

Philippe CARMONA

**NOTES DU COURS
PROCESSUS STOCHASTIQUES**

Philippe CARMONA

Laboratoire Jean Leray, UMR 6629, Université de Nantes, 92208, F-44322, Nantes cedex 03,
e-mail: philippe.carmona@math.univ-nantes.fr

NOTES DU COURS PROCESSUS STOCHASTIQUES

Philippe CARMONA

Résumé. — Ces notes sont un support pour le cours d'introduction aux processus stochastiques, de l'option mathématique de l'ECN Nantes. Nous conseillons la lecture d'ouvrages plus conséquents:

- “Théorèmes limites et processus de Poisson”, Grégory Miermont, <http://perso.ens-lyon.fr/gregory.miermont/thlim.pdf>
- “Essentials of stochastic processes”, R. Durrett , Springer
- “A first course in stochastic models”, H. Tijms, Wiley.

1

Processus de Poisson

SECTION 1

La loi exponentielle

Définition 1.1. — On dit que la variable aléatoire réelle X suit une loi exponentielle de paramètre $\lambda > 0$, et on note $X \sim \exp(\lambda)$ si elle admet pour densité $f_\lambda(x) = \lambda e^{-\lambda x} \mathbf{1}_{(x>0)}$.

On vérifie sans peine que la fonction de répartition de X est

$$F(t) = \mathbb{P}(X \leq t) = 1 - e^{-\lambda t} \quad (t \geq 0),$$

que X est de carré intégrable, $\mathbb{E}[X] = \frac{1}{\lambda}$, $\text{Var}(X) = \frac{1}{\lambda^2}$ et que $\frac{X}{\lambda} \sim \exp(1)$.

Proposition 1.1 (Manque de mémoire de la loi exponentielle)

Soit X une variable aléatoire réelle telle que $0 < X < +\infty$ p.s. Alors X satisfait l'équation fonctionnelle

$$\mathbb{P}(X > t + s \mid X > t) = \mathbb{P}(X > s) \quad (s, t \geq 0)$$

ssi il existe $\lambda > 0$ tel que $X \sim \exp(\lambda)$.

Démonstration. — Il est trivial de vérifier que la loi exponentielle vérifie l'équation fonctionnelle car $\mathbb{P}(X > s) = e^{-\lambda s}$, pour $s \geq 0$.

Réciproquement si X satisfait l'équation fonctionnelle $f(t+s) = f(s)f(t)$ avec $f(s) = \mathbb{P}(X > s)$, alors on a $f(s) > 0$ pour tout $s \geq 0$ car si $f(s) = 0$, alors $f(s/2) = \sqrt{f(s)} = 0$ et donc $1 = f(0) = \lim_{n \rightarrow +\infty} f(s/2^n) = 0$ ce qui est contradictoire. On peut donc considérer la fonction $g(t) = \ln f(t)$ qui vérifie l'équation de Cauchy

$$g(t+s) = g(t) + g(s) \quad (s, t \geq 0).$$

Il est bien connu que comme g est localement intégrable, il existe une constante réelle a telle que $g(x) = ax$, ce qui donne $f(t) = e^{at}$. Comme $X < +\infty$ ps, il existe t tel que $f(t) < 1$ ce qui montre que $a < 0$. \square

Proposition 1.2 (La course d'exponentielles). — Soit X, Y deux variables aléatoires indépendantes, $X \sim \exp(\lambda)$, $Y \sim \exp(\mu)$. Alors $Z = \inf(X, Y) \sim \exp(\lambda + \mu)$.

Démonstration. — Pour $t \geq 0$, $\mathbb{P}(Z > t) = \mathbb{P}(X > t, Y > t) = e^{-\lambda t}e^{-\mu t}$. \square

Si on fait une course, on désire non seulement connaître le temps du gagnant, mais aussi l'identité du gagnant. Soit donc X_1, \dots, X_n des variables aléatoires indépendantes qui suivent des lois exponentielles $X_i \sim \exp(\lambda_i)$. Alors $Z = \inf(X_1, \dots, X_n)$ est le temps du gagnant. Observons que par indépendance, si $i \neq j$, alors

$$\mathbb{P}(X_i = X_j) = \int \mathbf{1}_{(x=y)} f_{\lambda_i}(x) f_{\lambda_j}(y) dx dy = 0$$

et donc, par additivité,

$$\mathbb{P}(\exists i \neq j, X_i = X_j) \leq \sum_{i \neq j} \mathbb{P}(X_i = X_j) = 0.$$

On se placera donc sur un ensemble de probabilité 1,

$$\Omega' = \{\omega : \forall i \neq j, X_i(\omega) \neq X_j(\omega)\}$$

sur lequel la variable aléatoire numéro du gagnant est bien définie:

$$N = \inf \{i \leq n : X_i = Z\}.$$

Proposition 1.3 (Course d'exponentielles générale)

Les variables aléatoires N et Z sont indépendantes,

$$Z \sim \exp(\bar{\lambda}) \quad \text{et} \quad \mathbb{P}(N = i) = \frac{\lambda_i}{\bar{\lambda}} \quad (1 \leq i \leq n).$$

Démonstration. — On établit que pour $t \geq 0$ et $1 \leq i \leq n$, on a

$$\mathbb{P}(Z \geq t, N = i) = e^{-\bar{\lambda}t} \frac{\lambda_i}{\bar{\lambda}}.$$

On fait $t = 0$ pour obtenir la loi de N , puis on en déduit que, sous la loi conditionnelle $\mathbb{P}(\cdot | N = i)$ la fonction de répartition de Z est celle de la loi $\nu = \exp(\bar{\lambda})$. D'où le fait que la loi conditionnelle de Z sachant $N = i$ est ν . On conclut en écrivant que pour f, g mesurables positives:

$$\mathbb{E}[f(Z)h(N)] = \sum_i \mathbb{E}[f(Z) | N = i] h(i) \mathbb{P}(N = i) = \sum_i \nu(f) h(i) \mathbb{P}(N = i) = \nu(f) \mathbb{E}[h(N)].$$

\square

Exercice

Une autre histoire de bus. Pour se rendre de la gare à l'aéroport on a le choix entre deux lignes, la 1 et la 2. Les bus de la ligne 1 arrivent à l'arrêt suivant un processus de Poisson de paramètre λ_1 , ceux de la ligne 2 suivant un processus de Poisson de paramètre λ_2 . Les deux processus sont supposés indépendants. Les temps de parcours entre la gare et l'aéroport sont supposés déterministes, et valent respectivement t_1 et t_2 minutes. On supposera pour les applications numériques que $\lambda_1 = 1/5$ et $\lambda_2 = 1/10$.

John est un adepte de la ligne 1, car il trouve que les bus passent plus souvent : lorsqu'il arrive à la gare, il attend un bus de la ligne 1, puis se rend à l'aéroport. Sarah elle est une adepte de la ligne 2, car elle trouve que les bus vont plus vite : elle attend un bus de la ligne 2, puis se rend à l'aéroport.

1. Déterminer les temps de transport moyen de John et Sarah. En déduire une Condition nécessaire et Suffisante, portant sur t_1, t_2 pour que John ait la meilleure stratégie : sur un grand nombre de voyages il met en moyenne moins de temps que Sarah. Tracer, dans le plan (t_1, t_2) , la zone où John a raison.
2. Une troisième personne, Bill, a une stratégie différente : elle prend le premier bus disponible. Calculer son temps moyen de parcours, et tracer dans le plan les zones où John, Bill et Sarah ont la meilleure stratégie.
3. Application numérique $t_1 = 25$. Indiquer suivant les valeurs de t_2 , quelle est la meilleure stratégie. Interpréter vos résultats.

Remarque. — *Cet exercice utilise en premier lieu la loi des grands nombres pour se ramener à comparer des moyennes. La réponse à la première question est élémentaire pour une personne douée de bon sens : si $t_1 = 25$, comme les bus de la ligne 1 passent en moyenne toutes les 5 minutes, le temps moyen de transport de John est de 30 mn. Si $t_2 > 20$, alors Sarah mettra en moyenne plus de 30 mn, et donc John a la meilleure stratégie. Il est plus difficile de comparer la stratégie de Bill aux autres car son temps moyen d'attente est 3,33 mn, facile à calculer mais difficile à deviner!*

La loi Gamma On rappelle que la loi $\gamma(a, b)$ de paramètre d'échelle $a > 0$ et de paramètre de forme $b > 0$ admet pour densité:

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

avec $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx$ qui vérifie la relation $\Gamma(a+1) = a\Gamma(a)$ (on en déduit $\Gamma(n+1) = n!$).

Lemme 1.4 (additivité des lois gamma). —

$$\gamma(a, b) * \gamma(a', b) = \gamma(a + a', b).$$

Démonstration. — Il suffit de calculer la transformée de Laplace de la loi γ . Si $X \sim \gamma(a, b)$, alors

$$\mathbb{E}[e^{sX}] = \begin{cases} +\infty & \text{si } s \geq b \\ \left(\frac{b}{b-s}\right)^a & \text{sinon.} \end{cases}$$

□

Corollaire 1.5. — Soit X_1, \dots, X_n indépendantes de même loi $\exp(\lambda)$. Alors $S_n = X_1 + \dots + X_n$ suit la loi $\gamma(n, \lambda)$ et donc admet pour densité $\lambda^n \frac{x^{n-1}}{(n-1)!} e^{-\lambda x} \mathbf{1}_{(x>0)}$.

Démonstration. — On note que $\exp(\lambda) = \gamma(\lambda, 1)$ et on applique le Lemme précédent. □

La loi de Poisson Rappelons qu'une variable aléatoire entière N suit la loi de Poisson de paramètre $\lambda > 0$, notée $\mathcal{P}(\lambda)$, si

$$\mathbb{P}(N = k) = e^{-\lambda k} \frac{\lambda^k}{k!} \quad (k \in \mathbb{N}).$$

On montre aisément que $\mathbb{E}[N] = \text{Var}(N) = \lambda$, par exemple en calculant la fonction génératrice, la fonction caractéristique ou la transformée de Laplace: pour $|u| \leq 1; s, t \in \mathbb{R}$, on a :

$$G_N(u) = \mathbb{E}[u^N] = e^{\lambda(u-1)}, \quad \mathbb{E}[e^{sN}] = e^{\lambda(e^s-1)} \quad \mathbb{E}[e^{itN}] = e^{\lambda(e^{it}-1)}.$$

On déduit aussi immédiatement de la forme de la fonction caractéristique l'additivité des lois de Poisson:

$$\mathcal{P}(\lambda) * \mathcal{P}(\mu) = \mathcal{P}(\lambda + \mu).$$

Exercice

Soit N_1, N_2 deux variables aléatoires indépendantes qui suivent des lois de Poisson de paramètres respectifs λ_1 et λ_2 , et soit $N = N_1 + N_2$.

1. Déterminer la loi de N .
2. Montrer que conditionnellement à $N = k$, avec $k \in \mathbb{N}^*$, la loi de N_1 est une binomiale $\mathcal{B}(n, p)$ dont on déterminera les paramètres n et p .

SECTION 2

Le Processus de Poisson homogène

Le processus de Poisson est un processus de comptage : on compte le nombre d'occurrences au cours du temps d'un évènement spécifique (arrivée d'un appel téléphonique, entrée d'un client dans une boutique, etc ...). Il apparaît naturellement comme processus limite (voir section 4). Faire un dessin! Plus rigoureusement,

Définition 1.2. — *Un Processus de Poisson de paramètre λ est une famille de variables aléatoires $(N_t)_{t \geq 0}$ à valeurs dans \mathbb{N} telle que*

- (a) *la fonction $t \rightarrow N_t$ est croissante, continue à droite et ne croît que par sauts de 1.*
- (b) *Pour tous $s, t \geq 0$, $N_{t+s} - N_s \sim \mathcal{P}(\lambda t)$.*
- (c) *Si $t_0 < t_1 < \dots < t_n$, alors les variables aléatoires $(N_{t_{i+1}} - N_{t_i})_{0 \leq i \leq n-1}$ sont indépendantes.*

Remarque. — *Un processus qui vérifie (c) est dit à accroissements indépendants. La propriété (b) dit que les accroissements sont stationnaires (et poissoniens).*

N_t est donc le nombre d'évènements spécifiques qui se sont produits dans l'intervalle $[0, t]$.

On montrera plus tard comme conséquence du théorème de construction d'une mesure de Poisson générale, que si on se donne des variables aléatoires $(X_n)_{n \in \mathbb{N}}$ indépendantes de même loi $\exp(\lambda)$ et si on pose

$$S_0 = 0, S_n = X_1 + \dots + X_n, n \geq 1, \quad N_t = \sum_{n \geq 1} \mathbf{1}_{(S_n \leq t)}$$

alors $(N_t)_{t \geq 0}$ est un Processus de Poisson de paramètre λ . Les variables S_n sont appelées instants de saut du processus N , et on observe que par construction

$$\{N_t \geq n\} = \{S_n \leq t\} \quad \text{et donc} \quad \{N_t = n\} = \{S_n \leq t < S_{n+1}\}.$$

Proposition 1.6. — *Conditionnellement à $N_t = n$, (S_1, \dots, S_n) a même loi que (U'_1, \dots, U'_n) statistiques s'ordre de (U_1, \dots, U_n) IID de loi uniforme sur $[0, t]$. Cette loi nommée $\mathcal{D}_{n,t}$ est appelée loi de Dirichlet $\mathcal{D}_{n,t}$ et admet pour densité*

$$g_{n,t}(s_1, \dots, s_n) = \frac{n!}{t} \mathbf{1}_{(0 < s_1 < s_2 < \dots < s_n < t)}.$$

SECTION 3

Construction d'un processus de Poisson général

Définition 1.3. — On appelle mesure ponctuelle sur un espace mesurable (E, \mathcal{E}) une somme finie ou dénombrable de masses de Dirac.

On note $\mathcal{M}_p(E)$ l'ensemble des mesures ponctuelles sur (E, \mathcal{E}) . On notera $\mu = \sum_{x \in D} \delta_x$ une mesure $\mu \in \mathcal{M}_p(E)$: D est un ensemble fini ou dénombrable. On dit que la mesure μ est *simple* si chaque point x apparaît au plus une fois dans la somme.

On munit $\mathcal{M}_p(E)$ de la plus petite tribu rendant mesurables les fonctions $\mu \rightarrow \mu(f) = \sum_{x \in D} f(x)$ avec $f : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable positive.

Définition 1.4. — Etant donnée une mesure σ -finie λ sur un espace mesurable (E, \mathcal{E}) on dit que N est une mesure de Poisson d'intensité λ si N est une variable aléatoire à valeurs dans $\mathcal{M}_p(E)$ telle que

- si les $(A_i)_{i \in I}$ forme une famille d'ensembles mesurables deux à deux disjoints, alors les variables aléatoires $(N(A_i))_{i \in I}$ sont indépendantes.
- Si $A \in \mathcal{E}$ vérifie $\lambda(A) < +\infty$, alors la variable aléatoire $N(A)$ suit une loi de Poisson de paramètre $\lambda(A)$: $N(A) \sim \mathcal{P}(\lambda(A))$.

Un processus de Poisson peut également être considéré comme un nuage aléatoire de points D_ω , et alors $N(A) = \sum_{x \in D_\omega} \mathbf{1}_{(x \in A)}$ est la variable aléatoire qui compte le nombre de points qui tombent dans A .

Théorème 1.7. — Etant donnée une mesure σ -finie λ sur un espace mesurable (E, \mathcal{E}) , il existe un processus de Poisson d'intensité λ .

Démonstration. — Supposons dans un premier temps que λ est une mesure finie : on pose $a = \lambda(E) < +\infty$ et on considère :

- M une variable aléatoire de loi de Poisson de paramètre a : $M \sim \mathcal{P}(a)$.
- $(X_n)_{n \in \mathbb{N}}$ indépendantes de même loi $P_X = \frac{1}{a}\lambda$, et indépendantes de la variable aléatoire M .

