque l'on a $p_{i,i}^{(0)} = 1$. En revanche, s'il existe $n > 0$, tel $p_{i,i}^{(n)} > 0$, alors l'état i est dit **état de retour** et **état de non retour** dans le cas contraire.

On peut utiliser la relation d'équivalence qu'est la relation de communication pour répartir les différents états d'une chaîne de Markov en **classes d'équivalence**. On définit une classe d'équivalence comme l'ensemble des états qui communiquent entre eux. Deux états de deux classes différentes ne peuvent alors communiquer entre eux (sinon on aurait à faire à la même classe). En revanche, il se peut que l'on puisse aller d'une classe à une autre, mais le retour dans la classe précédente ne sera plus jamais possible (sinon les deux classes formeraient une même classe).

DÉFINITION 4.14. Une chaîne de Markov est dite **irréductible** si elle ne contient qu'une seule classe d'équivalence, autrement dit si tous les états de la chaîne communiquent entre eux.

EXERCICE 4.12. Déterminer quelles sont les chaînes irréductibles parmi les exemples précédents : la marche aléatoire simple (Cf. Exemple 2.1), la fortune du joueur A contre le joueur B (Cf. Exemple 2.4), le modèle de diffusion d'Ehrenfest (Cf. Exemple 2.5), la marche aléatoire multidimensionnelle (Cf. Section 2.3). Si certaines chaînes ne sont pas irréductibles donner leurs classes d'équivalence.

Preuve/Solution de l'exercice.

La marche aléatoire simple (Cf. Exemple 2.1), le modèle de diffusion d'Ehrenfest (Cf. Exemple 2.5), la marche aléatoire multidimensionnelle (Cf. Section 2) sont des exemples de chaînes irréductibles.

En revanche la fortune du joueur A contre le joueur B (Exemple 2.4) ne constitue pas une chaîne irréductible puisqu'elle est constituée de trois classes : $\{0\}$, $\{1, 2, \dots, a-1\}$ et $\{a\}$. \diamond

EXERCICE 4.13. *Considérons la chaîne de Markov à espace d'états $E = \{1, 2, 3, 4, 5\}$ de matrice de passage :*

$$P = \begin{pmatrix} 1/3 & 2/3 & 0 & 0 & 0 \\ 2/5 & 2/5 & 1/5 & 0 & 0 \\ 0 & 0 & 1/2 & 1/4 & 1/4 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

- 1) Donner son graphe de Markov.
- 2) Donner ses classes d'équivalence. Préciser l'accessibilité entre classes.

Solution de l'exercice.

1) Son graphe de Markov est donné par la Figure 8.

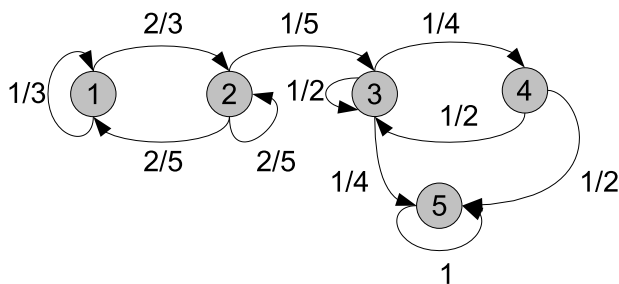


FIGURE 8. Graphe de Markov de l'Exercice 4.4.13.

2) Elle comporte trois classes : $C_1 = \{1, 2\}$, $C_2 = \{3, 4\}$ et $C_3 = \{5\}$. L'accessibilité entre les classes est donnée par : $C_1 \longrightarrow C_2 \longrightarrow C_3$. \diamond

DÉFINITION 4.15. *On dit qu'une classe d'équivalence C pour une chaîne de Markov est fermée si l'on a :*

$$(i \longrightarrow j \text{ et } i \in C) \implies j \in C,$$

*On dit aussi que la classe C est **absorbante**.*

Ainsi une fois que la chaîne atteint un des états de la classe fermée C , elle ne quitte plus cette classe.

6. Périodicité

DÉFINITION 4.16. *La période d'un état de retour i , que l'on note $d(i)$, est le P.G.C.D. de tous les entiers $n > 0$ pour lesquels on a : $p_{i,i}^{(n)} > 0$. Si l'on a $d(i) = 1$, alors l'état i est dit apériodique, sinon il est dit de période $d(i)$. Si, pour un état i , on a $p_{i,i}^{(n)} = 0$ pour tout n , alors on pose $d(i) = +\infty$.*

La période d'un l'état permet de savoir si le temps entre deux passage par cet état est, ou n'est pas, multiple d'un temps minimum. La plupart des chaînes sont apériodiques. Mais il est tout de même aisé de trouver des chaînes périodiques.

EXERCICE 4.14. *Étudier la périodicité des états des chaînes suivantes (en dissociant éventuellement entre les valeurs possibles des probabilités quand celles-ci ne sont pas spécifiées).*

- 1) Chaîne à deux états (Cf. Section 2.1)
 - 2) Marche aléatoire simple (Cf. Exemple 2.1).
 - 3) Joueur A contre joueur B (Cf. Exemple 2.4).
 - 4) Modèle de diffusion d'Ehrenfest (Cf. Exemple 2.5).
-

Solution de l'exercice.

1) Si $0 \leq p, q < 1$ alors $d(i) = 1$, pour $i = 1, 2$. De même, si $p = 1$ et $0 < q < 1$ ou si $q = 1$ et $0 < p < 1$. Si $p = 1$ et $q = 0$ alors $d(1) = +\infty$ et $d(2) = 1$. Réciproquement, Si $q = 1$ et $p = 0$, alors $d(2) = +\infty$ et $d(1) = 1$. Enfin, si $p = q = 1$, alors $d(i) = 2$ pour $i = 1, 2$.

2) Tous les états ont toujours même période. Cette dernière est égale à 2 si $0 < p < 1$. En revanche, si p est égal à 0 ou 1, alors la période est infinie.

3) On a toujours : $d(0) = d(a) = 1$. Si $0 < p < 1$ alors $d(i) = 2$, pour $i \in \{1, \dots, a-1\}$. Si $p(1-p) = 0$ alors $d(i) = +\infty$, pour $i \in \{1, \dots, a-1\}$.

4) $d(i) = 2$, pour tout $i = 0, \dots, a$. ◇

PROPOSITION 4.17. *Si la période d'un état est finie alors tout autre état qui communique avec celui-ci est de même période. La période est donc la même pour tous les éléments d'une même classe.*

Preuve. D'une part si $i \longleftrightarrow j$, alors il existe deux entiers n et m tels que : $p_{i,j}^{(n)} > 0$ et $p_{j,i}^{(m)} > 0$. D'autre part si l'on a $d(i) = d < +\infty$ alors il existe au moins un entier k tel que $p_{i,i}^{(k)} > 0$. Pour n'importe lequel de ces entiers k on a alors : $p_{j,j}^{(m+n+k)} \geq p_{j,i}^{(m)} p_{i,i}^{(k)} p_{i,j}^{(n)} > 0$. Mais ayant également $p_{i,i}^{(2k)} > 0$, on obtient aussi $p_{j,j}^{(m+n+2k)} > 0$. Ainsi la période $d(j)$ de l'état j divise à la fois $m+n+k$ et $m+n+2k$ et divise donc sa différence k . On a donc montré que tout k divisible par $d(i)$ l'est également par $d(j)$, donc $d(j)$ divise $d(i)$. Tout ceci étant parfaitement symétrique, on a aussi $d(i)$ qui divise $d(j)$. Ce qui prouve le résultat voulu. □

Donnons enfin un résultat que nous ne démontrerons pas. Sa démonstration repose principalement sur un résultat en théorie des nombres sur les pgcd.

LEMME 4.18. *Un état i d'une chaîne de Markov est apériodique si, et seulement si, on a pour tout entier n suffisamment grand*

$$p_{i,i}^{(n)} > 0.$$

7. Temps d'atteinte, états récurrents et transients

DÉFINITION 4.19. Soit i dans E un état quelconque de la chaîne de Markov (X_n) . On appelle temps d'atteinte de l'état i (ou temps de premier passage à l'état i), la v.a. T_i définie par

$$T_i = \inf\{n \geq 1 : X_n = i\},$$

avec la convention $\inf \emptyset = +\infty$.

Il est clair que cette v.a. est un temps d'arrêt par rapport à la filtration naturelle \mathcal{F} de la chaîne (X_n) (Cf. Chapitre ??).

Pour deux états i et j dans E , notons $f_{i,j}$ la probabilité que, partant de l'état i , la chaîne passe au moins une fois par l'état j , i.e.

$$f_{i,j} = P(T_j < +\infty | X_0 = i).$$

La probabilité $f_{i,i}$ est donc la probabilité que, partant de l'état i , la chaîne retourne à l'état i en un temps fini.

Notons également $f_{i,j}^{(n)}$ la probabilité que, partant de l'état i la chaîne aille, pour la première fois, à l'état j au temps n , i.e.

$$f_{i,j}^{(n)} = P(T_j = n | X_0 = i) = P(X_n = j, X_k \neq j, \forall k = 1, \dots, n-1 | X_0 = i),$$

avec la convention $f_{i,j}^{(0)} = 0$.

PROPOSITION 4.20. Pour tout états i et j et tout entier $n \geq 1$, on a :

$$p_{i,j}^{(n)} = \sum_{k=0}^n f_{i,j}^{(k)} p_{j,j}^{(n-k)}.$$

EXERCICE 4.15. Démonstration de cette proposition.

1) Montrer que l'on peut écrire :

$$p_{i,j}^{(n)} = \sum_{k=1}^{n-1} p_{j,j}^{(n-k)} f_{i,j}^{(k)} + f_{i,j}^{(n)}.$$

2) Conclure.

Preuve/Solution de l'exercice.

1) Pour tout (i, j) dans E^2 , on a :

$$\begin{aligned}
p_{i,j}^{(n)} &= P(X_n = j | X_0 = i) = P(\{X_n = j\}, \cup_{k=1}^n \{T_j = k\} | X_0 = i) \\
&= \sum_{k=1}^n P(\{X_n = j\}, \{T_j = k\} | X_0 = i) \\
&= \sum_{k=1}^{n-1} P(\{X_n = j\}, \{T_j = k\} | X_0 = i) + f_{i,j}^{(n)} \\
&= \sum_{k=1}^{n-1} P(\{X_n = j\} | \{T_j = k\}, X_0 = i) P(\{T_j = k\} | X_0 = i) + f_{i,j}^{(n)} \\
&= \sum_{k=1}^{n-1} P(\{X_n = j\} | X_k = j) P(\{T_j = k\} | X_0 = i) + f_{i,j}^{(n)} \\
&= \sum_{k=1}^{n-1} p_{j,j}^{(n-k)} f_{i,j}^{(k)} + f_{i,j}^{(n)},
\end{aligned}$$

où l'avant dernière égalité est justifiée par les formules vues en Section 4.

2) Comme on a : $p_{j,j}^{(0)} = 1$ et $f_{i,j}^{(0)} = 0$, on peut écrire

$$p_{i,j}^{(n)} = \sum_{k=0}^n f_{i,j}^{(k)} p_{i,j}^{(n-k)},$$

ce qu'il fallait démontrer. □

Définissons maintenant, pour tout n dans \mathbb{N} , les fonctions génératrices des suites de probabilités $(p_{i,j}^{(n)})_{n \in \mathbb{N}}$ et $(f_{i,j}^{(n)})_{n \in \mathbb{N}}$ respectivement par :

$$\begin{aligned}
P_{i,j}(s) &= \sum_{n=0}^{+\infty} p_{i,j}^{(n)} s^n, \quad \forall |s| \leq 1; \\
F_{i,j}(s) &= \sum_{n=0}^{+\infty} f_{i,j}^{(n)} s^n, \quad \forall |s| \leq 1.
\end{aligned}$$

Remarquons que ces séries sont absolument convergentes, pour $|s| < 1$, puisque les probabilités $p_{i,j}^{(n)}$ et $f_{i,j}^{(n)}$ sont majorées par 1.

PROPOSITION 4.21. *Pour tout couple (i, j) dans E^2 et pour tout $|s| < 1$, on a :*

$$P_{i,i}(s) = \frac{1}{1 - F_{i,i}(s)} \text{ et } P_{i,j}(s) = F_{i,j}(s) P_{j,j}(s),$$

ce qui peut être résumé par la formule :

$$P_{i,j}(s) = \delta_{i,j} + F_{i,j}(s) P_{j,j}(s).$$

Preuve. En utilisant la proposition précédente, on peut écrire, pour tout i dans E et tout $|s| < 1$:

$$\begin{aligned} P_{i,i}(s) &= \sum_{n=0}^{+\infty} p_{i,i}^{(n)} s^n = 1 + \sum_{n=1}^{+\infty} p_{i,i}^{(n)} s^n \\ &= 1 + \sum_{n=1}^{+\infty} \left(\sum_{k=0}^n f_{i,i}^{(k)} p_{i,i}^{(n-k)} \right) s^n = 1 + \sum_{n=0}^{+\infty} \left(\sum_{k=0}^n f_{i,i}^{(k)} p_{i,i}^{(n-k)} \right) s^n \\ &= 1 + P_{i,i}(s) F_{i,i}(s), \end{aligned}$$

où l'avant dernière égalité est justifiée par la nullité de $f_{i,i}^{(0)}$ et la dernière n'est que l'utilisation d'une formule classique de multiplication des séries entières¹.

L'autre formule se démontre exactement de la même manière. \square

DÉFINITION 4.22. On dit qu'un état i d'une chaîne de Markov est **récurrent** si, partant de l'état i , la probabilité que cette chaîne retourne à l'état i en un temps fini est égale à 1, i.e. si $f_{i,i} = 1$. Sinon l'état est dit **transient** ou **transitoire** (on a alors $f_{i,i} < 1$).

Autrement dit un état récurrent est tel que $P(T_i < +\infty | X_0 = i) = 1$. Le temps de retour à l'état i est donc *p.s.* fini. En revanche un état transitoire est tel que : $P(T_i < +\infty | X_0 = i) < 1 \iff P(T_i = +\infty | X_0 = i) > 0$, ce qui revient à dire qu'il y a une probabilité strictement positive que la chaîne ne repasse jamais par l'état transient i .

EXERCICE 4.16. Pour les chaînes de Markov suivantes, donner les états récurrents et transients.

- 1) Fortune du joueur A contre joueur B (Cf. Exemple 2.4).
 - 2) Chaîne de Markov de l'Exercice 4.13.
-

Solution de l'exercice.

1) Pour cette chaîne, les états 0 et a sont absorbants donc récurrents puisque l'on a :

$$1 = f_{0,0} = p_{0,0} = f_{a,a} = p_{a,a}.$$

Les autres états sont transitoires. En effet on a, par exemple, pour l'état 1 :

$$P(T_1 = +\infty | X_0 = 1) \geq P(X_1 = 0 | X_0 = 1) = q > 0,$$

puisque l'état 0 est absorbant.

2) L'état 5 est récurrent car absorbant (évident), tous les autres sont transitoires. En effet, par exemple, on a pour l'état 1 :

$$f_{1,1} \geq P(X_2 = 3, X_1 = 2 | X_0 = 1) = \frac{2}{3} \times \frac{1}{3} > 0.$$

Les autres se montrent de la même manière. \diamond

¹On rappelle que l'on a :

$$\left(\sum_{k=0}^{+\infty} a_k s^k \right) \left(\sum_{l=0}^{+\infty} b_l s^l \right) = \sum_{k=0}^{+\infty} \left(\sum_{l=0}^k a_k b_{k-l} \right) s^k.$$

PROPOSITION 4.23. *Un état i d'une chaîne de Markov est récurrent si, et seulement si, on a :*

$$\sum_{n=1}^{+\infty} f_{i,i}^{(n)} = 1.$$

Il est transient si, et seulement si,

$$\sum_{n=1}^{+\infty} f_{i,i}^{(n)} < 1.$$

EXERCICE 4.17. *Démontrer cette proposition.*

Solution de l'exercice. Il suffit de remarquer que l'on a :

$$\begin{aligned} f_{i,j} &= P(T_j < +\infty | X_0 = i) = P(\cup_{n=1}^{+\infty} \{T_j = n\} | X_0 = i) \\ &= \sum_{n=1}^{+\infty} P(\{T_j = n\} | X_0 = i) = \sum_{n=1}^{+\infty} f_{i,j}^{(n)} \end{aligned}$$

et d'appliquer cette formule à $i = j$. □

Afin de donner un autre critère, caractérisant la récurrence, introduisons un lemme technique.

LEMME 4.24. (**Lemme d'Abel**)

i) *Si une série de terme général a_n est convergente de somme égale à a , i.e.*

$$\sum_{n=0}^{+\infty} a_n = a < +\infty,$$

alors on a :

$$\lim_{s \rightarrow 1^-} \sum_{n=0}^{+\infty} a_n s^n = \sum_{n=0}^{+\infty} a_n = a.$$

ii) *Si une série de terme général a_n , tous positifs, est telle que*

$$\lim_{s \rightarrow 1^-} \sum_{n=0}^{+\infty} a_n s^n = a \leq +\infty,$$

alors on a :

$$\sum_{n=0}^{+\infty} a_n = a \leq +\infty.$$

Nous ne démontrons pas ce lemme, que l'on peut trouver dans n'importe quel bon livre d'analyse.

Nous sommes maintenant en mesure de donner une autre caractérisation de la récurrence.

THÉORÈME 4.25. *Un état i dans E l'espace des états d'une chaîne de Markov est récurrent si, et seulement si, l'on a :*

$$\sum_{n=0}^{+\infty} p_{i,i}^{(n)} = +\infty.$$

Il est donc transient si la série précédente est convergente.

EXERCICE 4.18. *Démontrer ce Théorème. (Ind. On utilisera, autant pour la partie nécessaire que pour la partie suffisante, le lemme d'Abel et la proposition précédente)*

Preuve/Solution de l'exercice. Montrons en premier lieu que cette condition est nécessaire. Supposons donc que l'état i est récurrent. D'après la proposition 4.23, on sait que :

$$\sum_{n=1}^{+\infty} f_{i,i}^{(n)} = 1.$$

En utilisant le *i)* du lemme d'Abel, on a alors :

$$\lim_{s \rightarrow 1^-} \sum_{n=0}^{+\infty} f_{i,i}^{(n)} s^n = \lim_{s \rightarrow 1^-} F_{i,i}(s) = 1.$$

Ainsi, grâce au lemme 4.21, il vient :

$$\lim_{s \rightarrow 1^-} P_{i,i}(s) = \lim_{s \rightarrow 1^-} \sum_{n=0}^{+\infty} p_{i,i}^{(n)} s^n = +\infty.$$

L'utilisation du point *ii)* du lemme d'Abel, nous permet de conclure que :

$$\sum_{n=0}^{+\infty} p_{i,i}^{(n)} = +\infty,$$

ce qui était le résultat désiré.

Pour démontrer que cette dernière égalité est suffisante, supposons que l'état i soit transient, c'est à dire que :

$$\sum_{n=1}^{+\infty} f_{i,i}^{(n)} < 1.$$

En raisonnant de manière similaire à précédemment on obtient que

$$\lim_{s \rightarrow 1^-} P_{i,i}(s) = \lim_{s \rightarrow 1^-} \sum_{n=0}^{+\infty} p_{i,i}^{(n)} s^n < +\infty,$$

ce qui, grâce au *ii)* du lemme d'Abel donne

$$\sum_{n=0}^{+\infty} p_{i,i}^{(n)} < +\infty.$$

Par contraposée, on obtient bien le caractère suffisant du critère. □

Remarquons, que d'après leurs définitions tout état de non retour est transient et tout état absorbant est récurrent. Ceci est facilement confirmé par le fait qu'un état de non retour est tel que $p_{i,i}^{(0)} = 1$ et $p_{i,i}^{(n)} = 0$, pour tout n . La série de la proposition 4.25 est donc trivialement convergente. Pour un état absorbant, cette série est bien sûr divergente puisque tous les termes de la série sont égaux à 1.

COROLLAIRE 4.26. *Si un état j dans E est récurrent alors on a, pour tout état i tel que $i \rightarrow j$:*

$$\sum_{n=0}^{+\infty} p_{i,j}^{(n)} = +\infty.$$

En revanche, si l'état j est transient, on a pour tout état i de E :

$$\sum_{n=0}^{+\infty} p_{i,j}^{(n)} < +\infty.$$

Il en découle, dans ce cas, que l'on a également, pour tout i dans E :

$$\lim_{n \rightarrow +\infty} p_{i,j}^{(n)} = 0.$$

EXERCICE 4.19. *Démonstration de ce Corollaire.*

1) *Supposons dans un premier temps j transient. Montrer, en utilisant en particulier le lemme d'Abel, que*

$$\lim_{s \rightarrow 1^-} P_{i,j}(s) = \frac{f_{i,j}}{1 - f_{j,j}} < +\infty.$$

En utilisant à nouveau le lemme d'Abel, montrer que le terme général de la série $\sum_{n=0}^{+\infty} p_{i,j}^{(n)}$ converge vers 0.

2) *On suppose maintenant j récurrent. Montrer, toujours en utilisant le lemme d'Abel, que :*

$$\lim_{s \rightarrow 1^-} P_{i,j}(s) = \frac{f_{i,j}}{1 - f_{j,j}}.$$

3) *Montrer que la limite précédente est infinie.*

4) *Achever en utilisant à nouveau le lemme d'Abel.*

Preuve/Solution de l'exercice.

1) Supposons dans un premier temps que l'état j soit transient. On sait, d'après le théorème 4.21, que l'on a

$$P_{i,j}(s) = \frac{F_{i,j}(s)}{1 - F_{j,j}(s)}.$$

Or les séries $\sum_{n=0}^{+\infty} f_{i,j}^{(n)}$ et $\sum_{n=0}^{+\infty} f_{j,j}^{(n)}$ sont convergentes de sommes respectivement égales à $f_{i,j}$ et $f_{j,j}$ toutes les deux majorées par 1 (puisque ce sont des probabilités). Le lemme d'Abel utilisé pour ces deux séries nous permet d'écrire que :

$$\lim_{s \rightarrow 1^-} P_{i,j}(s) = \frac{\lim_{s \rightarrow 1^-} F_{i,j}(s)}{1 - \lim_{s \rightarrow 1^-} F_{j,j}(s)} = \frac{f_{i,j}}{1 - f_{j,j}} < +\infty,$$

puisque le numérateur, comme probabilité, est majoré par un et le dénominateur $1 - f_{j,j}$ est strictement positif, par transience de l'état j . En utilisant le *ii*) du lemme d'Abel, il vient alors que :

$$\sum_{n=0}^{+\infty} p_{i,j}^{(n)} = \frac{f_{i,j}}{1 - f_{j,j}} < +\infty.$$

Le terme général de la série est alors bien sûr convergent vers 0.

2) Dans le cas où l'état j est récurrent, on fait un raisonnement tout à fait similaire. Ainsi, par les mêmes arguments, on montre que :

$$\lim_{s \rightarrow 1^-} P_{i,j}(s) = \frac{\lim_{s \rightarrow 1^-} F_{i,j}(s)}{1 - \lim_{s \rightarrow 1^-} F_{j,j}(s)} = \frac{f_{i,j}}{1 - f_{j,j}}.$$

3) Par récurrence de l'état j , on a : $f_{j,j} = 1$. De plus, puisqu'il s'agit d'une probabilité, le terme $f_{i,j}$ est majoré par 1. Soit maintenant un état i tel que $i \rightarrow j$. Il existe donc un entier n tel que $p_{i,j}^{(n)} > 0$. On a alors :

$$f_{i,j} = P(T_j < +\infty | X_0 = i) \geq p_{i,j}^{(n)} > 0.$$

Ayant donc en résumé $f_{j,j} = 1$ et $0 < f_{i,j} \leq 1$, la limite précédente est donc infinie.

4) En utilisant le *ii*) du lemme d'Abel, on en déduit que

$$\sum_{n=0}^{+\infty} p_{i,j}^{(n)} = \frac{f_{i,j}}{1 - f_{j,j}} = \infty,$$

ce qui achève notre démonstration. □

Le théorème 4.25 nous permet de voir que, partant d'un état i , l'espérance du nombre de retour à cet état est infinie si et seulement si l'état est récurrent. En effet, notons

$$N_i = \sum_{n=0}^{+\infty} \mathbb{1}_{\{X_n=i\}}.$$

Par le théorème de Tonelli, on peut inverser espérance et somme pour obtenir :

$$\mathbb{E}_i(N_i) = \sum_{n=0}^{+\infty} p_{i,i}^{(n)},$$

où \mathbb{E}_i est l'espérance pour la loi conditionnelle à $\{X_0 = i\}$. D'après la proposition précédente, la partie droite de l'égalité est infinie si, et seulement si, l'état i est récurrent.

Ce résultat est confirmé par la proposition suivante.

PROPOSITION 4.27. *Un état i est récurrent ou transient suivant que la probabilité*

$$P(X_n = i \text{ pour une infinité de } n | X_0 = i)$$

est égale, respectivement, à 1 ou 0.