On pose enfin :

$$N = N(\omega) = \sum_{i=1}^{M(\omega)} \delta_{X_i(\omega)}.$$

Montrons dans un premier temps que $N(A) \sim \mathcal{P}(\lambda(A))$. On a $N(A) = \sum_{i=1}^M \mathbf{1}_{(X_i \in A)}$. C'est un exercice classique : on peut par exemple calculer la fonction génératrice : si $u > 0$, en raison de l'indépendance de M et de la

suite $(X_n)_{n \in \mathbb{N}}$, on a

$$\begin{aligned}\mathbb{E} [u^{N(A)}] &= \sum_k \mathbb{E} [u^{N(A)} \mid M = k] \mathbb{P}(M = k) \\ &= \sum_k \mathbb{E} [u^{\mathbf{1}_{(X_1 \in A)} + \dots + \mathbf{1}_{(X_k \in A)}}] \mathbb{P}(M = k) \\ &= \sum_k \mathbb{E} [u^{\mathbf{1}_{(X_1 \in A)}}]^k \mathbb{P}(M = k) \\ &= \sum_k v^k \mathbb{P}(M = k) = \mathbb{E} [v^M] = e^{-a(1-v)} = e^{-\lambda(A)(1-u)}\end{aligned}$$

car $v = \mathbb{E} [u^{\mathbf{1}_{(X_1 \in A)}}] = u\mathbb{P}(X_1 \in A) + 1 - \mathbb{P}(X_1 \in A) = 1 - (1-u)\lambda(A)/a$. Pour démontrer que l'on a réellement une mesure de Poisson, on utilise la caractérisation suivante, appelée également “Master Formula”

Proposition 1.8. — La mesure ponctuelle aléatoire N est une mesure de Poisson d'intensité λ si et seulement si pour toute fonction $f : E \rightarrow [0, +\infty]$ mesurable,

$$\mathbb{E} [e^{-N(f)}] = \exp \left(- \int_E (1 - e^{-f(x)}) d\lambda(x) \right).$$

On peut alors utiliser pratiquement le même argument que précédemment:

$$\begin{aligned}\mathbb{E} [e^{-N(f)}] &= \sum_k \mathbb{E} [e^{-N(f)} \mid M = k] \mathbb{P}(M = k) \\ &= \sum_k \mathbb{E} [e^{-(f(X_1) + \dots + f(X_k))}] \mathbb{P}(M = k) \\ &= \sum_k \mathbb{E} [e^{-f(X_1)}]^k \mathbb{P}(M = k) \\ &= \exp(-a(1 - \mathbb{E} [e^{-f(X_1)}]))\end{aligned}$$

et on conclut car:

$$a(1 - \mathbb{E} [e^{-f(X_1)}]) = a\mathbb{E} [1 - e^{-f(X_1)}] = a \int \frac{1}{a} (1 - e^{-f(x)}) d\lambda(x) = \int_E (1 - e^{-f(x)}) d\lambda(x)$$

Traitons maintenant le cas où la mesure λ n'est pas finie. Comme elle est σ -finie, on peut partitionner $E = \cup_{p \in \mathbb{N}} E_p$ avec les E_p deux à deux disjoints et $\lambda(E_p) < +\infty$. On construit alors N_p mesure de Poisson, à valeurs dans E_p , d'intensité $\lambda_p = \lambda|_{E_p}$ restriction de λ à E_p . L'important est de construire ces mesures E_p de façon indépendante les unes des autres. Il suffit ensuite de

poser, pour f mesurable positive,

$$N(f) = \sum_p N_p(f|_{E_p})$$

Les règles d'additivité des paramètres pour des variables de Poisson indépendantes entraînent alors automatiquement que $N(f)$ suit une loi de Poisson de paramètre

$$\sum_p \lambda_p(f|_{E_p}) = \lambda(f).$$

□

de la Proposition 1.8. — \Leftarrow Si les $(A_i)_{i \in I}$ forme une famille finie d'éléments 2 à 2 disjoints, alors on pose $f = \sum_i a_i 1_{A_i}$, pour des réels $a_i \geq 0$ et on obtient:

$$\begin{aligned} \mathbb{E} [e^{-N(f)}] &= \mathbb{E} [e^{-\sum_i a_i N(A_i)}] = \exp - \int (1 - e^{-f(x)}) d\lambda(x) \\ &= \exp - \sum_i \int_{A_i} (1 - e^{-a_i}) d\lambda(x) \\ &= \exp - \sum_i (1 - e^{-a_i}) \lambda(A_i) = \prod_i \mathbb{E} [e^{-a_i N(A_i)}] \end{aligned}$$

ce qui prouve en considérant un seul indice que $N(A_i)$ suit bien une loi de Poisson de paramètre $\lambda(A_i)$, puis en considérant tous les indices que les variables aléatoires $(N(A_i))_{i \in I}$ sont indépendantes.

⇒ Pour la réciproque, en reprenant le calcul ci dessus à l'envers, on remarque que l'on a démontré la formule pour f étagée, puis on utilise le fait qu'une fonction mesurable positive f est limite croissante de fonctions simples f_n , et on passe à la limite dans les deux côtés de la formule:

$$N(f) = \lim \uparrow N(f_n), \quad (1 - e^{-f}) = \lim \uparrow (1 - e^{-f_n}).$$

□

En utilisant la construction de la preuve précédente ainsi que l'unicité en loi d'une mesure de Poisson d'intensité donnée, nous obtenons les propriétés suivantes.

Proposition 1.9. — Soit N une mesure de Poisson sur E d'intensité μ .

1. Soit $A \in \mathcal{E}$ tel que $\mu(A) < +\infty$. Alors $N(A) \sim \mathcal{P}(\mu(A))$ et conditionnellement à $N(A) = k$, $N|_A$ a même loi que $\sum_{i=1}^k \delta_{X_i}$ avec $(X_i)_{i \geq 1}$ IID de loi $\mu(\cdot | A)$.
2. Si $A_1, A_2, \dots, A_k \in \mathcal{E}$ sont disjoints alors les restrictions $M|_{A_i}$ sont des mesures de Poisson indépendantes d'intensités $\mu|_{A_i}$.
3. Tout mesure de Poisson peut s'écrire $N = \sum_{i \in I} \delta_{X_i}$ avec I un ensemble aléatoire fini ou dénombrable, et les X_i des variables aléatoires à valeurs dans E (pas nécessairement IID)
4. Si $f \geq 0$ mesurable ou $f \in L^1(\mu)$ alors

$$\mathbb{E}[N(f)] = \int f d\mu.$$

Démonstration. — Seule la dernière identité est à prouver. On suppose $f \geq 0$. On peut soit appliquer la formule exponentielle à αf et dériver par rapport à α en prenant $\alpha = 0$, soit écrire directement $N(f) = \sum_p N_p(f)$ et remarquer que par l'indentité de Wald

$$N_p(f) = \mathbb{E} \left[\sum_{i=1}^{M_p} f(X_i^{(p)}) \right] = \mathbb{E} [f(X_1^{(p)})] \mathbb{E}[M_p] = \mu_p(f) \mu(E_p) = \int_{E_p} f d\mu$$

et ensuite sommer ces identités. □

Il est maintenant facile de démontrer les proriétés de stabilité des mesures de Poisson.

Lemme 1.10 (Image). — Soit m une mesure de Poisson d'intensité μ sur (E, \mathcal{E}) et $f : (E, \mathcal{E}) \rightarrow (F, \mathcal{F})$ mesurable. Si la mesure image $f * \mu$ est σ -finie, lors l'image $f * M$ est une mesure de Poisson sur F d'intensité $f * \mu$.

Démonstration. — On a $M = \sum_{1 \leq i \leq N} \delta_{X_i}$ et donc $f * M = \sum_{1 \leq i \leq N} \delta_{f(X_i)}$. Pour tout ϕ mesurable positive

(1)

$$\mathbb{E} [e^{-f * M(\phi)}] = \mathbb{E} [e^{-M(\phi \circ f)}] = \exp \left(- \int (1 - e^{-\phi \circ f(x)}) d\mu(x) \right) = \exp \left(- \int (1 - e^{-\phi(y)}) df * \mu(y) \right)$$

□

Remarque. — On remarque que si $f * \mu$ n'est pas sigma-finie, cela entraîne que pour toute ϕ mesurable, $\phi > 0$ on a $\int \phi df * \mu = +\infty$, et donc il y a explosion car $f * M(\phi) = +\infty$ p.s. et donc ps $f * M$ n'est pas σ -finie.

Lemme 1.11 (Superposition). — Soit M_1, M_2 deux mesures de Poisson sur E indépendantes, d'intensités μ_1, μ_2 . Alors $M_1 + M_2$ est une mesure de Poisson d'intensité $\mu_1 + \mu_2$.

La preuve est triviale.

Lemme 1.12 (Marquage). — Soit $M = \sum_{i=1}^N \delta_{X_i}$ une mesure de Poisson d'intensité μ et soit $(Y_n)_{n \in \mathbb{N}}$ une suite IID de variable aléatoire de loi ν sur (F, \mathcal{F}) , indépendante de M c'est à dire indépendante de $\sigma(M) = \sigma(N, x_i, i \geq 1)$. Alors $M' := \sum_{1 \leq i \leq N} \delta_{(X_i, Y_i)}$ est une mesure de Poisson sur $E \times F$ d'intensité $\mu \otimes \nu$.

Démonstration. — Soit $f : E \times F \rightarrow \mathbb{R}$ mesurable positive

$$\begin{aligned} \mathbb{E} [e^{-M'(f)} | \sigma(M)] &= \mathbb{E} \left[\prod_{i=1}^N e^{-f(X_i, Y_i)} | \sigma(M) \right] \\ &= \prod_{i=1}^N \int e^{-f(X_i, y)} d\nu(y) = e^{-M(F)} \end{aligned}$$

avec $F(x) = \log(\int e^{-f(x, y)} d\nu(y))$. Donc,

$$\mathbb{E} [e^{-M'(f)}] = \mathbb{E} [e^{-M(F)}] = \exp(- \int (1 - e^{-F(x)}) d\mu(x)) = \exp \left(- \int (1 - e^{-f(x, y)}) d\mu(x) d\nu(y) \right).$$

□

Lemme 1.13 (Désintégration/Raréfaction/Thinning)

Soit $M = \sum_{1 \leq i \leq N} \delta_{X_i}$ une mesure de Poisson d'intensité μ . Soit $(Y_n)_{n \geq 1}$ une suite IID de variables de loi $\mathcal{B}(p)$, indépendante de M . Soit

$$M_1 = \sum_{1 \leq i \leq N, Y_i=1} \delta_{X_i} \quad M_0 = \sum_{1 \leq i \leq N, Y_i=0} \delta_{X_i}$$

Alors M_1 et M_0 sont deux mesures de Poisson indépendantes d'intensités $p\mu$ et $(1-p)\mu$.

Remarque. — Cela signifie que si l'on reçoit les clients à la porte d'une boutique, qu'il y en a en moyenne λ par unité de temps, et qu'on les dispatche avec la proba p sur le vendeur 1 et la proba $1-p$ sur le vendeur 2, alors le vendeur 1 verra arriver un flux poissonnien de clients, en moyenne $p\lambda$ par unité de temps (et que les flux arrivant sur les deux vendeurs sont indépendants).

Démonstration. — On considère $M' = \sum_{1 \leq i \leq N} \delta_{(X_i, Y_i)} \sim \mathcal{P}(\nu = \mu \times \text{Ber}(p))$. On le décompose suivant deux ensembles mesurables disjoints $A_1 = E \times \{1\}$, $A_0 = E \times \{0\}$. Cela donne deux mesures de Poisson indépendantes M'_i d'intensités $\nu|_{A_i}$. Puis on remarque que $M_i = f * M'_i$ avec $f(x, y) = x$. DOnc

les M_i sont des mesures de Poisson indépendantes d'intensité $f * \nu_{|A_1} = p\mu$ et $f * \mu_{|A_0} = (1-p)\mu$. \square

3.1 Construction du processus de Poisson homogène On prend pour μ la mesure de Lebesgue sur \mathbb{R}^+ et on pose $N(t) = M([0, t])$. Alors par la construction précédente la Proposition 1.6 est immédiate, comme sont immédiates les propriétés d'indépendance et de stationnarité des accroissements : si $A_i =]t_i, t_{i+1}]$ comme $t_1 < t_2 < \dots < t_n$ les A_i sont deux à deux disjoints donc les $N(A_i) = N(t_{i+1}) - N(t_i)$ sont indépendantes de loi de Poisson de paramètres $\lambda(A_i) = t_{i+1} - t_i$.

Pour montrer que cette construction donne bien un processus qui a la même loi que le processus de comptage construit à partir d'exponentielles IID, il faut montrer que les intervalles entre les temps de saut sont des exponentielles IID.

Lemme 1.14. — *Le premier temps de saut du processus N*

$$T = \inf \{t > 0 : N_t = 1\}$$

suit une loi exponentielle.

Démonstration. — Comme $\{T > t\} = \{N_t = 0\}$ avec $N_t \sim \mathcal{P}(\lambda t)$ on a

$$\mathbb{P}(T > t) = \mathbb{P}(N_t = 0) = e^{-\lambda t}$$

\square

Proposition 1.15 (Propriété de Markov). — *Soit N un Processus de Poisson de paramètre λ . Pour tout $s \geq 0$, $(\bar{N}_t = N_{t+s} - N_s)_{t \geq 0}$ est un Processus de Poisson de paramètre λ . En outre \bar{N} est indépendant du passé avant s : $(N_u, u \leq s)$.*

Démonstration. — Les accroissements de \bar{N} sont juste des translatés de s des accroissements de N : $\bar{N}_{u+v} - \bar{N}_u = N_{u+v+s} - N_{u+s}$. Ils sont donc indépendants et stationnaires. Tout comme N , le processus \bar{N} est croissant, continu à droite et ne croît que par sauts de 1. C'est donc un processus de Poisson de paramètre disons $\mu > 0$. Comme $\bar{N}_t = N_{t+s} - N_s \sim \mathcal{P}(\lambda t)$ on a bien $\mu = \lambda$.

Pour montrer que \bar{N} est indépendant du passé avant s : $(N_u, u \leq s)$, il suffit de montrer que pour tous $0 \leq u_1 \leq \dots \leq u_n \leq s$ et $0 \leq t_1 \leq \dots \leq t_p$ les vecteurs $(N_{u_1}, \dots, N_{u_n})$ et $(\bar{N}_{t_1}, \dots, \bar{N}_{t_p})$ sont indépendants. On écrit ces deux vecteurs en fonction des accroissements de N , avant et après s , ce qui montre immédiatement l'indépendance.

\square

Proposition 1.16 (Propriété de Markov forte). — *Soit N un Processus de Poisson de paramètre λ . Soit T un temps d'arrêt fini. Pour tout $s \geq 0$, $(\bar{N}_t = N_{t+T} - N_T)_{t \geq 0}$ est un Processus de Poisson de paramètre λ . En outre \bar{N} est indépendant du passé avant T : $(N_u, u \leq T)$.*

On peut maintenant montrer que les variables $\tau_i = T_{i+1} - T_i$ sont IID avec $T_0 = 0$ et $T_i = \inf \{t > 0 : N_t = i\}$. On sait déjà que $\tau_1 \sim \mathcal{E}(\lambda)$. On remarque que $\tau_2 = \tau_1(\tilde{N})$ avec $\tilde{N}_s = N_{t+\tau_1} - N_{\tau_1} = N_{t+\tau_1} - 1$. D'après la proposition précédente τ_2 est indépendant de $(N_u, u \leq \tau_1)$ donc de τ_1 et a même loi que τ_1 . On répète l'opération avec τ_3 indépendant de $\tau_1, \tau_2 \dots$

3.2 Le paradoxe de l'autobus On suppose que les intervalles de passage X_1, \dots, X_n, \dots des autobus à l'arrêt Morhonière sont des variables indépendantes de même loi $\exp(\lambda)$. Le nombre de bus passé avant t est donc $N_t = \sum_{n \geq 1} \mathbf{1}_{(S_n \leq t)}$ avec $S_n = X_1 + \dots + X_n$. C'est un Processus de Poisson de paramètre λ . Sachant que l'on arrive à l'arrêt à l'instant $t > 0$:

- Le temps d'attente du prochain bus est

$$R_t = \inf \{s > t : N_s = N_t + 1\} = S_{1+N_t} - t$$

- Le temps écoulé depuis le passage du dernier bus est

$$A_t = \sup \{s \leq t : N_s < N_t\} = t - S_{N_t}$$

Proposition 1.17. — *Les variables aléatoires R_t et A_t sont indépendantes, $R_t \sim \exp(\lambda)$ et $A_t \stackrel{d}{=} \inf(\mathbf{e}, t)$ avec $\mathbf{e} \sim \exp(\lambda)$.*

Démonstration. — A_t ne dépend que de $(N_s, s \leq t)$ et R_t ne dépend que de $(\bar{N}_s = N_{t+s} - N_t, s \geq 0)$. Par la propriété de Markov, A_t et R_t sont indépendantes.

$$R_t = \inf \{s \geq 0 : \bar{N}_s = 1\}$$

est le premier temps de saut de \bar{N} , donc $R_t \sim \exp(\lambda)$. On a $0 \leq A_t \leq t$ et si $0 < s < t$, alors

$$\{A_t \geq s\} = \{N'_s = 0\}$$

avec N' le Processus de Poisson de paramètre λ défini par $N'_u = N_{t-s+u} - N_{t-s}$. Donc,

$$\mathbb{P}(A_t \geq s) = \mathbb{P}(N'_s = 0) = e^{-\lambda s}.$$

D'autre part $\mathbb{P}(A_t = t) = \mathbb{P}(N_t = 0) = e^{-\lambda t}$ ce qui prouve que $A_t \stackrel{d}{=} \inf(\mathbf{e}, t)$. \square

Remarque. — *On a $\mathbb{E}[R_t] = \frac{1}{\lambda}$ et $\mathbb{E}[A_t] \rightarrow \mathbb{E}[\mathbf{e}] = \frac{1}{\lambda}$ par convergence monotone. Donc si $L_t = R_t + A_t$ est la longueur de l'intervalle dans lequel on tombe on obtient $\mathbb{E}[L_t] \rightarrow 2/\lambda$. C'est un paradoxe : si les autobus passent en moyenne toutes les 10 minutes, ie $\frac{1}{\lambda} = 10$, alors l'intervalle dans lequel on tombe a une longueur moyenne très proche de 20 minutes. En outre, on attendra en moyenne $\mathbb{E}[R_t] = \frac{1}{\lambda} = 10$ minutes le prochain bus, et ce indépendamment du temps écoulé depuis le passage du bus précédent.*

Le processus de Poisson composé

Exemple 1.1. — Le nombre de voitures qui s'arrêtent entre midi et treize heures devant le Mcdo de Rezé suit une loi de Poisson de paramètre $\lambda = 180$. Le nombre de passagers par voiture suit une loi binomiale $\mathcal{B}(n = 5, p = \frac{1}{2})$. Calculer la moyenne et la variance du nombre de clients X .

Le bon sens dit que $\mathbb{E}[X] = 180 \times 2,5 = 450$. Mais comment calculer la variance ? Il est naturel de supposer que si Y_i est le nombre de passagers de la i -ème voiture, alors les variables aléatoires Y_i sont indépendantes de même loi $\mathcal{B}(n = 5, p = \frac{1}{2})$. Si N_t est le nombre de voitures arrivées avant $t = 1$ heure, on commence l'échelle des temps à midi, on a

$$X = \sum_{i=1}^{N_t} Y_i.$$

Définition 1.5. — Un processus stochastique $(X_t)_{t \geq 0}$ est appelé Processus de Poisson composé s'il peut être représenté par $X_t = \sum_{i=1}^{N_t} Y_i$ avec $(N_t)_{t \geq 0}$ un Processus de Poisson indépendant de la suite IID $(Y_i)_{i \geq 1}$.

La proposition suivante permet de calculer aisément espérance et variance d'un processus de Poisson composé de paramètre λ .

Proposition 1.18. — 1. Si Y_1 est intégrable alors X_t est intégrable et on a $\mathbb{E}[X_t] = \lambda t \mathbb{E}[Y_1]$.
 2. SI Y_1 est de carré intégrable, alors X_t l'est aussi et $\text{Var}(X_t) = \lambda t \mathbb{E}[Y_1^2]$.

Pour l'exemple du Mcdo, cela donne $\text{Var}(X) = 180 \times 7,5 = 1350$. Ce qui donne une très forte probabilité d'observer entre 300 et 600 personnes. La Proposition est une conséquence immédiate du

Lemme 1.19. — Soit $X = \sum_{i=1}^N Y_i$ avec N une variable aléatoire entière indépendante de la suite IID $(Y_i)_{i \geq 1}$.

1. Si N et Y sont intégrables, alors on a l'identité de Wald: $\mathbb{E}[X] = \mathbb{E}[N]\mathbb{E}[Y]$.
2. Si N et Y sont dans \mathcal{L}^2 alors $X \in \mathcal{L}^2$ et

$$\text{Var}(X) = \mathbb{E}[N] \text{Var}(Y) + \text{Var}(N)\mathbb{E}[Y]^2.$$

Démonstration. — 1. Si Y_1 était une variable aléatoire entière on pourrait raisonner avec les fonctions génératrices. Cependant, le cas général n'est pas

plus compliqué:

$$\begin{aligned}
\mathbb{E}[|X|] &= \sum_n \mathbb{P}(N=n) \mathbb{E}[|X| \mid N=n] \\
&= \sum_n \mathbb{P}(N=n) \mathbb{E}[|Y_1 + \cdots + Y_n|] \\
&\leq \sum_n \mathbb{P}(N=n) (\mathbb{E}[|Y_1|] + \cdots + \mathbb{E}[|Y_1|]) \\
&= \mathbb{E}[|Y_1|] \sum_n \mathbb{P}(N=n) n = \mathbb{E}[|Y_1|] \mathbb{E}[N] < +\infty
\end{aligned}$$

donc X est intégrable et un calcul analogue, sans les valeurs absolues, montre que $\mathbb{E}[X] = \mathbb{E}[Y]\mathbb{E}[N]$.

2. On calcule la fonction caractéristique de X

$$\begin{aligned}
\phi_X(u) &= \mathbb{E}[e^{iuX}] = \sum_n \mathbb{P}(N=n) \mathbb{E}[e^{iuX} \mid N=n] \\
&= \sum_n \mathbb{P}(N=n) \mathbb{E}\left[e^{iu(Y_1+\cdots+Y_n)}\right] \\
&= \sum_n \mathbb{P}(N=n) \phi_{Y_1}(u)^n = \mathbb{E}[\phi_{Y_1}(u)^N] \\
&= G(\phi_{Y_1}(u)).
\end{aligned}$$

Comme Y_1 est de carré intégrable

$$\phi_{Y_1}(u) = 1 + u\mathbb{E}[Y_1] - (u^2/2)\mathbb{E}[Y_1^2] + o(u^2).$$

Comme N est de carré intégrable, $G'(s)$ et $G''(s)$ admettent des limites finies quand $s \uparrow 1$,

$$G(1-a) = 1 - aG'(1) + (a^2/2)G''(1) + o(a^2) = 1 - a\mathbb{E}[N] + (a^2/2)G''(1)\mathbb{E}[N(N-1)] + o(a^2).$$

et on en déduit un développement limité en 0 d'ordre 2 de $\phi_X(u)$ qui nous donne $\mathbb{E}[X^2]$.

On peut aussi utiliser la même technique qu'au 1 pour calculer $\mathbb{E}[X^2]$ directement :

$$\begin{aligned}
\mathbb{E}[X^2] &= \sum_n \mathbb{P}(N=n) \mathbb{E}[(Y_1 + \cdots + Y_n)^2] \\
&= \sum_n \mathbb{P}(N=n) (\text{var}(Y_1 + \cdots + Y_n) + \mathbb{E}[(Y_1 + \cdots + Y_n)]^2) \\
&= \sum_n \mathbb{P}(N=n) (n \text{Var}(Y_1) + n^2 \mathbb{E}[Y_1]^2) = \mathbb{E}[N] \text{Var}(Y_1) + \mathbb{E}[N^2] \mathbb{E}[Y_1]^2
\end{aligned}$$

□

Remarque. — On peut observer que si $M = \sum_{i=1}^N \delta_{\xi_i}$ est une mesure de Poisson d'intensité λ et $N_t = M([0, t])$ alors $X_t = M'(f_t)$ avec M' le processus marqué

$$M' = \sum_{1 \leq i \leq N} \delta_{(\xi_i, Y_i)}$$

et $f_t(\xi, y) = \mathbf{1}_{(\xi \leq t)} y$. Alors il vient, si ν est la loi de Y_1

$$\mathbb{E}[X_t] = \int f_t(\xi, y) \lambda d\xi d\nu(y) = \lambda t \int y d\nu(y) = \lambda t \mathbb{E}[Y]$$

De même on calcule la transformée de Laplace

$$\mathbb{E}[e^{-\alpha X_t}] = \mathbb{E}[e^{-M(\alpha f_t)}] = \exp\left(-\int (1 - e^{-\alpha f_t(\xi, y)}) \lambda d\xi d\nu(y)\right) = \exp(-\lambda t \mathbb{E}[1 - e^{-\alpha Y}])$$

et on en déduit avec $\psi(\alpha) = -\log \mathbb{E}[e^{-\alpha X_t}]$, $\mathbb{E}[X_t] = \psi'(0)$ et $\text{Var}(X_t) = \psi''(0) = \lambda t \mathbb{E}[Y^2]$.

SECTION 5

Le processus de Poisson inhomogène

C'est un Processus de Poisson d'intensité variable. Soit $\lambda : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ une fonction boréienne positive localement intégrable, i.e. pour tous $0 < a < b < +\infty$ $\int_a^b \lambda(t) dt < +\infty$.

Définition 1.6. — On dit que $(N_t)_{t \geq 0}$ est un Processus de Poisson inhomogène d'intensité λ si $(N_t)_{t \geq 0}$ est un processus à accroissements indépendants et pour tous $s, t \geq 0$

$$N_{t+s} - N_s \sim \mathcal{P}\left(\int_s^{t+s} \lambda(u) du\right).$$

Exemple 1.2. — On observe le taux d'arrivée, exprimé en clients par heure, à la boutique de harry. Il est 0 à 10 heures, 4 à midi, 6 à 14 heures, 2 à 16 heures et 0 à 18 heures. Entre ces valeurs il varie linéairement. Faire un dessin. On désire déterminer la loi du nombre de clients dans une journée et la probabilité que personne n'arrive avant midi.

Pour cela on suppose donc que le nombre de clients suit un Processus de Poisson inhomogène de taux λ linéaire par morceaux. Le nombre de clients arrivés dans la journée est donc

$$C = N_{18} - N_{10} \sim \mathcal{P}\left(\int_{10}^{18} \lambda(t) dt\right) = \mathcal{P}(20).$$

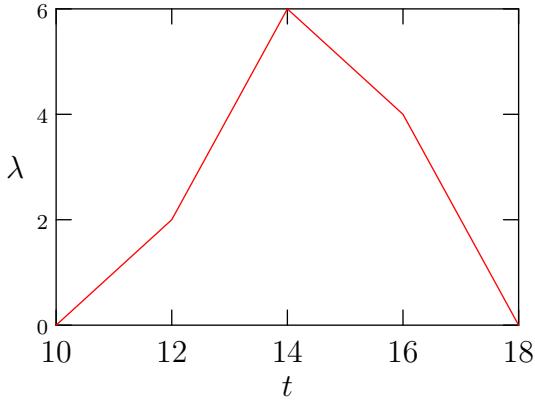


FIGURE 1. Intensité du processus d'arrivée des clients

On calcule l'intégrale en calculant l'aire sous la courbe, i.e. en additionnant l'aire des triangles ! De même le nombre de clients arrivés avant midi est $M = N_{12} - N_{10} \sim \mathcal{P}(4)$ et donc $\mathbb{P}(M = 0) = e^{-4}$.

En exercice, on peut, en supposant que Harry ferme exceptionnellement sa boutique à 17h30, calculer le nombre moyen de clients perdus et la probabilité qu'au moins un client trouve porte close (on trouve 0,5 et 0,39).

Observons que l'on ne peut pas savoir si on a affaire à un Processus de Poisson inhomogène. On fait une hypothèse de modélisation qui ne peut être validée que par une étude statistique.

Il est très facile de construire un Processus de Poisson inhomogène. Il suffit de changer l'horloge.

Proposition 1.20. — Si $(N_t)_{t \geq 0}$ est un Processus de Poisson d'intensité 1, et $F(t) = \int_0^t \lambda(s) ds$, alors $M_t = N_{F(t)}$ est un Processus de Poisson inhomogène d'intensité λ .

Une fois faite l'hypothèse de modélisation par un Processus de Poisson inhomogène, on identifie la fonction intensité à l'aide du

Lemme 1.21. — Soit $(N_t)_{t \geq 0}$ un Processus de Poisson inhomogène d'intensité λ , alors si λ est continue à droite en t :

$$\mathbb{P}(N_{t+h} - N_t \neq 0) = h\lambda(t) + o(h).$$

Démonstration. — EN effet, t étant fixé, la variable aléatoire $X = N_{t+h} - N_t$ suit une loi de Poisson de paramètre $f(h) = \int_t^{t+h} \lambda(s) ds = h\lambda(t) + o(h)$ et $\mathbb{P}(X \neq 0) = 1 - e^{-f(h)} = h\lambda(t) + o(h)$. \square

Universalité de la loi de Poisson

Nous allons généraliser le théorème de convergence de la loi binômiale vers la loi de Poisson :

$$\text{si } np_n \rightarrow \lambda > 0, \quad \text{alors} \quad \mathcal{B}(n, p_n) \Rightarrow \mathcal{P}(\lambda).$$

Théorème 1.22. — Soit $(X_{n,m})_{1 \leq m \leq n}$ des variables aléatoires entières indépendantes telles que si $p_{n,m} = \mathbb{P}(X_{n,m} = 1)$ et $\delta_{n,m} = \mathbb{P}(X_{n,m} \geq 2)$ alors

- (i) $\sum_{1 \leq m \leq n} p_{n,m} \xrightarrow[n \rightarrow \infty]{} \lambda > 0$
- (ii) $\sup_{1 \leq m \leq n} p_{n,m} \xrightarrow[n \rightarrow \infty]{} 0$
- (iii) $\sum_{1 \leq m \leq n} \delta_{n,m} \rightarrow 0$

Alors la suite $S_n = X_{n,1} + \dots + X_{n,n}$, converge en loi vers une variable aléatoire Z qui suit une loi de Poisson de paramètre $\lambda > 0$.

Application : un système de marcheurs aléatoires Dans une rue, supposée très grande, on place des clients, qui se promènent au hasard. Si on compte le nombre de clients qui sont à un instant donné, assez grand, devant une boutique précise, on observe une variable aléatoire qui suit une loi de Poisson.

Le modèle mathématique est le suivant. La rue est \mathbb{Z} . A l'origine des temps $n = 0$, on place un client en chaque site x de \mathbb{Z} : S_n^x est la position à l'instant n du marcheur qui à l'origine des temps était en x , $S_x^0 = x$. La boutique est au point 0 de \mathbb{Z} . Le nombre de clients devant la boutique à l'instant n est donc

$$Y_n = \sum_{x \in \mathbb{Z}} \xi_{n,x} \quad \text{avec} \quad \xi_{n,x} = \mathbf{1}_{(S_n^x=0)}.$$

On désire montrer que $Y_n \rightarrow Z \sim \mathcal{P}(1)$. En fait la somme est restreinte à une somme finie de variables aléatoires indépendantes qui prennent les valeurs 0 et 1 : $Y_n = \sum_{|x| \leq n} \xi_{n,x}$. Pour appliquer le théorème 1.22, il reste à montrer que

$$\sum_{|x| \leq n} \mathbb{P}(\xi_{n,x} = 1) \rightarrow 1 \quad \text{et} \quad \sup_{|x| \leq n} \mathbb{P}(\xi_{n,x} = 1) \rightarrow 0.$$

Qu'est ce donc qu'un marcheur issu de x ? Sa position à l'instant n est

$$S_n^x = x + \eta_1 + \dots + \eta_n$$

avec les variables aléatoires η_i indépendantes de même loi $\mathbb{P}(\eta_i = \pm 1) = \frac{1}{2}$. En observant que $\frac{1}{2}(1 + \eta_i) \sim \mathcal{B}(1, \frac{1}{2})$ on voit que $S_n^x = x + 2B_n - n$ avec

$B_n \sim \mathcal{B}(n, \frac{1}{2})$ et donc que

$$\mathbb{P}(\xi_{n,x} = 1) = \mathbb{P}(x + 2B_n - n = 0) = \mathbb{P}\left(B_n = \frac{n-x}{2}\right).$$

Donc

$$\sum_{|x| \leq n} \mathbb{P}(\xi_{n,x} = 1) = \sum_{|x| \leq n} \mathbb{P}\left(B_n = \frac{n-x}{2}\right) = \sum_k \mathbb{P}(B_n = k) = 1$$

et évidemment

$$\sup_{|x| \leq n} \mathbb{P}(\xi_{n,x} = 1) = \sup_k \mathbb{P}(B_n = k) \rightarrow 0.$$

Remarque. — On peut même démontrer que si on a une boutique à chaque site $x \in \mathbb{Z}$, alors pour n grand, devant chaque boutique il y a un nombre de clients suivant une loi $\mathcal{P}(1)$ et que ces variables aléatoires sont toutes indépendantes et de moyenne 1.

2

Renouvellement

SECTION 1

Processus de renouvellement

On considère un processus de comptage $(N(t), t \geq 0)$ dans lequel les intervalles de temps entre les arrivées sont encore IID mais pas forcément exponentiels. Soit $(X_n, n \geq 1)$ IID positives de moyenne $\mu \in (0, +\infty)$, $S_0 = 0$ et $S_n = X_1 + \dots + X_n$ l'époque de la n ème arrivée.

Définition 2.1. — $N(t) = \sum_{n \geq 1} \mathbf{1}_{(S_n \leq t)} = \sup \{n \geq 1 : S_n \leq t\}$ avec $\sup \emptyset = 0$) est le processus de renouvellement engendré par la suite $(X_i, i \geq 1)$.

Exemple 2.1. — Inventaire. La demande du produit pendant la semaine $t = 1, 2, \dots$ est une suite IID $(X_t, t \geq 1)$. Alors $1 + N(u)$ est le nombre de semaines avant l'épuisement du stock $u \geq 0$.

Théorème 2.1 (loi des grands nombres). —

$$\lim_{t \rightarrow +\infty} \frac{N(t)}{t} = \frac{1}{\mu} \quad p.s.$$

Démonstration. — Il suffit de remarquer que $\frac{S_n}{n} \rightarrow \mu$ et d'utiliser la monotonie de $n \rightarrow S_n$ pour avoir l'encadrement

$$\frac{S_{N(t)}}{N(t)} \leq \frac{t}{N(t)} \leq \frac{S_{1+N(t)}}{N(t)}$$

qui donne le résultat car $N(t) \rightarrow +\infty$. □

Proposition 2.2. — La fonction de renouvellement $M(t) = \mathbb{E}[N(t)] = \sum_{n \geq 1} \mathbb{P}(S_n \leq t)$ vérifie

$$\lim_{t \rightarrow +\infty} \frac{M(t)}{t} = \frac{1}{\mu}$$

Lemme 2.3 (Identité de Wald). — Soit τ un temps d'arrêt, i.e. pour tout n , $\{\tau \leq n\}$ est \mathcal{F}_n mesurable. Si τ est intégrable, alors

$$\mathbb{E}[S_\tau] = \mu \mathbb{E}[\tau].$$

Démonstration. — Par conditionnement,

$$\mathbb{E}[S_\tau] = \sum_n \mathbb{E}[S_n \mathbf{1}_{(\tau=n)}] = \mathbb{E}\left[\sum_{m \geq 1} X_m \mathbf{1}_{(\tau \geq m)}\right]$$

Or $\{\tau \geq m\}^C = \{\tau < m\} \in \mathcal{F}_{m-1}^X$ est indépendant de X_m donc

$$\mathbb{E}[S_\tau] = \sum_{m \geq 1} \mathbb{E}[X_m] \mathbb{E}[\mathbf{1}_{(\tau < m)}] = \mu \mathbb{E}[\tau].$$

Remarquons que nous n'utilisons pas l'intégrabilité de τ car les X_i sont positives. Si les X_i sont signées, alors il faut utiliser l'intégrabilité. \square

Lemme 2.4. — Si X_1 est majorée par A , alors, $\frac{t}{\mu} \leq \mathbb{E}[N(t)] + 1 \leq \frac{t+A}{\mu}$.

Démonstration. — On applique l'identité de Wald à $\tau = \inf\{n : S_n > t\} = 1 + N(t)$ et on obtient

$$\mathbb{E}[S_\tau] = \mu(1 + \mathbb{E}[N(t)]) \geq t.$$

Comme les X_i sont majorées par A on a $1 + N(t) \leq t + A$ d'où la seconde inégalité. \square

Démonstration. — (Preuve de la proposition 2.2) Si X_1 est bornée par A on passe à la limite dans l'encadrement. Dans le cas général on applique le résultat à $\bar{X}_i = X_i \wedge A$ et on a $\bar{S}_n \leq S_n$ donc $\bar{N}(t) \geq N(t)$ et on obtient en faisant $A \rightarrow +\infty$)

$$\limsup \frac{M(t)}{t} \leq \lim \frac{\bar{M}(t)}{t} = \frac{1}{\mathbb{E}[X_1 \wedge A]} \rightarrow \frac{1}{\mu}$$

\square

Exemple 2.2. — Bill a une console. La durée d'un pack de piles est uniformément répartie entre 30 et 60 jours. Combien de pack de piles utilisera-t-il en 10 ans ? (Une année veut 360 jours).