EXERCICE 4.20. *Démonstration de cette Proposition.*

1) Soit $Q_{i,i}^N$ la probabilité que, partant de l'état i , la chaîne repasse au moins N fois par cet état. En "découpant" suivant les valeurs de T_i , montrer que l'on a :

$$Q_{i,i}^N = f_{i,i} Q_{i,i}^{N-1}.$$

En déduire que :

$$Q_{i,i}^N = (f_{i,i})^N.$$

2) Conclure en considérant la limite, quand $N \rightarrow +\infty$, de ces probabilités $Q_{i,i}^N$.

Preuve/Solution de l'exercice. Notons $Q_{i,i}^N$ la probabilité que, partant de l'état i , la chaîne repasse au moins N fois par cet état. On a :

$$\begin{aligned} Q_{i,i}^N &= P(\{N_i \geq N+1\} | X_0 = i) \\ &= P(\{N_i \geq N+1\}, \cup_{k=1}^{+\infty} \{T_i = k\} | X_0 = i) \\ &= \sum_{k=1}^{+\infty} P(\{N_i \geq N+1\}, \{T_i = k\} | X_0 = i) \\ &= \sum_{k=1}^{+\infty} P(N_i \geq N+1 | X_k = i) P(T_i = k | X_0 = i) \\ &= \sum_{k=1}^{+\infty} Q_{i,i}^{N-1} f_{i,i}^{(k)} = f_{i,i} Q_{i,i}^{N-1}. \end{aligned}$$

En réitérant cette formule, on obtient :

$$Q_{i,i}^N = f_{i,i} Q_{i,i}^{N-1} = (f_{i,i})^2 Q_{i,i}^{N-2} = \dots = (f_{i,i})^{N-1} Q_{i,i}^1.$$

Comme on a : $Q_{i,i}^1 = f_{i,i}$, il vient :

$$Q_{i,i}^N = (f_{i,i})^N.$$

En remarquant que les événements considérés dans les probabilités $Q_{i,i}^N$ forment une suite décroissante, on peut écrire :

$$\lim_{N \rightarrow +\infty} Q_{i,i}^N \equiv Q_{i,i} = P(X_n = i \text{ pour une infinité de } n | X_0 = i).$$

Ainsi, la probabilité $Q_{i,i}$ est égale à 1 si, et seulement si, $f_{i,i} = 1$, ou encore, d'après la proposition 4.23, si, et seulement si, l'état i est récurrent. La transience de l'état i est nous l'avons vu, équivalente à $f_{i,i} < 1$, ce qui d'après ce qui précède est équivalent à $Q_{i,i} = 0$. \square

THÉORÈME 4.28. *Pour une chaîne de Markov irréductible et récurrente, on a pour tout état (i, j) dans E^2 :*

$$f_{i,j} = P(T_j < +\infty | X_0 = i) = 1,$$

ce qui implique que pour tout j de E on a :

$$P(T_j < +\infty) = 1.$$

Preuve. La classe étant récurrente, d'après la proposition 4.27, on peut écrire, pour m quelconque :

$$\begin{aligned} 1 &= P(X_n = j \text{ pour une infinité de } n | X_0 = j) \\ &\leq P(X_n = j \text{ pour un } n \geq m + 1 | X_0 = j) \leq 1, \end{aligned}$$

où la première inégalité est dû à :

$$\begin{aligned} \{X_n = j \text{ pour une infinité de } n\} &= \{\forall m, \exists n \geq m : X_n = j\} \\ &= \bigcap_{m \in \mathbb{N}} \bigcup_{n \geq m} \{X_n = j\} \subset \bigcup_{n \geq m} \{X_n = j\}, \end{aligned}$$

pour m quelconque dans \mathbb{N} . Les deux probabilités précédentes sont ainsi égales. On en déduit que, en prenant m tel que $p_{j,i}^{(m)} > 0$, on a :

$$\begin{aligned} 1 &= P(X_n = j \text{ pour une infinité de } n | X_0 = j) \\ &= P(X_n = j \text{ pour un } n \geq m + 1 | X_0 = j) \\ &= \sum_{k \in E} P(X_n = j \text{ pour un } n \geq m + 1 | X_m = k) P(X_m = k | X_0 = i) \\ &= \sum_{k \in E} P(T_j < +\infty | X_0 = k) p_{j,k}^{(m)}, \end{aligned}$$

où l'on a utilisé la propriété de Markov pour obtenir la dernière égalité. Mais la matrice $P^{(m)}$ étant stochastique, on a déjà

$$\sum_{k \in E} p_{j,k}^{(m)} = 1,$$

ce qui implique nécessairement que $f_{i,j} = P(T_j < +\infty | X_0 = k) = 1$, pour tout état k dans E .

Quand au dernier résultat du théorème, il s'obtient rapidement en décomposant suivant les valeurs possibles pour X_0 , :

$$P(T_j < +\infty) = \sum_{i \in E} P(T_j < +\infty | X_0 = i) P(X_0 = i) = \sum_{i \in E} f_{i,j} P(X_0 = i) = 1.$$

puisque la loi initiale est forcément une probabilité. □

PROPOSITION 4.29. *La récurrence est une propriété de classe, i.e. :*

$$(i \longleftrightarrow j \text{ et } i \text{ récurrent}) \implies j \text{ récurrent.}$$

EXERCICE 4.21. *Démontrer cette Proposition. (On pourra utiliser la proposition 4.25)*

Preuve/Solution de l'exercice. Soient i et j deux états qui communiquent. Par définition, il existe donc m et n dans \mathbb{N} tels que :

$$p_{i,j}^{(m)} > 0 \text{ et } p_{j,i}^{(n)} > 0.$$

On a alors, en utilisant la minoration classique $p_{j,j}^{(m+n+k)} \geq p_{j,i}^{(n)} p_{i,i}^{(k)} p_{i,j}^{(m)}$:

$$\sum_{k=0}^{+\infty} p_{j,j}^{(m+n+k)} \geq \sum_{k=0}^{+\infty} p_{j,i}^{(n)} p_{i,i}^{(k)} p_{i,j}^{(m)} = p_{j,i}^{(n)} p_{i,j}^{(m)} \sum_{k=0}^{+\infty} p_{i,i}^{(k)} = +\infty,$$

puisque l'état i est supposé récurrent. L'état j est donc lui aussi récurrent. \square

EXERCICE 4.22. *Étude de la marche aléatoire simple (Cf. Exemple 2.1).*

- 1) *Considérer le cas de l'état 0. Calculer les valeurs de $p_{0,0}^n$, pour $n \in \mathbb{N}$.*
 - 2) *On pourra utiliser la proposition 4.25 et la formule Stirling pour conclure sur l'état 0. (Ind. La réponse pourra varier en fonction des valeurs de p)*
 - 3) *Qu'en est-il des autres états ?*
-

Solution de l'exercice. D'après sa définition la marche aléatoire simple est telle que, pour tout n dans \mathbb{N} , on a :

$$p_{0,0}^{(2n+1)} = 0 \text{ et } p_{0,0}^{(2n)} = C_{2n}^n p^n (1-p)^n.$$

En utilisant la formule de Stirling

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n,$$

quand $n \rightarrow +\infty$, on a :

$$p_{0,0}^{(2n)} \sim \frac{2\sqrt{\pi n} \left(\frac{2n}{e}\right)^{2n}}{2\pi n \left(\frac{n}{e}\right)^{2n}} p^n (1-p)^n = \frac{[4p(1-p)]^n}{\sqrt{\pi n}}.$$

Or il est aisé de vérifier que l'on a $p(1-p) \leq 1/4$ avec égalité seulement dans le cas où $p = 1/2$. Si l'on est dans cette dernière situation, alors la série

$$\sum_{n=0}^{+\infty} p_{0,0}^{(n)} = \sum_{n=0}^{+\infty} p_{0,0}^{(2n)}$$

est de même nature que la série

$$\sum_{n=0}^{+\infty} \frac{1}{\sqrt{\pi n}}$$

qui est divergente. Ainsi d'après la proposition 4.25, l'état 0 est récurrent. Ayant vu qu'une marche aléatoire simple est irréductible, i.e. tous les états communiquent entre eux, et que la récurrence est une propriété de classe, tous les états de la marche aléatoire simple sont récurrents.

Supposons maintenant que l'on ait $p \neq 1/2$. Notons alors $r = 4p(1-p) < 1$. La série

$$\sum_{n=0}^{+\infty} p_{0,0}^{(n)}$$

est alors de même nature que la série

$$\sum_{n=0}^{+\infty} r^n$$

que l'on sait être convergente pour $r < 1$. La proposition 4.25 nous permet de conclure que l'état 0 est transient. Tous les états de la marche aléatoire simple sont alors transients dans le cas où $p \neq 1/2$. \diamond

8. Propriété de Markov forte

Donnons la définition d'un temps d'arrêt associé à une filtration, notion qui sera à nouveau étudiée dans le Chapitre ?? sur les martingales.

DÉFINITION 4.30. Soit (\mathcal{F}_n) une filtration sur un espace probabilisé (Ω, \mathcal{A}, P) . Une application $T : \Omega \rightarrow \bar{\mathbb{N}}$ est dite temps d'arrêt relativement à la filtration (\mathcal{F}_n) si

$$\forall n \in \mathbb{N} : \{T \leq n\} \in \mathcal{F}_n$$

ou encore de manière équivalente si

$$\forall n \in \mathbb{N} : \{T = n\} \in \mathcal{F}_n$$

Définissons la tribu des événements antérieurs à un temps d'arrêt T .

DÉFINITION 4.31. Soit T un temps d'arrêt relativement à une filtration (\mathcal{F}_n) . On appelle tribu des événements antérieurs à T la tribu \mathcal{F}_T définie par :

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : \forall n \in \mathbb{N} \text{ on ait } A \cap \{T \leq n\} \in \mathcal{F}_n\},$$

où la tribu \mathcal{F}_∞ est définie par

$$\mathcal{F}_\infty = \bigvee_n \mathcal{F}_n.$$

On peut montrer que la tribu \mathcal{F}_T est également telle que

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : \forall n \in \mathbb{N} \text{ on ait } A \cap \{T = n\} \in \mathcal{F}_n\}.$$

On peut maintenant donner la propriété de Markov forte associée à une chaîne de Markov. L'idée, simple, est de montrer que l'on a toujours la propriété de Markov quand on ne conditionne plus en un temps fixé, disons n quand on conditionne par rapport à $\{X_n = i\}$, mais quand on conditionne en un temps aléatoire donné par un temps d'arrêt T , donc quand on conditionne par rapport à un événement du type $\{X_T = i\}$. Nous formulons cette propriété dans sa forme plus générale à mettre en relation avec les formules de conditionnement obtenues en section 4.

THÉORÈME 4.32. (**Propriété de Markov forte**) Soit T un temps d'arrêt relativement à la filtration naturelle associée à une chaîne de Markov (X_n) . Pour tout événement A de la tribu \mathcal{F}_T tel que $P(T < +\infty, A, X_T = i) > 0$ on a :

$$\begin{aligned} & P(X_{T+1} = j_1, \dots, X_{T+k} = j_k | T < +\infty, A, X_T = i) \\ &= P(X_1 = j_1, \dots, X_k = j_k | X_0 = i) \end{aligned}$$

EXERCICE 4.23. Démonstration de ce Théorème.

1) Montrer, en justifiant bien les étapes de votre démonstration, que l'on peut écrire :

$$\begin{aligned} & P(X_{T+1} = j_1, \dots, X_{T+k} = j_k, T < +\infty, A, X_T = i) \\ &= \sum_{n=0}^{+\infty} P(X_{n+1} = j_1, \dots, X_{n+k} = j_k | X_n = i) P(T = n, A, X_n = i). \end{aligned}$$

2) En déduire que l'on a :

$$\begin{aligned} & P(X_{T+1} = j_1, \dots, X_{T+k} = j_k, T < +\infty, A, X_T = i) \\ &= P(X_1 = j_1, \dots, X_k = j_k | X_0 = i) P(T < +\infty, A, X_n = i) \end{aligned}$$

et conclure.

Solution de l'exercice. On a en effet

$$\begin{aligned}
& P(X_{T+1} = j_1, \dots, X_{T+k} = j_k, T < +\infty, A, X_T = i) \\
&= \sum_{n=0}^{+\infty} P(X_{T+1} = j_1, \dots, X_{T+k} = j_k, T = n, A, X_T = i) \\
&= \sum_{n=0}^{+\infty} P(X_{n+1} = j_1, \dots, X_{n+k} = j_k, T = n, A, X_n = i) \\
&= \sum_{n=0}^{+\infty} P(X_{n+1} = j_1, \dots, X_{n+k} = j_k | T = n, A, X_n = i) P(T = n, A, X_n = i).
\end{aligned}$$

Dans cette dernière somme les termes non nuls sont seulement ceux pour lesquels on a : $P(T = n, A, X_n = i) > 0$. Or sous cette condition et en utilisant une formule vue en section 4, on peut écrire :

$$\begin{aligned}
& P(X_{n+1} = j_1, \dots, X_{n+k} = j_k | T = n, A, X_n = i) \\
&= P(X_{n+1} = j_1, \dots, X_{n+k} = j_k | X_n = i)
\end{aligned}$$

puisque l'événement $\{T = n\} \cap A$ est dans la tribu \mathcal{F}_n par définition de la tribu arrêtée \mathcal{F}_T . L'homogénéité en temps de la chaîne de Markov nous donne alors :

$$\begin{aligned}
& P(X_{n+1} = j_1, \dots, X_{n+k} = j_k | T = n, A, X_n = i) \\
&= P(X_1 = j_1, \dots, X_k = j_k | X_0 = i).
\end{aligned}$$

On a donc

$$\begin{aligned}
& P(X_{T+1} = j_1, \dots, X_{T+k} = j_k, T < +\infty, A, X_T = i) \\
&= \sum_{n=0}^{+\infty} P(X_1 = j_1, \dots, X_k = j_k | X_0 = i) P(T = n, A, X_n = i) \\
&= P(X_1 = j_1, \dots, X_k = j_k | X_0 = i) \sum_{n=0}^{+\infty} P(T = n, A, X_n = i) \\
&= P(X_1 = j_1, \dots, X_k = j_k | X_0 = i) P(T < +\infty, A, X_n = i),
\end{aligned}$$

ce qui, en divisant par $P(T < +\infty, A, X_n = i)$ supposée strictement positive, nous donne le résultat voulu. \square

On tire de ce résultat que, conditionnellement à $\{T < \infty\}$ et $\{X_T = i\}$, le processus (X_{T+n}) est encore une chaîne de Markov avec pour matrice de passage la même matrice P que la chaîne (X_n) et pour loi initiale la loi de dirac en i .

9. Récurrence positive et récurrence nulle

Rappelons qu'un état i est dit récurrent si $f_{i,i} = 1$ et transient si $f_{i,i} < 1$. Nous avons vu que dans ces deux cas le temps de retour à l'état i est respectivement p.s. fini et infini avec une probabilité non nulle. Intéressons nous maintenant à l'espérance de ce temps de retour.

DÉFINITION 4.33. Pour tout état i d'une chaîne de Markov, on définit le **temps moyen de récurrence** μ_i par :

$$\mu_i = E_i(T_i) = E(T_i | X_0 = i) = \begin{cases} +\infty & \text{si } i \text{ est transient;} \\ \sum_{n \geq 1} n f_{i,i}^{(n)} & \text{si } i \text{ est récurrent.} \end{cases}$$

Puisque, comme nous venons de le rappeler, pour un état transient on a $P(T_i = +\infty | X_0 = i) > 0$, l'espérance du temps de retour μ_i est infinie, i.e. $\mu_i = \mathbb{E}_i(T_i) = +\infty$. En revanche si l'état i est récurrent, on peut avoir un temps moyen de récurrence μ_i fini ou infini.

DÉFINITION 4.34. Un état récurrent d'une chaîne de Markov est dit **récurrent nul** si $\mu_i = +\infty$ et **récurrent positif** (ou non nul) si $\mu_i < +\infty$.

L'espérance du temps de retour à un état transient étant infinie, on rencontre dans certains ouvrages le terme redondant d'état transient nul. Nous n'utiliserons pas cette terminologie car elle pourrait laisser sous-entendre qu'il existe des états transients positifs, ce qui est impossible comme nous venons de le voir. Tout état transient est forcément nul.

Les états récurrents nuls sont donc entre les états transients et les états récurrents positifs. Ils sont récurrents mais leur temps moyen de récurrence est infini comme les états transients.

Le critère suivant, que nous ne démontrerons pas, permet de caractériser la positivité d'un état récurrent.

THÉORÈME 4.35. (**Critère de nullité**) Un état récurrent d'une chaîne de Markov est nul si, et seulement si,

$$\lim_{n \rightarrow +\infty} p_{i,i}^{(n)} = 0.$$

THÉORÈME 4.36. La positivité, ou la nullité, est une propriété de classe.

EXERCICE 4.24. Démontrer ce Théorème. (Ind. On pourra utiliser le Théorème 4.35)

Solution de l'exercice. Soit i un état récurrent nul et j un autre état, appartenant à la même classe que i . Puisque i et j communiquent entre eux, il existe des entiers n et m tels que :

$$p_{i,j}^{(m)} > 0 \text{ et } p_{j,i}^{(n)} > 0.$$

De l'inégalité maintenant classique

$$p_{i,i}^{(m+n+k)} \geq p_{i,j}^{(m)} p_{i,i}^{(k)} p_{j,i}^{(n)},$$

vraie pour tout k dans E , on tire que la nullité de l'état i implique celle de l'état j . \square

Ainsi une classe d'équivalence d'une chaîne de Markov est soit transiente, soit récurrente nulle, soit récurrente positive. En revanche, si l'espace des états E de la chaîne de Markov est fini, on ne peut trouver d'état, et donc bien sûr de classe, récurrente nulle.

PROPOSITION 4.37. *Soit (X_n) une chaîne de Markov sur un espace d'états E fini. Elle possède alors au moins un état récurrent et les états récurrents sont tous positifs.*

Preuve. Supposons que tous les états de la chaîne soient transients. D'après le corollaire 4.26 on a pour tout (i, j) dans E^2 :

$$\lim_{n \rightarrow +\infty} p_{i,j}^{(n)} = 0.$$

Ainsi, puisque E est fini, on peut écrire :

$$\lim_{n \rightarrow +\infty} \sum_{j \in E} p_{i,j}^{(n)} = \sum_{j \in E} \lim_{n \rightarrow +\infty} p_{i,j}^{(n)} = 0,$$

ce qui est bien sûr en contradiction avec l'égalité :

$$\sum_{j \in E} p_{i,j}^{(n)} = 1,$$

pour tout n dans \mathbb{N} .

Il existe donc au moins un état récurrent pour cette chaîne et soit i cet état. De deux choses l'une : soit il est seul dans sa classe, soit il n'est pas seul. Dans le premier cas, il communique uniquement avec lui même, alors nécessairement on $p_{i,i}^{(n)} = 1$ pour tout n et l'état i donc, d'après le critère de nullité, est récurrent positif.

Dans le second cas, il communique avec au moins un autre état j différent de lui même. Supposons que cette classe récurrente, notée C , soit de plus nulle. L'état i étant accessible de l'état j , il existe au moins un entier m tel que $p_{j,i}^{(m)} > 0$. On a alors, pour tout entier n :

$$p_{i,i}^{(m+n)} \geq p_{i,j}^{(n)} p_{j,i}^{(m)}.$$

Puisque l'état i est supposé récurrent nul, le critère de nullité assure la convergence vers 0 du terme à gauche dans l'inégalité, quand n tend vers $+\infty$ et donc également celle de $p_{i,j}^{(n)}$, toujours quand n tend vers $+\infty$. Or la classe C étant de cardinal fini (puisque l'ensemble des états E de la chaîne est supposé fini), on en déduit que :

$$\lim_{n \rightarrow +\infty} \sum_{j \in C} p_{i,j}^{(n)} = 0.$$

Or ceci est en contradiction avec le fait que l'on ait pour tout n :

$$1 = \sum_{j \in E} p_{i,j}^{(n)} = \sum_{j \in C} p_{i,j}^{(n)},$$

où la dernière égalité est justifiée par l'impossibilité de quitter une classe récurrente. La classe récurrente C ne peut donc être nulle et est donc positive, ce qui est bien le résultat cherché. \square

Il n'est pas nécessaire de démontrer le corollaire suivant tout à fait évident.

COROLLAIRE 4.38. *Une chaîne de Markov irréductible sur un espace fini est forcément récurrente positive.*

10. Loi stationnaire et théorèmes limites

À partir de maintenant, on va se poser la question de la convergence de la chaîne de Markov. Bien sûr, sauf dans des cas très particuliers, la chaîne de Markov (X_n) ne convergera pas vers un état donné. Mais en revanche, il n'est pas injustifié d'espérer que, sous de bonnes conditions, la suite (X_n) converge en loi vers une loi limite.

On verra que l'existence d'une telle loi limite est très liée à la notion de loi stationnaire que nous allons maintenant définir.

10.1. Loi stationnaire.

DÉFINITION 4.39. Soit (X_n) une chaîne de Markov à valeurs dans une espace d'état discret E . Un vecteur colonne de probabilités $\pi = (\pi_n)_{n \in E}$, donc tel que

$$\pi_n \geq 0, \forall n \in E \text{ et } \sum_{n \in E} \pi_n = 1,$$

est dit **loi (ou distribution) stationnaire** si l'on a :

$$\pi = P^* \pi \iff \pi^* = \pi^* P,$$

où A^* est la matrice transposée de la matrice A . Autrement dit cette loi est dite stationnaire si, pour tout j dans E on a

$$\pi_j = \sum_{i \in E} \pi_i p_{i,j}.$$

On parle également de loi invariante ou de loi d'équilibre.

Remarquons que l'on aurait pu dire que π est un vecteur propre à droite de la matrice P^* avec 1 pour valeur propre à droite.

Une telle loi est dite stationnaire car si la loi initiale de la chaîne, i.e. la loi de la v.a. X_0 , est π alors celle de X_n , pour tout n , est encore π . En effet, constatons déjà que pour la loi de la v.a. X_1 on a, pour tout k dans E :

$$P(X_1 = k) = \sum_{n \in E} P(X_1 = k, X_0 = n) = \sum_{n \in E} p_{n,k} \pi_n = \pi_k,$$

où la dernière égalité est justifiée par l'invariance de la loi π . La loi de la v.a. X_1 est donc également π . La loi de X_1 est donc toujours donnée par $\pi^* P$ et, dans le cas d'une loi initiale stationnaire, est égale encore à π^* .

Maintenant, constatons que l'on a :

$$(P^*)^2 \pi = P^*(P^* \pi) = P^* \pi = \pi$$

et, par récurrence évidente, pour tout n dans \mathbb{N} :

$$(P^*)^n \pi = \pi.$$

Or en faisant le même raisonnement que précédemment, la loi de la chaîne à l'instant n , i.e. la loi de la v.a. X_n , est $\pi^* P^n$ qui est donc encore égale à π . La loi des X_n n'évolue donc pas avec le temps. Et bien sûr dans un tel cas, la loi π est également la loi limite de la chaîne.

En fait on peut montrer bien plus. Dans le cas où la loi initiale est une loi stationnaire, alors le processus (X_n) est stationnaire au sens fort, comme définie dans le cours "Introduction aux Processus Stochastiques". En effet, pour tout k dans \mathbb{N} , pour tout vecteur d'entiers (n_1, \dots, n_k)

tels que $n_1 < \dots < n_k$ et tout vecteur (i_1, \dots, i_k) dans E^k , on a, par homogénéité de la chaîne et pour tout n :

$$\begin{aligned} P(X_{m+n_1} = i_1, \dots, X_{m+n_k} = i_k) &= p_{i_{k-1}, i_k}^{(n_k - n_{k-1})} \dots p_{i_1, i_2}^{(n_2 - n_1)} \pi_{i_1} \\ &= P(X_{n_1} = i_1, \dots, X_{n_k} = i_k). \end{aligned}$$

Ainsi on a, pour tout k dans \mathbb{N} , pour tout vecteur d'entiers (n_1, \dots, n_k) tels que $n_1 < \dots < n_k$ et tout entier m :

$$\mathcal{L}(X_{n_1}, \dots, X_{n_k}) = \mathcal{L}(X_{m+n_1}, \dots, X_{m+n_k}),$$

ce qui prouve bien que le processus est stationnaire au sens fort.

Le théorème qui suit formule différemment le résultat que nous venons d'obtenir.

THÉORÈME 4.40. *Soit $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov de matrice de passage P et de loi initiale π stationnaire. Alors, pour tout entier m , le processus translaté $(X_{m+n})_{n \in \mathbb{N}}$ est une chaîne de Markov de matrice de passage P (rien de neuf ici !) et de loi initiale π . Autrement dit, les deux chaînes de Markov ont même loi.*

L'appellation loi d'équilibre peut être déjà pressentie par théorème qui suit. Elle sera encore plus claire quand nous verrons les théorèmes limites.

THÉORÈME 4.41. *Soit (X_n) une chaîne de Markov à espace d'états E fini. Supposons que pour un état i de E on ait, quand n tend vers $+\infty$:*

$$p_{i,j}^{(n)} \rightarrow \pi_j, \text{ pour tout } j \text{ dans } E.$$

Alors la loi π est nécessairement une loi stationnaire.