On a $\mu = \mathbb{E}[X_1] = 45$ jours car $X_1 \sim \mathcal{U}([30, 60])$. On obtient pour $t = 3600$ jours a peu près

$$80 = t/\mu \leq 1 + M(t) \leq (t + A)/\mu = 81.37$$

donc il utilise entre 80 et 81 packs en moyenne (et 80 asymptotiquement)

Le temps qu'il reste à vivre à l'instant t , temps d'attente de la prochaine arrivée d'un événement important, est $R_t = S_{1+N(t)} - t$. On prouve, à l'aide de l'équation de renouvellement, le résultat suivant sur l'asymptotique de R_t

Proposition 2.5. — La suite R_t converge en loi vers R_∞ de densité $\frac{\mathbb{P}(X_1 > t)}{\mathbb{E}[X_1]}$. En particulier on a

$$(2) \quad \mathbb{E}[R_\infty] = \frac{\mathbb{E}[X_1^2]}{2\mathbb{E}[X_1]}, \quad \mathbb{E}[R_\infty^2] = \frac{\mathbb{E}[X_1^3]}{3\mathbb{E}[X_1]}.$$

Pour une loi exponentielle $X_1 \sim \mathcal{E}(\lambda)$ on a $R_\infty \stackrel{d}{=} X_1$ car c'est un processus de Poisson. Pour une loi uniforme sur $[2a, 3a]$ on a $\mathbb{E}[X] = 2a$ et $\text{Var}(X) = \text{Var}(a(1 + 2U)) = 4a^2 \text{Var}(U) = \frac{a^2}{3}$ et donc $\mathbb{E}[R_\infty] = \frac{13}{12}a$ est bien plus petite que $\mathbb{E}[X] = 2a$.

SECTION 2

Renewal reward processes

C'est le terme anglais pour un processus de renouvellement composé.

Définition 2.2. — Soit $(N(t), t \geq 0)$ un processus de renouvellement et soit $(R_i, i \geq 1)$ une suite IID de variables réelles intégrables. Alors

$$R(t) := \sum_{i=1}^{N_t} R_i$$

est un renewal reward process.

Exemple 2.3. — Une machine est successivement en marche et en panne. $(U_n, n \geq 1)$ est la suite des durées des périodes de marche et $(D_n, n \geq 1)$ est la suite des durées des périodes de panne. On note $D(t)$ le temps total passé en panne avant t .

Le renewal reward process est $R(t)$ associé au processus de renouvellement engendré par $X_i = U_i + D_i$ et de processus de reward $R_i = D_i$. On suppose que $\mathbb{E}[U_1] = 5\mathbb{E}[D_1]$, et on va voir que cela entraîne que $R(t)/t \rightarrow \frac{1}{6}$ p.s.

Il doit être clair que quelle que soit la façon de mesurer $D(t)$ on a alors $D(t)/t \rightarrow \frac{1}{6}$ également car on a

$$R(t)eD(t) \leq R(t) + D_{1+N(t)}$$

Théorème 2.6 (renewal reward theorem). — Si $\mathbb{E}[R_1] < +\infty$ alors

$$\lim_{t \rightarrow +\infty} \frac{R(t)}{t} = \frac{\mathbb{E}[R_1]}{\mathbb{E}[X_1]} \quad p.s.$$

Ce théorème reste valable quelle que soit la façon dont on récupère les récompenses (en début, en fin de cycle ou continûment). On l'écrit *la récompense moyenne en temps long égale la récompense moyenne sur un cycle de renouvellement divisée par la durée d'un cycle de renouvellement*

Démonstration. — Par la loi forte des grands nombres $\frac{1}{n}(R_1 + \dots + R_n) \rightarrow \mathbb{E}[R_1]$ et donc comme $N(t) \rightarrow +\infty$ ps, on a

$$\lim_{t \rightarrow +\infty} \frac{R(t)}{N(t)} = \mathbb{E}[R_1]$$

Comme $N(t)/t \rightarrow \frac{1}{\mu}$ p.s., on a par encadrement,

$$\frac{N(t)}{t} \frac{R(t)}{N(t)} = \frac{R(t)}{t} \rightarrow \frac{\mathbb{E}[R_1]}{\mu}.$$

□

Exemple 2.4 (Coût réel d'une voiture). — Supposons que la durée de vie d'une voiture est une variable aléatoire positive $V \sim \mathcal{E}(\lambda)$. Mr Brown achète une voiture neuve dès que la précédente est cassée ou atteint T années. Une voiture neuve coûte A euros, et un coût additionnel de B euros est payé si la voiture tombe définitivement en panne avant T (par exemple pour payer le dépannage, etc ...).

1. Quel est le coût moyen par unité de temps de la stratégie de Mr Brown ? Quelle valeur de T doit-il choisir ? A.N. $A = 20000$ euros, $B = 200$ euros, $\lambda = 1/4$.
2. On propose un modèle plus réaliste dans lequel entretenir sa voiture pendant n années après l'achat coûte $\alpha n + \beta n^2$ avec $\alpha > 0, \beta > 0$ donnés. Quelle est la valeur optimale de T ? A.N. $\alpha = 300$, $\beta = 1000$.

On va donner un résultat asymptotique (en temps long). Ici la durée d'un renouvellement/cycle est

$$(3) \quad \mathbb{E}[X_1] = \mathbb{E}[V \wedge T] = \int_0^\infty \mathbb{P}(V \wedge T > t) dt = \int_0^T e^{-\lambda t} dt = \frac{1 - e^{-\lambda T}}{\lambda}$$

et le coût du cycle est $R_1 = 1 + B \mathbf{1}_{(T < V_1)}$ donc

$$(4) \quad \mathbb{E}[R_1] = A + B \mathbb{P}(V > T) = A + Be^{-\lambda T}$$

On obtient finalement comme coût annuel sur le long terme

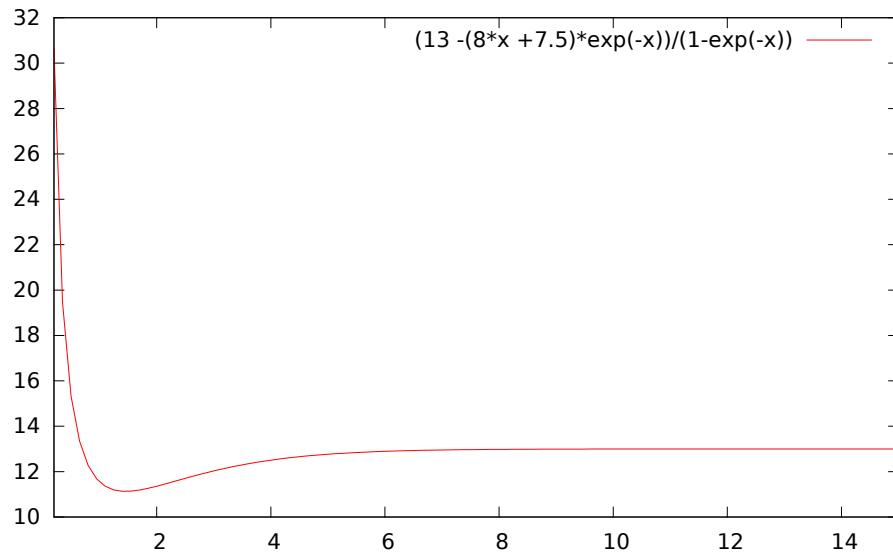
$$c(T) = \lim_{t \rightarrow +\infty} \frac{R(t)}{t} = \frac{\mathbb{E}[R_1]}{\mathbb{E}[X_1]} = \lambda \frac{A + Be^{-\lambda T}}{1 - e^{-\lambda T}} = \lambda(-B + \frac{A + B}{1 - e^{-\lambda T}}).$$

En conséquence la valeur optimale est $T^* = +\infty$ quelles que soient les valeurs de A, B : il ne faut changer la voiture qu'à la première panne, et pas avant.

Pour la seconde question, on voit que l'on obtient comme coût

$$d(T) = c(T) + \alpha + \beta \frac{\mathbb{E}[X_1^2]}{\mathbb{E}[X_1]} = -\lambda B + \alpha + \frac{\lambda(A+B) + \frac{2}{\lambda}(1 - (1+\lambda T)e^{-\lambda T})}{1 - e^{-\lambda T}}$$

Avec les valeurs numériques on voit qu'il faut minimiser la fonction suivante de $x = \lambda T$. On voit que $x^* \approx 1$ et donc $T \approx \frac{1}{\lambda} = 4 = \mathbb{E}[V]$. Dans ce cas il vaut mieux donc attendre le temps moyen de vie de la voiture!



3

Espérance conditionnelle et loi conditionnelle

SECTION 1

Le cas discret

Rappelons que étant donné un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et un évènement B de probabilité non nulle, la fonction d'ensemble $\mathbb{P}(\cdot | B) : \mathcal{F} \rightarrow [0, 1]$ définie par

$$(5) \quad \mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

est une probabilité sur (Ω, \mathcal{F}) appelée *probabilité conditionnelle sachant B* . Il est élémentaire de vérifier alors que si X est une variable aléatoire définie sur (Ω, \mathcal{F}) , \mathbb{P} -intégrable, alors X est $\mathbb{P}(\cdot | B)$ intégrable et

$$(6) \quad \mathbb{E}[X | B] = \frac{\mathbb{E}[X 1_B]}{\mathbb{P}(B)}.$$

(il est clair que cette relation est vérifiée par les variables aléatoires positives).

Définition 3.1. — *Un système complet d'évènements est une famille finie ou dénombrable $(B_i)_{i \in I}$ d'évènements deux à deux disjoints, de probabilité non nulle, tels que $\mathbb{P}(\cap_{i \in I} B_i) = 1$.*

Étant donné un système complet d'évènements, il est naturel de définir *la probabilité conditionnelle de A sachant $(B_i)_{i \in I}$* comme étant la *variable aléatoire* qui vaut $\mathbb{P}(A | B_i)$ sur l'évènement B_i . Plus précisément si $B = \sigma(B_i, i \in I)$ est la tribu engendrée par un système complet, alors pour tout évènement A :

$$(7) \quad \mathbb{P}(A | B)(\omega) = \sum_{i \in I} \mathbb{P}(A | B_i) 1_{B_i}(\omega) p.s.$$

Remarque. — C'est un exercice classique de théorie de la mesure que de vérifier que cette définition de la variable aléatoire $\mathbb{P}(A | \mathcal{B})$, à un ensemble de mesure nulle près, ne dépend que de la tribu \mathcal{B} et pas du système complet d'évènements qui l'engendre.

En effet si $(B_i)_{i \in I}$ est une partition alors $\mathcal{B} = \mathcal{P}$ avec

$$(8) \quad \mathcal{P} = \{A : \exists J \subset I, A = \cap_{i \in J} B_i\}.$$

Dans le cas général, si $N = (\cap_{i \in I} B_i)^C$ alors $\mathbb{P}(N) = 0$ et

$$(9) \quad \mathcal{B} = \mathcal{P} \cap \{A \cap N, A \in \mathcal{P}\}.$$

Si X est une variable aléatoire positive ou intégrable, on définit de façon analogue

$$(10) \quad \mathbb{E}[X | \mathcal{B}](\omega) = \sum_{i \in I} \mathbb{E}[X | B_i] 1_{B_i}(\omega)$$

Le cas le plus classique est celui où l'on dispose d'une variable aléatoire Y à valeurs dans un espace discret E ; alors, pour X positive ou intégrable,

$$(11) \quad \mathbb{E}[X | Y] = \sum_y \mathbb{E}[X | Y = y] \mathbf{1}_{(Y=y)}.$$

Hélas tout se complique lorsque la variable aléatoire Y prend ses valeurs dans un ensemble continu (non discret), car alors $\{Y = y\}$ est un évènement de probabilité nulle. Il est naturel de penser qu'il suffit alors de définir $\mathbb{E}[X | Y] = \phi(Y)$ avec $\phi(y) = \mathbb{E}[X | Y = y] = \lim \mathbb{E}[X | A_n]$ avec A_n une suite d'évènement de probabilité non nulle convergeant vers l'évènement $\{Y = y\}$. Hélas, mille fois hélas, le résultat dépend fortement du choix de la suite A_n , et il n'y a pas de choix canonique d'une telle suite. Un exercice classique sur les couples de gaussiennes dans le plan, que vous trouverez en fin de chapitre, devraitachever de vous convaincre.

Nous allons donc utiliser une généralisation de la caractérisation suivante de l'espérance conditionnelle. La variable X étant intégrable et la variable aléatoire Y discrète, $\mathbb{E}[X | Y]$ est l'unique, presque sûrement, variable aléatoire $\phi(Y)$ intégrable telle que pour toute fonction mesurable bornée f :

$$(12) \quad \mathbb{E}[Xf(Y)] = \mathbb{E}[\phi(Y)f(Y)].$$

SECTION 2

Espérance conditionnelle

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et \mathcal{G} une sous tribu de \mathcal{F} .

Théorème 3.1 (Définition et Théorème). — Soit X une variable aléatoire positive ou intégrable. Alors, il existe une unique, presque sûrement, variable aléatoire notée $\mathbb{E}[X | \mathcal{G}]$ telle que

1. $\omega \rightarrow \mathbb{E}[X | \mathcal{G}](\omega)$ est \mathcal{G} mesurable.
2. pour tout $B \in \mathcal{G}$:

$$(13) \quad \mathbb{E}[X 1_B] = \mathbb{E}[\mathbb{E}[X | \mathcal{G}] 1_B].$$

En particulier, on a :

$$(14) \quad \mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X | \mathcal{G}]].$$

Démonstration. — **existence** Supposons d'abord X positive. La mesure positive

$$(15) \quad \mu(A) = \mathbb{E}[X 1_A] \quad (A \in \mathcal{F}),$$

est par construction absolument continue par rapport à \mathbb{P} sur la tribu \mathcal{F} (si $\mathbb{P}(A) = 0$, alors $\mu(A) = 0$). Sa restriction à la sous tribu \mathcal{G} définit donc également une mesure absolument continue par rapport à \mathbb{P} , sur \mathcal{G} . On peut donc considérer sa dérivée de Radon-Nykodim, $Z = \frac{d\mu}{d\mathbb{P}}|_{\mathcal{G}}$ qui est par construction une variable aléatoire positive \mathcal{G} mesurable telle que

$$(16) \quad \mu(B) = \mathbb{E}[X 1_B] = \int Z 1_B d\mathbb{P} = \mathbb{E}[Z 1_B] \quad (B \in \mathcal{G}).$$

Maintenant si X est une variable aléatoire intégrable, alors on décompose $X(\omega)$ en parties positive et négative : $X = X^+ - X^-$. On définit les espérances conditionnelles $\mathbb{E}[X^\pm | \mathcal{G}]$ qui sont bien des variables aléatoires \mathcal{G} mesurables, positives et intégrables. Il est immédiat que $Z = \mathbb{E}[X^+ | \mathcal{G}] - \mathbb{E}[X^- | \mathcal{G}]$ satisfait les hypothèses 1 et 2.

unicité Si Y et Z conviennent alors $B = \{Y \geq Z\}$ est dans \mathcal{G} , donc :

$$(17) \quad \mathbb{E}[X 1_B] = \mathbb{E}[Y 1_B] = \mathbb{E}[Z 1_B]$$

ce qui entraîne $0 = \mathbb{E}[(Y - Z) 1_B]$ et donc $(Y - Z) \mathbf{1}_{(Y \geq Z)} = 0$ presque sûrement. De même, $(Z - Y) \mathbf{1}_{(Z \geq Y)} = 0$ presque sûrement, et donc $Y = Z$ presque sûrement. \square

Remarque. — Lorsque X est intégrable, la condition 2 peut être remplacée par la condition suivante : pour tout variable aléatoire bornée \mathcal{G} mesurable U ,

$$(2') \quad \mathbb{E}[X U] = \mathbb{E}[\mathbb{E}[X | \mathcal{G}] U].$$

Lorsque X est simplement positive, on demande juste à U d'être positive et \mathcal{G} mesurable.

Il est immédiat de vérifier que lorsque \mathcal{G} est engendrée par un système complet d'événements, alors cette définition coïncide avec la définition (10).

Définition 3.2. — On pose, pour X positive où intégrable,

$$(18) \quad \mathbb{E}[X | Y] = \mathbb{E}[X | \sigma(Y)].$$

D'après le Lemme de Doob, c'est donc une fonction mesurable de Y , $\mathbb{E}[X | Y] = \phi(Y)$ caractérisée par

$$(19) \quad \mathbb{E}[Xf(Y)] = \mathbb{E}[\phi(Y)g(Y)]$$

pour toute fonction mesurable g , bornée si X est intégrable, positive si X est positive.

Proposition 3.2. — On se place sous les conditions d'application du Théorème 3.1 pour définir correctement toutes les espérances conditionnelles ci-dessous pour les variables aléatoires X, Y, X_n, \dots . Les égalités et inégalités ne sont valables que presque sûrement.

1. (linéarité) $\mathbb{E}[aX + bY | \mathcal{G}] = a\mathbb{E}[X | \mathcal{G}] + b\mathbb{E}[Y | \mathcal{G}]$
2. (monotonie) si $X \leq Y$, alors $\mathbb{E}[X | \mathcal{G}] \leq \mathbb{E}[Y | \mathcal{G}]$.
3. (convergence monotone) si $0 \leq X_n$ et $X_n \uparrow X$ presque sûrement, alors $\mathbb{E}[X_n | \mathcal{G}] \uparrow \mathbb{E}[X | \mathcal{G}]$.
4. (convergence dominée) si $X_n \rightarrow X$ presque sûrement et si $\sup_n |X_n| \leq Y$ et Y est intégrable, alors $\mathbb{E}[X_n | \mathcal{G}] \rightarrow \mathbb{E}[X | \mathcal{G}]$ presque sûrement.
5. (Jensen conditionnelle) Si ϕ est convexe et si $\phi(X)$ est positive ou intégrable, alors $\phi(\mathbb{E}[X | \mathcal{G}]) \leq \mathbb{E}[\phi(X) | \mathcal{G}]$. En particulier, si X est intégrable

$$|\mathbb{E}[X | \mathcal{G}]| \leq \mathbb{E}[|X| | \mathcal{G}]$$

6. (projection) Si $\mathcal{H} \subset \mathcal{G}$ est une sous tribu de \mathcal{G} alors,

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

7. si X est indépendante de \mathcal{G} alors $\mathbb{E}[X | \mathcal{G}] = \mathbb{E}[X]$ (l'espérance conditionnelle est constante).
8. si Y est \mathcal{G} mesurable, alors

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

Démonstration. — Les propriétés 1 et 2 sont triviales. Pour la convergence monotone conditionnelle, soit Z la limite croissante de $\mathbb{E}[X_n | \mathcal{G}]$. Alors Z est positive \mathcal{G} mesurable et pour tout $B \in \mathcal{G}$, par convergence monotone $\mathbb{E}[X_n 1_B] = \mathbb{E}[\mathbb{E}[X_n | \mathcal{G}] 1_B]$ converge vers $\mathbb{E}[X 1_B]$ et vers $\mathbb{E}[Z 1_B]$. Donc $Z = \mathbb{E}[X | \mathcal{G}]$. La convergence dominée conditionnelle se démontre comme la convergence dominée classique en utilisant un lemme de Fatou conditionnel ! Pour l'inégalité de Jensen conditionnelle, supposons par exemple que X et $\phi(X)$ soient intégrables, et posons $Y = \mathbb{E}[X | \mathcal{G}]$, $Z = \mathbb{E}[\phi(X) | \mathcal{G}]$. Alors,

pour tout variable aléatoire U bornée \mathcal{G} mesurable, et toute fonction linéaire $\psi(x) = ax + b$ qui minore ϕ , on a:

$$\begin{aligned}\mathbb{E}[ZU] &= \mathbb{E}[\phi(X)U] \geq \mathbb{E}[\psi(X)U] = \mathbb{E}[(aX + b)U] \\ &= \mathbb{E}[(aY + b)U] = \mathbb{E}[\psi(Y)U].\end{aligned}$$

En conséquence, en prenant $U = \mathbf{1}_{(\psi(Y) \geq Z)}$ on obtient

$$\mathbb{E}[(\psi(Y) - Z)\mathbf{1}_{(\psi(Y) \geq Z)}] \leq 0$$

et donc $Z \geq \psi(Y)$ presque sûrement. On note Ω_ψ un ensemble de probabilité 1, $\mathbb{P}(\Omega_\psi) = 1$, tel que pour tout $\omega \in \Omega_\psi$ on ait $Z(\omega) \geq \psi(Y(\omega))$.

Observons que $\phi(x) = \sup\{\psi(x) : \psi \in \Lambda\}$ avec Λ l'ensemble des fonctions $\psi(x) = ax + b$ linéaires, qui minorent ϕ et à coefficients a, b rationnels. En conséquence si $\Omega' = \cap_{\psi \in \Lambda} \Omega_\psi$, on a $\mathbb{P}(\Omega') = 1$ et pour ω dans Ω' , $Z(\omega) \geq \phi(Y(\omega))$. C'est à dire $Z \geq \phi(Y)$ presque sûrement.

6. On pose $Y = \mathbb{E}[X | \mathcal{G}]$ et $Z = \mathbb{E}[Y | \mathcal{H}]$. Soit $B \in \mathcal{H}$. Comme $B \in \mathcal{H}$, on a:

$$\mathbb{E}[X1_B] = \mathbb{E}[Y1_B] = \mathbb{E}[Z1_B]$$

d'où le résultat.

7. Si $B \in \mathcal{G}$ et $\mu = \mathbb{E}[X]$, alors, par indépendance de X et 1_B ,

$$\mathbb{E}[X1_B] = \mathbb{E}[X]\mathbb{E}[1_B] = \mu\mathbb{E}[1_B] = \mathbb{E}[\mu 1_B]$$

Comme la variable constante égale à μ est bien \mathcal{G} mesurable, on a bien $\mathbb{E}[X | \mathcal{G}] = \mu$.

8. Supposons X et Y positives. Si $B \in \mathcal{G}$, alors, $Y1_B$ est \mathcal{G} mesurable positive, donc

$$\mathbb{E}[XY1_B] = \mathbb{E}[\mathbb{E}[X | \mathcal{G}](Y1_B)] = \mathbb{E}[(\mathbb{E}[X | \mathcal{G}]Y)1_B],$$

d'où le résultat.

Si X et XY sont intégrables, on décompose X et Y en parties positive et négative et on applique la linéarité. \square

Rappelons que $L^2(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de Hilbert, muni du produit scalaire $(X, Y) \rightarrow \mathbb{E}[XY]$ et que l'on peut voir $L^2(\Omega, \mathcal{G}, \mathbb{P})$ comme un sous espace fermé formé des variables de carré intégrable, qui sont \mathcal{G} mesurables.

Proposition 3.3. — Si X est de carré intégrable, alors $\mathbb{E}[X | \mathcal{G}]$ est la projection orthogonale de X sur le sous espace fermé $L^2(\Omega, \mathcal{G}, \mathbb{P})$.

Démonstration. — En effet, si Z est cette projection orthogonale et si U est \mathcal{G} mesurable bornée, alors $U \in L^2(\Omega, \mathcal{G}, \mathbb{P})$ et donc U est orthogonale à $X - Z$, $0 = \mathbb{E}[U(X - Z)]$, ce qui entraîne $\mathbb{E}[XU] = \mathbb{E}[ZU]$. Comme Z est bien \mathcal{G} mesurable, c'est l'espérance conditionnelle $\mathbb{E}[X | \mathcal{G}]$. \square

Loi conditionnelle

Si on n'avait pas des identités presque sûres, mais de vraies identités, alors à ω fixé, l'application $f \rightarrow \mathbb{E}[f(X) | \mathcal{G}]$ aurait toutes les propriétés d'une mesure de probabilité. Il nous faut un argument topologique supplémentaire pour rendre cela rigoureux, et la définition d'un noyau.

Définition 3.3. — Étant donnés deux espaces mesurés (Ω, \mathcal{F}) et (E, \mathcal{E}) , un noyau de transition de (Ω, \mathcal{F}) sur (E, \mathcal{E}) est une application $q : \Omega \times \mathcal{E} \rightarrow [0, 1]$ qui vérifie :

1. pour tout $A \in \mathcal{E}$, la fonction $\omega \rightarrow q(\omega, A) = q_\omega(A)$ est \mathcal{F} mesurable ;
2. pour tout $\omega \in \Omega$, la fonction $A \rightarrow q(\omega, A) = q_\omega(A)$ est une probabilité sur (E, \mathcal{E}) .

Pour toute fonction $f : (E, \mathcal{E}) \rightarrow \mathbb{R}$ borélienne positive ou bornée, on notera

$$\int f(x) dq_\omega(x) = \int f(x) q(\omega, dx)$$

la variable aléatoire espérance de f par rapport à la probabilité dépendant de ω .

Il est immédiat de vérifier que si $q : (\Omega, \mathcal{F}) \rightarrow \mathcal{M}_1(E)$ est une application mesurable, avec $\mathcal{M}_1(E)$ espace des mesures de probabilité sur E métrique séparable, muni de la topologie de la convergence faible, alors q induit un noyau de (Ω, \mathcal{F}) sur (E, \mathcal{E}) , par $q(\omega, A) = q(\omega)(A)$.

Théorème 3.4. — Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E, \mathcal{E})$ une variable aléatoire à valeurs dans un espace polonais E muni de sa tribu des boréliens et soit \mathcal{G} une sous tribu de \mathcal{F} . Alors il existe une unique application mesurable $q : (\Omega, \mathcal{F}) \rightarrow \mathcal{M}_1(E)$ telle que

$$\mathbb{E}[f(X) | \mathcal{G}] = \int f(x) q(\omega, dx)$$

pour tout fonction $f : (E, \mathcal{E}) \rightarrow \mathbb{R}$ borélienne positive ou bornée. Ce noyau est appelé loi conditionnelle régulière de X étant donné \mathcal{G} .

Démonstration. — Voir Borkar [?].

□

Une conséquence de ce théorème est l'existence de lois conditionnelles pour les couples de variables aléatoires.

Proposition 3.5. — Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E, \mathcal{E})$ une variable aléatoire à valeurs dans un espace polonais E muni de sa tribu des boréliens, et soit $Y : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E', \mathcal{E}')$ une variable aléatoire à valeurs dans un espace métrique

séparable. Alors il existe un noyau K de (E', \mathcal{E}') sur (E, \mathcal{E}) tel que pour toute fonction $\phi : (E, \mathcal{E}) \rightarrow \mathbb{R}$ borélienne bornée ou positive,

$$(20) \quad \mathbb{E}[\phi(X) | Y](\omega) = \int \phi(x) dK_{Y(\omega)}(x) \quad ps.$$

La mesure $K_y(dx) = K(y, dx)$ est appelée loi conditionnelle de X sachant $Y = y$ et notée parfois $\mathcal{L}(X | Y = y)$ ou $P_X^{Y=y}$.

Démonstration. — On applique le théorème précédent à la tribu $\mathcal{G} = \sigma(Y)$. Comme l'application q est $\sigma(Y)$ mesurable, et que Y est à valeurs dans un espace métrique séparable, le Lemme de Doob ?? assure l'existence d'une application mesurable $K : (E', \mathcal{E}') \rightarrow \mathcal{M}_1(E)$ telle que $q(\omega) = K(Y(\omega))$, d'où le résultat. \square

Le cadre d'application le plus courant est celui des variables aléatoires à valeurs dans \mathbb{R} ou \mathbb{R}^d . L'existence d'une densité pour le couple permet de donner une forme explicite à la loi conditionnelle.

Proposition 3.6. — Soit (X, Y) un couple de variables aléatoires réelles admettant la densité f . Alors la loi conditionnelle de X sachant $Y = y$, $\mathcal{L}(X | Y = y)$, admet pour densité

$$(21) \quad f_{Y=y}(x) = \frac{f(x, y)}{\int f(x', y) dx'} \mathbf{1}_{(y \in A)} \quad \text{avec } A = \left\{ y : \int f(x', y) dx' > 0 \right\}.$$

SECTION 4

Conditionnement et indépendance

On se donne Y et Z deux variables aléatoires définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans des espaces polonais F et G , et $f : F \times G \rightarrow E$ une application mesurable à valeurs dans un espace polonais E .

Proposition 3.7. — Soit \mathcal{G} une sous tribu de \mathcal{F} telle que Y soit \mathcal{G} mesurable et Z soit indépendante de \mathcal{G} . Alors la loi conditionnelle régulière de $X = f(Y, Z)$ sachant \mathcal{G} est donnée par (ϕ étant borélienne positive ou bornée)

$$(22) \quad \mathbb{E}[\phi(X) | \mathcal{G}] = \int \phi(f(Y(\omega), z)) dP_Z(z) = \mathbb{E}[\phi(f(y, Z))]_{|y=Y(\omega)}$$

En particulier, la loi conditionnelle régulière de $X = f(Y, Z)$ sachant $Y = y$ est la loi de la variable aléatoire $f(y, Z)$.

Démonstration. — Le cas particulier s'obtient en prenant $\mathcal{G} = \sigma(Y)$. Pour le cas général soit $\phi : E \rightarrow \mathbb{R}$ borélienne, disons bornée. Soit $B \in \mathcal{G}$. Alors, l'espace \mathcal{H} des fonctions $g : F \times G \rightarrow \mathbb{R}$ boréliennes bornées telles que

$$\mathbb{E}[g(Y, Z)1_B] = \mathbb{E}\left[\left(\int g(Y, z) dP_Z(z)\right)1_B\right]$$

est un espace vectoriel monotone, qui contient la classe \mathcal{C} des fonctions de la forme $g(y, z) = h(y)k(z)$, h, k boréliennes bornées. En effet, d'après le théorème de Fubini-Tonelli, et l'indépendance de $k(Z)$ et $h(Y)1_B$, on a

$$\begin{aligned} \mathbb{E}[g(Y, Z)1_B] &= \mathbb{E}[h(Y)1_B k(Z)] = \mathbb{E}[h(Y)1_B] \mathbb{E}[k(Z)] \\ &= \mathbb{E}[h(Y)1_B] \int k(z) dP_Z(z) \\ &= \mathbb{E}\left[\left(\int h(Y)k(z) dP_Z(z)\right)1_B\right] \\ &= \mathbb{E}\left[\left(\int g(Y, z) dP_Z(z)\right)1_B\right]. \end{aligned}$$

En vertu du théorème des classes monotones, \mathcal{H} contient toutes les fonction $\sigma(\mathcal{C})$ bornées. Comme $\sigma(\mathcal{C})$ est la tribu des boréliens, on a

$$\mathbb{E}[\phi(f(Y, Z))1_B] = \mathbb{E}\left[\left(\int \phi(f(Y, z)) dP_Z(z)\right)1_B\right]$$

ce qui prouve (22) puisque B est arbitraire. □

4

Chaîne de Markov à temps discret

SECTION 1

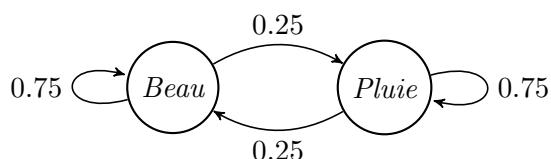
Introduction

Exemple 4.1. — On dispose de statistiques, de tableaux de chiffres, indiquant qu'il fait beau 50% du temps et mauvais 50% du temps. Une première façon naïve de prédire le temps est la suivante

$$\mathbb{P}(\text{beau}) = \mathbb{P}(\text{mauvais}) = 0.5$$

La justification est la loi des grands nombres : l'hypothèse de base est que le climat du lendemain est indépendant du climat des autres jours.

Malheureusement, on constate en examinant les chiffres qu'il y a 3 fois plus de chances que le climat du lendemain reste le même que celui d'aujourd'hui (plutôt qu'il ne change). On représente ce modèle par le diagramme



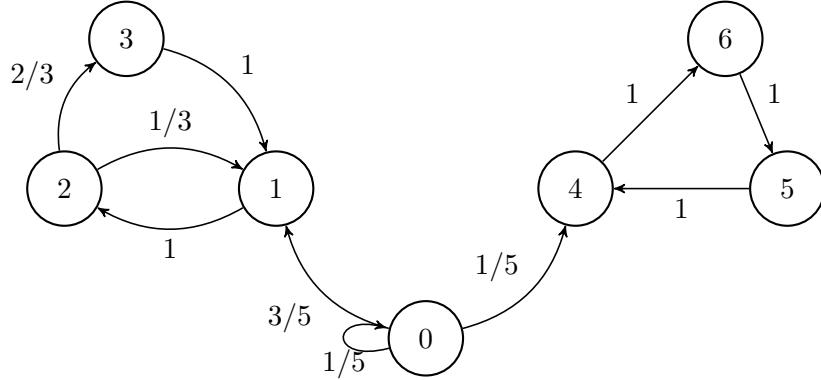
ou encore par la matrice $P = \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}$.

Pour prédire le temps de demain nous avons besoin de

- un mécanisme de transition (décris par un graphe ou une matrice)
- savoir quel temps il fait aujourd'hui

Considérons un exemple plus évolué

Exemple 4.2. —



On aimerait répondre à la liste de questions suivantes:

- Partant de 0 quelle est la probabilité de toucher 6 ?
- Partant de 1 quelle est la probabilité de toucher 3 ?
- partant de 1 combien de temps en moyenne faut-il pour toucher 3 . (réponse 3)
- Partant de 1, la proportion de temps passé en 2 est, en temps long, 3/8
- Partant de 0, la probabilité que je sois en 1 à l'instant n , est pour n grand, proche de 9/32.

SECTION 2

Définitions et propriétés de base

Soit I un ensemble fini ou dénombrable, appelé espace d'états (nommé aussi S ou E). Une mesure (positive) sur I est une famille $(\lambda_i, i \in I)$ de nombres positifs. C'est une loi de probabilité si $\sum_i \lambda_i = 1$.

Une matrice $P = (p_{ij})$ est stochastique (on dit aussi matrice de transition) si toutes ses lignes sont des probabilités i.e.

$$\forall i, j \ p_{ij} \geq 0 \quad \text{et} \quad \forall i, \ \sum_j p_{ij} = 1$$

Il y a une bijection évidente entre les matrices de transition et les graphes valués donnée par il y a une flèche de i à j valuée par p_{ij} ssi $p_{ij} > 0$.

Définition 4.1. — Une chaîne de Markov homogène à valeurs dans I de matrice de transition P et de loi initiale λ est une famille de variables aléatoires $(X_n, n \in \mathbb{N})$ telle que

- $\mathbb{P}(X_0 = i) = \lambda(i)$
- $\mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n, \dots, X_0 = i_0) = \mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n) = p_{i_n, i_{n+1}}$

On dit alors que $(X_n)_{n \in \mathbb{N}}$ est markov (λ, P) .

Exemple 4.3. — $I = \{1, 2, 3\}$ est l'espace des ‘états avec 1 qui signifie en panne, 2 en fonctionnement erratique et en bon fonctionnement. Les transitions sont décrites par la matrice

$$P \begin{pmatrix} 0 & 2 & 0 \\ 0 & 2/3 & 1/3 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

Question Si je pars d'une machine en bon état, quelle est la probabilité α (resp. β) qu'elle soit en bon état (resp. erratique) dans 4 jours ?

Réponse il y a deux chemins de même probabilité pour obtenir une machine en bon état $\alpha = 2 \frac{1}{2}^2 \frac{1}{3}^2$

2.1 Caractérisation

Proposition 4.1. — $(X_n)_{0 \leq n \leq N}$ est Markov λ, P ssi pour tous i_0, \dots, i_N

$$\mathbb{P}(X_0 = i_0, \dots, X_N = i_N) = \lambda(i_0)p_{i_0, i_1} \dots p_{i_{N-1}, i_N}$$

Démonstration. — conditionnements successifs et récurrence. \square

On note δ_i la masse de Dirac en i et \mathbb{P}_i la loi d'une chaîne (δ_i, P) (on dit que la chaîne est issue de i car $X_0 = i$ ps).

Proposition 4.2. — Si $(X_n)_{n \geq 0}$ est Markov (λ, P) , alors conditionnellement à $X_m = i$ le processus $(X_{m+n}, n \geq 0)$ est Markov (δ_i, P) et est indépendant de (X_0, \dots, X_m) .