Preuve. Les sommes qui suivent étant finies, on peut écrire :

$$\sum_{j \in E} \pi_j = \sum_{j \in E} \lim_{n \rightarrow +\infty} p_{i,j}^{(n)} = \lim_{n \rightarrow +\infty} \sum_{j \in E} p_{i,j}^{(n)} = 1,$$

ce qui prouve que π est bien une loi de probabilité sur E . Par ailleurs, ayant pour tout (i, j) dans E^2 et tout n dans \mathbb{N}

$$p_{i,j}^{(n)} = \sum_{k \in E} p_{i,k}^{(n-1)} p_{k,j},$$

il vient :

$$\pi_j = \lim_{n \rightarrow +\infty} p_{i,j}^{(n)} = \lim_{n \rightarrow +\infty} \sum_{k \in E} p_{i,k}^{(n-1)} p_{k,j} = \sum_{k \in E} \lim_{n \rightarrow +\infty} p_{i,k}^{(n-1)} p_{k,j} = \sum_{k \in E} \pi_k p_{k,j},$$

ce qui prouve que la loi π est bien stationnaire. \square

Il est important de noter qu'il n'existe pas toujours une loi stationnaire. Il en est ainsi par exemple dans le cas d'une marche aléatoire simple avec $p = 1$ (et donc $q = 0$). En effet dans ce cas, comme $p_{i,j}$ est non nul seulement si $j = i + 1$, l'équation que doit vérifier une loi stationnaire

$$\pi_j = \sum_{i \in E} \pi_i p_{i,j}, \quad \forall j \in \mathbb{Z},$$

implique que $\pi_j = \pi_{j-1}$ et que donc tous les π_j sont égaux à une constante. La mesure discrète ainsi obtenue sur \mathbb{Z} n'est bien sûr pas une probabilité. Constatons également que pour une marche aléatoire simple générale (i.e sur \mathbb{Z} avec $p \in [0, 1]$) on a, pour tout $(i, j) \in \mathbb{Z}^2$,

la convergence de $p_{i,j}^{(n)}$ vers 0 quand n tend vers $+\infty$. Mais la mesure ainsi obtenue est effectivement invariante mais n'est toujours pas une probabilité sur \mathbb{Z} . Le théorème précédent est donc uniquement vrai dans le cas d'un espace des états fini.

À l'inverse il existe des chaînes de Markov possédant une infinité de loi stationnaire.

EXERCICE 4.25. *Démontrer que pour la chaîne modélisant le jeu entre A et B (Cf. Exemple 2.4), il existe une infinité de loi stationnaires et donner leurs formes.*

Solution de l'exercice.

On vérifie aisément que n'importe quelle probabilité sur $\{0, 1, \dots, a\}$ de la forme $\pi = (\rho, 0, \dots, 0, 1 - \rho)^*$ est stationnaire pour cette chaîne. \diamond

Cas des matrices bistochastiques.

Une matrice stochastique P est dite **bistochastiques** si la somme de ses colonnes vaut aussi 1. C'est par exemple le cas pour une matrice de transition P qui serait symétrique.

Notons $\mathbb{1}$ le vecteur colonne sur E ayant toutes ses composantes égales à 1. Si P est bistochastique, ce vecteur $\mathbb{1}$ est alors naturellement vecteur propre à droite de P^* i.e.

$$P^* \mathbb{1} = \mathbb{1}.$$

Ainsi, le vecteur $\mathbb{1}$ est une mesure invariante pour la chaîne mais ce n'est bien sûr pas une probabilité. Dans le cas où E est infini, tout vecteur de la forme $c\mathbb{1}$ où $c \in \mathbb{R}^+$ est toujours une mesure invariante mais pas une probabilité.

En revanche, si E est fini, alors en prenant $c = 1/\text{Card}(E)$, on obtient une loi invariante. Ainsi par exemple, toute chaîne de Markov sur E fini ayant une matrice symétrique admet une loi invariante.

EXERCICE 4.26. *Étude de la marche à deux états.*

1) *Montrer que l'on peut écrire*

$$\begin{aligned} p_{1,1}^{(n+1)} &= (1-p)p_{1,1}^{(n)} + q(1-p_{1,1}^{(n)}) = p_{1,1}^{(n)}(1-p-q) + q \\ p_{1,2}^{(n+1)} &= p(1-p_{1,2}^{(n)}) + (1-q)p_{1,2}^{(n)} = p_{1,2}^{(n)}(1-p-q) + p \\ p_{2,1}^{(n+1)} &= (1-p)p_{2,1}^{(n)} + q(1-p_{2,1}^{(n)}) = p_{2,1}^{(n)}(1-p-q) + q \\ p_{2,2}^{(n+1)} &= p(1-p_{2,2}^{(n)}) + (1-q)p_{2,2}^{(n)} = p_{2,2}^{(n)}(1-p-q) + p \end{aligned}$$

2) *Résoudre ce système d'équation de récurrence du premier ordre (dans les cas non triviaux, c'est à dire quand p et q sont ni tous les deux égaux à 0, ni tous les deux égaux à 1).*

3) *Faire tendre $n \rightarrow +\infty$ pour obtenir une loi stationnaire.*

Solution de l'exercice.

1) Rappelons que sa matrice de transition est

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

En utilisant la relation $P^{n+1} = P^n P$, on obtient les relations

$$\begin{aligned} p_{1,1}^{(n+1)} &= (1-p)p_{1,1}^{(n)} + q p_{1,2}^{(n)} \\ p_{1,2}^{(n+1)} &= p p_{1,1}^{(n)} + (1-q)p_{1,2}^{(n)} \\ p_{2,1}^{(n+1)} &= (1-p)p_{2,1}^{(n)} + q p_{2,2}^{(n)} \\ p_{2,2}^{(n+1)} &= p p_{2,1}^{(n)} + (1-q)p_{2,2}^{(n)} \end{aligned}$$

D'autre part, on sait que l'on a pour tout n on a :

$$p_{1,1}^{(n)} + p_{1,2}^{(n)} = p_{2,1}^{(n)} + p_{2,2}^{(n)} = 1.$$

On déduit de ces deux systèmes d'équations que l'on a :

$$\begin{aligned} p_{1,1}^{(n+1)} &= (1-p)p_{1,1}^{(n)} + q(1-p_{1,1}^{(n)}) = p_{1,1}^{(n)}(1-p-q) + q \\ p_{1,2}^{(n+1)} &= p(1-p_{1,2}^{(n)}) + (1-q)p_{1,2}^{(n)} = p_{1,2}^{(n)}(1-p-q) + p \\ p_{2,1}^{(n+1)} &= (1-p)p_{2,1}^{(n)} + q(1-p_{2,1}^{(n)}) = p_{2,1}^{(n)}(1-p-q) + q \\ p_{2,2}^{(n+1)} &= p(1-p_{2,2}^{(n)}) + (1-q)p_{2,2}^{(n)} = p_{2,2}^{(n)}(1-p-q) + p \end{aligned}$$

2) Effectuons maintenant un rappel sur la résolution d'une équation de récurrence du premier ordre. Soit donc une équation de récurrence de la forme :

$$x_{n+1} = ax_n + b$$

à résoudre. Résolvons d'abord l'équation sans terme constant, i.e.

$$y_{n+1} = a y_n.$$

La solution est bien évidemment

$$y_n = A a^n.$$

Cherchons maintenant une constante solution de la première équation. Elle doit vérifier l'équation $c = a c + b$ et est donc égale à $b/(1-a)$. Ainsi la solution générale de l'équation est

$$x_n = A a^n + b/(1-a).$$

Ce rappel fait, revenons maintenant à la résolution de nos quatre équations de récurrence et déterminons leurs solutions : $p_{1,1}^{(n)}$, $p_{1,2}^{(n)}$, $p_{2,1}^{(n)}$, $p_{2,2}^{(n)}$. Nous nous intéressons seulement aux cas non dégénérés où les réels p et q sont ni tous les deux égaux à 0, ni tous les deux égaux à 1. La première équation est donc :

$$p_{1,1}^{(n+1)} = p_{1,1}^{(n)}(1-p-q) + q,$$

ce qui, d'après le rappel que nous venons de faire, est de solution de forme générale :

$$p_{1,1}^{(n)} = A(1-p-q)^n + \frac{q}{p+q}.$$

La condition initiale $p_{1,1}^{(0)} = 1$ nous impose d'avoir :

$$1 = A + \frac{q}{p+q}, \text{ et donc } A = \frac{p}{p+q}.$$

Finalement la solution de l'équation cherchée est :

$$p_{1,1}^{(n)} = \frac{p}{p+q} (1-p-q)^n + \frac{q}{p+q}.$$

De la même manière on obtient les solutions des autres équations de récurrence :

$$\begin{aligned} p_{1,2}^{(n)} &= -\frac{p}{p+q} (1-p-q)^n + \frac{p}{p+q} \\ p_{2,1}^{(n)} &= \frac{p}{p+q} (1-p-q)^n + \frac{q}{p+q} \\ p_{2,2}^{(n)} &= -\frac{p}{p+q} (1-p-q)^n + \frac{p}{p+q} \end{aligned}$$

3) Ayant bien sûr $|1-p-q| < 1$, les limites de ces probabilités quand n tend vers $+\infty$ sont données par la matrice :

$$P^\infty = \begin{pmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ \frac{q}{p+q} & \frac{p}{p+q} \end{pmatrix}$$

Le théorème précédent nous permet d'affirmer que la loi de probabilité sur $E = \{1, 2\}$ définie par $\pi = (q/(p+q), p/(p+q))$ est nécessairement une mesure invariante sur E . Constatons au passage que la limite des $p_{i,j}^{(n)}$ ne dépend effectivement pas de i . \diamond

EXERCICE 4.27. *Considérons une chaîne de Markov (X_n) à valeurs dans $E = \{1, 2, 3\}$, dont le graphe de Markov est donné dans la Figure 9 :*

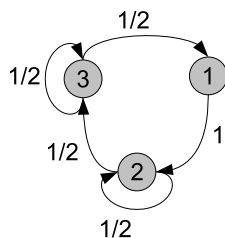


FIGURE 9. Graphe de Markov de l'Exercice 4.27

- 1) Donner la matrice de transition de cette chaîne.
- 2) Trouver une loi stationnaire (Ind. Ici on pourra partir directement de la définition). Est-elle unique ?

Solution de l'exercice.

1) La matrice de passage est

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix}.$$

2) Cherchons s'il existe une ou des lois stationnaires pour une telle chaîne. Une telle loi π doit par définition vérifier l'équation $\pi^*P = \pi^*$, ce qui nous donne le système d'équations :

$$\begin{cases} \pi_1 = \frac{1}{2}\pi_3 \\ \pi_2 = \pi_1 + \frac{1}{2}\pi_2 \\ \pi_3 = \frac{1}{2}\pi_2 + \frac{1}{2}\pi_3 \end{cases} \iff \begin{cases} \pi_1 = \frac{1}{2}\pi_3 \\ \pi_2 = \pi_3 \end{cases} .$$

Comme on doit avoir $\pi_1 + \pi_2 + \pi_3 = 1$, l'unique solution du système est $(1/5, 2/5, 2/5)$. Ainsi il existe une unique loi stationnaire π donnée par $\pi^* = (1/5, 2/5, 2/5)^*$ pour une telle chaîne de Markov sur le triangle. \diamond

10.2. Loi stationnaire et chaîne irréductible.

THÉORÈME 4.42. *Une chaîne de Markov irréductible admet une distribution stationnaire π si, et seulement si, elle est récurrente positive. Dans ce cas, la distribution π est unique et est définie, pour tout i dans E , par :*

$$\pi_i = \frac{1}{\mu_i} .$$

Une chaîne irréductible ne peut donc à la fois être transiente et posséder une distribution stationnaire. Or on sait que la marche aléatoire simple est irréductible et transiente dans le cas où $p \neq 1/2$. Pour une telle chaîne il n'existe donc pas de loi stationnaire.

COROLLAIRE 4.43. *Une chaîne de Markov irréductible sur un espace d'état fini E admet une unique loi stationnaire π . Celle-ci est telle que, pour tout i dans E , on ait :*

$$\pi_i = \frac{1}{\mu_i} .$$

Soit k un état fixé dans E et, pour chaque état i on définit l'espérance du nombre de visites de l'état i entre deux visites de l'état k (on dira plus simplement le nombre moyen de visites... même si ce n'est pas mathématiquement la même chose) par :

$$\rho_i(k) = \mathbb{E}_k \left(\sum_{n=0}^{T_k-1} \mathbb{1}_{\{X_n=i\}} \right) = \mathbb{E} \left(\sum_{n=0}^{T_k-1} \mathbb{1}_{\{X_n=i\}} | X_0 = k \right) .$$

On peut alors montrer que l'on a, pour tout couple (i, k) dans E^2 :

$$\rho_i(k) = \frac{\mu_k}{\mu_i} = \frac{\pi_i}{\pi_k} .$$

10.3. théorèmes limites. On peut maintenant s'attaquer au problème de la convergence de la chaîne de Markov vers une situation d'équilibre. On s'intéresse donc à déterminer les situations où l'on a la convergence, pour tout i dans E , des $p_{i,j}^{(n)}$, pour $j \in E$, quand n tend vers $+\infty$ ou encore à la convergence des $P(X_n = j)$ quand n tend vers $+\infty$.

Il est important dans un premier temps de souligner que la périodicité de la chaîne peut être source de problème dans cet objectif de convergence vers l'équilibre. En effet considérons le cas très simple de la chaîne de Markov à deux états, i.e. à valeurs dans $E = \{1, 2\}$, avec $p_{1,2} = p_{2,1} = 1$. Le graphe d'une telle chaîne est donné dans la Figure 10 :

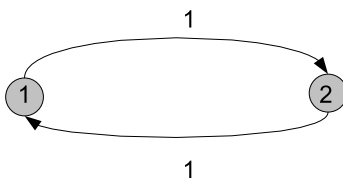


FIGURE 10. Graphe de la Chaîne de Markov à deux états de période 2

Il est évident que l'on a

$$p_{1,1}^{(n)} = p_{2,2}^{(n)} = \begin{cases} 0 & \text{si } n \text{ est impair} \\ 1 & \text{si } n \text{ est pair} \end{cases}$$

et

$$p_{1,2}^{(n)} = p_{2,1}^{(n)} = \begin{cases} 1 & \text{si } n \text{ est impair} \\ 0 & \text{si } n \text{ est pair} \end{cases}$$

On ne peut donc pas espérer une convergence des $p_{i,i}^{(n)}$ vers des limites π_i , pour $i = 1, 2$, et donc une convergence de la chaîne de Markov vers une situation d'équilibre. Et pourtant cette chaîne admet une loi stationnaire : $\pi = (1/2, 1/2)'$.

C'est pourquoi nous allons ici nous restreindre aux chaînes de Markov apériodiques. Mais cela dit, il existe des résultats pour les chaînes périodiques.

DÉFINITION 4.44. *On dit qu'un état d'une chaîne de Markov est ergodique s'il est récurrent positif et apériodique.*

On note que l'ergodicité est une propriété de classe, puisque la récurrence positive et la périodicité le sont. Une chaîne de Markov irréductible récurrente positive et apériodique est dite simplement chaîne de Markov ergodique.

THÉORÈME 4.45. *Soit $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov ergodique (i.e. irréductible récurrente positive et apériodique) de loi initiale π_0 quelconque. On a alors, pour tout j dans E :*

$$P(X_n = j) \rightarrow \pi_j = \frac{1}{\mu_j}, \text{ quand } n \rightarrow +\infty.$$

En particulier, on a pour tout (i, j) dans E^2

$$p_{i,j}^{(n)} \rightarrow \pi_j, \text{ quand } n \rightarrow +\infty.$$

10.4. Inversion du temps. D'après sa définition, une chaîne de Markov est telle que le passé et le futur sont indépendants, conditionnellement au présent. Pour s'en convaincre soient (i, j, k) un triplet d'états d'une chaîne de Markov à valeurs dans une espace E dénombrable. Pour tout entier n , on a bien :

$$\begin{aligned} & P(X_{n+1} = k, X_{n-1} = j | X_n = i) \\ &= P(X_{n+1} = k | X_{n-1} = j, X_n = i) P(X_{n-1} = j | X_n = i) \\ &= P(X_{n+1} = k | X_n = i) P(X_{n-1} = j | X_n = i). \end{aligned}$$

Cette propriété étant symétrique, on se pose naturellement la question de ce que l'on obtiendrait en inversant le temps. Soit donc $(X_n)_{n \in \mathbb{Z}}$ une chaîne de Markov, de matrice de transition P , irréductible, récurrente positive et de loi stationnaire π (on sait qu'il en existe

une et une seule). Supposons que, pour tout n , la loi de X_n soit π . Notons que cette hypothèse est plus forte que de demander que la loi de X_0 soit π (cette dernière ne permettant d'avoir la loi π pour les X_n pour uniquement $n \geq 0$). On définit alors la chaîne inversée en temps $(Y_n)_{n \in \mathbb{Z}}$ par :

$$Y_n = X_{-n},$$

pour $n \in \mathbb{Z}$. Montrons que (Y_n) est alors encore une chaîne de Markov et que pour tout n la loi de Y_n est également π . Que la loi de Y_n soit π pour tout n est trivial. Pour le caractère markovien, on a en effet, pour tout $n \in \mathbb{Z}$, tout $k \in \mathbb{N}$ et tout $(i_{n+1}, i_n, i_{n-1}, \dots, i_{n-k}) \in E^{k+1}$:

$$\begin{aligned} & P(Y_{n+1} = i_{n+1} | Y_n = i_n, Y_{n-1} = i_{n-1}, \dots, Y_{n-k} = i_{n-k}) \\ = & \frac{P(Y_{n+1} = i_{n+1}, Y_n = i_n, Y_{n-1} = i_{n-1}, \dots, Y_{n-k} = i_{n-k})}{P(Y_n = i_n, Y_{n-1} = i_{n-1}, \dots, Y_{n-k} = i_{n-k})} \\ = & \frac{P(X_{-n-1} = i_{n+1}, X_{-n} = i_n, X_{-n+1} = i_{n-1}, \dots, X_{-n+k} = i_{n-k})}{P(X_{-n} = i_n, X_{-n+1} = i_{n-1}, \dots, X_{-n+k} = i_{n-k})} \\ = & \frac{\pi_{i_{n+1}} p_{i_{n+1}, i_n} p_{i_n, i_{n-1}} \cdots p_{i_{n-k+1}, i_{n-k}}}{\pi_{i_n} p_{i_n, i_{n-1}} \cdots p_{i_{n-k+1}, i_{n-k}}} \\ = & \frac{\pi_{i_{n+1}} p_{i_{n+1}, i_n}}{\pi_{i_n}} = \frac{P(X_{-n-1} = i_{n+1}, X_{-n} = i_n)}{P(X_{-n} = i_n)} \\ = & P(Y_{n+1} = i_{n+1} | Y_n = i_n), \end{aligned}$$

et (Y_n) est donc bien une chaîne de Markov.

DÉFINITION 4.46. *On dit que la chaîne de Markov $(X_n)_{n \in \mathbb{Z}}$ est réversible si la chaîne inversée en temps $(Y_n)_{n \in \mathbb{Z}}$ associée à (X_n) a la même matrice de passage que cette dernière.*

THÉORÈME 4.47. *La chaîne de Markov $(X_n)_{n \in \mathbb{Z}}$ est réversible si, et seulement si il existe une probabilité π telle que, pour tout couple (i, j) de E , on a :*

$$\pi_i p_{i,j} = \pi_j p_{j,i}.$$

Une loi de probabilité π , sur E , vérifiant cette propriété est dite loi réversible pour la chaîne (X_n) de matrice de passage P .

Preuve. Notons $Q = (q_{i,j})_{(i,j) \in E^2}$ la matrice de passage de la chaîne $(Y_n)_{n \in \mathbb{Z}}$. On a :

$$\begin{aligned} q_{i,j} &= P(Y_{n+1} = j | Y_n = i) \\ &= P(X_{-n-1} = j | X_{-n} = i) \\ &= \frac{p_{j,i} \pi_j}{\pi_i}, \end{aligned}$$

en utilisant le même genre de raisonnement que ci-dessus. On a alors clairement $q_{i,j} = p_{i,j}$ si, et seulement si : $\pi_i p_{i,j} = \pi_j p_{j,i}$. \square

Un des intérêts de cette propriété de réversibilité est de donner une manière aisée (mais qui ne marche pas toujours) pour déterminer une loi stationnaire associée à une chaîne de Markov.

THÉORÈME 4.48. *Si une chaîne de Markov irréductible (X_n) possède une loi réversible π alors elle est récurrente positive avec pour unique loi stationnaire π .*

Preuve. Supposons que la chaîne irréductible (X_n) possède une loi réversible π . On a alors :

$$\begin{aligned} \sum_{i \in E} \pi_i P_{i,j} &= \sum_{i \in E} \pi_j P_{j,i} \\ &= \pi_j \sum_{i \in E} P_{j,i} = \pi_j, \end{aligned}$$

et on a bien $\pi^* P = \pi^*$. La chaîne possède donc une loi stationnaire. Comme elle est irréductible, on sait qu'elle est récurrente positive et que la loi stationnaire est unique, ce qui achève la démonstration. \square

Voyons maintenant quelques exemples.

EXERCICE 4.28. On considère la chaîne de Markov à valeurs dans $E = \{1, 2, 3\}$ et de matrice de passage

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

- 1) Donner le graphe de Markov.
 - 2) Existe-t-il une loi stationnaire ? Si oui, est-elle unique ?
 - 3) Déterminer la loi stationnaire si cette dernière existe.
 - 4) La chaîne est-elle réversible ?
 - 5) Conclure.
-

Solution de l'exercice.

- 1) Son graphe est donné par la Figure 11.

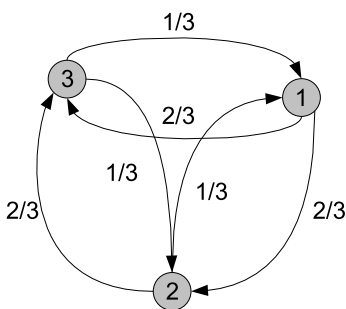


FIGURE 11. Graphe de la chaîne de Markov de l'Exercice 4.28

- 2) La chaîne est irréductible sur un espace E fini. D'après le cours, il existe donc une unique loi stationnaire π .

- 3) Comme la matrice est bistochastique, on a

$$\pi = \frac{1}{\text{Card}E} \mathbb{1} = (1/3, 1/3, 1/3)^*.$$

4) En utilisant la formule obtenue dans le théorème 4.47, la chaîne de Markov (Y_n) inversion en temps de la chaîne (X_n) est de matrice de passage :

$$Q = P^*$$

Comme la matrice P n'est pas symétrique, la chaîne (X_n) n'est pas réversible ($Q \neq P$).

5) Toutes les chaînes ne sont pas réversibles, même celles possédant une loi stationnaire. \diamond

EXERCICE 4.29. Reprenons la chaîne de Markov modélisant la fortune du joueur A contre B (Cf. Exemple 2.4).

1) Déterminer une loi réversible.

2) Conclure.

Solution de l'exercice.

1) Rappelons que l'on a pour une telle chaîne :

$$\begin{cases} p_{i,i+1} = p, \text{ pour } i = 1, \dots, a-1 \\ p_{i,i-1} = q, \text{ pour } i = 1, \dots, a-1 \\ p_{0,0} = p_{a,a} = 1 \\ p_{0,1} = p_{a,a-1} = 0 \end{cases}$$

Si une telle loi réversible π existe, elle doit, d'après le théorème 4.47, vérifier le système d'équations :

$$\begin{cases} \pi_i p = \pi_{i+1} q \iff \pi_{i+1} = \left(\frac{p}{q}\right) \pi_i, \text{ pour } i = 1, \dots, a-2 \\ 0 = \pi_1 q = \pi_{a-1} p \end{cases},$$

ce qui nous donne $\pi = (\rho, 0, \dots, 0, 1 - \rho)^*$, où $\rho \in [0, 1]$.

2) On retrouve bien la loi stationnaire calculée précédemment. \diamond

EXERCICE 4.30. Intéressons nous maintenant à une chaîne de Markov, très proche de la précédente, toujours à valeurs dans $\{0, 1, \dots, a-1, a\}$ mais cette fois-ci avec matrice de transition :

$$P = \begin{pmatrix} q & p & 0 & \cdots & \cdots & \cdots & 0 \\ q & 0 & p & 0 & \cdots & \cdots & \vdots \\ 0 & q & 0 & p & 0 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & q & 0 & p & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & q & 0 & p \\ 0 & \vdots & \vdots & \vdots & 0 & q & p \end{pmatrix}$$

On a donc dans ce cas :

$$\begin{cases} p_{i,i+1} = p, \text{ pour } i = 0, \dots, a-1 \\ p_{i,i-1} = q, \text{ pour } i = 1, \dots, a \\ p_{0,0} = q \text{ et } p_{a,a} = p \end{cases}$$

Déterminer une loi stationnaire si elle existe. (Ind. On pourra dissocier les résultats suivant les valeurs de p)

Solution de l'exercice.

Une loi réversible pour une telle chaîne, si elle existe, doit vérifier le système d'équations :

$$\pi_i p = \pi_{i+1} q, \text{ pour } i = 0, \dots, a-1,$$

ce qui nous donne

$$\pi_i = \pi_0 \left(\frac{p}{q} \right)^i.$$

Cherchant une loi de probabilité, il nous faut prendre π_0 de manière à avoir

$$\sum_{i=0}^a \pi_i = 1,$$

i.e

$$\begin{cases} \pi_0 \frac{1 - (\frac{p}{q})^{a+1}}{1 - \frac{p}{q}} = 1 \text{ si } p \neq q \\ (a+1)\pi_0 = 1 \text{ si } p = q = 1/2 \end{cases}$$

ce qui nous donne

$$\begin{cases} \pi_0 = \frac{1 - \frac{p}{q}}{1 - (\frac{p}{q})^{a+1}} \text{ si } p \neq q \\ \pi_0 = \frac{1}{a+1} \text{ si } p = q = 1/2 \end{cases}$$

et donc la loi réversible, et également stationnaire d'après le théorème précédent, est :

$$\pi = \frac{1 - \frac{p}{q}}{1 - (\frac{p}{q})^{a+1}} (1, p/q, (p/q)^2, \dots, (p/q)^a)^*,$$

dans le cas où $p \neq q$ et

$$\pi = \frac{1}{a+1} (1, \dots, 1)^*,$$

dans le cas où $p = q$.

On serait tenté de dire que, si p est plus grand que q , la chaîne se décale progressivement vers a et que par conséquent la chaîne inversée en temps irait, elle, plutôt vers 0. Mais seulement la première affirmation est justifiée. En effet, la chaîne inversée n'est considérée qu'en état d'équilibre pour la chaîne initiale, qui est, dans le cas où p est largement plus grand que q , concentrée autour de la valeur a (et par symétrie évidente autour de 0 si p est largement plus petit que q). La chaîne reste ainsi le plus souvent proche de l'état a , tout en faisant de brèves excursions vers la gauche (i.e. en direction de 0). Ce comportement est bien sûr symétrique en temps, ce qui signifie que la chaîne inversée reste elle aussi le plus souvent proche de a . \diamond

Donnons une interprétation intuitive de la stationnarité et de la réversibilité. Soit (X_n) une chaîne de Markov à espace d'états E . Supposons que l'on observe un (très !) grand nombre de trajectoires de cette chaîne en état d'équilibre. À un instant n , la proportion de

trajectoires étant dans l'état i est donc π_i où π est la loi d'équilibre. À l'instant suivant, soit $n + 1$, la proportion de ceux qui partent de i (ce qui ne veut pas forcément dire qu'il le quitte) est donc π_i . La proportion de ceux qui arrivent à l'état i (même si ils y étaient déjà) est :

$$\sum_j \pi_j p_{j,i}.$$

Cette chaîne est réellement en équilibre ou en état stationnaire si ces deux proportions sont égales. C'est à dire si l'on a bien :

$$\pi_i = \sum_j \pi_j p_{j,i}.$$

La propriété de réversibilité est, elle, plus exigeante. Elle demande que la proportion de ceux qui quittent l'état i pour aller vers l'état j soit la même que celle de ceux qui font le trajet inverse, i.e. $\pi_i p_{i,j} = \pi_j p_{j,i}$. On pourrait parler d'équilibre local, ce qui est bien plus fort que l'équilibre (global) considéré précédemment.

10.5. Théorie ergodique. La notion de théorie ergodique peut être utilisée en référence à un comportement limite de la chaîne mais aussi en référence à un comportement limite des moyennes sur une période. Plus précisément on s'intéresse pour une chaîne en régime stationnaire, à la proportion de temps passé dans chaque état. C'est ce que nous voulons étudier dans cette section.

Considérons $N_i(n)$ le nombre de visites de l'état i qu'effectue la chaîne (X_n) avant le temps n , i.e.

$$N_i(n) = \sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=i\}}.$$

et $N_i(n)/n$ la proportion de temps passé par la chaîne dans l'état i avant le temps n .

THÉORÈME 4.49. *Soit (X_n) une chaîne de Markov à valeurs dans un espace dénombrable E , irréductible et de loi initiale π_0 quelconque. On a alors, pour tout état i de E , on a p.s. :*

$$\lim_{n \rightarrow +\infty} \frac{N_i(n)}{n} = \frac{1}{\mu_i}$$

où $\mu_i = \mathbb{E}_i(T_i)$ est toujours le temps moyen de retour à l'état i .

De plus, si la chaîne est récurrente positive, alors pour toute fonction f bornée de E vers \mathbb{R} , on a p.s.

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = \bar{f},$$

où

$$\bar{f} = \sum_{i \in E} \pi_i f(i) = \int f d\pi$$

et π est l'unique distribution stationnaire associée à la chaîne.

11. Bibliographie

- Dominique Foata et Aimé Fuchs. *Processus Stochastiques*. Dunod 2004.
- Valérie Girardin et Nikolaos Limnios. *Probabilités*. Vuibert 2001.
- Karlyn, S. and H. Taylor. *A First Course in Stochastic Processes*. Academic Press, San Diego, 1975.

CHAPITRE 5

Espérances conditionnelles

1. Introduction

Soit (Ω, \mathcal{A}, P) un espace probabilisé et B un événement de la tribu \mathcal{A} tel que : $p(B) > 0$. On sait définir la probabilité conditionnelle à B , notée $P(\cdot|B)$, par :

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Si X est une v.a. positive, on peut naturellement définir son espérance conditionnelle à B par :

$$\mathbb{E}(X|B) = \int X(\omega) dP(\omega|B),$$

c'est à dire l'espérance de X prise par rapport à la probabilité conditionnelle $P(\cdot|B)$. On peut montrer que cette espérance conditionnelle vérifie l'égalité (parfois utile pour les calculs) :

$$(1) \quad \mathbb{E}(X|B) = \frac{\mathbb{E}(X \mathbb{1}_B)}{P(B)} = \frac{1}{P(B)} \int_B X(\omega) dP(\omega).$$

En effet cette formule est trivialement vérifiée pour toute v.a. indicatrice de la forme $X = \mathbb{1}_A$. Ce n'est rien d'autre que la formule classique des probabilités conditionnelles rappelée ci-dessus. Grâce à la linéarité de l'espérance, cette formule est donc vérifiée également pour toute fonction étagée de la forme

$$X_n = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i},$$

où les α_i sont des réels positifs et les A_i des événements disjoints de la tribu \mathcal{A} . On sait que toute fonction mesurable positive peut être écrite comme limite croissante de fonctions étagées positives. Ainsi, toujours par linéarité de l'espérance et avec le soutien supplémentaire du théorème de convergence monotone, la formule (1) est bien vraie pour toute v.a. positive.

Pour toute v.a. intégrable (et non nécessairement positive), on peut alors utiliser l'équation (1) pour définir son espérance conditionnelle à B . On écrit en effet $X = X^+ - X^-$ avec $X^+ = \max(X, 0)$ et $X^- = \max(-X, 0)$. D'après ce qui précède, on sait définir et les espérances conditionnelles $\mathbb{E}(X^+|B)$ et $\mathbb{E}(X^-|B)$. Elles vérifient l'équation (1), ce qui, puisque X est supposée intégrable, assure qu'elles sont toutes les deux finies. On peut alors définir $\mathbb{E}(X|B)$ par $\mathbb{E}(X|B) = \mathbb{E}(X^+|B) - \mathbb{E}(X^-|B)$ et avoir la formule (1) également vérifiée pour une v.a. intégrable.

Considérons maintenant une v.a. Y discrète, i.e. à valeurs dans un ensemble $E = \{y_1, y_2, \dots, y_k, \dots\}$ au plus dénombrable. D'après ce qui précède on peut définir, pour toute variable X intégrable, l'espérance conditionnelle

$$\mathbb{E}(X|\{Y = y_k\}),$$

pour tout y_k de E . On a ici aussi, comme précédemment :

$$(2) \quad \mathbb{E}(X|\{Y = y_k\}) = \frac{\mathbb{E}(X \mathbb{1}_{\{Y=y_k\}})}{P(Y = y_k)} = \frac{1}{P(Y = y_k)} \int_{\{Y=y_k\}} X(\omega) dP(\omega).$$

Notons alors h la fonction de E vers \mathbb{R} définie par :

$$h(y_k) = \mathbb{E}(X|\{Y = y_k\})$$

et considérons $h(Y)$ la composée de h avec la v.a. Y . On peut alors écrire, pour tout événement B sur la tribu \mathcal{E} mise sur E :

$$\begin{aligned}\mathbb{E}(h(Y)\mathbb{1}_{\{Y \in B\}}) &= \int_{\{\omega: Y(\omega) \in B\}} h(Y(\omega)) dP(\omega) = \int_{\{y_k \in B\}} h(y_k) dP_Y(y_k) \\ &= \sum_{y_k \in B} h(y_k) P(Y = y_k) = \sum_{y_k \in B} \mathbb{E}(X \mathbb{1}_{\{Y=y_k\}}) = \mathbb{E}(X \mathbb{1}_{\{Y \in B\}}),\end{aligned}$$

où la seconde égalité est due au théorème du transport et la quatrième à l'équation (2).

Ainsi, cette fonction $h(Y)$, que l'on note $\mathbb{E}(X|Y)$, est d'une part, d'après le lemme de Doob, une fonction $\sigma(Y)$ mesurable (où $\sigma(Y)$ est la tribu engendrée par la v.a. Y) et d'autre part telle que :

$$\mathbb{E}(\mathbb{E}(X|Y)\mathbb{1}_A) = \mathbb{E}(X\mathbb{1}_A),$$

pour tout A de $\sigma(Y)$. Nous verrons que cette propriété est caractéristique de l'espérance conditionnelle que nous construirons justement par cette double propriété.

2. Espérance conditionnelle

Sur un espace probabilisé (Ω, \mathcal{A}, P) , on considère une tribu \mathcal{B} incluse dans \mathcal{A} telle que les négligeables de \mathcal{A} soient aussi ceux de \mathcal{B} (i.e. ceux de \mathcal{A} sont inclus dans la tribu \mathcal{B}).

Soit $L^2(\Omega, \mathcal{A}, P)$ la classe des v.a.r. \mathcal{A} -mesurables et de carré intégrables. Rappelons que les éléments de cette classe sont définis à une égalité *p.s.* près. Rappelons également que cet espace est un espace de Hilbert pour le produit scalaire défini par :

$$\langle X, Y \rangle_{L^2} = \mathbb{E}(XY),$$

pour X et Y dans $L^2(\Omega, \mathcal{A}, P)$.

L'espace $L^2(\Omega, \mathcal{B}, P)$ est un sous espace de Hilbert fermé dans $L^2(\Omega, \mathcal{A}, P)$. On peut ainsi définir l'espérance conditionnelle pour des v.a.r. de carré intégrables grâce à la notion de projection orthogonale dans les espaces de Hilbert.

DÉFINITION 5.1. *Soit X une v.a.r. de $L^2(\Omega, \mathcal{A}, P)$. On appelle espérance conditionnelle de X conditionnelle à \mathcal{B} , notée $\mathbb{E}(X|\mathcal{B})$, la projection orthogonale de X sur $L^2(\Omega, \mathcal{B}, P)$.*

On peut caractériser l'espérance conditionnelle grâce au théorème suivant.

THÉORÈME 5.2. *L'espérance $\mathbb{E}(X|\mathcal{B})$ est l'unique v.a.r. Z de carré intégrable et \mathcal{B} -mesurable telle que*

$$\forall Y \in L^2(\Omega, \mathcal{B}, P) : \mathbb{E}(XY) = \mathbb{E}(ZY).$$

Preuve. Les propriétés de la projection sur les espaces de Hilbert permettent d'affirmer que la projection Z de X sur $L^2(\Omega, \mathcal{B}, P)$ est l'unique élément de $L^2(\Omega, \mathcal{B}, P)$ tel que

$$X - Z \perp_{L^2} L^2(\Omega, \mathcal{B}, P)$$

i.e. Z est l'unique élément de $L^2(\Omega, \mathcal{B}, P)$ tel que

$$\begin{aligned}\langle X - Z, Y \rangle_{L^2} &= 0, \quad \forall Y \in L^2(\Omega, \mathcal{B}, P) \\ \iff \mathbb{E}(XY) &= \mathbb{E}(ZY), \quad \forall Y \in L^2(\Omega, \mathcal{B}, P),\end{aligned}$$

ce qui achève la preuve. □

Il est important de bien noter que l'espérance conditionnelle $\mathbb{E}(X|\mathcal{B})$ n'est définie qu'à une égalité *p.s.* près (plus précisément l'unicité n'est vraie qu'à une égalité *p.s.* près).

Considérons maintenant un cas particulier important. Supposons que la tribu \mathcal{B} soit la tribu $\sigma(Y)$ engendrée par une v.a. Y à valeurs dans un espace E . On sait, d'après le Lemme de Doob, que toute v.a. $\sigma(Y)$ -mesurable s'écrit sous la forme $h(Y)$ où h une fonction mesurable de E vers \mathbb{R} . Ainsi $\mathbb{E}(X|\sigma(Y))$, que l'on note plus simplement $\mathbb{E}(X|Y)$, s'écrit aussi sous la forme $h(Y)$.

PROPOSITION 5.3. (*Propriétés élémentaires de l'espérance conditionnelle*)

- (1) *L'espérance conditionnelle est linéaire ;*
- (2) *Si on a $X = x$ p.s., pour x fixé de \mathbb{R} , alors $\mathbb{E}(X|\mathcal{B}) = x$;*
- (3) *Si la v.a.r. X est indépendante de la tribu \mathcal{B} , alors $\mathbb{E}(X|\mathcal{B}) = E(X)$;*
- (4) *Si Z est une variable \mathcal{B} -mesurable, alors on a p.s. : $\mathbb{E}(ZX|\mathcal{B}) = Z\mathbb{E}(X|\mathcal{B})$;*
- (5) *Pour toute v.a.r. X positive, on a p.s. : $\mathbb{E}(X|\mathcal{B}) \geq 0$;*
- (6) *On a p.s. : $|\mathbb{E}(X|\mathcal{B})| \leq \mathbb{E}(|X||\mathcal{B})$*
- (7) *On a : $\mathbb{E}(\mathbb{E}(X|\mathcal{B})) = \mathbb{E}(X)$.*

EXERCICE 5.1. Démontrer cette proposition. (*Ind. On utilisera très souvent la caractérisation donnée par le Théorème 5.2. Pour le point 5, on pourra, en utilisant la propriété de l'espérance conditionnelle, montrer que $P(\mathbb{E}(X|\mathcal{B}) < -1/n) = 0$, pour tout $n \in \mathbb{N}$. Pour le point 6, on pourra considérer les parties positives X^+ et négatives X^- de X*)

Preuve/solution de l'exercice.

1) Évident puisque l'espérance conditionnelle est une projection. On peut également le retrouver en utilisant le Théorème 5.2.

2) La v.a.r. Z égale p.s. à x est bien dans $L^2(\Omega, \mathcal{B}, P)$ et telle que $\mathbb{E}(XY) = \mathbb{E}(ZY)$ pour tout Y de $L^2(\Omega, \mathcal{B}, P)$. Elle convient donc comme espérance conditionnelle et, par unicité, c'est uniquement elle.

3) On prend ici $Z = \mathbb{E}(X)$ qui est aussi une variable de $L^2(\Omega, \mathcal{B}, P)$. Par indépendance, on a bien :

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(X)Y),$$

pour tout Y de $L^2(\Omega, \mathcal{B}, P)$. La v.a.r. Z ainsi définie convient bien et par unicité, c'est bien elle.

4) On note $Z_1 = Z\mathbb{E}(X|\mathcal{B})$. Cette v.a.r. est bien \mathcal{B} -mesurable et de carré intégrable puisque Z est bornée et que $\mathbb{E}(X|\mathcal{B})$ est dans $L^2(\Omega, \mathcal{B}, P)$. Pour tout Y de $L^2(\Omega, \mathcal{B}, P)$, on a :

$$\mathbb{E}(Z_1Y) = \mathbb{E}(YZ\mathbb{E}(X|\mathcal{B})) = \mathbb{E}(YZX),$$

puisque YZ est \mathcal{B} -mesurable et de carré intégrable. Par unicité on a alors :

$$\mathbb{E}(ZX|\mathcal{B}) = Z\mathbb{E}(X|\mathcal{B}).$$

Cette propriété souligne que, conditionnellement à \mathcal{B} , la v.a.r. Z est connue et se comporte alors comme une constante dans l'espérance conditionnelle.

5) Notons $Z = \mathbb{E}(X|\mathcal{B})$ et $A_n = \{Z < -1/n\}$ qui est un événement de \mathcal{B} . La fonction indicatrice de cet événement est donc \mathcal{B} -mesurable et on peut écrire :

$$0 \leq \mathbb{E}(X\mathbb{1}_{A_n}) = \mathbb{E}(Z\mathbb{1}_{A_n}) \leq -\frac{1}{n}P(A_n) \leq 0.$$

On a donc nécessairement $P(A_n) = 0$, pour tout n de \mathbb{N} . Par croissance de la suite des événements (A_n) on a :

$$P(\cup_n A_n) = 0 \iff P(Z < 0) = 0$$

et la v.a.r. Z est donc bien *p.s.* positive.

6) On sait que l'on peut écrire : $X = X^+ - X^-$ et $|X| = X^+ + X^-$. Ainsi, il vient :

$$|\mathbb{E}(X|\mathcal{B})| = |\mathbb{E}(X^+|\mathcal{B}) - \mathbb{E}(X^-|\mathcal{B})| \leq |\mathbb{E}(X^+|\mathcal{B})| + |\mathbb{E}(X^-|\mathcal{B})| = \mathbb{E}(X^+ + X^-|\mathcal{B}) = \mathbb{E}(|X||\mathcal{B}).$$

7) On sait que l'on doit avoir pour tout Y de $L^2(\Omega, \mathcal{B}, P)$:

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

Il suffit de prendre pour Y la v.a.r. *p.s.* égale à 1 (qui est bien dans $L^2(\Omega, \mathcal{B}, P)$) pour obtenir le résultat. \square

On peut maintenant étendre la notion d'espérance conditionnelle au cas des v.a.r. positives ou intégrables et non nécessairement dans $L^2(\Omega, \mathcal{A}, P)$. Commençons par le cas des variables positives.

THÉORÈME 5.4. Notons $\bar{L}_+(\Omega, \mathcal{A}, P)$ l'espace des classes d'équivalence des v.a. de (Ω, \mathcal{A}, P) vers $\bar{\mathbb{R}}^+ = [0, +\infty]$ et $\bar{L}_+(\Omega, \mathcal{B}, P)$ le sous espace de ces classes d'équivalence qui soient \mathcal{B} -mesurables.

Pour tout X de $\bar{L}_+(\Omega, \mathcal{A}, P)$, il existe un élément unique de $\bar{L}_+(\Omega, \mathcal{B}, P)$ tel que :

$$\mathbb{E}(ZY) = \mathbb{E}(XY),$$

pour tout Y de $\bar{L}_+(\Omega, \mathcal{B}, P)$. La v.a. Z est alors notée $\mathbb{E}(X|\mathcal{B})$.

On est obligé de considérer des v.a. non nécessairement *p.s.* finies car on sait que même si une v.a. est *p.s.* finie, son espérance conditionnelle ne l'est pas forcément.

EXERCICE 5.2. Démonstration de ce théorème.

On considère la suite de v.a. positives (X_n) définies pour tout n par $X_n = X \wedge n = \inf\{X, n\}$.

1) Montrer que l'on peut définir, pour tout n , la v.a. Z_n telle que : $Z_n = \mathbb{E}(X_n|\mathcal{B})$.

2) Montrer que la suite (Z_n) est une suite de v.a. positives, croissante vers une v.a. Z positive et \mathcal{B} -mesurable.

3) Montrer que l'on a, pour toute v.a. Y positive et \mathcal{B} -mesurable :

$$\mathbb{E}(X_n Y) = \mathbb{E}(Z_n Y).$$

(Ind. on pourra considérer la suite (Y_p) où $Y_p = Y \wedge p$)

4) En déduire que l'on a, pour tout Y positive et \mathcal{B} -mesurable :

$$\mathbb{E}(XY) = \mathbb{E}(ZY).$$

5) On s'attache maintenant à prouver l'unicité de Z ainsi définie. Supposons qu'il existe Z_1 et Z_2 vérifiant le point 4 précédent. Montrer, en utilisant ce dernier point, que l'on a nécessairement $P(Z_1 \leq a < b \leq Z_2) = 0$, pour tout couple de réels (a, b) tel que $0 \leq a < b$.

6) En déduire que l'on a nécessairement : $P(Z_1 \neq Z_2) = 0$ et donc bien l'unicité de l'espérance conditionnelle.

Preuve/solution de l'exercice.

1) La v.a. X_n est, pour tout n , positive et bornée. Elle est donc dans $L^2(\Omega, \mathcal{A}, P)$. Par définition de l'espérance conditionnelle dans L^2 , il existe donc, pour tout n , une unique v.a. $Z_n = \mathbb{E}(X_n|\mathcal{B})$ de $L^2(\Omega, \mathcal{B}, P)$.

2) Comme (X_n) est une suite de v.a. positives *p.s.* croissante, par la propriété 5 de la Proposition 5.3, la suite (Z_n) est également une suite de v.a. positives, \mathcal{B} -mesurables. Elle est croissante vers une v.a. Z positive et \mathcal{B} -mesurable (comme suite de v.a. \mathcal{B} -mesurables).

3) Les v.a. (Y_p) sont dans $L^2(\Omega, \mathcal{B}, P)$. D'après la définition de l'espérance conditionnelle, on a donc, pour tout n et tout p : $\mathbb{E}(X_n Y_p) = \mathbb{E}(Z_n Y_p)$. La croissance de Y_p vers Y quand p tend vers $+\infty$ (et la suite X_n étant bornée) assure la convergence monotone croissante de $X_n Y_p$ vers $X_n Y$ et celle de $Z_n Y_p$ vers $Z_n Y$. Par le théorème de convergence monotone on a donc : $\mathbb{E}(X_n Y) = \mathbb{E}(Z_n Y)$, pour tout n .

4) On fait exactement le même raisonnement qu'au point précédent en faisant tendre cette fois-ci n vers $+\infty$.

5) La v.a. $\mathbb{1}_{\{Z_1 \leq a < b \leq Z_2\}}$ étant \mathcal{B} -mesurable (puisque Z_1 et Z_2 le sont), on peut écrire :

$$aP(Z_1 \leq a < b \leq Z_2) \geq \mathbb{E}(Z_1 \mathbb{1}_{\{Z_1 \leq a < b \leq Z_2\}}) = \mathbb{E}(Z_2 \mathbb{1}_{\{Z_1 \leq a < b \leq Z_2\}}) \geq bP(Z_1 \leq a < b \leq Z_2),$$

l'égalité du centre étant assurée par la propriété de l'espérance conditionnelle. On aboutit à une contradiction sauf si $P(Z_1 \leq a < b \leq Z_2) = 0$.

6) On peut écrire :

$$\{Z_1 < Z_2\} = \bigcup_{(a,b) \in \mathbb{Q}^2, a < b} \{Z_1 \leq a < b \leq Z_2\}.$$

L'événement $\{Z_1 < Z_2\}$ est donc négligeable par union dénombrable de négligeables. Par symétrie, on a $P(Z_1 > Z_2) = 0$ et donc $P(Z_1 \neq Z_2) = 0$. \square

COROLLAIRE 5.5. *Avec les mêmes notations qu'au Théorème 5.4, pour avoir $Z = \mathbb{E}(X|\mathcal{B})$ il suffit d'avoir l'une des conditions équivalentes suivantes :*

- l'égalité $\mathbb{E}(ZY) = \mathbb{E}(XY)$, pour toute v.a. r Y bornée et \mathcal{B} -mesurable ;
- l'égalité $\mathbb{E}(Z\mathbb{1}_B) = \mathbb{E}(X\mathbb{1}_B)$, pour tout événement B de \mathcal{B} .

EXERCICE 5.3. *Démonstration de ce Corollaire.***Preuve/solution de l'exercice.**

1) Soit Y une v.a. quelconque de $\bar{L}_+(\Omega, \mathcal{B}, P)$ et définissons $Y_n = Y \wedge n$. Par hypothèse, on a donc :

$$\mathbb{E}(ZY_n) = \mathbb{E}(XY_n)$$

Par le théorème de convergence monotone on obtient, en faisant tendre n vers $+\infty$, la relation $\mathbb{E}(ZY) = \mathbb{E}(XY)$.

2) Supposons que Y soit une v.a. r \mathcal{B} -mesurable, positive et bornée. On sait que Y peut être approchée par une suite croissante de fonctions étagées de la forme (avec des notations qui se comprennent aisément) :

$$Y_n = \sum_{i=1}^{p_n} \alpha_i^n \mathbb{1}_{B_i^n}.$$

L'hypothèse et la linéarité de l'espérance conditionnelle assure que l'on a pour tout n l'égalité $\mathbb{E}(ZY_n) = \mathbb{E}(XY_n)$. Une nouvelle application du théorème de convergence monotone nous remet dans le cas 1) et achève donc la preuve. \square

Enfin on peut définir l'espérance conditionnelle pour des v.a.r. intégrables (et non nécessairement positives).