Démonstration. — Il suffit de montrer que pour tout évènement $A \in \mathcal{F}_m^X := \sigma(X_0, \dots, X_m)$ et tout évènement $B = \{X_m = i_m, \dots, X_{m+n} = i_{m+n}\}$ de $\sigma(X_p, p \geq m)$ on a

$$(23) \quad \mathbb{P}(A \cap B \mid X_m = i) = \mathbb{P}(A \mid X_m = i)\mathbb{P}_i(X_0 = i_m, \dots, X_n = i_{m+n})$$

En effet les évènements B de cette forme constituent un π système qui engendre $\sigma(X_p, p \geq m)$. On démontre (23) en prenant A de la forme $\{X_0 = i_0, \dots, X_m = i_m\}$ car tout évènement de \mathcal{F}_m^X est une réunion disjointe de tels ensembles. \square

Etant donné une variable aléatoire $Z \in \mathcal{F}_\infty^X$, par le Lemme de Doob, $Z = F(X_n, n \geq 0)$ avec F une fonction mesurable, et donc on peut considérer l'opérateur de shift $Z \circ \theta^m = F(X_{m+n}, n \geq 0)$ (et on note $\theta = \theta^1$). Alors, la proposition précédente s'écrit

Proposition 4.3. — Pour tout variable aléatoire Z , \mathcal{F}_∞^X mesurable, positive ou bornée, et tout $m \geq 0$,

$$\mathbb{E}[Z \circ \theta^m \mid \mathcal{F}_m] = \mathbb{E}_{X_m}[Z] \quad p.s.$$

Remarque. — En appliquant la proposition précédente à $m = 0$ on obtient que quelle que soit la loi initiale λ de X_0 , du moment qu'elle charge le point i , ($\lambda(i) > 0$), on a pour tout évènement A , $\mathbb{P}_i(A) = \mathbb{P}(A | X_0 = i)$. En bref, conditionner la chaîne à valoir i en 0 revient à la faire partir de i .

2.2 Calculs algébriques On note une mesure positive λ comme un vecteur ligne et donc λP est la mesure $(\lambda P)(j) = \sum_i \lambda(i)p_{i,j}$. On note $p_{i,j}^{(n)}$ l'élément i, j de la matrice P^n et donc on a $p_{i,j}^{(2)} = \sum_k p_{i,k}p_{k,j}$ et en général

$$p_{i;j}^{(n)} = \sum_{i_0=i, i_1, \dots, i_n=j} p_{i_0, i_1} \cdots p_{i_{n-1}, i_n}$$

Proposition 4.4. — Soit $(X_n)_{n \geq 0}$ Markov (λ, P) . Alors pour toutn la loi de X_n est λP^n i.e.

$$\mathbb{P}(X_n = j) = (\lambda P^n)(j)$$

En particulier, pour $\lambda = \delta_i$ et tout $m \geq 0$, on obtient

$$\mathbb{P}_i(X_n = j) = \mathbb{P}(X_{m+n} = j | X_m = i) = p_{i,j}^{(n)}$$

Démonstration. — Formule des probabilités totales :

$$\mathbb{P}(X_n = j) = \sum_{i_0, \dots, i_n=j} \mathbb{P}(X_0 = i_0, \dots, X_n = i_n)$$

□

Exemple 4.4. — Pour le climat à Nantes, s'il fait beau aujourd'hui, la probabilité qu'il fasse beau dans trois jours est

$$\mathbb{P}(X_3 = 1 | X_0 = 1) = p_{i,i}^{(3)} = 9/16$$

car on calcule

$$P^2 = \begin{pmatrix} 10/16 & 6/16 \\ 6/16 & 10/16 \end{pmatrix}$$

2.3 Autres propriétés

Définition 4.2. — On appelle filtration une suite croissante de tribus $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ sur Ω . On dit que la va $T : \omega \rightarrow \bar{N}$ est un temps d'arrêt si pour tout n , $\{T \leq n\} \in \mathcal{F}_n$. On dit qu'un processus Z_n est adapté si pour tout n la va Z_n est \mathcal{F}_n mesurable.

Pour une chaîne de Markov $(X_n, n \in \mathbb{N})$ on considère la filtration naturelle $\mathcal{F}_n = \sigma(X_k, k \leq n)$ et $\mathcal{F}_\infty = \sigma X_n, n \geq 0$. Un des temps d'arrêt le plus utilisé est le premier temps de retour en i

$$T_i = \inf n \geq 1 : X_n = i$$

avec la convention $\inf \emptyset = +\infty$. On note $\mathcal{F}_T = \sigma\{Z_T, A \text{ processus adapté}\}$ la filtration arrêtée en T .

Proposition 4.5. — Pour toute va Z \mathcal{F}_∞ mesurable, positive ou bornée, et tout temps d'arrêt T , sur $T < +\infty$ on a p.s.

$$\mathbb{E}[Z \circ \theta^T | \mathcal{F}_T] = \mathbb{E}_{X_T}[Z]$$

— SECTION 3 —

Classification des états

3.1 Structure de classe

Définition 4.3. — On dit que i mène à j et on le note $i \rightarrow j$ si

$$\mathbb{P}_i(\exists n \geq 0 : X_n = j) > 0.$$

On dit que i et j communiquent et on le note $i \leftrightarrow j$ si $i \rightarrow j$ et $j \rightarrow i$.

Proposition 4.6. — Si $i \neq q$ alors sont équivalentes

- (a) $i \rightarrow j$
- (b) $\exists i_0 = i, i_1, \dots, i_n = j$ tel que $p_{i_0, i_1} \cdots p_{i_{n-1}, i_n} > 0$
- (c) Il existe un chemin dans le graphe de transition menant de i à j .
- (d) $\exists n \geq 0$ tel que $p_{i,j}^{(n)} > 0$.

Démonstration. — (b) et (c) sont identiques; Comme une somme de termes positifs est > 0 ssi il existe un term > 0 , on en déduit que (b) \iff (d). Enfin l'équivalence entre (a) et (d) découle de la double inégalité

$$p_{i,j}^{(n)} = \mathbb{P}_i(X_n = j) \leq \mathbb{P}_i(\exists m, X_m = j) \leq \sum_m \mathbb{P}_i(X_m = j)$$

En effet, si (a), alors $\mathbb{P}_i(\exists m, X_m = j) > 0$ donc $\sum_m \mathbb{P}_i(X_m = j) > 0$ et il existe m tel que $\mathbb{P}_i(X_m = j) > 0$. \square

Corollaire 4.7. — la relation $i \leftrightarrow j$ est une relation d'équivalence. Les classes d'équivalence sont appelées classes de communication.

Démonstration. — Pour la transitivité si $i \rightarrow j$ et $j \rightarrow k$ alors il existe m, n tels que $p_{i,j}^{(n)} > 0$ et $p_{j,k}^{(m)} > 0$ donc $p_{i,k}^{(n+m)} \geq p_{i,j}^{(n)} p_{j,k}^{(m)} > 0$. \square

Définition 4.4. — La chaîne est dite irréductible s'il y a une seule classe. La classe C est dite ouverte s'il existe $i \in C$ et $j \notin C$ tels que $i \rightarrow j$.

Proposition 4.8. — La classe C est ouvertessi il existe $i \in C$ et $j \notin C$ tels que $p_{i,j} > 0$.

On suppose la classe fermée et que p.s. $X_0 \in C$. Alors presque sûrement $\forall n, X_n \in C$. En d'autres termes une chaîne reste toujours p.s. dans une classe fermée.

Démonstration. — Pour la première partie il existe un chemin fini de C à l'extérieur de C . Il y a bien un moment où il quitte C .

Pour la seconde partie on fait une récurrence et on voit que l'héritage découle de la remarque suivante. COMME C est fermée, si $i \in C$, alors pour tout $j \notin C$, $p_{i,j}^{(n)} = 0$ et donc en sommant sur tous les $j \notin C$, $\mathbb{P}_i(X_n \in C^C) = 0$ et donc

$$\mathbb{P}(X_n \notin C) = \sum_i \mathbb{P}(X_0 = i) \mathbb{P}_i(X_n \notin C) = 0$$

□

3.2 Récurrence et transience

Définition 4.5. — On dit que i est transient si $\sum_n p_{ii}^{(n)} < +\infty$ et récurrent si $\sum_n p_{ii}^{(n)} = +\infty$.

Proposition 4.9. — La récurrence est une propriété de classe. Soit C une classe de communication. Alors soit tous les éléments de C sont récurrents, soit ils sont tous transients.

Démonstration. — soit $i, j \in C$ on a $i \leftrightarrow j$ donc il existe k, m tels que $p_{ij}^{(m)} > 0$ et $p_{ji}^{(k)} > 0$ On déduit de l'inégalité

$$p_{ii}^{(n+m+k)} \geq p_{ij}^{(m)} p_{jj}^{(n)} p_{ji}^{(k)}$$

que les séries $\sum_n p_{ii}^{(n)}$ et $\sum_n p_{jj}^{(n)}$ sont de même nature. □

Le nombre de retours en i (ou temps passé en i à partir de 1) est $N_i = \sum_{n \geq 1} \mathbf{1}_{(X_n=i)}$ et la probabilité de retour en i est $f_i := \mathbb{P}_i(N_i > 0) = \mathbb{P}_i(T < +\infty)$ avec $T = \inf \{n \geq 1 : X_n = i\}$.

Lemme 4.10. — pour tout $r \in \mathbb{N}$, $\mathbb{P}_i(N_i > r) = f_i^{r+1}$.

Démonstration. — SOit $T_0 = 0$ et $T_{k+1} = \inf n > T_k : X_n = 1$ le $k+1$ ième temps de passage en i (avec la convention $\inf \emptyset = +\infty$). Alors $\{N_i > r\} =$

$\{T_r < +\infty\}$ et on a $T_{k+1} = T_k + T_1 \circ \theta^{T_k}$, donc par la propriété de Markov forte, comme $X_{T_k} = i$ lorsque $T_k < +\infty$,

$$\begin{aligned}\mathbb{P}_i(N_i > k+1) &= \mathbb{P}_i(T_k + T_1 \circ \theta^{T_k} < +\infty) = \mathbb{P}(T_k < +\infty, T_1 \circ \theta^{T_k} < +\infty) \\ &= \mathbb{E}_i \left[\mathbf{1}_{(T_k < +\infty)} \mathbb{E}_{X_{T_k}} [\mathbf{1}_{(T_1 < +\infty)}] \right] \\ &= \mathbb{P}_i(T_1 < +\infty) \mathbb{P}_i(T_k < +\infty) = f_i f_i^r.\end{aligned}$$

□

Théorème 4.11 (dichotomie). — Si $f_i = 1$ alors presque sûrement $X_n = i$ pour une infinité d'entiers n et i est un état récurrent.

Si $f_i < 1$, alors presque sûrement $X_n = i$ pour seulement un nombre fini d'entiers et l'état i est transient.

Démonstration. — On observe que d'une part

$$\mathbb{E}_i [N_i] = \sum_{k \in N} \mathbb{P}_i(N_i > k) = \sum_k f_i^{k+1} = \begin{cases} \frac{f_i}{1-f_i} & \text{si } f_i < 1; \\ +\infty & \text{si } f_i = 1; \end{cases}$$

et d'autre part

$$\mathbb{E}_i [N_i] = \sum_n \mathbb{E}_i [\mathbf{1}_{(X_n=i)}] = \sum_n p_{ii}^{(n)}$$

d'où la dichotomie. Si $f_i = 1$ alors $N_i = +\infty$ ps (car pour tout k , $\mathbb{P}_i(N_i > k) = 1$) et i est transient, alors que si $f_i < 1$ on a $N_i < +\infty$ ps et i est récurrent. □

Remarque. — On montre de la même façon que si C est une classe récurrente et $i, j \in C$ alors \mathbb{P}_i ps, on a $N_j = +\infty$ et $\mathbb{E}_i [N_j] = \sum_n p_{ij}^{(n)} = +\infty$. Si C est une classe transiente et $i, j \in C$ alors \mathbb{P}_i ps, on a $N_j < +\infty$ et $\mathbb{E}_i [N_j] = \sum_n p_{ij}^{(n)} < +\infty$.

Proposition 4.12. — Soit C une classe de communication.

1. Si C est ouverte alors C est transiente.
2. Si C est ouverte et C est finie, alors p.s. il existe $n_0 = n_0(\omega)$ tel que $\forall n \geq n_0, X_n \notin C$ (au bout d'un moment on sort définitivement d'une classe ouverte finie).
3. Si C est fermée et C est finie alors C est récurrente.

Démonstration. — 1) Soit $i \in C, j \notin C$ tels que $p_{ij} > 0$. On a $j \not\rightarrow i$ donc, $1 = \mathbb{P}_j(\forall n \geq 0, X_n \neq i)$. Donc, par Markov,

$$\mathbb{P}_i(N_i = 0) \geq \mathbb{P}_i(X_1 = j, \forall n \geq 1 X_n \neq i) = p_{ij} \mathbb{P}_j(\forall n \geq 0, X_n \neq i) = p_{ij} > 0$$

i.e. $f_i = \mathbb{P}_i(N_i > 0) < 1$ et i est transient.

2) Si C est finie, on note $N_C = \sum_{j \in C} N_j$ le temps passé en C , et on a $\mathbb{E}_i [N_C] = \sum_{j \in C} \mathbb{E}_i [N_j]$ qui est une somme finie. Donc si C est en outre ouverte, alors

elle est transiente et pour tout $j \in C$ on a $\mathbb{E}_i [N_j] < +\infty$ donc $\mathbb{E}_i [N_C] < +\infty$ et donc $N_C < +\infty$ ps.

3) Si C est finie et fermée, et que $i \in C$, alors $X_0 \in C$ ps, et on a vu que ps $\forall n, X_n \in C$ et donc $N_C = +\infty$ ps, donc $\mathbb{E}_i [N_c] = +\infty$ et il existe j tel que $\mathbb{E}_i [N_j] = +\infty$, donc i est récurrent. \square

Corollaire 4.13. — *Une chaîne de Markov irréductible sur un espace d'états fini est récurrente.*

Remarque. — *Sur un espace d'états infini tout peut arriver. Sur \mathbb{Z} la marche simple est récurrente (TCL) alors que la marche biaisée est transiente (LGN : $S_n \rightarrow +\infty$)*

SECTION 4

Temps d'atteinte et probabilités d'absorption

On considère une chaîne de Markov, à espace d'états fini, qui a des états transients et des états absorbants, et on met la matrice de transition sous la forme canonique en plaçant les états transients au début.

$$P = \left(\begin{array}{c|c} Q & R \\ \hline 0 & I \end{array} \right)$$

Proposition 4.14. — *Pour une chaîne de Markov avec un état absorbant, presque sûrement la chaîne sera absorbée, i.e. $Q^n \rightarrow 0$*

Proposition 4.15. — *Pour une chaîne de Markov avec un état absorbant, la matrice $I - Q$ est inversible d'inverse $N = \sum Q^n$ et n_{ij} est le nombre moyen de fois où la chaîne est dans l'état j , si elle part de l'état i .*

N est la matrice fondamentale.

Démonstration. — On appelle A l'ensemble des états absorbants. Alors, si $i, j \notin A$ $n_{ij} = \mathbb{E}_i [\sum_k \mathbf{1}_{(X_k=j)}] = \sum_k q_{ij}^{(k)} = (I - Q)_{ij}^{(-1)}$. \square

Théorème 4.16. — *Soit t_i le temps moyen d'absorption de la chaîne si elle part dans l'état i . Alors, $t = N1$. En particulier, si $i \notin A$, alors $t_i = 1 + \sum_{j \notin A} p_{ij} t_j = 1 + (Qt)_i$.*

Démonstration. — $(N1)_i = \sum_j n_{ij}$ \square

Théorème 4.17. — *Soit b_{ij} la probabilité que la chaîne partant de i soit absorbée en j . Alors, $B = NR$.*

Démonstration. — $B_{ij} = \sum_{n,k} \mathbb{P}(X_n = k, X_{n+1} = j \mid X_0 = i) = \sum_{nk} q_{ik}^{(n)} r_{kj} = (NR)_{ij}$.

□

Exemple 4.5. — espace d'états $\{1, 2, 3, 4\}$, $p_{i,i+1} = 0.4 = 1 - p_{i,i-1}$ si $i = 2, 3$, 1 et 4 absorbants. Alors on réordonne les états $E = \{2, 3, 1, 4\}$ et on calcule $Q = \begin{pmatrix} 0 & 0.4 \\ 0.6 & 0 \end{pmatrix}$ et $R = \begin{pmatrix} 0.6 & 0 \\ 0 & 0.4 \end{pmatrix}$. On trouve $N = \begin{pmatrix} 1.32 & 0.526 \\ 0.789 & 1.32 \end{pmatrix}$ et $B = NR = \begin{pmatrix} 0.789 & 0.21 \\ 0.473 & 0.52 \end{pmatrix}$. Si 1 est gagnant et 4 perdant et que l'on me donne deux fois ma mise quand je pars de 3, et une seule fois quand je pars de 2, quelle doit être ma stratégie ? (on part de 3 car $20.473 > 0.789$).

On a parfois besoin de connaître la loi du temps de sortie d'un point.

Lemme 4.18. — On suppose que la chaîne est issue de i et que i est non absorbant. Soit $T = \inf n \geq 1 : X_n \neq i$. Alors sous \mathbb{P}_i , les variables aléatoires X_T et T sont indépendantes, T de loi géométrique de paramètre $1 - p_{ii}$ et X_T de loi

$$\mathbb{P}(X_T = j) = \frac{p_{ij}}{1 - p_{ii}}$$

Exemple 4.6. — On revient sur l'exemple 3; Partant de 0 la proba de toucher 6 est donc la proba $\mathbb{P}_0(X_T = 4) = 1/4$ et la probabilité d'aller dans la classe fermée $\{1, 2, 3\}$ est $3/4$. Le temps pour toucher 3 peut être $+\infty$ si je suis absorbé dans la classe fermée $\{4, 5, 6\}$. Donc la moyenne de ce temps est $+\infty$.

SECTION 5

Mesures Invariantes

On dit que la mesure positive λ est invariante (ou encore stationnaire) si $\lambda P = \lambda$.

Proposition 4.19. — Si $(X_n)_{n \in \mathbb{N}}$ est Markov (λ, P) et λ est invariante, alors la loi de X_n est constante et vaut λ . En outre pour tout m , $(X_{m+n}, n \geq 0)$ est Markov (λ, P)

Démonstration. — En effet $X_n \sim \lambda P^n = \lambda$ et

$$\mathbb{P}((X_{m+n}, n \geq 0) \in A \mid \mathcal{F}_n) = \mathbb{P}(\{(X_n, n \geq 0) \in A\} \circ \theta^m) = \mathbb{P}_{X_m}(\{(X_n, n \geq 0) \in A\})$$

et on prend les espérances. □

Les limites des lois de X_n sont automatiquement des probabilités invariantes sur un espace fini.