THÉORÈME 5.6. *Soit X une v.a.r. de $L^1(\Omega, \mathcal{A}, P)$. Il existe une unique v.a.r. Z de $L^1(\Omega, \mathcal{B}, P)$ telle que pour tout v.a.r. \mathcal{B} -mesurable bornée Y on ait : $\mathbb{E}(XY) = \mathbb{E}(ZY)$. La v.a.r. Z ainsi définie est appelée espérance conditionnelle à \mathcal{B} et est notée $\mathbb{E}(X|\mathcal{B})$.*

EXERCICE 5.4. *Démonstration de ce théorème.*

1) On considère la suite de v.a.r. (X_n) définie pour tout n par :

$$X_n = (X \wedge n) \vee (-n),$$

où $a \wedge b = \min\{a, b\}$ et $a \vee b = \max\{a, b\}$. Montrer que l'on peut définir pour tout n la v.a.r. $Z_n = \mathbb{E}(X_n|\mathcal{B})$.

2) Montrer que l'on a l'inégalité $\mathbb{E}|Z_n - Z_m| \leq \mathbb{E}|X_n - X_m|$ et en déduire que la suite (Z_n) converge dans L^1 vers une v.a.r. Z de $L^1(\Omega, \mathcal{B}, P)$

3) Vérifier que la v.a.r. Z vérifie la propriété de l'espérance conditionnelle, i.e. que l'on a : $\mathbb{E}(XY) = \mathbb{E}(ZY)$ pour toute v.a. \mathcal{B} -mesurable bornée Y .

4) Établir l'unicité (Ind. On pourra utiliser des ensemble de la forme $A_n = \{Z_1 + 1/n < Z_2\}$, où Z_1 et Z_2 sont deux candidats à être l'espérance conditionnelle $\mathbb{E}(X|\mathcal{B})$).

Preuve/solution de l'exercice.

1) Les v.a.r. de la suite (X_n) sont bornées et donc toutes dans $L^2(\Omega, \mathcal{A}, P)$. On peut donc définir $Z_n = \mathbb{E}(X_n|\mathcal{B})$, pour tout n .

2) On peut écrire, pour tout n et m :

$$\begin{aligned} \mathbb{E}|\mathbb{E}(X_n|\mathcal{B}) - \mathbb{E}(X_m|\mathcal{B})| &= \mathbb{E}|\mathbb{E}(X_n - X_m|\mathcal{B})| \\ &\leq \mathbb{E}(\mathbb{E}(|X_n - X_m||\mathcal{B})) = \mathbb{E}|X_n - X_m|. \end{aligned}$$

De plus, la convergence *p.s.* de X_n vers X et l'inégalité $|X_n - X| \leq 2|X|$ permettent d'appliquer le théorème de convergence dominée pour obtenir la convergence dans L^1 de X_n vers X . La suite (X_n) est donc de Cauchy dans L^1 et par l'inégalité précédente, la suite (Z_n) l'est aussi. L'espace L^1 étant complet (puisque de Banach), on a également la convergence dans L^1 de Z_n vers une limite que nous noterons Z .

3) On sait que les v.a.r. X_n sont dans $L^2(\Omega, \mathcal{A}, P)$. De plus, toute v.a.r. \mathcal{B} -mesurable bornée Y est également dans $L^2(\Omega, \mathcal{A}, P)$. Par propriété de l'espérance conditionnelle dans L^2 , on a donc : $\mathbb{E}(X_n Y) = \mathbb{E}(Z_n Y)$.

Maintenant, la v.a. Y étant supposée bornée, les convergences vues précédemment entraînent celles de $X_n Y$ vers XY et de $Z_n Y$ vers ZY , toujours dans L^1 . On en déduit que

$$\mathbb{E}(X_n Y) \rightarrow \mathbb{E}XY \text{ et } \mathbb{E}(Z_n Y) \rightarrow \mathbb{E}ZY,$$

quand $n \rightarrow +\infty$. Les termes des deux séries étant égaux pour tout n , on a bien le résultat annoncé.

4) Soient Z_1 et Z_2 deux v.a.r. de $L^1(\Omega, \mathcal{B}, P)$ telles que :

$$\mathbb{E}(XY) = \mathbb{E}(Z_1Y) = \mathbb{E}(Z_2Y),$$

pour tout Y de $L^1(\Omega, \mathcal{B}, P)$. En définissant les événements $A_n = \{Z_1 + 1/n < Z_2\}$ pour tout n , les v.a.r. $\mathbb{1}_{A_n}$ sont \mathcal{B} -mesurables bornées pour tout n et on a donc

$$\mathbb{E}(Z_2\mathbb{1}_{A_n}) = \mathbb{E}(Z_1\mathbb{1}_{A_n}) \leq \mathbb{E}((Z_2 - 1/n)\mathbb{1}_{A_n}) = \mathbb{E}Z_2 - \frac{1}{n}P(A_n).$$

Mais comme la v.a.r. Z_2 est dans $L^1(\Omega, \mathcal{B}, P)$, l'espérance $\mathbb{E}(Z_2\mathbb{1}_{A_n})$ est finie. L'inégalité précédente implique alors que l'on a : $P(A_n) = 0$, pour tout n . On en déduit que l'événement $\cup_n A_n$ est également négligeable et donc que l'on a $P(Z_1 < Z_2) = 0$. Par symétrie on a $P(Z_1 > Z_2) = 0$ et donc $P(Z_1 \neq Z_2) = 0$, ce qui prouve bien l'unicité de l'espérance conditionnelle dans L^1 . \square

PROPOSITION 5.7. *Soit X une v.a.r. de $L^1(\Omega, \mathcal{A}, P)$ et \mathcal{C} une classe de fonctions \mathcal{B} -mesurables bornées stable par multiplication qui engendre la tribu \mathcal{B} (i.e. telle que $\sigma(\mathcal{C}) = \mathcal{B}$).*

Alors, pour qu'un élément Z de $L^1(\Omega, \mathcal{A}, P)$ soit égal à l'espérance conditionnelle $\mathbb{E}(X|\mathcal{B})$, il suffit que l'on ait la relation $\mathbb{E}(XY) = \mathbb{E}(ZY)$ pour tout Y de \mathcal{C} .

EXERCICE 5.5. *Démontrer ce théorème.*

Preuve/solution de l'exercice. Ce n'est qu'une utilisation du théorème des classes monotones (version fonctionnelle) rappelé au Chapitre 1. En effet, notons

$$\mathcal{H} = \{Y \text{ v.a.r. bornée} : \mathbb{E}(XY) = \mathbb{E}(ZY)\}.$$

Il s'agit d'un espace vectoriel qui contient les constantes et qui est stable par limite croissante bornée puisque, si (Y_n) est une suite de v.a. croissante vers Y bornée, le théorème de convergence dominée nous assure les convergences

$$\mathbb{E}(XY_n) \rightarrow \mathbb{E}(XY) \text{ et } \mathbb{E}(ZY_n) \rightarrow \mathbb{E}(ZY),$$

quand $n \rightarrow +\infty$ et la limite Y est alors également bien dans \mathcal{H} .

Par hypothèse l'ensemble \mathcal{H} contient l'ensemble \mathcal{C} . Par le théorème des classes monotones, l'espace \mathcal{H} contient toutes les fonctions $\sigma(\mathcal{C})$ -mesurables bornées, c'est à dire toute les v.a. \mathcal{B} -mesurables bornées, et Z est donc bien l'espérance conditionnelle $\mathbb{E}(X|\mathcal{B})$ grâce au Théorème 5.6. \square

Cette proposition permet souvent de simplifier les calculs en vérifiant la propriété caractéristique pour les espérances conditionnelles en considérant seulement une partie (génératrice) des fonctions \mathcal{B} -mesurables bornées. Il en est ainsi par exemple dans les cas suivants :

- Si \mathcal{C} est un ensemble d'événements de \mathcal{A} qui engendre \mathcal{B} (i.e. $\mathcal{B} = \sigma(\mathcal{C})$), alors il suffit de vérifier $\mathbb{E}(X\mathbb{1}_C) = \mathbb{E}(Z\mathbb{1}_C)$ pour tout C de \mathcal{C} .

- Si $\mathcal{B} = \sigma(U)$, où U est une v.a.r sur (Ω, \mathcal{A}, P) , alors il suffit de vérifier $\mathbb{E}(XY) = \mathbb{E}(ZY)$ pour toutes les fonctions Y de la forme $Y = \exp(i\alpha U)$ pour α dans \mathbb{R} , ou bien pour toutes les fonctions Y de la forme $Y = \mathbb{1}_{U \in [a, b]}$, pour tout intervalle $[a, b]$, ou bien encore pour toutes fonctions Y de la forme $Y = \mathbb{1}_{\{U \geq a\}}$, pour a dans \mathbb{R} (ou seulement dans \mathbb{Q})
- Si $\mathcal{B} = \sigma(U)$, où U est une v.a.r bornée (resp. p.s. positive) sur (Ω, \mathcal{A}, P) , il suffit de vérifier la propriété pour toutes les fonctions Y de la forme $Y = \exp(-\alpha U)$ pour α dans \mathbb{R} (resp. dans \mathbb{R}^+).

Notons que les propriétés élémentaires mentionnées dans la Proposition 5.3 au sujet de l'espérance conditionnelle dans L^2 restent également vraies pour les v.a. positive ou dans L^1 .

Nous allons maintenant voir une série de nouvelles propriétés pour l'espérance conditionnelle, dont certaines montrent que l'espérance conditionnelle se comporte souvent de manière très analogue à l'espérance "classique".

PROPOSITION 5.8. *L'espérance conditionnelle vérifie les propriétés suivantes.*

- (1) (**Inégalité de Jensen**) Pour toute v.a.r. X de L^1 , toute fonction φ , convexe, de \mathbb{R} vers lui-même et telle que $\varphi(X)$ soit dans L^1 , on a

$$\varphi(\mathbb{E}(X|\mathcal{B})) \leq \mathbb{E}(\varphi(X)|\mathcal{B}), \text{ p.s.}$$

On a la même inégalité si X est positive et φ convexe de \mathbb{R}^+ vers lui-même.

- (2) (**Variance conditionnelle**) Pour une v.a.r. de L^2 , on a p.s. :

$$\mathbb{E}(X^2|\mathcal{B}) - (\mathbb{E}(X|\mathcal{B}))^2 = \mathbb{E}[(X - \mathbb{E}(X|\mathcal{B}))^2|\mathcal{B}].$$

Cette v.a.r. est appelée variance conditionnelle et est notée $\sigma^2(X|\mathcal{B})$.

- (3) (**Inégalité de Hölder**) Soient p et q des réels de $]1, +\infty[$ tels que $1/p + 1/q = 1$ et soient X et Y des v.a.r. respectivement dans $L^p(\Omega, \mathcal{A}, P)$ et $L^q(\Omega, \mathcal{A}, P)$. On a sur tout Ω :

$$|\mathbb{E}(XY|\mathcal{B})| \leq (\mathbb{E}(|X|^p|\mathcal{B}))^{1/p} (\mathbb{E}(|Y|^q|\mathcal{B}))^{1/q}.$$

- (4) (**Convergence monotone**) Soit (X_n) une suite de v.a. positives qui converge en croissant vers une v.a. X . On a alors p.s. la convergence croissante suivante :

$$\mathbb{E}(X_n|\mathcal{B}) \longrightarrow \mathbb{E}(X|\mathcal{B}), \text{ quand } n \rightarrow +\infty.$$

- (5) (**Inégalité de Fatou**) Pour toute suite (X_n) de v.a. positives on a p.s. :

$$\mathbb{E}(\liminf_n X_n|\mathcal{B}) \leq \liminf_n \mathbb{E}(X_n|\mathcal{B}).$$

- (6) (**Convergence dominée**) Pour toute suite (X_n) de v.a. positives telle que

$$X_n \xrightarrow{\text{p.s.}} X, \text{ quand } n \rightarrow +\infty \text{ et } |X_n| \leq Y, \text{ pour tout } n, \text{ avec } \mathbb{E}(Y|\mathcal{B}) < +\infty,$$

on a :

$$\mathbb{E}(X_n|\mathcal{B}) \xrightarrow{\text{p.s.}} \mathbb{E}(X|\mathcal{B}).$$

- (7) (**Conditionnement successifs**) Pour deux sous tribus \mathcal{C} et \mathcal{B} de \mathcal{A} telles que $\mathcal{C} \subset \mathcal{B} \subset \mathcal{A}$ on a

$$\mathbb{E}(X|\mathcal{C}) = \mathbb{E}(\mathbb{E}(X|\mathcal{B})|\mathcal{C}) \equiv \mathbb{E}(X|\mathcal{B}|\mathcal{C}) = \mathbb{E}(X|\mathcal{C}|\mathcal{B}).$$

En revanche si l'on n'a pas une relation d'inclusion dans un sens ou dans un autre entre les tribus \mathcal{C} et \mathcal{B} , les espérances conditionnelles $\mathbb{E}(X|\mathcal{B}|\mathcal{C})$ et $\mathbb{E}(X|\mathcal{C}|\mathcal{B})$ sont en général différentes.

Si de plus l'espérance conditionnelle $\mathbb{E}(X|\mathcal{B})$ est \mathcal{C} -mesurable, alors on a : $\mathbb{E}(X|\mathcal{B}) = \mathbb{E}(X|\mathcal{C})$.

EXERCICE 5.6. Démonstration de ce théorème.

1) Prouver l'inégalité de Jensen en se souvenant qu'une fonction φ est convexe si et seulement si elle est telle que :

$$\varphi(x) = \sup_{(a,b) \in A} (ax + b),$$

où $A = \{(a,b) : \forall y, ay + b \leq \varphi(y)\}$.

2) Montrer la propriété des variances conditionnelles.

3) Montrer en premier lieu que les espérances conditionnelles présentes dans la formule sont bien définies dans L^1 . Appliquer ensuite l'inégalité

$$xy \leq \frac{1}{p}x^p + \frac{1}{q}y^q,$$

vraie pour tout x et y de \mathbb{R}^+ , aux v.a.

$$\frac{|X|}{[\mathbb{E}(|X|^p|\mathcal{B})]^{1/p}} \text{ et } \frac{|Y|}{[\mathbb{E}(|Y|^q|\mathcal{B})]^{1/q}}.$$

pour établir l'inégalité :

$$\frac{\mathbb{E}(|XY||\mathcal{B})}{[\mathbb{E}(|X|^p|\mathcal{B})]^{1/p}[\mathbb{E}(|Y|^q|\mathcal{B})]^{1/q}} \leq 1,$$

sur l'ensemble $B = \{\mathbb{E}(|X|^p|\mathcal{B}) > 0 \text{ et } \mathbb{E}(|Y|^q|\mathcal{B}) > 0\}$. Terminer en montrant que l'inégalité est trivialement vérifiée sur le complémentaire de B . On pourra utiliser les ensembles de la forme : $B_X = \{\mathbb{E}(|X|^p|\mathcal{B}) = 0\}$ et $B_Y = \{\mathbb{E}(|Y|^q|\mathcal{B}) = 0\}$.

4) Démontrer la propriété de convergence monotone pour l'espérance conditionnelle.

5) Pour établir le résultat du type "Lemme de Fatou", on pourra considérer les suites (Y_n) et (Z_n) définies par

$$Y_n = \inf_{p \geq n} X_p \text{ et } Z_n = \inf_{p \geq n} \mathbb{E}(X_p|\mathcal{B}).$$

6) a) Considérer une suite (X_n) de v.a. positives telle que $X_n \leq Y$ où Y est une v.a. dont l'espérance conditionnelle à \mathcal{B} est bornée par un entier N , i.e. telle que $\mathbb{E}(Y|\mathcal{B}) \leq N < +\infty$. Montrer alors que la convergence p.s. de X_n vers 0, entraîne la convergence :

$$\mathbb{E}(X_n|\mathcal{B}) \xrightarrow{p.s.} 0, \text{ quand } n \rightarrow +\infty.$$

On pourra considérer la suite (U_n) définie par

$$U_n = \sup_{p \geq n} \mathbb{E}(X_p|\mathcal{B}),$$

et montrer qu'elle est convergente vers une v.a. U d'espérance nulle.

b) Dédurre de ce qui précède que l'on a la convergence annoncée dans le théorème si l'on rajoute l'hypothèse que la v.a. majorante Y est d'espérance conditionnelle $\mathbb{E}(Y|\mathcal{B})$ bornée.

c) Utiliser le résultat précédent pour établir que cette convergence des espérances conditionnelles est en fait obtenue dès que l'espérance conditionnelle $\mathbb{E}(Y|\mathcal{B})$ est finie (et non nécessairement bornée). On pourra utiliser une suite (X'_n) définie, pour tout n , par $X'_n = X_n \mathbb{1}_{A_N}$, où $A_N = \{\mathbb{E}(Y|\mathcal{B}) < N\}$ et N est un entier quelconque.

7) Démontrer directement ces formules de conditionnement successifs.

Preuve/solution de l'exercice.

1) D'après la propriété rappelée pour une fonction convexe, on peut écrire :

$$\begin{aligned} \varphi(\mathbb{E}(X|\mathcal{B})) &= \sup_{(a,b) \in A} (a\mathbb{E}(X|\mathcal{B}) + b) = \sup_{(a,b) \in A} \mathbb{E}(aX + b|\mathcal{B}) \\ &\leq \sup_{(a,b) \in A} \mathbb{E}(\varphi(X)|\mathcal{B}) = \mathbb{E}(\varphi(X)|\mathcal{B}). \end{aligned}$$

2) On peut écrire :

$$\begin{aligned} \mathbb{E} \left[(X - \mathbb{E}(X|\mathcal{B}))^2 | \mathcal{B} \right] &= \mathbb{E} \left[(X^2 - 2X\mathbb{E}(X|\mathcal{B}) + \mathbb{E}^2(X|\mathcal{B})) | \mathcal{B} \right] \\ &= \mathbb{E}(X^2|\mathcal{B}) + \mathbb{E}^2(X|\mathcal{B}) - 2\mathbb{E}^2(X|\mathcal{B}) \\ &= \mathbb{E}(X^2|\mathcal{B}) - \mathbb{E}^2(X|\mathcal{B}). \end{aligned}$$

3) Puisque les v.a. X et Y sont respectivement dans L^p et L^q , les v.a. X^p et Y^q sont dans L^1 . De plus, d'après l'inégalité de Hölder pour les espérances classiques, le produit XY est dans L^1 . On peut donc bien définir les espérances conditionnelles de ces v.a. dans L^1 .

Maintenant, sur l'ensemble $B = \{\mathbb{E}(|X|^p|\mathcal{B}) > 0 \text{ et } \mathbb{E}(|Y|^q|\mathcal{B}) > 0\}$, les v.a.

$$\frac{|X|}{[\mathbb{E}(|X|^p|\mathcal{B})]^{1/p}} \text{ et } \frac{|Y|}{[\mathbb{E}(|Y|^q|\mathcal{B})]^{1/q}}$$

sont bien définies et positives. On peut alors leur appliquer l'inégalité rappelée dans l'énoncé pour obtenir, sur l'ensemble B :

$$\frac{|X|}{[\mathbb{E}(|X|^p|\mathcal{B})]^{1/p}} \frac{|Y|}{[\mathbb{E}(|Y|^q|\mathcal{B})]^{1/q}} \leq \frac{|X|^p}{p\mathbb{E}(|X|^p|\mathcal{B})} + \frac{|Y|^q}{q\mathbb{E}(|Y|^q|\mathcal{B})}.$$

L'événement B étant \mathcal{B} -mesurable, en prenant l'espérance conditionnelle de l'inégalité précédente on obtient, toujours sur B :

$$\frac{\mathbb{E}(|XY||\mathcal{B})}{[\mathbb{E}(|X|^p|\mathcal{B})]^{1/p}[\mathbb{E}(|Y|^q|\mathcal{B})]^{1/q}} \leq \frac{1}{p} + \frac{1}{q} = 1,$$

ce qui, grâce à la propriété 6) de la Proposition 5.3, prouve sur l'événement B l'inégalité de Hölder désirée.

Considérons maintenant l'événement $B_X = \{\mathbb{E}(|X|^p|\mathcal{B}) = 0\}$. Cet événement est dans \mathcal{B} et par propriété des espérances conditionnelles, on a :

$$\mathbb{E}(|X|^p \mathbb{1}_{B_X}) = \mathbb{E}(\mathbb{E}(|X|^p|\mathcal{B}) \mathbb{1}_{B_X}) = 0,$$

ce qui implique que la v.a. $|X|^p \mathbb{1}_{B_X}$ est *p.s.* nulle. On en tire que

$$\mathbb{E}(|XY| \mathbb{1}_{B_X} | \mathcal{B}) = 0$$

et l'inégalité de Hölder est alors trivialement vérifiée sur B_X , les deux cotés de l'inégalité étant nuls. Il en est par symétrie évidente de même sur $B_Y = \{\mathbb{E}(|Y|^q|\mathcal{B}) = 0\}$. Comme on a $\Omega = B \cup B_X \cup B_Y$, l'inégalité de Hölder est toujours vraie.

4) Soit (X_n) une suite définie comme dans le théorème et notons (Y_n) la suite définie, pour tout n , par $Y_n = \mathbb{E}(X_n|\mathcal{B})$. Par propriétés de l'espérance conditionnelle, cette dernière constitue une suite de v.a. positives, \mathcal{B} -mesurables et croissante vers une v.a. Y qui est \mathcal{B} -mesurable et à valeurs dans $[0, +\infty]$. Or, pour tout n et tout B de \mathcal{B} , on a : $\mathbb{E}(X_n \mathbb{1}_B|\mathcal{B}) = \mathbb{E}(Y_n \mathbb{1}_B|\mathcal{B})$. Par le théorème de convergence monotone pour les espérances classiques, le terme de gauche converge en croissant vers $\mathbb{E}(X \mathbb{1}_B|\mathcal{B})$ et celui de droite vers $\mathbb{E}(Y \mathbb{1}_B|\mathcal{B})$, ce qui prouve que Y n'est autre que l'espérance conditionnelle $\mathbb{E}(X|\mathcal{B})$ et la démonstration est achevée.

5) Les suites (Y_n) et (Z_n) , définies dans l'énoncé de l'exercice, convergent en croissant respectivement vers $\liminf_n X_n$ et $\liminf_n \mathbb{E}(X_n|\mathcal{B})$. Par ailleurs, la monotonie de l'espérance conditionnelle nous assure que l'on a

$$\mathbb{E}(X_p|\mathcal{B}) \geq \mathbb{E}(Y_n|\mathcal{B}),$$

pour tout $p \geq n$ et donc :

$$Z_n \geq \mathbb{E}(Y_n|\mathcal{B}),$$

pour tout n . Par le théorème de convergence monotone que nous venons de démontrer au point 4), le terme de droite converge en croissant vers $\mathbb{E}(\liminf_n Y_n|\mathcal{B})$. En faisant tendre n vers $+\infty$ des deux cotés de l'inégalité précédente, on obtient l'inégalité de Fatou pour les espérances conditionnelles.

6) a) Considérons la suite (U_n) définie, pour tout n , par

$$U_n = \sup_{p \geq n} \mathbb{E}(X_p|\mathcal{B}).$$

Comme cette suite est décroissante et positive elle est convergente vers une v.a. U également positive. Or on a :

$$\mathbb{E}(U_n) = \mathbb{E}(\sup_{p \geq n} \mathbb{E}(X_p|\mathcal{B})) \leq \mathbb{E}(\mathbb{E}(\sup_{p \geq n} X_p|\mathcal{B})) = \mathbb{E}(\sup_{p \geq n} X_p).$$

Mais la convergence *p.s.* de la suite (X_n) vers 0 entraîne celle de $(\sup_{p \geq n} X_p)$ vers 0. Comme chaque élément de cette dernière suite est majorée par la v.a. Y qui est telle que $\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|\mathcal{B})) \leq N$, le théorème de convergence dominée classique assure alors la convergence :

$$\mathbb{E}(\sup_{p \geq n} X_p) \rightarrow 0, \text{ quand } n \rightarrow +\infty.$$

Cette dernière convergence associée à l'inégalité précédemment établie, montre que l'on a la convergence de $\mathbb{E}(U_n)$ vers 0. La variable limite U de la suite positive (U_n) est donc positive et d'espérance nulle. Elle ne peut être différente de 0. On a donc bien la convergence :

$$\mathbb{E}(X_n|\mathcal{B}) \xrightarrow{p.s.} 0, \text{ quand } n \rightarrow +\infty.$$

b) Soit maintenant une suite (X_n) de v.a. de signe quelconque mais telle que :

$$X_n \xrightarrow{p.s.} X, \text{ quand } n \rightarrow +\infty \text{ et } |X_n| \leq Y, \text{ pour tout } n, \text{ avec } \mathbb{E}(Y|\mathcal{B}) \leq N < +\infty,$$

où N est un entier fixé.

On a naturellement la convergence

$$X_n - X \xrightarrow{p.s.} 0$$

et les majorations *p.s.*

$$|X_n - X| \leq 2Y \text{ et } \mathbb{E}(2Y|\mathcal{B}) \leq 2N.$$

D'après le point a) précédent nous avons donc la convergence

$$|\mathbb{E}(X_n - X|\mathcal{B})| \leq \mathbb{E}(|X_n - X||\mathcal{B}) \xrightarrow{p.s.} 0,$$

ce qui prouve bien le résultat attendu.

c) Soit enfin une suite (X_n) de v.a. qui vérifie les propriétés du point b) précédent à l'exception de l'espérance conditionnelle qui est ici supposée finie (et non bornée), i.e. $\mathbb{E}(Y|\mathcal{B}) < +\infty$.