Proposition 4.20. — On suppose I fini et que pour un $i_0 \in I$ on ait

$$\forall j, \quad p_{i_0,j}^{(n)} \rightarrow \pi_j$$

Alors π est une probabilité invariante.

Démonstration. — Par convergence dominée

$$\pi_j = \lim p_{i_0,j}^{(n)} = \lim p_{i_0,j}^{(n+1)} = \lim \sum_k p_{i_0,k}^{(n)} p_{kj} = \sum_k \pi_k p_{kj} = (\pi P)_j$$

□

On fixe i et on considère $T = T_i = \inf \{n \geq 1 : X_n = i\}$ et

$$\gamma(j) = \gamma^i(j) = \mathbb{E}_i \left[\sum_{0 \leq k < T_i} \mathbf{1}_{(X_k=j)} \right]$$

Proposition 4.21. — On suppose la chaîne irréductible récurrente. Alors

1. La mesure γ^i est invariante et de support I , $\forall j, \gamma^i(j) > 0$
2. Toute mesure invariante est un multiple de γ^i .

En conséquence, la chaîne admet une probabilité invariante ssi γ^i est de masse finie, et alors l'unique mesure de probabilité invariante est $\pi(j) = \gamma^i(j)/\mathbb{E}_i[T_i]$. En particulier,

$$\pi(i) = \frac{1}{\mathbb{E}_i[T_i]}$$

. On dit que la chaîne est positive récurrente

Démonstration. — 1) Soit $f : I \rightarrow \mathbb{R}$ positive. Alors, comme $T_i \leq \infty$ et $X_{T_i} = i$, on a

$$\begin{aligned} \gamma(f) &= \sum_j f(j) \gamma(j) = \mathbb{E}_i \left[\sum_{0 \leq k < T_i} f(X_k) \right] = \mathbb{E}_i \left[\sum_{1 \leq k \leq T_i} f(X_k) \right] \\ &= \mathbb{E}_i \left[\sum_{1 \leq k < +\infty} f(X_k) \mathbf{1}_{(k \leq T_i)} \right] \\ &= \sum_{1 \leq k < +\infty} \mathbb{E}_i [f(X_k) \mathbf{1}_{(k \leq T_i)}] \end{aligned}$$

Or $\{k \leq T - i\} \in \mathcal{F}_{k-1}$ donc

$$\mathbb{E}_i [f(X_k) \mathbf{1}_{(k \leq T_i)}] = \mathbb{E}_i [\mathbb{E}[f(X_k) | \mathcal{F}_{k-1}] \mathbf{1}_{(k \leq T_i)}] = \mathbb{E}_i [Pf(X_{k-1}) \mathbf{1}_{(k \leq T_i)}]$$

et donc

$$\begin{aligned} & \gamma(f) \sum_{1 \leq k < +\infty} \mathbb{E}_i [Pf(X_{k-1}) \mathbf{1}_{(k \leq T_i)}] \\ &= \sum_{0 \leq k < +\infty} \mathbb{E}_i [Pf(X_m) \mathbf{1}_{(m < T_i)}] = \gamma(Pf) \end{aligned}$$

ce qui prouve que $\gamma = \gamma P$.

2) En décomposant suivant les temps de retour successifs en i , on obtient

$$\mathbb{E}_i \left[\sum_n f(X_n) \right] = \sum_k \gamma^i(f) = +\infty$$

si $\gamma^i(f) > 0$ et 0 sinon. Or pour $f(x) = \mathbf{1}_{(x=j)}$ on sait que $\sum_n f(X_n) = +\infty$ ps, donc $\gamma^i(f) = \gamma^i(j) > 0$.

3) la preuve est pénible et demande des arguments de martingale listés ci-dessous \square

Remarque. — Si une chaîne rec irr admet une proba invariante π , alors elle est positive récurrente, p_i et γ^i sont proportionnelles, donc γ^i est de masse finie.

Corollaire 4.22. — Soit X une chaîne de markov irréductible sur un espace d'états fini. Alors la chaîne est positive récurrente et admet une unique probabilité invariante.

Lemme 4.23. — Soit une chaîne irréductible récurrente. Si $Pf = f$ et $f \geq 0$, alors f est constante.

Démonstration. — $f(X_n)$ est une martingale positive, donc converge ps donc ps il existe $c = c(\omega)$ et n_0 tq pour tout $n \geq n_0$, $f(X_n) = c$. Or ps X_n visite tous les points infiniment souvent, donc $\forall i \in I$, $f(i) = c(\omega)$ ce qui prouve que f est constante. \square

Lemme 4.24. — Soit μ une mesure invariante qui charge tout l'espace. Alors $Q(y, x) = \frac{\mu(x)}{\mu(y)} p(x, y)$ définit une matrice stochastique. Soit Y_n une chaîne de Markov de matrice Q . Alors si P est récurrente irréductible, il en est de même de Q . Si π est une mesure invariante pour P et $f = \pi/\mu$, alors $Qf = f$.

Démonstration. — L'irréductibilité de Y est triviale, il suffit de trouver un chemin de y à x dans le graphe de P .

On a évidemment $q^{(n)}(y, x) = \frac{\mu(x)}{\mu(y)} p^{(n)}(x, y)$ et donc $\sum_n q^{(n)}(x, x) = +\infty$ ssi $\sum_n p_{xx}^{(n)} = +\infty$. \square

Fin de la preuve de la proposition On a si $f = \pi/\mu$ qui vérifie $Qf = f$ donc Q étant irréductible récurrente, f est constante, ie les mesures invariantes sont toutes proportionnelles.

5.1 Théorème ergodique

Théorème 4.25. — Soit X une chaîne irréductible positive récurrente. Alors, pour toute fonction f positive ou bornée, presque sûrement

$$\frac{1}{n} \sum_{k=1}^n f(X_k) \rightarrow \pi(f)$$

Démonstration. — C'est le renewal reward theorem avec pour temps de régénération T_i avec i fixé. On a ps

$$\frac{1}{n} \sum_{k=1}^n f(X_k) \rightarrow \gamma^i(f)/\mathbb{E}_i[T_i] = \pi(f)$$

□

Exemple 4.7. — Pour l'exemple de l'introduction, la proportion de temps passé en temps long dans l'état 2, en partant de 1, est de $3/8$. Elle vaut $\pi(2) = 3/8$ avec π unique proba invariante de la chaîne réduite à $\{1, 2, 3\}$.

5.2 Mesures réversibles Il est parfois difficile de calculer la mesure de proba invariante même quand on sait qu'elle existe et est unique.

Définition 4.6. — On dit que la mesure π est réversible si elle vérifie les équations de bilan détaillé

$$\pi(x)p_{xy} = \pi(y)p_{yx}$$

Lemme 4.26. — Une mesure réversible est invariante.

Démonstration. —

$$\pi P(j) = \sum_i \pi(i)p_{ij} = \sum_i \pi(j)p_{ji} = \pi(j) \sum_i p_{ji} = \pi(j)$$

□

Remarque. — Si la matrice de transition est symétrique, alors la mesure $\pi(i) = 1$ est réversible donc invariante et si on est sur un espace fini, cela entraîne que l'unique mesure invariante est la loi uniforme.

Remarque. — On montre facilement que si $0 \leq N$, et $Y_N = X_{N-n}, 0 \leq n \leq N$ est la chaîne retournée dans le temps (à l'instant N), si $X_0 \sim \pi$ loi réversible, alors Y est une chaîne de Markov (π, P) en montrant que

$$\mathbb{P}(Y_0 = i_0, \dots, Y_N = i_N) = \mathbb{P}(X_0 = i_0, \dots, X_N = i_N)$$

SECTION 6

Convergence vers l'équilibre

La matrice de transition $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ vérifie $P^{2n} = I$ et $P^{2n+1} = P$. En conséquence $p_{ij}^{(n)}$ ne converge pas

Définition 4.7. — On dit que l'état i est apériodique si $\text{pgcd}\{n \geq 1, p_{ii}^{(n)} > 0\} = 1$.

Proposition 4.27. — On considère une chaîne irréductible. Si i est apériodique, alors tous les états sont apériodiques. En outre, si I est fini, il existe n_0 tel que pour tout $n \geq n_0$, et tous j, k $p^{(n)}(j, k) > 0$.

Démonstration. — On pose $S = \{n \geq 1, p_{ii}^{(n)} > 0\}$ Alors S est un semi groupe de $(\mathbb{N}, +)$ i.e. $S + S \subset S$ et $\text{pgcd}(S) = 1$, donc si $I = S - S$, alors I est un idéal de \mathbb{Z} donc de la forme $d\mathbb{Z}$ et comme $\text{pgcd}(S) = 1$ on a $d = 1$ donc $1 \in S - S$: il existe $n_0 \in S$ tel que $n_0 + 1 \in S$ Alors si $k \geq n_0^2 + n_0$, quand on fait la division de k par n_0 , $k = qn_0 + r$, on a $q \geq n_0$ donc $k = r(n_0 + 1) + (q - r)n_0 \in S$. Comme $i \rightarrow j$ on en déduit qu'il existe $n(i, j)$ tel que pour tout $n \geq n(i, j)$ $p_{i,j}^{(n)} > 0$. \square

Théorème 4.28. — Soit une chaîne de Markov irréductible positive récurrente de proba invariante π et apériodique. Alors, pour toute loi initiale

$$\mathbb{P}(X_n = i) \rightarrow \pi(i)$$

en particulier pour tout i ,

$$p_{ij}^{(n)} \rightarrow \pi(j)$$

Démonstration. — Preuve par couplage. Dans le cas d'un espace d'états fini on peut utiliser Perron-Frobenius. \square

SECTION 7

Le processus de Galton Watson

5

Martingales

Les martingales sont les invariants forts des systèmes aléatoires. Elles sont le pendant des lois de conservation des systèmes physiques.

SECTION 1

Filtrations et Martingales

1.1 définition

Définition 5.1. — *Un espace probabilisé filtré $(\Omega, \mathcal{F}, (\mathcal{F}_n)_n, \mathbb{P})$ est la donnée*

- *d'un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$;*
- *d'une filtration $(\mathcal{F}_n)_n$, c'est à dire qu'une famille croissante de sous tribus de \mathcal{F} :*

$$\mathcal{F}_0 \subset \cdots \subset \mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \cdots \subset \mathcal{F}.$$

La tribu \mathcal{F}_n représente la quantité d'information connue à l'instant n . Habituellement $\mathcal{F}_n = \mathcal{F}_n^W = \sigma(W_0, \dots, W_n)$ est la *filtration naturelle* d'un processus W . On note également $\mathcal{F}_\infty = \vee_n \mathcal{F}_n$ la plus petite tribu qui contient toutes les tribus \mathcal{F}_n .

Définition 5.2. — *Un processus $(X_n)_n$ est adapté (à la filtration $(\mathcal{F}_n)_n$ si pour tout n , la variable aléatoire X_n est \mathcal{F}_n -mesurable. Un processus $(X_n)_n$ est dit intégrable si pour tout n la variable aléatoire X_n est intégrable : $\forall n, \mathbb{E}[|X_n|] < +\infty$.*

Définition 5.3. — *Une martingale est un processus adapté intégrable $(X_n)_n$ tel que*

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

Une sousmartingale est un processus adapté intégrable $(X_n)_n$ tel que

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] \geq X_n \quad (\forall n).$$

Une surmartingale est un processus adapté intégrable $(X_n)_n$ tel que

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] \leq X_n \quad (\forall n).$$

Remarque. — — Une surmartingale décroît en moyenne.

- $(X_n)_n$ est une surmartingalessi $(-X_n)_n$ est une sousmartingale.
- Si $(X_n)_n$ est une sousmartingale, alors $(X_n - X_0)_n$ est une sousmartingale issue de 0.

1.2 Somme de variables aléatoires indépendantes centrées Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes, intégrables, de moyenne nulle : $\mathbb{E}[X_k] = 0$. On pose

$$S_0 = 0, \quad S_n = X_1 + \cdots + X_n, \quad \mathcal{F}_n = \sigma(X_1, \dots, X_n), \quad \mathcal{F}_0 = \{\emptyset, \Omega\}.$$

Le processus $(S_n)_n$ est évidemment adapté et intégrable (une somme de variables aléatoires intégrables est intégrable), et

$$\begin{aligned} \mathbb{E}[S_{n+1} - S_n | \mathcal{F}_n] &= \mathbb{E}[X_{n+1} | \mathcal{F}_n] \\ &= \mathbb{E}[X_{n+1}] \quad (\text{car } X_{n+1} \text{ est indépendante de } \mathcal{F}_n) \\ &= 0. \end{aligned}$$

Une question intéressante est de savoir quand la suite S_n converge.

Si $(X_i)_{i \geq 1}$ IID et $\mathbb{P}(X_i = \pm 1) = \frac{1}{2}$ c'est la marche aléatoire simple sur \mathbb{Z} et S_n est une martingale.

Si X_i est IID et $\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = -1)$ alors S_n est la marche aléatoire biaisée et c'est $M_n = S_n - np$ qui est une martingale.

1.3 Produit de variables aléatoires indépendantes intégrables de moyenne 1 Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes, intégrables, de moyenne 1 : $\mathbb{E}[X_k] = 1$. On pose

$$M_0 = 1, \quad M_n = X_1 \cdots X_n, \quad \mathcal{F}_n = \sigma(X_1, \dots, X_n), \quad \mathcal{F}_0 = \{\emptyset, \Omega\}.$$

Alors le processus $(M_n)_n$ est évidemment adapté, intégrable (le produit de variables aléatoires indépendantes intégrables est intégrable) et

$$\begin{aligned} \mathbb{E}[M_{n+1} \mathcal{F}_n] &= \mathbb{E}[M_n X_{n+1} \mathcal{F}_n] \\ &= M_n \mathbb{E}[X_{n+1} | \mathcal{F}_n] \quad (\text{car } M_n \text{ est } \mathcal{F}_n \text{ mesurable}) \\ &= M_n \mathbb{E}[X_{n+1}] \quad (\text{car } X_{n+1} \text{ est indépendante de } \mathcal{F}_n) \\ &= M_n. \end{aligned}$$

Dans le cas positif, le théorème de convergence des surmartingales entraîne que la martingale positive M_n converge presque sûrement vers une va intégrable M_∞ . Nous verrons sous quelles conditions on a $\mathbb{E}[M_\infty] = 1$.

En mathématiques financières, X_n positive représente le changement de valeur (multiplicatif ± 10 pour cent par exemple) en un période. On prend $X = e^N/\mathbb{E}[e^N]$ avec $N \sim \mathcal{N}(0, 1)$ our un modèle continu, mais le modèle le plus utilisé est le modèle binomial pour lequel $X = (1+a)e^{-v}$ avec la proba p et $X = (1+a)^{-1}e^{-v}$ avec la proba $1-p$ avec donc $e^v = p(1+a) + (1-p)/(1+a)$. Alors M_n est binomiale:

$$\mathbb{P}(M_n = X_0(1+a)^m e^{-vn}) = \binom{n}{\frac{m+n}{2}} p^{(m+n)/2} (1-p)^{(n-m)/2}$$

Pour la marche aléatoire biaisée $M_n = \rho^{S_n}$ avec $\rho = \frac{1-p}{p}$ est une martingale de type multiplicatif car $\xi_1 = \rho^{X_1}$ vérifie $\mathbb{E}[\xi_1] = p\rho + (1-p)\frac{1}{\rho} = 1$.

1.4 Accumulation de l'information sur une variable aléatoire Soit $(\mathcal{F}_n)_n$ une filtration et $\xi \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ une variable aléatoire intégrable. On pose $X_n = \mathbb{E}[\xi | \mathcal{F}_n]$. C'est évidemment un processus adapté et intégrable. En outre, grâce à la propriété de projection de l'espérance conditionnelle (Proposition 3.2),

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

Dans ce cas, nous serons capable de montrer que $M_n \rightarrow M_\infty = \mathbb{E}[\xi | \mathcal{F}_\infty]$, c'est à dire que l'information sur ξ connue à l'instant n converge vers l'information maximale sur ξ .

On verra que toute martingale UI est de ce type

1.5 Le processus de GW Soit $(\xi_i^{(k)}, i \geq 1, k \geq 1)$ iid de même loi que N variable aléatoire entière dont la loi est appelée loi de reproduction : $\xi_i^{(k)}$ est le nombre d'enfants de la personne numéro i de la génération k . Alors le nombre d'individus à l'instant $n+1$ est

$$Z_{n+1} = \sum_{i=1}^{Z_n} \xi_i^{(n+1)}$$

On montre que $W_n = m^{-n} Z^n$ est une martingale avec $m = \mathbb{E}[N] <= \infty$ nombre d'enfants moyen et $\mathcal{F}_n = \sigma(\xi_i^{(k)}, i \geq 1, k \leq n)$.

1.6 Jeu équilibré et stratégie de mise

Définition 5.4. — Un processus $(C_n)_{n \geq 1}$ est prévisible si pour tout $n \geq 1$, la variable aléatoire C_n est \mathcal{F}_{n-1} mesurable. Si $(X_n)_{n \in \mathbb{N}}$ est une sous(sur)martingale,

alors le processus transformé $(Y_n = (C.X)_n)_{n \in \mathbb{N}}$ est l'intégrale stochastique discrète:

$$Y_0 = 0, \quad Y_n = \sum_{k=1}^n C_k (X_k - X_{k-1}) \quad (n \geq 1).$$

La martingale X représente un jeu équilibré, et C la stratégie de mise : on mise C_n à l'instant n , C_n ne dépendant que de l'information à l'instant précédent $n-1$. Alors Y_n représente le gain à l'instant n . Le théorème suivant dit que si le jeu est équilibré, alors il n'y a pas de stratégie de mise gagnante (en moyenne).

Théorème 5.1. — (a) Soit $(C_n)_{n \geq 1}$ un processus prévisible positif borné (pour tout $n \geq 1$, il existe $K_n < +\infty$ telle que presque sûrement $|C_n(\omega)| \leq K_n$). Soit $(X_n)_{n \in \mathbb{N}}$ une surmartingale. Alors $C.X$ est une surmartingale nulle en zéro.
(b) Soit $(C_n)_{n \geq 12}$ un processus prévisible borné et X une martingale. Alors, $C.X$ est une martingale nulle en zéro.