Introduisons la suite (X'_n) définie, pour tout n , par $X'_n = X_n \mathbb{1}_{A_N}$, où $A_N = \{\mathbb{E}(Y|\mathcal{B}) < N\}$ pour un entier N quelconque. On a bien sûr

$$X'_n \xrightarrow{p.s.} X' = X \mathbb{1}_{A_N}.$$

De plus, pour tout n , la v.a. $|X'_n|$ est majorée par la v.a. $Y' = Y \mathbb{1}_{A_N}$ qui est telle que, p.s. :

$$\mathbb{E}(Y'|\mathcal{B}) = \mathbb{1}_{A_N} \mathbb{E}(Y|\mathcal{B}) \leq N,$$

puisque l'événement A_N est élément de \mathcal{B} . Le résultat du point b) précédent permet d'assurer la convergence :

$$\mathbb{E}(X_n|\mathcal{B}) \mathbb{1}_{A_N} \xrightarrow{p.s.} \mathbb{E}(X|\mathcal{B}) \mathbb{1}_{A_N}.$$

L'égalité

$$\Omega = \cup_{N \in \mathbb{N}} A_N$$

et la linéarité de l'espérance conduisent alors à la conclusion du théorème.

7) Notons $Y = \mathbb{E}(X|\mathcal{B})$ et $Z = \mathbb{E}(Y|\mathcal{C})$. Il s'agit de montrer que l'on a $Z = \mathbb{E}(X|\mathcal{C})$.

Soit C un élément quelconque de \mathcal{C} . Par définition de Z puis par définition de Y on peut écrire :

$$\mathbb{E}(Z \mathbb{1}_C) = \mathbb{E}(Y \mathbb{1}_C) = \mathbb{E}(\mathbb{E}(X|\mathcal{B}) \mathbb{1}_C) = \mathbb{E}(\mathbb{E}(X \mathbb{1}_C|\mathcal{B})) = \mathbb{E}(X \mathbb{1}_C),$$

ce qui prouve que $Z = \mathbb{E}(X|\mathcal{C})$.

Enfin la dernière affirmation est aisément obtenue en remarquant que l'espérance conditionnelle $\mathbb{E}(X|\mathcal{B})$ sort de l'espérance conditionnelle à \mathcal{C} si elle \mathcal{C} -mesurable. \square

Voyons maintenant le lien avec les lois conditionnelles vues en cours de base de Probabilités. Supposons que Z soit une v.a. de Ω dans \mathbb{R} telle que sa loi soit absolument continue par rapport à une mesure μ_1 sur \mathbb{R} . Soit également X une v.a. de Ω vers un ensemble E de loi absolument continue par rapport à une mesure μ_2 sur E . Supposons enfin que la loi du couple aléatoire (Z, X) soit absolument continue par rapport à la mesure produit $\mu_1 \otimes \mu_2$ et de densité $f(z, x)$.

La densité de la v.a. X par rapport à la mesure μ_2 est alors donnée par

$$g(x) = \int_{\mathbb{R}} f(z, x) \mu_1(dz)$$

On sait définir la densité conditionnelle de Z sachant X par

$$h(z, x) = \frac{f(z, x)}{g(x)}$$

pour x tel que $g(x)$ ne soit pas nul (on peut définir cette densité conditionnelle sur tout E en donnant la valeur 1, par exemple, quand $g(x) = 0$).

LEMME 5.9. Avec les hypothèses précédentes sur le couple de v.a. (Z, X) on peut écrire, pour toute v.a. φ telle que $\varphi(Z)$ soit intégrable :

$$\mathbb{E}(\varphi(Z)|X) = k_\varphi(X)$$

avec

$$k_\varphi(x) = \int_{\mathbb{R}} \varphi(z) h(z, x) \mu_1(dz),$$

pour tout x de E .

EXERCICE 5.7. *Démontrer ce résultat.*

Preuve/solution de l'exercice. D'après le Lemme de Doob, la v.a. $k_\varphi(X)$ est $\sigma(X)$ -mesurable. Il nous faut donc seulement montrer que l'on a pour toute v.a. $\sigma(X)$ -mesurable Y :

$$\mathbb{E}(k_\varphi(X)Y) = \mathbb{E}(\varphi(Z)Y).$$

Or, Y étant $\sigma(X)$ -mesurable, il existe d'après le lemme de Doob une fonction ψ telle que $Y = \psi(X)$.

On en est donc déduit à montrer que l'on a pour toute fonction ψ mesurable bornée :

$$\mathbb{E}(k_\varphi(X)\psi(X)) = \mathbb{E}(\varphi(Z)\psi(X)).$$

Or, on peut écrire :

$$\begin{aligned} \mathbb{E}(k_\varphi(X)\psi(X)) &= \int k_\varphi(x)\psi(x)g(x)\mu_2(dx) = \int_x \left(\int_z \varphi(z)h(z,x)\mu_1(dz) \right) \psi(x)g(x)\mu_2(dx) \\ &= \int_x \int_z \psi(x)\varphi(z)h(z,x)g(x)\mu_1(dz)\mu_2(dx) \\ &= \iint \psi(x)\varphi(z)f(z,x)\mu_1(dz)\mu_2(dx) \\ &= \mathbb{E}(\varphi(Z)\psi(X)). \end{aligned}$$

La variable $k_\varphi(X)$ convient. Elle est donc l'unique (à une égalité *p.s.* près) solution. \square

CHAPITRE 6

Martingales

1. Martingales

1.1. Filtrations et temps d'arrêt.

DÉFINITION 6.1. Une **filtration** sur un espace probabilisé (Ω, \mathcal{F}, P) est une suite croissante (\mathcal{F}_n) de sous tribus de \mathcal{F} , i.e., pour tout n , les sous tribus \mathcal{F}_n et \mathcal{F}_{n+1} de \mathcal{F} sont telles que :

$$\mathcal{F}_n \subset \mathcal{F}_{n+1}.$$

De plus on définit :

$$\mathcal{F}_\infty = \bigvee_n \mathcal{F}_n \equiv \sigma(\bigcup_n \mathcal{F}_n),$$

Dans cette définition on utilise donc la notation $\mathcal{F} \vee \mathcal{G}$ pour signifier la tribu engendrée par l'union de deux tribus \mathcal{F} et \mathcal{G} (on sait que l'union de deux tribus ne donne pas forcément une tribu).

On supposera dans la suite que la tribu \mathcal{F}_0 contient tous les négligeables de \mathcal{F}_∞ .

Exemple fondamental de filtration : la filtration naturelle associée à un processus. Soit (X_n) une suite de v.a.r. sur (Ω, \mathcal{F}, P) . Définissons, pour tout n , la tribu

$$\mathcal{F}_n = \sigma(X_0, \dots, X_n).$$

La filtration ainsi construite est appelée filtration naturelle associée au processus (X_n) .

DÉFINITION 6.2. Soit (\mathcal{F}_n) une filtration sur un espace probabilisé (Ω, \mathcal{F}, P) . Un processus $(X_n)_{n \geq 0}$ est dit **adapté** à la filtration (\mathcal{F}_n) si la v.a. X_n est \mathcal{F}_n -mesurable, pour tout n .

Naturellement la filtration naturelle associée au processus (X_n) est la plus petite filtration le rendant adapté.

DÉFINITION 6.3. Une application T de Ω vers $\bar{\mathbb{N}} \equiv \mathbb{N} \cup \{+\infty\}$ est appelée **(\mathcal{F}_n) -temps d'arrêt (t.d.a.)** si on a :

$$1) \forall n \in \mathbb{N} : \{T \leq n\} = \{\omega : T(\omega) \leq n\} \in \mathcal{F}_n,$$

ou encore de manière équivalente si :

$$2) \forall n \in \mathbb{N} : \{T = n\} \in \mathcal{F}_n.$$

EXERCICE 6.1. Démontrer l'équivalence énoncée dans la Définition 6.3.

Solution de l'exercice. Pour montrer que 1) implique 2), il suffit de remarquer que l'on a, pour tout n :

$$\{T = n\} = \{T \leq n\} \setminus \{T \leq n-1\}.$$

Le premier terme à droite de l'égalité est dans \mathcal{F}_n par définition d'un t.d.a. et le second dans \mathcal{F}_{n-1} . Par croissance des tribus dans une filtration et par propriété des tribus, l'événement de gauche est donc bien dans \mathcal{F}_n .

Pour la réciproque on utilise à nouveau la croissance des tribus dans une filtration et l'égalité

$$\{T \leq n\} = \{T = 0\} \cup \{T = 1\} \cup \dots \cup \{T = n\},$$

vraie pour tout n . ◇

Il est important de bien se convaincre que si T est un t.d.a. alors on a

$$\forall n : \{T \geq n\} \in \mathcal{F}_n$$

mais que cette propriété n'est pas suffisante pour que T soit un t.d.a.

Intuitivement un t.d.a. est une horloge, un temps où l'on décide de continuer ou d'arrêter un jeu par exemple. Naturellement, si la triche n'est pas possible, la décision de s'arrêter à la n ème partie ne peut se baser que sur la connaissance des résultats des parties précédentes. De même, en bourse, la décision de vendre ou d'acheter une action ne peut dépendre que du passé du cours de cette action, et pas du futur sous peine d'être accusé de délit d'initié !

EXERCICE 6.2. Soit (X_n) un processus (\mathcal{F}_n) -adapté de (Ω, \mathcal{F}, P) vers un espace probabilisable (E, \mathcal{E}) . Soit A un événement de la tribu \mathcal{E} et les v.a. T_1 et T_2 définies par :

$$\begin{aligned} T_1 &= \inf\{n \in \mathbb{N} : X_n \in A\}, \\ T_2 &= \sup\{n \in \mathbb{N} : X_n \in A\}. \end{aligned}$$

Les v.a. T_1 et T_2 sont-elles des t.d.a. ?

Solution de l'exercice. On peut écrire

$$\{T_1 \leq n\} = \{X_0 \in A\} \cup \{X_1 \in A\} \cup \dots \cup \{X_n \in A\},$$

et en déduire facilement que T_1 est un t.d.a. En revanche, on peut seulement écrire

$$\{T_1 = n\} = \bigcap_{p \geq 1} \{X_{n+p} \in \bar{A}\},$$

ce qui n'a aucune raison, a priori, d'être dans la tribu \mathcal{F}_n . ◇

EXERCICE 6.3. Soit T un (\mathcal{F}_n) -t.d.a.

1) Montrer que l'on a :

$$\{T = \infty\} \in \mathcal{F}_\infty.$$

2) Montrer que $T + 1$ est également un t.d.a. En est-il de même pour $T - 1$?

3) Montrer que si S est un autre (\mathcal{F}_n) -t.d.a., alors les v.a. $\max(T, S)$, $\min(T, S)$ et $T + S$ sont encore des t.d.a.

Solution de l'exercice.

1) Il suffit de noter que l'on peut écrire

$$\{T < \infty\} = \bigcup_n \{T = n\}.$$

Chaque événement de droite étant dans \mathcal{F}_∞ , le complémentaire de l'événement $\{T = \infty\}$ est donc également dans \mathcal{F}_∞ .

Ceci nous permet de comprendre pourquoi, dans certains ouvrages, on peut trouver la condition

$$\forall n \in \bar{\mathbb{N}} : \{T \leq n\} \in \mathcal{F}_n.$$

2) On a :

$$\{T + 1 = n\} = \{T = n - 1\} \in \mathcal{F}_{n-1} \subset \mathcal{F}_n$$

ce qui montre $T + 1$ est bien un t.d.a. En revanche, on a :

$$\{T - 1 = n\} = \{T = n + 1\} \in \mathcal{F}_{n+1}$$

et cet événement n'a donc lui, a priori, aucune raison d'être dans la tribu \mathcal{F}_n .

3) Les égalités

$$\begin{aligned} \{\max(T, S) \leq n\} &= \{T \leq n\} \cap \{S \leq n\} \\ \{\min(T, S) > n\} &= \{T > n\} \cap \{S > n\} \\ \{T + S = n\} &= \cup_{k=0}^n \{T = k\} \cap \{S = n - k\} \end{aligned}$$

permettent aisément d'établir ce résultat. ◇

LEMME 6.4. *La v.a. T est un (\mathcal{F}_n) -t.d.a. si, et seulement si, le processus (X_n) , défini en tout n par $X_n = \mathbb{1}_{\{T \leq n\}}$, est (\mathcal{F}_n) -adapté. Dans ce cas, on a*

$$T = \inf\{n \in \mathbb{N} : X_n = 1\}.$$

La démonstration de ce lemme est triviale et laissée au lecteur.

DÉFINITION 6.5. *On appelle tribu des événements antérieurs à un (\mathcal{F}_n) -t.d.a., la tribu \mathcal{F}_T définie par*

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : \forall n \in \mathbb{N}, A \cap \{T \leq n\} \in \mathcal{F}_n\}.$$

C'est une exercice facile que de vérifier que \mathcal{F}_T est effectivement une tribu.

PROPOSITION 6.6.

- Si T est un t.d.a. p.s. constant égal à k , alors la tribu \mathcal{F}_T n'est autre que la tribu \mathcal{F}_k .
- Un événement A est dans \mathcal{F}_T si, et seulement si, on a, pour tout n , l'événement $A \cap \{T = n\}$ dans \mathcal{F}_n .

EXERCICE 6.4. *Démontrer cette proposition.*

Preuve/Solution de l'exercice.

1) Si T est p.s. égal à k , on a pour tout A de \mathcal{F} les égalités :

$$\begin{aligned} A \cap \{T \leq n\} &= \emptyset, \text{ si } k > n \\ A \cap \{T \leq n\} &= A, \text{ si } k \leq n. \end{aligned}$$

Ainsi, on peut écrire :

$$A \in \mathcal{F}_T \iff \forall n : A \cap \{T \leq n\} \in \mathcal{F}_n \iff \forall n \geq k : A \in \mathcal{F}_n \iff A \in \bigcap_{n \geq k} \mathcal{F}_n = \mathcal{F}_k.$$

2) Pour la condition nécessaire, il suffit de remarquer que l'on a :

$$A \cap \{T = n\} = A \cap \{T \leq n\} \cap \overline{\{T \leq n-1\}}$$

qui est dans \mathcal{F}_n puisque les événements $A \cap \{T \leq n\}$ et $\overline{\{T \leq n-1\}}$ le sont.

Pour la condition suffisante, on utilise les égalités

$$A \cap \{T \leq n\} = A \cap \left(\bigcup_{p=1}^n \{T = p\} \right) = \bigcup_{p=1}^n (A \cap \{T = p\}),$$

qui montrent que l'événement à gauche de l'égalité est dans \mathcal{F}_n puisque chaque élément dans l'union de droite est dans \mathcal{F}_p , pour $p \leq n$. \square

PROPOSITION 6.7. *On a les propriétés suivantes :*

(1) *Si T et S sont deux t.d.a., on a :*

$$S \underset{p.s.}{\leq} T \Rightarrow \mathcal{F}_S \subset \mathcal{F}_T.$$

(2) *Pour tout t.d.a. S et T , on a :*

$$\mathcal{F}_S \cap \mathcal{F}_T = \mathcal{F}_{S \wedge T}.$$

(3) *Pour tout t.d.a. S et T , on a :*

$$\mathcal{F}_S \vee \mathcal{F}_T = \mathcal{F}_{S \vee T}.$$

(4) *Si (T_n) est une suite de t.d.a. décroissante vers T , alors T est un t.d.a. et on a :*

$$\mathcal{F}_T = \bigcap_n \mathcal{F}_{T_n}.$$

(5) *Si (T_n) est une suite de t.d.a. croissante vers T , alors T est un t.d.a. et on a :*

$$\mathcal{F}_T = \bigvee_n \mathcal{F}_{T_n}.$$

EXERCICE 6.5. *Démontrer les points 1, 2 et 4 de cette proposition (Ind. pour les points 2 et 4 on pourra procéder par double inclusion). Les autres points sont un peu plus délicats à démontrer et leurs preuves ne seront pas détaillées dans ce polycopié.*

Preuve/Solution de l'exercice.

1) Soit N le négligeable en dehors duquel on a $S(\omega) \leq T(\omega)$. On peut écrire, pour tout événement A de \mathcal{F}_S :

$$A \cap \{T \leq n\} = (A \cap \{T \leq n\} \cap N) \cup (A \cap \{T \leq n\} \cap \bar{N}).$$

Le premier terme dans l'union de droite est négligeable, donc par hypothèse dans \mathcal{F}_0 et, par croissance de la filtration, dans \mathcal{F}_n . Le second terme peut s'écrire sous la forme :

$$A \cap \{S \leq n\} \cap \{T \leq n\} \cap \bar{N}$$

ce qui prouve qu'il est dans \mathcal{F}_n , puisque A est dans \mathcal{F}_S , que T est un t.d.a. et que N est dans \mathcal{F}_n .

2) D'après le point précédent, la tribu $\mathcal{F}_{S \wedge T}$ est incluse dans \mathcal{F}_S aussi bien que dans \mathcal{F}_T . Elle est donc incluse dans $\mathcal{F}_S \cap \mathcal{F}_T$.

Considérons maintenant un événement A de $\mathcal{F}_S \cap \mathcal{F}_T$. On peut écrire

$$A \cap \{S \wedge T \leq n\} = (A \cap \{T \leq n\}) \cup (A \cap \{S \leq n\}).$$

Il s'agit bien d'un événement de \mathcal{F}_n puisque chaque terme de l'union est dans \mathcal{F}_n par définition d'une tribu des événements antérieurs à un t.d.a. L'événement A est donc bien dans $\mathcal{F}_{S \wedge T}$, ce qui prouve l'inclusion inverse et donc l'égalité

$$\mathcal{F}_S \cap \mathcal{F}_T = \mathcal{F}_{S \wedge T}.$$

4) Soit (T_n) une suite de t.d.a. décroissante vers T . L'égalité

$$\{T \leq k\} = \bigcup_n \{T_n \leq k\},$$

vraie pour tout k , montre bien que T est un t.d.a. puisque chaque terme de droite est dans \mathcal{F}_k . De plus, l'inégalité $T \leq T_n$, vraie pour tout n , implique l'inclusion de \mathcal{F}_T dans \mathcal{F}_{T_n} et donc

$$\mathcal{F}_T \subset \bigcap_n \mathcal{F}_{T_n}.$$

Réciproquement, soit A un événement de $\bigcap_n \mathcal{F}_{T_n}$. L'égalité, pour tout k ,

$$A \cap \{T \leq k\} = \bigcup_n (A \cap \{T_n \leq k\})$$

montre bien que A est un élément de \mathcal{F}_T puisque chaque terme dans l'union de droite est dans \mathcal{F}_k . \square

PROPOSITION 6.8. *Une v.a. Z est \mathcal{F}_T mesurable si, et seulement si, pour tout entier k , la v.a. $Z \mathbb{1}_{\{T \leq k\}}$ est \mathcal{F}_k mesurable.*

EXERCICE 6.6. *Démonstration de cette proposition. Autant pour la condition nécessaire que pour la condition suffisante, on pourra utiliser le théorème des classes monotones en se restreignant dans un premier temps aux variables bornées. L'extension aux variables quelconques se fait ensuite par des arguments classiques déjà utilisés dans la Chapitre 5 sur les espérances conditionnelles.*

Preuve/Solution de l'exercice.

Condition nécessaire. L'ensemble

$$\mathcal{H} = \{\text{v.a. } Z \text{ bornées} : Z \mathbb{1}_{\{T \leq k\}} \text{ est } \mathcal{F}_k\text{-mesurable, pour tout } k \in \mathbb{N}\}$$

constitue un espace vectoriel de fonctions bornées, contenant les constantes et stable par limite monotone bornée. De plus l'ensemble

$$\mathcal{C} = \{Z = \mathbb{1}_A : A \in \mathcal{F}_T\}$$

est une classe de fonctions réelles bornées stable par multiplication. Par définition de \mathcal{F}_T , l'ensemble \mathcal{C} est inclus dans \mathcal{H} et, d'après le théorème des classes monotones, il en est de même pour l'ensemble de toutes les fonctions $\sigma(\mathcal{C})$ -mesurables bornées. Ainsi toute v.a. Z qui est \mathcal{F}_T -mesurable bornée est telle que, pour tout k dans \mathbb{N} , la v.a. $Z \mathbb{1}_{\{T \leq k\}}$ est \mathcal{F}_k -mesurable.

Considérons maintenant une v.a. positive, \mathcal{F}_T -mesurable, non nécessairement bornée, et la suite (Z_p) de v.a. définies par : $Z_p = Z \wedge p$, pour tout p de \mathbb{N} . Les v.a. Z_p sont \mathcal{F}_T -mesurables bornées et donc telles que, pour tout k et tout p de \mathbb{N} , la v.a. $Z_p \mathbb{1}_{\{T \leq k\}}$ est \mathcal{F}_k -mesurable. Ceci et la convergence de la suite (Z_p) vers Z implique que, pour tout k , la v.a. $Z \mathbb{1}_{\{T \leq k\}}$ est \mathcal{F}_k -mesurable. Pour terminer la démonstration et étendre le résultat aux v.a. \mathcal{F}_T -mesurables Z de signe quelconque, il suffit de considérer les parties positives Z^+ et négatives Z^- de Z et d'appliquer le résultat précédent pour constater que

$$Z \mathbb{1}_{\{T \leq k\}} = Z^+ \mathbb{1}_{\{T \leq k\}} - Z^- \mathbb{1}_{\{T \leq k\}}$$

est \mathcal{F}_k -mesurable, pour tout k de \mathbb{N} .

Condition suffisante. Réciproquement, on considère les ensembles

$$\mathcal{H} = \{Z : Z \text{ } \mathcal{F}_T\text{-mesurable bornée}\}$$

et

$$\mathcal{C} = \{Z = \mathbb{1}_A : \forall k \in \mathbb{N}, A \cap \{T \leq k\} \in \mathcal{F}_k\}$$

et on raisonne de manière très similaire à celle utilisée pour établir la condition nécessaire. \square

PROPOSITION 6.9. *Si (X_n) est une suite (\mathcal{F}_n) -adaptée, alors, pour tout t.d.a. T , la v.a. X_T définie, pour tout ω de Ω par $X_T(\omega) = X_{T(\omega)}(\omega)$, est \mathcal{F}_T -mesurable.*

EXERCICE 6.7. *Démontrer cette proposition (Ind. On pourra découper suivant les valeurs prises par le t.d.a. T).*

Preuve/Solution de l'exercice.

On peut écrire

$$X_T = \sum_{k \in \mathbb{N}} X_k \mathbb{1}_{\{T=k\}} + X_\infty \mathbb{1}_{\{T=\infty\}}.$$

Ainsi, pour tout k de \mathbb{N} , la v.a.

$$X_T \mathbb{1}_{\{T \leq k\}} = \sum_{p=0}^k X_p \mathbb{1}_{\{T=p\}},$$

apparaît clairement comme \mathcal{F}_k -mesurable, ce qui prouve, grâce à la Proposition 6.8, que X_T est \mathcal{F}_T -mesurable. \square

1.2. Martingales, Sousmartingales et Surmartingales.

1.2.1. Définition.

DÉFINITION 6.10. *Un suite de v.a. (X_n) de Ω vers \mathbb{R} est dite une **martingale** relativement à une filtration (\mathcal{F}_n) (ou (\mathcal{F}_n) -martingale) si elle vérifie les propriétés suivantes :*

- (1) (X_n) est un processus (\mathcal{F}_n) -adapté ;
- (2) $\mathbb{E}|X_n| < +\infty$, pour tout n dans \mathbb{N} ;
- (3) (**propriété de martingale**) $\mathbb{E}(X_n | \mathcal{F}_{n-1}) = X_{n-1}$, pour tout $n \geq 1$.

Le processus (X_n) est appelé sousmartingale (resp. surmartingale) si les propriétés 1) et 2) sont respectées conjointement avec l'inégalité $\mathbb{E}(X_n | \mathcal{F}_{n-1}) \geq X_{n-1}$ (resp. $\mathbb{E}(X_n | \mathcal{F}_{n-1}) \leq X_{n-1}$), pour tout $n \geq 1$.

On constate que, dans le cas d'une martingale, la condition 1) est automatiquement assurée dès que la condition 3) l'est. Il y a donc redondance. En revanche, dans le cas d'une sousmartingale ou d'une surmartingale, la condition 1) est nécessaire car non assurée par la condition 3). Par ailleurs on note que bien évidemment (X_n) est une surmartingale si, et seulement si, le processus $(-X_n)$ est une sousmartingale.

Il apparaît clairement que la suite des espérances d'une martingale est constante alors qu'elle est croissante (resp. décroissante) pour une sousmartingale (resp. surmartingale).

Les martingales sont utilisées pour modéliser les jeux de hasard équitables. Supposons en effet qu'à un jeu de hasard, la v.a. X_n représente la fortune d'un joueur à la n ème partie. La propriété de martingale stipule que, si le joueur a la connaissance de l'évolution passée de sa fortune (c'est à dire jusqu'au temps $n - 1$) alors l'espérance de la valeur de sa fortune après la partie suivante (la n ème) est égale à celle actuelle (i.e. à $n - 1$). En moyenne ses gains restent inchangés.

PROPOSITION 6.11. *Si (X_n) est une (\mathcal{F}_n) -martingale, on a alors, pour tout $m < n$:*

$$\mathbb{E}(X_n | \mathcal{F}_m) = X_m.$$

La preuve est immédiate.

Remarquons que si $X = (X_n)$ est une martingale et si U est une v.a. \mathcal{F}_0 -mesurable et dans L^1 , alors le processus $X - U = (X_n - U)$ est encore une martingale. C'est pourquoi, on pourra se restreindre à l'étude des martingales nulles en 0 (quitte à soustraire sa valeur initiale X_0).

1.2.2. Exemples classiques de martingales.

EXERCICE 6.8.

1) (Somme de v.a. indépendantes et centrées) Soit (Y_n) une suite de v.a.r. indépendantes, centrées et dans L^1 . Montrer que le processus (S_n) , défini pour tout n par

$$S_n = \sum_{k=0}^n Y_k,$$

est une martingale par rapport à une filtration que l'on précisera.

2) (Produit de v.a. positives, indépendantes et d'espérance 1) Soit (X_n) une suite de v.a.r. indépendantes, positives et telles que $\mathbb{E}(X_n) = 1$, pour tout n de \mathbb{N} . Montrer que le processus (M_n) défini, pour tout n , par :

$$M_n = \prod_{k=1}^n X_k$$

est une martingale par rapport à la filtration naturelle associée au processus (X_n) . Que se passe-t-il si l'on a pour tout n de \mathbb{N} : $\mathbb{E}(X_n) \geq 1$? Même question si l'on a : $\mathbb{E}(X_n) \leq 1$, pour tout n de \mathbb{N} .

3) (Accumulation d'information sur une v.a.) Soit (\mathcal{F}_n) une filtration sur un espace probabilisé (Ω, \mathcal{F}, P) et X une v.a. intégrable sur cet espace. Pour tout n dans \mathbb{N} , définissons $M_n = \mathbb{E}(X | \mathcal{F}_n)$, qui est, au sens de la projection dans L^2 , la meilleure estimation de X à partir de \mathcal{F}_n . Montrer que (M_n) est une (\mathcal{F}_n) -martingale. On parle de **Martingale de Doob**.

Solution de l'exercice.

1) Prenons pour filtration (\mathcal{F}_n) , la filtration naturelle associée au processus (Y_n) , i.e. telle que, pour tout n , on ait $\mathcal{F}_n = \sigma(Y_0, Y_1, \dots, Y_n)$. Le processus (S_n) est naturellement (\mathcal{F}_n) -adapté et dans L^1 . Par ailleurs,

$$\mathbb{E}(S_n | \mathcal{F}_{n-1}) = \mathbb{E}(S_{n-1} + Y_n | \mathcal{F}_{n-1}) = S_{n-1} + \mathbb{E}(Y_n) = S_{n-1},$$

où l'avant dernière égalité est due à la \mathcal{F}_{n-1} -mesurabilité de S_{n-1} et à l'indépendance entre Y_n et \mathcal{F}_{n-1} . Ainsi, (S_n) est bien une (\mathcal{F}_n) -martingale.

2) La suite (M_n) est évidemment (\mathcal{F}_n) -adaptée et, grâce à l'indépendance, dans L^1 . De plus, on peut écrire :

$$\mathbb{E}(M_n | \mathcal{F}_{n-1}) = \mathbb{E}(M_{n-1} X_n | \mathcal{F}_{n-1}) = M_{n-1} \mathbb{E}(X_n | \mathcal{F}_{n-1}) = M_{n-1},$$

puisque la v.a. X_n est indépendante de \mathcal{F}_{n-1} et d'espérance 1.

En revanche, si l'on a $\mathbb{E}(X_n) \geq 1$ (resp. $\mathbb{E}(X_n) \leq 1$), pour tout n de \mathbb{N} , la suite (M_n) est une (\mathcal{F}_n) -sousmartingale (resp. surmartingale).

3) La suite (M_n) est clairement (\mathcal{F}_n) -adaptée. Grâce à l'inégalité de Jensen pour les espérances conditionnelles il vient :

$$\mathbb{E}|M_n| = \mathbb{E}(|\mathbb{E}(X | \mathcal{F}_n)|) \leq \mathbb{E}(\mathbb{E}(|X| | \mathcal{F}_n)) = \mathbb{E}|X|,$$

ce qui prouve que M_n est dans L^1 puisque X l'est. Enfin on a :

$$\mathbb{E}(M_n | \mathcal{F}_{n-1}) = \mathbb{E}(\mathbb{E}(X | \mathcal{F}_n) | \mathcal{F}_{n-1}) = \mathbb{E}(X | \mathcal{F}_{n-1}) = M_{n-1},$$

et la propriété de martingale est bien vérifiée. \square

1.2.3. *Processus prévisibles et version discrète de l'intégrale stochastique.*

DÉFINITION 6.12. Soit $H = (H_n)$ un processus à temps discret et (\mathcal{F}_n) une filtration sur l'espace (Ω, \mathcal{F}, P) . Le processus H est dit (\mathcal{F}_n) -**prévisible** si, pour tout n , la v.a. H_n est \mathcal{F}_{n-1} -mesurable.

Intuitivement, un processus prévisible H est tel que H_n est connu au temps $n - 1$.

DÉFINITION 6.13. Soit (\mathcal{F}_n) une filtration sur l'espace (Ω, \mathcal{F}, P) . Soit (H_n) et (X_n) deux processus sur cet espace. La **transformée de martingale** du processus X par le processus H est le processus noté $H \cdot X$ et défini, pour tout n dans \mathbb{N} , par :

$$H \cdot X_n = \sum_{k=1}^n H_k (X_k - X_{k-1}).$$

THÉORÈME 6.14. On se donne une filtration (\mathcal{F}_n) sur l'espace (Ω, \mathcal{F}, P) , un processus (\mathcal{F}_n) -prévisible (H_n) et (X_n) une (\mathcal{F}_n) -martingale (resp. sousmartingale, surmartingale)

Si le processus (H_n) est borné, i.e. si il existe K dans \mathbb{R}^+ tel que $|H_n(\omega)| \leq K$, pour tout ω de Ω et tout n de \mathbb{N} , alors la transformée de martingale $(H \cdot X)$ est une martingale (resp. sousmartingale, surmartingale) nulle en 0.

On a le même résultat si (H_n) n'est pas borné mais dans L^2 ainsi que (X_n) .

On peut voir dans ce cas la transformée de martingale comme une version discrète de l'intégrale stochastique, notion qui sera étudiée en temps continu en Master 2.

EXERCICE 6.9. *Démontrer ce théorème.*

Preuve/Solution de l'exercice.

Le processus $(H \cdot X)$ est adapté puisque, pour tout n ,

$$H \cdot X_n = \sum_{k=1}^n H_k(X_k - X_{k-1})$$

est clairement \mathcal{F}_n -mesurable. De plus, on a :

$$\mathbb{E}|H \cdot X_n| \leq K \sum_{k=1}^n \mathbb{E}|X_k - X_{k-1}| < +\infty,$$

puisque H est borné et (X_n) dans L^1 . On remarque ici que si, H n'est plus borné mais dans L^2 ainsi que X , alors le processus $H \cdot X$ reste dans L^1 .

Enfin, on peut écrire :

$$\mathbb{E}(H \cdot X_n - H \cdot X_{n-1} | \mathcal{F}_{n-1}) = \mathbb{E}(H_n(X_n - X_{n-1}) | \mathcal{F}_{n-1}) = H_n \mathbb{E}(X_n - X_{n-1} | \mathcal{F}_{n-1}) = 0, \text{ (resp. } \geq, \leq),$$

ce qui prouve la propriété de martingale pour la transformée de martingale. \square

Les transformées de martingales et la propriété de martingale qu'elles vérifient trouvent une interprétation intuitive en théorie des jeux. Supposons en effet que $X_n - X_{n-1}$ représente le gain net d'un joueur par unité mise à la n ième partie d'un jeu de hasard. Comme on l'a dit précédemment, une partie équilibrée revient à supposer que (X_n) est une martingale. En revanche, s'il s'agit d'une sousmartingale (resp. surmartingale), i.e. telle que $\mathbb{E}(X_n - X_{n-1} | \mathcal{F}_{n-1}) \geq 0$ (resp. $\mathbb{E}(X_n - X_{n-1} | \mathcal{F}_{n-1}) \leq 0$), alors le jeu est favorable (resp. défavorable) au joueur qui a des gains d'espérance positive (resp. négative).

Soit maintenant H_n la mise du joueur pour la n ième partie. Il est naturel d'autoriser qu'il choisisse sa mise en se basant sur l'histoire des parties précédentes (on ne peut lui reprocher... même si l'on verra que cela ne modifiera pas ses gains en espérance). En revanche, sans triche possible, il ne pourra utiliser l'information qu'apportera la n ième partie au moment de faire sa mise. C'est pourquoi, on supposera naturellement que le processus des mises (H_n) est prévisible.

Le gain de ce joueur à la n ième partie est donc modélisé par la quantité $H_n(X_n - X_{n-1})$ et ses gains cumulés depuis le début de la partie par $H \cdot X_n$. Le théorème précédent nous assure qu'en espérance ses gains sont constants et nuls.

1.2.4. *Martingale arrêtée et Théorème d'arrêt pour les martingales.* Soit (X_n) une (\mathcal{F}_n) -martingale, nulle en 0, et T un (\mathcal{F}_n) -t.d.a. Considérons le processus (H_n) défini par :

$$H_n = \mathbb{1}_{\{n \leq T\}}.$$

Ce processus est (\mathcal{F}_n) -prévisible puisque l'on a

$$\{n \leq T\} = \overline{\{T \leq n-1\}}.$$

Dans ce cas, la transformée de martingale de (X_n) par (H_n) , d'expression pour tout n

$$H \cdot X_n = \mathbb{1}_{\{1 \leq T\}}(X_1 - X_0) + \cdots + \mathbb{1}_{\{n \leq T\}}(X_n - X_{n-1}),$$

est égale, sur $\{T(\omega) = k\}$, à $X_n(\omega)$ si $n \leq k$ et $X_k(\omega) = X_{T(\omega)}(\omega)$ si $n > k$. Ainsi, on peut écrire :

$$H \cdot X_n = X_{T \wedge n}.$$

On parle de martingale arrêtée et, d'après le Théorème 6.14, il s'agit aussi d'une martingale. D'où la définition et le corollaire au Théorème 6.14 qui suivent.

DÉFINITION 6.15. Soit (X_n) une (\mathcal{F}_n) -martingale et T un (\mathcal{F}_n) -t.d.a. Le processus (X_n^T) défini en tout n de \mathbb{N} par $X_n^T = X_{T \wedge n}$ est appelé **martingale arrêtée au temps T** .

Notons que l'on peut également parler de processus arrêté à un t.d.a. même si le processus n'est pas une martingale.

COROLLAIRE 6.16. Si T est un (\mathcal{F}_n) -t.d.a. et (X_n) une (\mathcal{F}_n) -martingale (resp. sousmartingale, surmartingale) nulle en 0 alors le processus arrêté (X_n^T) est encore une martingale (resp. sousmartingale, surmartingale). De plus, on a :

$$\mathbb{E}X_{T \wedge n} = \mathbb{E}(X_0) = 0, \text{ (resp. } \geq, \leq),$$

pour tout n .

C'est une application directe du Théorème 6.14.

PROPOSITION 6.17. Soit T est un (\mathcal{F}_n) -t.d.a. et (X_n) une (\mathcal{F}_n) -martingale. La (\mathcal{F}_n) -martingale (X_n^T) est aussi une $(\mathcal{F}_{T \wedge n})$ -martingale.

EXERCICE 6.10. Démontrer cette proposition.

Preuve/Solution de l'exercice. Puisque (X_n^T) est une (\mathcal{F}_n) -martingale, on a :

$$\mathbb{E}(X_{T \wedge n} | \mathcal{F}_p) = X_{T \wedge p},$$

pour tout $p \leq n$. L'inclusion $\mathcal{F}_{T \wedge p} \subset \mathcal{F}_p$ et la $\mathcal{F}_{T \wedge p}$ -mesurabilité de $X_{T \wedge p}$ implique (Cf. Proposition 5.8 sur les propriétés de l'espérance conditionnelle) que l'on a :

$$\mathbb{E}(X_{T \wedge n} | \mathcal{F}_{T \wedge p}) = X_{T \wedge p},$$

pour tout $p \leq n$. □

Revenons maintenant sur l'un des résultats donné par le Corollaire précédent, à savoir que pour une martingale arrêtée (X_n^T) , on a $\mathbb{E}X_{T \wedge n} = \mathbb{E}(X_0) = 0$. On peut naturellement se demander si l'on peut avoir le même style de résultat en ne considérant pas la v.a. $X_{T \wedge n}$ mais la v.a. X_T . C'est à dire a-t-on

$$\mathbb{E}X_T = \mathbb{E}(X_0) = 0 ?$$

Et éventuellement, si ce n'est pas toujours le cas, sous quelles conditions peut avoir ce résultat ?

On va d'abord aisément se convaincre qu'un tel résultat n'est pas toujours vrai. Considérons en effet (X_n) une marche aléatoire symétrique sur \mathbb{Z} , démarrant à 0. On sait que l'on peut représenter cette marche aléatoire sous la forme

$$X_n = \sum_{k=1}^n \xi_k,$$

où (ξ_n) est une suite de v.a. i.i.d. de loi de Rademacher, c'est à dire de loi $\frac{1}{2}(\delta_{-1} + \delta_1)$. Le processus (X_n) apparaît alors clairement comme une martingale, nulle en 0. D'après le Corollaire précédent on sait que, pour tout (\mathcal{F}_n) -t.d.a. T , où (\mathcal{F}_n) est la filtration naturelle associée au processus (X_n) , on a

$$\mathbb{E}X_{T \wedge n} = \mathbb{E}(X_0) = 0,$$

pour tout n . Considérons le t.d.a. $T = \inf\{n : X_n = 1\}$ (c'est bien un t.d.a. d'après une propriété vue dans la première section de ce chapitre). L'égalité précédente est vérifiée pour ce t.d.a., mais en revanche elle ne l'est pas si on remplace $T \wedge n$ par T . En effet la v.a. X_T étant *p.s.* égale à 1, on a :

$$1 = \mathbb{E}(X_T) \neq \mathbb{E}(X_0) = 0.$$

Nous allons voir que dans ce cas le Théorème d'arrêt ne s'applique pas.

THÉORÈME 6.18. Théorème d'arrêt

Soit (\mathcal{F}_n) une filtration sur un espace probabilisé (Ω, \mathcal{F}, P) , un (\mathcal{F}_n) -t.d.a. T et (X_n) une (\mathcal{F}_n) -martingale (resp. sousmartingale, surmartingale).

La v.a. X_T est intégrable et on a

$$\mathbb{E}X_T = \mathbb{E}X_0, \text{ (resp. } \geq, \leq),$$

si l'une des conditions suivantes est vérifiée :

- T est un t.d.a. borné, i.e. il existe k tel que $P(T \leq k) = 1$;
- X est bornée, i.e. il existe K dans \mathbb{R}^+ tel que $|X_n(\omega)| \leq K$ pour tout n et tout ω , et T est *p.s.* fini ;
- T est d'espérance finie et il existe K de \mathbb{R}^+ tel que $|X_n(\omega) - X_{n-1}(\omega)| \leq K$, pour tout n et tout ω .

EXERCICE 6.11. Démontrer ce Théorème (Ind. on pourra utiliser le théorème de convergence dominée pour établir le résultat dans le cas des deux dernières conditions).

Preuve/Solution de l'exercice.

1) On sait que l'on a, pour tout n :

$$\mathbb{E}X_{T \wedge n} = \mathbb{E}X_0.$$

En prenant $n = k$, on obtient le résultat.

2) Puisque T est *p.s.* fini, on a

$$X_{T \wedge n} \xrightarrow{p.s.} X_T,$$

quand n tend vers $+\infty$. La suite de v.a. $(X_{T \wedge n})$ étant intégrable puisque bornée, on peut appliquer le théorème de convergence dominée pour obtenir

$$\mathbb{E}X_{T \wedge n} \rightarrow \mathbb{E}X_T,$$

quand $n \rightarrow +\infty$. cette convergence et l'égalité, pour tout n , de $\mathbb{E}X_{T \wedge n}$ avec $\mathbb{E}X_0$ permet d'établir le résultat annoncé.

3) En remarquant que l'on a

$$|X_{T \wedge n} - X_0| = \left| \sum_{k=1}^{T \wedge n} (X_k - X_{k-1}) \right| \leq KT$$

et que T est intégrable par hypothèse, on obtient à nouveau le résultat en appliquant le théorème de convergence dominée. \square

THÉORÈME 6.19. Autre version du théorème d'arrêt

Si (X_n) est une (\mathcal{F}_n) -martingale et T un (\mathcal{F}_n) -t.d.a. borné par k , on a alors :

$$\mathbb{E}(X_k | \mathcal{F}_T) = X_T.$$

EXERCICE 6.12. Démontrer ce Théorème (Ind. En "découpant" suivant les valeurs de T , on pourra montrer directement que l'on a $\mathbb{E}(X_k | \mathcal{F}_T) = X_T$).

Preuve/Solution de l'exercice.

Puisque T est p.s. borné par k , on peut écrire :

$$X_T \stackrel{p.s.}{=} \sum_{p=0}^k X_p \mathbb{1}_{\{T=p\}}.$$

Soit alors A un événement de la tribu \mathcal{F}_T , il vient :

$$\mathbb{E}(X_T \mathbb{1}_A) = \sum_{p=0}^k \mathbb{E}(X_p \mathbb{1}_{\{T=p\}} \mathbb{1}_A) = \sum_{p=0}^k E(X_p \mathbb{1}_{\{T=p\}} \mathbb{1}_A) = E(X_k \mathbb{1}_A),$$

où la seconde égalité est justifiée par l'appartenance de $\{T=p\} \cap A$ à la tribu \mathcal{F}_p (par définition de \mathcal{F}_T) et l'égalité $X_p = \mathbb{E}(X_k | \mathcal{F}_p)$. On a donc bien montré que l'on a $E(X_k | \mathcal{F}_T) = X_T$. \square

1.2.5. Décomposition de Doob d'une sousmartingale.

THÉORÈME 6.20. (Décomposition de Doob)

1) Soit $X = (X_n)$ un processus (\mathcal{F}_n) -adapté et intégrable. Il admet alors la décomposition suivante

$$X = X_0 + M + A,$$

où M est une martingale nulle en 0 et A un processus prévisible également nul en 0. Cette décomposition est unique à une indistinguabilité près (Cf Chapitre 2), au sens où si il existe \tilde{M} et \tilde{A} vérifiant la même équation, on a $P(\forall n : M_n = \tilde{M}_n, A_n = \tilde{A}_n) = 1$.

2) Le processus $X = (X_n)$ est une sousmartingale si, et seulement si, le processus A est un processus croissant, i.e. tel que : $P(\forall n : A_n \leq A_{n+1}) = 1$.

EXERCICE 6.13. *Démontrer ce Théorème (Ind. Dans la première question on pourra construire le processus (A_n) à partir des quantités $\mathbb{E}(X_n - X_{n-1}|\mathcal{F}_{n-1})$).*

Preuve/Solution de l'exercice.

1) Définissons le processus (A_n) , nul en zéro et tel que

$$A_n - A_{n-1} = \mathbb{E}(X_n - X_{n-1}|\mathcal{F}_{n-1}).$$

Ce processus est bien (\mathcal{F}_n) -prévisible et intégrable. Soit maintenant (M_n) le processus défini pour tout n par : $M_n = X_n - X_0 - A_n$. Il est naturellement adapté et intégrable. On a :

$$\mathbb{E}(M_n - M_{n-1}|\mathcal{F}_{n-1}) = \mathbb{E}(X_n - X_{n-1}|\mathcal{F}_{n-1}) - (A_n - A_{n-1}) = 0,$$

ce qui prouve qu'il s'agit bien d'une martingale (évidemment nulle en 0).

Pour démontrer l'unicité de la décomposition, supposons qu'il existe deux autres processus (A'_n) et (M'_n) avec les mêmes propriétés et tels que l'on ait également : $X_n = X_0 + M'_n + A'_n$. On a alors :

$$A'_n - A'_{n-1} = X_n - X_{n-1} - (M'_n - M'_{n-1}).$$

En prenant l'espérance conditionnelle à \mathcal{F}_{n-1} de cette égalité, on obtient :

$$A'_n - A'_{n-1} = \mathbb{E}(A'_n - A'_{n-1}|\mathcal{F}_{n-1}) = E(X_n - X_{n-1}|\mathcal{F}_{n-1}) - E(M'_n - M'_{n-1}|\mathcal{F}_{n-1}) = A_n - A_{n-1},$$

puisque (M'_n) est également une martingale. Les deux processus (A_n) et (A'_n) sont donc égaux, puisqu'ils ont des accroissements égaux et la même valeur initiale 0. Les deux processus (M_n) et (M'_n) sont alors également égaux.

2) Si le processus (X_n) est une sousmartingale, la définition des accroissements du processus (A_n) montre qu'ils sont *p.s.* positifs. Réciproquement, si l'on peut écrire : $X_n = X_0 + M_n + A_n$ avec un processus prévisible (A_n) croissant, il vient :

$$\mathbb{E}(X_n - X_{n-1}|\mathcal{F}_{n-1}) = \mathbb{E}(M_n - M_{n-1}|\mathcal{F}_{n-1}) + \mathbb{E}(A_n - A_{n-1}|\mathcal{F}_{n-1}) = 0 + A_n - A_{n-1} \geq 0,$$

et le processus (X_n) est bien une sousmartingale. \square

On peut trouver un exemple d'application de la décomposition de Doob dans ce que l'on appelle le crochet d'une martingale. Prenons en effet (M_n) une (\mathcal{F}_n) -martingale dans L^2 , nulle en 0 et considérons le processus (M_n^2) . L'inégalité de Jensen nous donne

$$\mathbb{E}(M_n^2|\mathcal{F}_{n-1}) \geq \mathbb{E}^2(M_n|\mathcal{F}_{n-1}) = M_{n-1}^2,$$

et montre ainsi que le processus (M_n^2) est une (\mathcal{F}_n) -sousmartingale. Grâce à la décomposition de Doob, on sait qu'il existe une martingale N nulle en zéro et un processus prévisible croissant A , également nul en zéro, tel que : $M^2 = N + A$. Le processus A est appelé crochet de la martingale M et est noté $\langle M \rangle$.

1.2.6. *Filtration inverse et martingale inverse.*

DÉFINITION 6.21. Une suite décroissante de tribus (\mathcal{F}_n) sur un espace (Ω, \mathcal{F}, P) et telle que $\mathcal{F}_\infty = \bigcap_n \mathcal{F}_n$ est dite **filtration inverse**.

On supposera toujours que \mathcal{F}_∞ contient tous les négligeables de \mathcal{F}_0 .

DÉFINITION 6.22. Soit (\mathcal{F}_n) une filtration inverse sur un espace (Ω, \mathcal{F}, P) probabilisé. Un processus (M_n) est dit **(\mathcal{F}_n) -martingale inverse** si l'on a :

- (1) $\mathbb{E}(M_n | \mathcal{F}_{n+1}) = M_{n+1}$, pour tout n ;
- (2) (M_n) est (\mathcal{F}_n) -adapté ;
- (3) $\mathbb{E}|M_n| < +\infty$, pour tout n .

2. Martingales bornées dans L^1

2.1. Définition.

DÉFINITION 6.23. Soit (\mathcal{F}_n) une filtration sur un espace probabilisé (Ω, \mathcal{F}, P) . Une (\mathcal{F}_n) -martingale (ou sousmartingale ou encore surmartingale) est dite **bornée dans L^1** si l'on a :

$$\sup_n \mathbb{E}|X_n| < +\infty.$$

EXERCICE 6.14.

- 1) Montrer qu'une martingale positive est bornée dans L^1 .
- 2) Montrer qu'une surmartingale positive est également bornée dans L^1 .

Solution de l'exercice.

1) On a

$$\mathbb{E}|X_n| = \mathbb{E}X_n = \mathbb{E}X_0 < +\infty,$$

pour tout n de \mathbb{N} .

2) Comme on a

$$\mathbb{E}X_n = \mathbb{E}(\mathbb{E}(X_n | \mathcal{F}_0)) \leq \mathbb{E}X_0,$$

il vient

$$\mathbb{E}|X_n| = \mathbb{E}X_n \leq \mathbb{E}X_0 < +\infty,$$

pour tout n . □

2.2. Convergences. Pour établir un résultat de convergence *p.s.*, il faut en premier introduire un lemme de Doob sur les montées effectuées par une martingale. Plusieurs méthodes de preuve sont possibles pour ce résultat. Nous choisissons celle utilisant l'interprétation des martingales en théorie des jeux.

Soit donc (X_n) un processus tel que $X_n - X_{n-1}$ représente le gain par unité de temps misee pour la n ème partie. Et considérons la stratégie de jeu suivante :

Choisir deux réels a et b tels que $a < b$.

Répéter sans arrêt :

début de boucle

Ne pas miser tant que X n'est pas passé en dessous de a .

Une fois que X est passé en dessous de a , miser une unité tant que l'on est en dessous de b .

Une fois que X est passé en dessous de b , arrêter de miser.

fin de boucle

Cette stratégie peut être modélisée par le processus prévisible $H = (H_n)$ défini par :

$$H_1 = \mathbb{1}_{\{X_0 < a\}} \text{ et, pour tout } n, H_n = \mathbb{1}_{\{H_{n-1}=1\}} \mathbb{1}_{\{X_{n-1} \leq b\}} + \mathbb{1}_{\{H_{n-1}=0\}} \mathbb{1}_{\{X_{n-1} < a\}}.$$

Comme nous l'avons déjà dit, les gains cumulés sont donnés par la transformée de martingale de X par H , noté $H \cdot X$.