Démonstration. — Soit $Y_n = (C.X)_n$. Par construction, comme les variables aléatoires C_k sont bornées et les variables aléatoires X_k intégrables, le processus Y est intégrable ; de même il est évidemment adapté. En outre, comme C_{n+1} est \mathcal{F}_n mesurable,

$$\mathbb{E}[Y_{n+1} - Y_n | \mathcal{F}_n] = \mathbb{E}[C_{n+1}(X_{n+1} - X_n) | \mathcal{F}_n] = C_{n+1}\mathbb{E}[(X_{n+1} - X_n) | \mathcal{F}_n].$$

- (a) On a $\mathbb{E}[(X_{n+1} - X_n) | \mathcal{F}_n] \leq 0$ et $C_{n+1} \geq 0$, donc $\mathbb{E}[Y_{n+1} - Y_n | \mathcal{F}_n] \leq 0$.
- (b) On a $\mathbb{E}[(X_{n+1} - X_n) | \mathcal{F}_n] = 0$, donc $\mathbb{E}[Y_{n+1} - Y_n | \mathcal{F}_n] = 0$. \square

Examinons la stratégie de doublement dans un jeu de pile ou face équilibré. Au début je mise 1, si je perds je mise 2, si je perds encore je mise 4, etc On a $C_n = 2^k$ si $X_{n-k} = -1 = \dots = X_{n-1}$ j'ai perdu les k fois précédentes, et $C_n = 1$ sinon. On a toujours $\mathbb{E}[Y_n] = 0$.

Mais on dit qu'en fait on s'arrête de jouer la première fois que l'on gagne et que le gain est $1 = 2^{k+1} - (1 + 2 + \dots + 2^k)$. C'est vrai mais là on ne s'arrête pas à un instant fixé d'avance.

SECTION 2

Temps d'arrêt et Théorèmes d'Arrêt

Définition 5.5. — Un temps d'arrêt est une application $T : \Omega \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ telle que :

$$\forall n, \quad \{T \leq n\} \in \mathcal{F}_n.$$

Remarque. — T est un temps d'arrêt si pour tout n , $\{T = n\} \in \mathcal{F}_n$. En effet $\{T = 0\} = \{T \leq 0\}$, $\{T = n\} = \{T \leq n\} \setminus \{T \leq n-1\}$ et $\{T \leq n\} = \bigcup_{k=1}^n \{T = k\}$.

Intuitivement, la décision d'arrêter à l'instant n est prise uniquement en fonction des informations connues à l'instant n .

Exemple 5.1. — Étant donnée une suite adaptée $(X_n)_{n \in \mathbb{N}}$, sont des temps d'arrêt :

$$\begin{aligned} T &= \inf \{n \geq 0 : X_n \in A\} \quad (\inf \emptyset = +\infty) \quad \text{temps d'entrée dans } A \\ T &= \inf \{n \geq 1 : X_n \in A\} \quad \text{temps de passage dans } A \\ T &= \inf \{n \geq 1 : |X_n - X_{n-1}| \geq 2\} \end{aligned}$$

Définition 5.6. — Étant donnés un processus $(X_n)_{n \in \mathbb{N}}$ et un temps d'arrêt T , le processus arrêté en T X^T est défini par : $X_n^T = X_{T \wedge n}$.

Si $(X_n)_{n \in \mathbb{N}}$ est adapté, alors pour tout n , $X_{T \wedge n}$ est $\mathcal{F}_{T \wedge n}$ mesurable, et donc \mathcal{F}_n mesurable : le processus arrêté est adapté.

Proposition 5.2. — Une sousmartingale arrêtée est une sousmartingale.

Démonstration. — Soit donc $(X_n)_{n \in \mathbb{N}}$ une sousmartingale, T un temps d'arrêt et le processus $C_n = \mathbf{1}_{(T \geq n)}$. Alors,

$$(C.X)_n = \sum_{k=1}^n \mathbf{1}_{(T \leq k)} (X_k - X_{k-1}) = X_{T \wedge n} - X_0$$

c'est à dire $C.X = X^T - X_0$. Le processus $(C_n)_{n \geq 1}$ est positif borné et prévisible car $\{T \geq n\}^C = \{T \leq n-1\} \in \mathcal{F}_{n-1}$. Donc $C.X$, et par conséquence X^T , est une sousmartingale. \square

Théorème 5.3 (Théorème de convergence des sousmartingales)

Soit $(X_n)_{n \in \mathbb{N}}$ une sousmartingale telle que $\sup_n \mathbb{E}[X_n^+] < +\infty$. Alors, $(X_n)_{n \in \mathbb{N}}$ est bornée dans L^1 , i.e. $\sup_n \mathbb{E}[|X_n|] < +\infty$, et X_n converge presque sûrement vers une variable aléatoire X_∞ qui est intégrable.

Corollaire 5.4. — Une surmatingale positive converge presque sûrement vers une variable aléatoire intégrable.

Corollaire 5.5. — Soit $(X_n)_{n \in \mathbb{N}}$ une sousmartingale bornée dans L^p avec $p > 1$. Alors $\sup_n |X_n|$ est dans L^p et X_n converge presque sûrement et dans L^p vers une variable aléatoire X_∞ .

Remarque. — Ce résultat est faux pour $p = 1$. On peut montrer, à l'aide du théorème de Kakutani sur les martingales produits, que si $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires indépendantes positives de même loi non dégénérée,

telle que $\exp X_1 = 1$, alors $M_n = X_1 \dots X_n$ est une martingale, bornée dans L^1 car $\mathbb{E}[M_n] = 1$, et M_n converge presque sûrement vers $M_\infty = 0$.

Théorème 5.6 (Théorème de convergence des martingales)

Soit $(X_n)_{n \in \mathbb{N}}$ une martingale. Sont équivalentes,

- (i) $(X_n)_{n \in \mathbb{N}}$ est Uniformément intégrable.
- (ii) $(X_n)_{n \in \mathbb{N}}$ converge dans L^1 .
- (iii) Il existe une variable aléatoire intégrable X_∞ , \mathcal{F}_∞ mesurable, telle que $X_n = \mathbb{E}[X_\infty | \mathcal{F}_n]$.

Dans ces conditions X_n converge aussi vers X_∞ presque sûrement et dans L^1 . On dit que la martingale est fermée, ou régulièrre.

On obtient des énoncés semblables pour les sur et sous martingales UI.

Le Théorème d'arrêt suivant signifie que pour une martingale Uniformément intégrable, on peut remplacer les temps fixes par des temps d'arrêt quelconques, même infinis.

Théorème 5.7. — Soit $(X_n)_{n \in \mathbb{N}}$ une martingale Uniformément intégrable. Pour tout temps d'arrêt T on définit: $M_T = M_n$ si $T(\omega) = n$, $M_T = M_\infty$, si $T(\omega) = +\infty$. Alors,

- (i) $X_T = \mathbb{E}[X_\infty | \mathcal{F}_T]$
- (ii) $\mathbb{E}[X_0] = \mathbb{E}[X_T]$.

Kolmogorov a établi des conditions nécessaires et suffisantes pour la convergence d'une série de variables aléatoires indépendantes $\sum_n X_n$.

Théorème 5.8 (Théorème des trois séries). — Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes. Alors $\sum_n X_n$ converge presque sûrement si et seulement si il existe une constante $K > 0$ (et alors ce sera vrai pour toutes les constantes $K > 0$), telle que

- (i) $\sum_n \mathbb{P}(|X_n| > K) < +\infty$
- (ii) $\sum_n \mathbb{E}[X_n^K]$ converge
- (iii) $\sum_n \text{Var}(X_n^K) < +\infty$

avec la notation : $X^K(\omega) = X(\omega) \mathbf{1}_{(|X(\omega)| \leq K)}$.

6

Appendices

SECTION 1

Le lemme de Doob

Théorème 6.1. — Soit $f : E \rightarrow (F, \mathcal{F})$. On note $\sigma(f) = f^{-1}(\mathcal{F})$ la plus petite tribu sur E qui rend f mesurable. Soit $g : (E, \sigma(f)) \rightarrow (G, \mathcal{G})$ avec \mathcal{G} la tribu des boréliens de G , avec G espace métrique séparable. Alors il existe une fonction $h : (F, \mathcal{F}) \rightarrow (G, \mathcal{G})$ mesurable telle que $g = h \circ f$.

Démonstration. — On traite le cas $G = [0, 1]$ puis on utilise le fait qu'un espace métrique séparable est mesurablement isomorphe à une partie de $[0, 1]$ muni de sa tribu des boréliens.

Si $g = 1_A$ est une indicatrice, alors $A \in \sigma(f)$ donc $A = f^{-1}(B)$ avec $B \in \mathcal{F}$ et $g = 1_B \circ f$.

Par linéarité, si g est une fonction étagée $g = \sum_i \alpha_i 1_{A_i}$ avec $A_i = f^{-1}(B_i)$ alors, $g = h \circ f$ avec $h = \sum_i \alpha_i 1_{B_i}$.

Si g est $\sigma(f)$ mesurable positive, alors il existe une suite de fonctions croissantes étagées positives g_n convergeant vers g . Il existe h_n mesurable positive telle que $g_n = h_n \circ f$. L'ensemble D des x pour lesquels la suite h_n converge vers une limite finie est mesurable, et on pose $h(x) = (\limsup_n h_n(x)) \mathbf{1}_{(x \in D)}$. C'est bien une fonction mesurable positive et par construction $f(x) \in D$, car $h_n(f(x)) \rightarrow g(x)$, donc $g = h \circ f$. \square

Le théorème des classes monotones

Une introduction à la théorie de la mesure et à l'intégrale de Lebesgue a été vue en licence. Pour aller plus loin en probabilités nous aurons besoin de résultats complémentaires que nous regroupons dans ce chapitre. Pour faciliter le travail du lecteur nous avons rappelé dans l'appendice de ce chapitre les principaux résultats vus en licence concernant l'intégration et l'analyse de Fourier également très utile en probabilités.

2.1 Classes monotones et Tribus

2.1.1 Classes monotones Ω étant un ensemble, un ensemble de parties \mathcal{M} de Ω est appelée *classe monotone* si les propriétés suivantes sont vérifiées

- i) $\Omega \in \mathcal{M}$
- ii) si $A, B \in \mathcal{M}$ et si $B \subseteq A$ alors $A \setminus B \in \mathcal{M}$
- iii) si $A_j \in \mathcal{M}, \forall j \in \mathbb{N}$ et si $A_j \subseteq A_{j+1}$ alors $\bigcup_{j \in \mathbb{N}} A_j \in \mathcal{M}$.

Proposition 6.2. — Si \mathcal{E} est un ensemble de parties de Ω , l'intersection des classes monotones contenant \mathcal{E} est une classe monotone, notée $\mathcal{M}(\mathcal{E})$, appelée classe monotone engendrée par \mathcal{E} .

L'intérêt de cette notion est dans le résultat suivant

Théorème 6.3. — Si \mathcal{E} est une famille de parties de Ω stable par intersection finie, alors la classe monotone engendrée par \mathcal{E} coïncide avec la tribu engendrée par \mathcal{E} .

Démonstration. — Soit $\sigma(\mathcal{E})$ la tribu engendrée par \mathcal{E} . Une tribu étant une classe monotone on a l'inclusion $\mathcal{M}(\mathcal{E}) \subseteq \sigma(\mathcal{E})$.

Suivant l'Exercice.1 il suffit alors de montrer que $\mathcal{M}(\mathcal{E})$ est stable par intersection finie.

Définissons

$$\mathcal{M}_1 = \{A \in \mathcal{M}(\mathcal{E}), \forall B \in \mathcal{E}, A \cap B \in \mathcal{M}(\mathcal{E})\}$$

\mathcal{M}_1 est une classe monotone contenant \mathcal{E} donc $\mathcal{M}(\mathcal{E}) \subseteq \mathcal{M}_1$.

Ensuite définissons

$$\mathcal{M}_2 = \{A \in \mathcal{M}(\mathcal{E}), \forall B \in \mathcal{M}(\mathcal{E}), A \cap B \in \mathcal{M}(\mathcal{E})\}$$

On vérifie de même que \mathcal{M}_2 est une classe monotone. Pour conclure il suffit de montrer qu'elle contient \mathcal{E} . Soient $A \in \mathcal{E}$ et $B \in \mathcal{M}(\mathcal{E})$. Or on sait que $B \in \mathcal{M}_1$, on a donc $B \cap C \in \mathcal{M}(\mathcal{E})$ pour tout $C \in \mathcal{E}$. Il suffit de choisir $C = A$. \square

Le théorème des classes monotones admet une version fonctionnelle qu'il est utile de connaître et qui s'énonce comme suit.

Définition 6.1. — *Un ensemble \mathcal{F} de fonctions réelles sur Ω est dit stable par convergence monotone bornée si pour toute suite bornée et monotone $\{f_n\}$ éléments de \mathcal{F} la limite simple sur Ω , $f(x) = \lim_{n \rightarrow +\infty} f_n(x)$ est dans \mathcal{F} . Un espace vectoriel monotone est un espace vectoriel de fonctions réelles et bornées sur Ω , contenant les constantes et stable par convergence monotone bornée*

Il est facile de voir que si μ, ν sont deux probabilités, l'ensemble des fonctions mesurables bornées f telles que $\int f d\mu = \int f d\nu$ est un e.v. monotone.

Théorème 6.4 (classes monotones fonctionnelles)

Soit \mathcal{F} un espace vectoriel monotone. Si \mathcal{C} est une partie de \mathcal{F} stable par multiplication et contenant les constantes alors \mathcal{F} contient toutes les fonctions bornées et $\sigma(\mathcal{C})$ -mesurables.

Démonstration. Introduisons \mathcal{F}_0 le plus petit sous-espace vectoriel de \mathcal{F} contenant \mathcal{C} et stable par convergence monotone bornée.

Montrons que \mathcal{F}_0 est stable par multiplication. Soient $f \in \mathcal{C}$ et $g \in \mathcal{F}_0$. Montrons alors que $fg \in \mathcal{F}_0$. Il suffit de supposer que $f \geq 0$. L'argument qui suit est semblable à celui utilisé dans 6.3.

Introduisons :

$$\mathcal{F}_1 = \{g \in \mathcal{F}_0, fg \in \mathcal{F}_0\}.$$

\mathcal{F}_1 est un espace vectoriel, contenant \mathcal{C} (et les constantes) et stable par convergence monotone bornée. Donc $\mathcal{F}_0 \subseteq \mathcal{F}_1$ et on a bien montré que $fg \in \mathcal{F}_0$. Montrons maintenant que si $g, h \in \mathcal{F}_0$ alors $gh \in \mathcal{F}_0$.

Pour cela on introduit

$$\mathcal{F}_2 = \{h \in \mathcal{F}_0, fg \in \mathcal{F}_0\}.$$

\mathcal{F}_2 est un espace vectoriel contenant les constantes et quitte à faire une translation on peut supposer que $g \geq 0$. \mathcal{F}_2 est également stable par convergence monotone bornée et contient \mathcal{C} . On a donc $\mathcal{F}_0 \subseteq \mathcal{F}_2$.

Montrons maintenant que \mathcal{F}_0 coïncide avec l'ensemble des fonctions bornées et $\sigma(\mathcal{C})$ -mesurable sur Ω .

Soit $f \in \mathcal{F}_0$. Montrons que $|f| \in \mathcal{F}_0$. Il en résultera que si $f, g \in \mathcal{F}_0$ alors $\min(f, g) \in \mathcal{F}_0$ et $\max(f, g) \in \mathcal{F}_0$.

On obtient cette propriété en utilisant une suite croissante de polynômes $p_n(x)$ sur $[-1, 1]$ telle $\lim_{n \rightarrow +\infty} p_n(x) = |x|$. (Exercice)

On pose $\mathcal{A} = \{A \subseteq \Omega, 1_A \in \mathcal{F}_0\}$. On vérifie facilement grâce aux propriétés précédentes que \mathcal{A} est une tribu. Notons par $\mathcal{B}(\mathcal{A})$ l'espace des fonctions réelles, bornées et \mathcal{A} -mesurables. Soit $f \in \mathcal{B}(\mathcal{A})$. Si de plus $f \geq 0$, on sait (voir

appendice, (M6)) qu'il existe une suite croissante de fonctions \mathcal{A} -étagées telle que $\lim_{n \rightarrow +\infty} f_n = f$ simplement sur Ω . On en déduit donc que $f \in \mathcal{F}_0$. D'où il en résulte que $\mathcal{B}(\mathcal{A}) \subseteq \mathcal{F}_0$. Montrons que l'on a l'égalité. Soit $f \in \mathcal{F}_0$. \mathcal{F}_0 étant un espace vectoriel contenant les constantes, il suffit de supposer que $f \geq 0$ et de montrer que $[f \geq 1] \in \mathcal{A}$. Or la suite $g_n = (\min\{1, f\})^n$ est une suite monotone de \mathcal{F}_0 qui converge vers $1_{[f \geq 1]} \in \mathcal{F}_0$.

On en déduit que $\sigma(\mathcal{C}) \subseteq \mathcal{A}$ (car si $f \in \mathcal{C}$ alors $f \in \mathcal{F}_0$ est donc \mathcal{A} -mesurable). Par suite on a

$$\mathcal{B}(\sigma(\mathcal{C})) \subseteq \mathcal{B}(\mathcal{A}) = \mathcal{F}_0 \subseteq \mathcal{F}.$$

Donnons une application du Théorème précédent.

Corollaire 6.5. — *Soient E un espace métrique muni de sa tribu borélienne et μ, μ' deux mesures finies sur E . Alors $\mu = \mu'$ si et seulement si pour toute fonction réelle f continue et bornée sur E on a $\int f d\mu = \int f d\mu'$.*

Démonstration. — On utilise pour \mathcal{C} l'algèbre $\mathcal{C}_b(E)$ des fonctions continues et bornées sur E . Il est facile de montrer que la tribu engendrée par $\mathcal{C}_b(E)$ coïncide avec la tribu borélienne de E . Ce qui donne le résultat. \square

Remarque. — *Lorsque $E = \mathbb{R}^n$ on peut également utiliser $\mathcal{C} = S(\mathbb{R}^n)$, $S(\mathbb{R}^n)$ étant l'espace de Schwartz (voir appendice).*