Un exemple de réalisation d'une telle partie et de la stratégie adoptée est donnée dans la Figure 1.

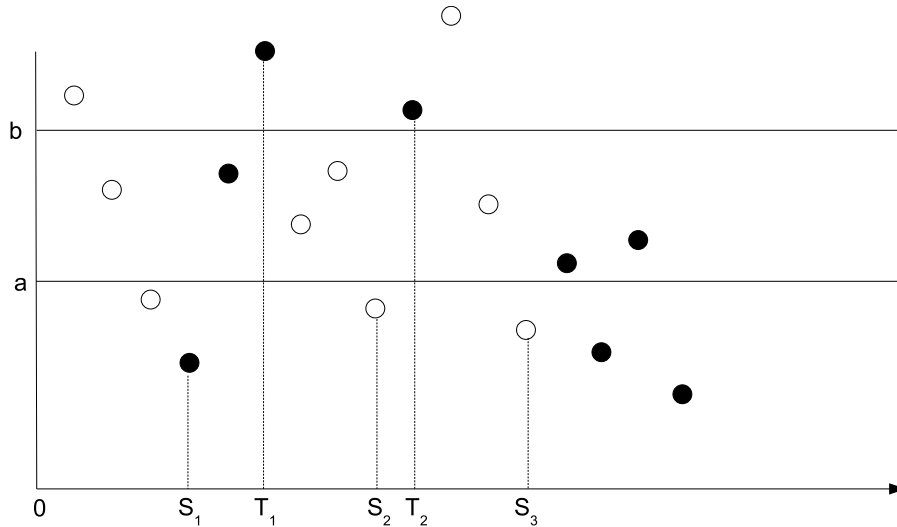


FIGURE 1. Exemple de partie de hasard avec stratégie de montées. Les points blancs représentent les parties sans mise et les noirs avec mise.

Introduisons la suite des t.d.a. définis par les relations :

$$\begin{aligned} S_1 &= \inf\{n : X_n < a\}, \\ T_1 &= \inf\{n > S_1 : X_n > b\}, \\ &\vdots \\ S_p &= \inf\{n > T_{p-1} : X_n < a\}, \\ T_p &= \inf\{n > S_p : X_n > b\}, \end{aligned}$$

et notons $U_N[a, b]$ le nombre de montées entre a et b réalisées par (X_n) au temps N . On a :

$$U_N[a, b] = \sum_p \mathbb{1}_{\{T_p \leq N\}}.$$

En consultant éventuellement la Figure 1, il apparaît clairement que, grâce à la stratégie choisie par le joueur, chaque montée augmente ses gains d'au moins $b - a$. De plus, si au temps N le joueur se trouve dans une période de mise, deux situations peuvent se produire : soit l'on a $X_N > a$, soit $X_N < a$. Dans la première situation, le gain du joueur est positif et donc supérieur à $(X_N(\omega) - a)^-$ qui est nul. Dans la seconde, son gain peut être négatif, mais est de toute façon majoré en valeur absolue par $(X_N(\omega) - a)^-$. Ainsi la perte du joueur, due à l'interruption de la partie au temps N avant qu'il n'ait achevé "sa montée", est de toute façon inférieure à $-(X_N(\omega) - a)^-$. On a donc la formule :

$$Y_N(\omega) = H \cdot X_N(\omega) \geq (b - a)U_N[a, b] - (X_N(\omega) - a)^-.$$

LEMME 6.24. Lemme de Doob sur les montées

Soit (X_n) une (\mathcal{F}_n) -surmartingale et $U_N[a, b]$ le nombre de montées entre a et b jusqu'au temps N pour cette surmartingale. Alors :

$$(b - a)\mathbb{E}(U_N[a, b]) \leq \mathbb{E}[(X_n(\omega) - a)^-].$$

EXERCICE 6.15.

Démontrer ce lemme.

Preuve/Solution de l'exercice. Le processus (H_n) est (\mathcal{F}_n) -prévisible, borné et positif. On sait que la transformée de martingale $H \cdot X$ est une surmartingale. On a donc $\mathbb{E}Y_N \leq 0$ et en utilisant la minoration sur Y_N vue précédemment, on obtient le résultat. \square

COROLLAIRE 6.25. Soit (X_n) une surmartingale bornée dans L^1 et $a < b$ deux réels. Notons

$$U_\infty[a, b] = \lim_{N \rightarrow +\infty} U_N[a, b].$$

On a alors

$$(b - a)\mathbb{E}(U_\infty[a, b]) \leq |a| + \sup_n \mathbb{E}|X_n|$$

et donc $P(U_\infty[a, b] = \infty) = 0$.

EXERCICE 6.16.

Démontrer ce lemme.

Preuve/Solution de l'exercice. D'après le lemme précédent, on a :

$$(b - a)\mathbb{E}(U_N[a, b]) \leq |a| + \mathbb{E}|X_N| \leq |a| + \sup_n \mathbb{E}|X_n|.$$

Puisque la suite $(U_N[a, b])_N$ est positive et croissante vers $U_\infty[a, b]$, le théorème de convergence monotone nous donne :

$$(b - a)\mathbb{E}(U_\infty[a, b]) \leq |a| + \sup_n \mathbb{E}|X_n|$$

ce qui entraîne que $U_\infty[a, b]$ est *p.s.* fini. \square

THÉORÈME 6.26. *Une martingale, sousmartingale, surmartingale bornée dans L^1 est p.s. convergente, i.e.*

$$p.s. : \lim_{n \rightarrow +\infty} X_n = X_\infty.$$

EXERCICE 6.17.

Démonstration de ce théorème.

1) *Montrer dans un premier temps que*

$$p.s. : \lim_{n \rightarrow +\infty} X_n = X_\infty,$$

avec X_∞ dans $\bar{\mathbb{R}}$. On pourra pour cela utiliser les événements de la forme

$$\Lambda_{a,b} = \{\omega : \liminf_n X_n(\omega) < a < b < \limsup_n X_n(\omega)\}.$$

2) *En utilisant le Lemme de Fatou, on montrera que la v.a. limite X_∞ est p.s. fini.*

Preuve/Solution de l'exercice.

1) En notant Λ l'ensemble des ω de Ω pour lesquels $X_n(\omega)$ n'est pas convergent dans $\bar{\mathbb{R}}$, on peut écrire :

$$\Lambda = \{\omega : \liminf_n X_n(\omega) < \limsup_n X_n(\omega)\} = \bigcup_{(a,b) \in \mathbb{Q}^2 : a < b} \Lambda_{a,b}.$$

Mais on a :

$$\Lambda_{a,b} \subseteq \{\omega : U_\infty[a, b] = \infty\}$$

et on a vu dans le corollaire précédent que la probabilité de ce dernier ensemble est nulle. Par dénombrabilité des rationnels, l'ensemble Λ est donc également de probabilité nulle.

2) En utilisant le Lemme de Fatou, on peut écrire :

$$\mathbb{E}|X_\infty| = \mathbb{E}(\liminf_n |X_n|) \leq \liminf_n \mathbb{E}|X_n| \leq \sup_n \mathbb{E}|X_n|,$$

qui est fini puisque, par hypothèse, la martingale (ou surmartingale ou sousmartingale) est bornée dans L^1 . La v.a. X_∞ est donc bien p.s. fini. \square

Notons bien que l'on n'a pas dit que la martingale (sousmartingale ou surmartingale) est convergente dans L^1 .

COROLLAIRE 6.27. *Une surmartingale (ou une martingale) positive est p.s. convergente.*

Preuve. On a déjà vu qu'une surmartingale (ou martingale) positive est bornée dans L^1 , d'où le résultat grâce au théorème précédent. \square

3. Inégalité de Doob et conséquences

THÉORÈME 6.28. *Inégalité de Doob*

Soit (X_n) une sousmartingale positive et notons $X_n^* = \sup_{p \leq n} X_p$.

Alors, pour tout $\lambda > 0$, on a :

$$\lambda P(X_n^* \geq \lambda) \leq \int_{X_n^* \geq \lambda} X_n dP.$$

EXERCICE 6.18.

Démonstration de ce théorème.

1) En considérant les t.d.a. $T = \inf\{p : X_p \geq \lambda\}$ et $S = T \wedge n$, montrer que l'on a

$$X_S \geq \lambda \mathbb{1}_{\{T \leq n\}} + X_n \mathbb{1}_{\{T > n\}}.$$

2) En appliquant convenablement un théorème d'arrêt, en déduire le résultat.

Preuve/Solution de l'exercice.

1) On a, d'après la définition du t.d.a. T :

$$X_S = X_T \mathbb{1}_{\{T \leq n\}} + X_n \mathbb{1}_{\{T > n\}} \geq \lambda \mathbb{1}_{\{T \leq n\}} + X_n \mathbb{1}_{\{T > n\}}.$$

2) En appliquant le théorème d'arrêt au t.d.a. borné S , il vient : $\mathbb{E}X_S \leq \mathbb{E}X_n$. En prenant l'espérance de l'inégalité établie à la question précédente, on obtient :

$$\begin{aligned} \mathbb{E}X_n &\geq \lambda P(T \leq n) + \mathbb{E}(X_n \mathbb{1}_{\{T > n\}}) \\ \iff \mathbb{E}(X_n \mathbb{1}_{\{T \leq n\}}) &\geq \lambda P(T \leq n), \end{aligned}$$

ce qui nous donne le résultat, en remarquant que l'on a : $\{T \leq n\} = \{X_n^* \geq \lambda\}$. □

COROLLAIRE 6.29. *Inégalité de Kolmogorov*

Soit (X_n) une suite de v.a. de carré intégrables, centrées et indépendantes. Définissons

$$S_n = X_1 + \dots + X_n \text{ et } S_n^* = \sup_{p \leq n} S_p.$$

Alors, pour tout $a > 0$:

$$P(S_n^* \geq a) \leq \frac{1}{a^2} \sum_{i=1}^n \mathbb{E}(X_i^2).$$

EXERCICE 6.19.

Démontrer ce Corollaire.

Preuve/Solution de l'exercice. Le processus (S_n) est une martingale par rapport à la filtration naturelle associées au processus (X_n) . Comme on l'a vu en fin de Section 6.20, le processus (S_n^2) est donc une sousmartingale positive. On peut alors appliquer l'inégalité de Doob pour obtenir :

$$P(S_n^* \geq a) \leq P((S_n^*)^2 \geq a^2) \leq \frac{1}{a^2} \mathbb{E}S_n^2 = \frac{1}{a^2} \sum_{i=1}^n \mathbb{E}X_i^2,$$

la dernière égalité étant justifiée par indépendance des (X_n) . \square

COROLLAIRE 6.30.

1) Soit (X_n) une sousmartingale positive et bornée dans L^1 . La v.a. $X_\infty^* = \sup_n X_n$ est alors p.s. finie.

2) Soit (X_n) une sousmartingale positive et dans L^p , pour un réel $p > 1$. On a alors :

$$\|X_n^*\|_{L^p} \leq q \|X_n\|_{L^p},$$

où q est tel que : $1/q + 1/p = 1$.

EXERCICE 6.20.

Démonstration de ce Corollaire.

1) Montrer qu'en notant $M = \sup_n \mathbb{E}X_n$ qui est fini par hypothèse, on a pour tout $\lambda > 0$:

$$P(X_\infty^* \geq \lambda) \leq \frac{M}{\lambda}.$$

En déduire le résultat annoncé dans le 1) du Corollaire.

2) Montrer que l'on peut écrire, pour tout n :

$$\mathbb{E}(X_n^*)^p \leq p \mathbb{E}(X_n \int_0^{X_n^*} \lambda^{p-2} d\lambda).$$

(Ind. On pourra utiliser le Théorème de Tonelli et l'inégalité de Doob).

En supposant dans un premier temps que la sousmartingale positive (X_n) est bornée, montrer que l'on peut en déduire que l'on a :

$$(\mathbb{E}(X_n^*)^p)^{1/p} \leq \frac{p}{p-1} (\mathbb{E}(X_n)^p)^{1/p}.$$

Étendre ce résultat au cas où (X_n) n'est pas bornée.

Preuve/Solution de l'exercice.

1) L'inégalité de Doob nous donne, pour tout $\lambda > 0$:

$$P(X_n^* \geq \lambda) \leq \frac{\mathbb{E}X_n}{\lambda} \leq \frac{M}{\lambda}.$$

La croissance de la suite (X_n^*) vers X_∞^* , implique la croissance des événements de la forme $\{X_n^* \geq \lambda\}$ vers $\{X_\infty^* \geq \lambda\}$ quand $n \rightarrow +\infty$. On a donc bien :

$$P(X_\infty^* \geq \lambda) \leq \frac{M}{\lambda}.$$

Mais l'égalité

$$\{X_\infty^* = \infty\} = \bigcap_n \{X_\infty^* \geq n\}$$

permet d'obtenir

$$P(X_\infty^* = \infty) = \lim_{n \rightarrow +\infty} P(X_\infty^* \geq n) \leq \lim_{n \rightarrow +\infty} \frac{M}{n} = 0.$$

La v.a. X_∞^* est donc bien *p.s.* finie. Autrement dit la suite (X_n) est *p.s.* bornée.

2) On peut écrire :

$$\begin{aligned} \mathbb{E}(X_n^*)^p &= \mathbb{E} \left(\int_0^{X_n^*} p \lambda^{p-1} d\lambda \right) = p \int_0^{+\infty} \mathbb{E}(\mathbb{1}_{\{\lambda \leq X_n^*\}}) \lambda^{p-1} d\lambda \\ &\leq p \int_0^{+\infty} \mathbb{E}(\mathbb{1}_{\{\lambda \leq X_n^*\}} X_n) \lambda^{p-2} d\lambda = p \mathbb{E} \left(X_n \int_0^{X_n^*} \lambda^{p-2} d\lambda \right). \end{aligned}$$

De plus, on a :

$$\mathbb{E} \left(X_n \int_0^{X_n^*} \lambda^{p-2} d\lambda \right) = \frac{1}{p-1} \mathbb{E}(X_n (X_n^*)^{p-1}).$$

L'inégalité précédente et celle de Hölder, nous donnent alors :

$$\mathbb{E}(X_n^*)^p \leq \frac{p}{p-1} (\mathbb{E}(X_n)^p)^{\frac{1}{p}} (\mathbb{E}(X_n^*)^p)^{1-\frac{1}{p}}.$$

Comme l'on suppose dans un premier temps que X_n^* est bornée, on peut diviser cette dernière inégalité par son dernier terme à droite pour obtenir :

$$(\mathbb{E}(X_n^*)^p)^{\frac{1}{p}} \leq \frac{p}{p-1} (\mathbb{E}(X_n)^p)^{\frac{1}{p}}.$$

Si la sousmartingale (X_n) n'est pas bornée, on applique le résultat précédent à la sousmartingale $Y_n^k = X_n \wedge k$ qui est, elle, bornée. On a donc, pour tout n :

$$\left(\mathbb{E}(Y_n^{k,*})^p \right)^{\frac{1}{p}} \leq \frac{p}{p-1} \left(\mathbb{E}(Y_n^k)^p \right)^{\frac{1}{p}}.$$

La convergence croissante de (Y_n^k) vers X_n , quand $k \rightarrow +\infty$ et le Théorème de convergence monotone, permet d'achever la démonstration. \square

4. Martingales bornées dans L^p , pour $p > 1$

4.1. Définition et convergences.

DÉFINITION 6.31. Une martingale (sousmartingale ou surmartingale) (X_n) est dite **bornée dans L^p** si :

$$\sup_n \|X_n\|_{L^p} < +\infty.$$

THÉORÈME 6.32. *Une martingale (resp. sousmartingale, surmartingale) bornée dans L^p , pour un réel $p > 1$, est convergente p.s. et dans L^p , i.e.*

$$X_n \xrightarrow[p.s.]{L^p} X_\infty, \text{ quand } n \rightarrow +\infty.$$

De plus, on a :

$$X_n = \mathbb{E}(X_\infty | \mathcal{F}_n) \text{ (resp. } \leq, \geq).$$

EXERCICE 6.21. *Démonstration de ce Théorème dans le cas d'une martingale.*

1) Montrer que (X_n) converge p.s. vers une v.a. X_∞ , quand n tend vers $+\infty$.

2) En notant $X_n^* = \sup_{p \leq n} |X_p|$, montrer que

$$\|X_\infty^*\|_{L^p} \leq q \sup_n \|X_n\|_{L^p},$$

où q est tel que $1/p + 1/q = 1$.

3) Dédurre du 1) et du 2) que l'on a bien également la convergence dans L^p de (X_n) vers X_∞ .

4) Montrer que si une suite (U_n) converge vers U_∞ dans L^1 alors, on a pour toute sous tribu \mathcal{B} de \mathcal{F} :

$$\mathbb{E}(U_n | \mathcal{B}) \xrightarrow{L^1} \mathbb{E}(U_\infty | \mathcal{B}), \text{ quand } n \rightarrow +\infty.$$

En déduire que $X_n = \mathbb{E}(X_\infty | \mathcal{F}_n)$.

Preuve/Solution de l'exercice.

1) Puisque la martingale (X_n) est bornée dans L^p pour $p > 1$, elle est bornée dans L^1 et donc p.s. convergente vers une v.a. X_∞ d'après un résultat de la section précédente.

2) Par l'inégalité de Jensen, on a déjà vu que le processus $|X_n|$ est une sousmartingale positive. Le Corollaire 6.30 de l'inégalité de Doob, nous donne

$$\|X_n^*\|_{L^p} \leq q \|X_n\|_{L^p} \leq q \sup_n \|X_n\|_{L^p} < +\infty.$$

Comme la suite (X_n^*) converge en croissant vers X_∞^* , l'inégalité précédente et le théorème de convergence monotone conduisent à

$$\|X_\infty^*\|_{L^p} \leq q \sup_n \|X_n\|_{L^p}.$$

3) D'après le 1), on a la convergence

$$|X_n - X_\infty|^p \xrightarrow{p.s.} 0, \text{ quand } n \rightarrow +\infty.$$

Par ailleurs, l'inégalité $|X_n - X_\infty| \leq 2X_\infty^*$ montre que la suite $(|X_n - X_\infty|^p)$ est majorée par une v.a. dans L^1 . Par le théorème de convergence dominée, on a donc :

$$\mathbb{E}|X_n - X_\infty|^p \rightarrow 0, \text{ quand } n \rightarrow +\infty.$$

4) L'inégalité de Jensen conduit à :

$$\mathbb{E}(|\mathbb{E}(U_n - U_\infty | \mathcal{B})|) \leq \mathbb{E}(\mathbb{E}(|U_n - U_\infty| | \mathcal{B})) = \|U_n - U_\infty\|_{L^1}.$$

La convergence dans L^1 de (U_n) entraîne donc celle $(\mathbb{E}(U_n | \mathcal{B}))$, toujours dans L^1 , vers $\mathbb{E}(U_\infty | \mathcal{B})$.

En appliquant ce résultat à la suite $(\mathbb{E}(X_m|\mathcal{F}_n))_m$, pour n fixé, on obtient

$$X_n = \mathbb{E}(X_\infty|\mathcal{F}_n)$$

puisque chaque terme de la suite est égal à X_n par la propriété de martingales. \square

COROLLAIRE 6.33. Soit (\mathcal{F}_n) une filtration et

$$\mathcal{F}_\infty = \bigvee_n \mathcal{F}_n$$

la limite croissante des \mathcal{F}_n . Soit X une v.a. dans L^p , pour $p > 1$.

On a alors :

$$\mathbb{E}(X|\mathcal{F}_n) \xrightarrow[p.s.]{L^p} \mathbb{E}(X|\mathcal{F}_\infty).$$

EXERCICE 6.22. Démonstration de ce Corollaire.

1) Notons $X_n = \mathbb{E}(X|\mathcal{F}_n)$ et $Z = \mathbb{E}(X|\mathcal{F}_\infty)$. Montrer que la suite (X_n) converge p.s. et dans L^p vers une v.a. X_∞ .

2) Montrer que pour tout A dans $\cup_n \mathcal{F}_n$, il existe un rang n_0 tel que pour tout $n \geq n_0$ on ait :

$$\mathbb{E}(X_n \mathbb{1}_A) = \mathbb{E}(X \mathbb{1}_A).$$

En déduire que, pour tout A de $\cup_n \mathcal{F}_n$, on a :

$$\mathbb{E}(X \mathbb{1}_A) = \mathbb{E}(X_\infty \mathbb{1}_A)$$

et achever la démonstration du Corollaire.

Preuve/Solution de l'exercice.

1) On sait que (X_n) est une martingale. Par l'inégalité de Jensen pour les espérances conditionnelles, on montre facilement qu'elle est bornée dans L^p . Le théorème précédent assure alors sa convergence p.s. et dans L^p vers une v.a. terminale X_∞ .

2) Puisque A est dans $\cup_n \mathcal{F}_n$, il existe un rang n_0 tel que, pour tout $n \geq n_0$, l'événement A est dans \mathcal{F}_n . L'égalité $X_n = \mathbb{E}(X|\mathcal{F}_n)$, permet alors d'écrire pour tout $n \geq n_0$:

$$\mathbb{E}(X_n \mathbb{1}_A) = \mathbb{E}(X \mathbb{1}_A).$$

De plus la convergence dans L^1 de (X_n) vers X_∞ assure celle de $(X_n \mathbb{1}_A)$ vers $X_\infty \mathbb{1}_A$ et donc la convergence

$$\mathbb{E}(X_n \mathbb{1}_A) \rightarrow \mathbb{E}(X_\infty \mathbb{1}_A), \text{ quand } n \rightarrow +\infty.$$

De ces deux derniers résultats, on tire l'égalité

$$\mathbb{E}(X \mathbb{1}_A) = \mathbb{E}(X_\infty \mathbb{1}_A),$$

pour tout A dans $\cup_n \mathcal{F}_n$. Le théorème des classes monotones permet de terminer la démonstration puisque l'on sait que \mathcal{F}_∞ est engendré par $\cup_n \mathcal{F}_n$. \square

4.2. Le cas particuliers des martingales dans L^2 . Considérons (M_n) une martingale dans L^2 . Remarquons en premier lieu qu'une telle martingale est à accroissements orthogonaux dans L^2 . En effet, pour des entiers $s \leq t \leq u \leq v$, on peut écrire :

$$\begin{aligned} \langle M_v - M_u, M_t - M_s \rangle_{L^2} &= \mathbb{E}((M_v - M_u)(M_t - M_s)) = \mathbb{E}(\mathbb{E}((M_v - M_u)(M_t - M_s)|\mathcal{F}_u)) \\ &= \mathbb{E}((M_t - M_s)\mathbb{E}((M_v - M_u)|\mathcal{F}_u)) = 0. \end{aligned}$$

Ainsi l'égalité

$$M_n = M_0 + \sum_{k=1}^n (M_k - M_{k-1})$$

exprime M_n comme une somme de termes orthogonaux dans L^2 et, d'après le théorème de Pythagore, on a :

$$(3) \quad \mathbb{E}(M_n^2) = \mathbb{E}M_0^2 + \sum_{k=1}^n \mathbb{E}(M_k - M_{k-1})^2.$$

On a alors le théorème suivant.

THÉORÈME 6.34. *Soit (M_n) une martingale dans L^2 . Cette martingale est bornée dans L^2 si, et seulement si,*

$$\sum_{k=1}^{+\infty} \mathbb{E}(M_k - M_{k-1})^2 < +\infty,$$

ce qui est encore équivalent à dire que :

$$M_n \xrightarrow{L^2} M_\infty, \text{ quand } n \rightarrow +\infty.$$

Dans ce cas, la convergence est également p.s.

Preuve.

En utilisant l'égalité (3), il est évident que la martingale est bornée dans L^2 si, et seulement si,

$$\sum_{k=1}^{+\infty} \mathbb{E}(M_k - M_{k-1})^2 < +\infty.$$

Le reste est une application du Théorème 6.34. □

COROLLAIRE 6.35. Somme de v.a. indépendantes, centrées et dans L^2

Soit (X_n) une suite de v.a. dans L^2 , centrées et indépendantes. Notons $\sigma_k^2 = \text{Var}X_k$.

La série $\sum_{k=1}^{+\infty} X_k$ est p.s. convergente dès que

$$\sum_{k=1}^{+\infty} \sigma_k^2 < +\infty.$$

EXERCICE 6.23. *Démontrer ce Corollaire.*

Preuve/Solution de l'exercice. Soit (\mathcal{F}_n) la filtration naturelle associée à la suite (X_n) et définissons $M_n = \sum_{i=1}^n X_i$, pour tout $n \geq 1$ et $M_0 = 0$. On sait que (M_n) est une martingale et les hypothèses du Corollaire assurent qu'elle est dans L^2 . On a alors :

$$\mathbb{E}(M_n^2) = \sum_{k=1}^n \mathbb{E}(M_k - M_{k-1})^2.$$

En remarquant que

$$\mathbb{E}(M_k - M_{k-1})^2 = \mathbb{E}(X_k)^2 = \sigma_k^2,$$

le Théorème 6.34 permet d'obtenir le résultat. □