COMPLÉMENTS DE PROBABILITÉS

Aldéric JOULIN

Polycopié de cours

GMM4 - INSA de Toulouse

Année universitaire 2013-2014

Table des matières

Chapitre 0

Espérance conditionnelle

0.1 Quelques éléments sur les tribus

Avant de rentrer dans le coeur du sujet, faisons quelques rappels sur la notion de tribu.

Définition 0.1.1. *Soit Ω un ensemble et $\mathcal{A}$ une famille de parties de Ω. On dit que $\mathcal{A}$ est une tribu (sur Ω) si:*

(i) $\Omega \in \mathcal{A}$.

(ii) pour tout ensemble $A \in \mathcal{A}$, on a $A^c \in \mathcal{A}$, où A^c désigne le complémentaire de A dans Ω (stabilité par passage au complémentaire).

(iii) pour toute famille $(A_n)_{n\in\mathbb{N}}$ de Ω satisfaisant $A_n \in \mathcal{A}$ pour tout $n \in \mathbb{N}$, alors $\cup_{n\in\mathbb{N}} A_n \in \mathcal{A}$ (stabilité par union dénombrable).

En particulier, on appelle tribu engendrée par une classe de parties $\mathcal{C}$ de Ω la plus petite tribu sur Ω contenant $\mathcal{C}$, c'est-à-dire l'intersection de toutes les tribus contenant $\mathcal{C}$. On la note $\sigma(\mathcal{C})$. Un exemple classique est la tribu borélienne sur $\Omega = \mathbb{R}^d$, où $\mathcal{C}$ désigne l'ensemble des ouverts (ou fermés) de $\mathbb{R}^d$.

Dans la suite de ce cours, l'espace de probabilité sur lequel les variables aléatoires (v.a.) sont définies est noté $(\Omega, \mathcal{A}, \mathbb{P})$.

Définition 0.1.2. *Étant donnée une v.a. X à valeurs dans un espace mesurable $(E, \mathcal{E})$, on appelle tribu engendrée par X, et on la note $\sigma(X)$, la sous-tribu de $\mathcal{A}$ engendrée par l'ensemble des images réciproques de X. Autrement dit,*

$$\begin{aligned} \sigma(X) &:= \sigma\left(X^{-1}(B) : B \in \mathcal{E}\right) \\ &= \sigma\left(\{\omega \in \Omega : X(\omega) \in B\}\,; B \in \mathcal{E}\right) \\ &= \sigma\left(\{X \in B\}\,; B \in \mathcal{E}\right), \end{aligned}$$

où la dernière égalité est simplement une notation. Il s'agit donc de la plus petite tribu sur Ω rendant X mesurable. De même, si $X_1, \ldots, X_n$ est une suite de v.a. à valeurs dans $(E, \mathcal{E})$, alors on définit la tribu engendrée par cette suite comme

$$\sigma\left(X_1, \ldots, X_n\right) := \sigma\left(\{\cap_{i=1}^n \{X_i \in B_i\}\,; B_1, \ldots, B_n \in \mathcal{E}\right).$$

À présent, nous allons énoncer un résultat très utile en pratique, le lemme de Doob, dû à un célèbre probabiliste américain du milieu de 20ème siècle. En particulier, ce lemme nous donne un critère simple pour établir la mesurabilité d'une v.a. Y par rapport à la tribu engendrée par une autre v.a. X, sous réserve qu'elles sont toutes 2 à valeurs dans $\mathbb{R}$ (ou éventuellement dans $\mathbb{R}^d$).

Lemme 0.1.3 (Cas réel)**.** *Étant donnée une variable aléatoire réelle (v.a.r.) X, une autre v.a.r. Y est $\sigma(X)$-mesurable si et seulement s'il exite une fonction borélienne $h : \mathbb{R} \to \mathbb{R}$ telle que $Y = h(X)$.*

Démonstration. Admise. ■

0.2 Espérance conditionnelle

Maintenant, nous sommes en mesure d'introduire la notion d'espérance conditionnelle. Dans la suite et sauf mention du contraire, toutes nos v.a. seront à valeurs réelles. De plus, lorsque l'on aura une égalité entre (ou convergence de) v.a.r., on sous-entendra qu'elles seront vérifiées presque sûrement, c'est-à-dire pour tout ω en dehors d'un sous-ensemble (inclus dans un événement) de Ω de probabilité 0.

Définition 0.2.1. *Soit $\mathcal{F}$ une sous-tribu de $\mathcal{A}$, engendrée par une partition $(A_n)_{n\in\mathbb{N}_*}$ de Ω, à savoir*

$$\begin{cases} \cup_{n\in\mathbb{N}_*} A_n &= \Omega \\ A_i \cap A_j &= \emptyset \quad \textit{pour} \quad i \neq j. \end{cases}$$

Notons $\mathcal{N} := \{n \in \mathbb{N}_* : \mathbb{P}(A_n) > 0\}$. On appelle probabilité conditionnelle d'un événement $A \in \mathcal{A}$ sachant la tribu $\mathcal{F}$ la quantité:

$$\mathbb{P}(A \mid \mathcal{F})(\omega) := \sum_{n\in\mathcal{N}} \mathbb{P}(A \mid A_n)\, 1_{A_n}(\omega), \quad \omega \in \Omega,$$

où $1_{A_n}(\omega)$ est l'indicatrice valant 1 si $\omega \in A_n$ et 0 sinon.

Notez que cette probabilité conditionnelle est une v.a.r., le conditionnement ayant lieu par rapport à une tribu et non un événement. De surcroît, elle est mesurable par rapport à la tribu $\mathcal{F}$ car c'est une fonction des A_n. Enfin, elle est constante sur les A_n et vaut alors $\mathbb{P}(A \mid A_n)$.

En passant à l'espérance dans la définition précédente et en permutant somme et espérance grâce au théorème de convergence monotone, on a

$$\begin{aligned} \mathbb{E}\left[\mathbb{P}(A \mid \mathcal{F})\right] &= \sum_{n\in\mathcal{N}} \mathbb{P}(A \mid A_n)\, \mathbb{E}\left[1_{A_n}\right] \\ &= \sum_{n\in\mathcal{N}} \mathbb{P}(A \mid A_n)\, \mathbb{P}(A_n) \\ &= \mathbb{P}(A), \end{aligned}$$

la dernière égalité étant obtenue par la FPT, la célèbre Formule des Probabilités Totales. Ainsi, la probabilité conditionnelle par rapport à une tribu vaut, en moyenne, la probabilité initiale.

En utilisant une reformulation avec les indicatrices, la définition de la probabilité conditionnelle devient:

$$\mathbb{E}\left[1_A \mid \mathcal{F}\right] := \sum_{n\in\mathcal{N}} \mathbb{E}\left[1_A \mid A_n\right] 1_{A_n}.$$

Bien entendu, la quantité intervenant dans le membre de droite doit être comprise comme

$$\mathbb{E}\left[1_A \mid A_n\right] := \mathbb{P}(A \mid A_n) = \frac{\mathbb{P}(A\cap A_n)}{\mathbb{P}(A_n)} = \frac{\mathbb{E}\left[1_{A\cap A_n}\right]}{\mathbb{P}(A_n)} = \frac{\mathbb{E}\left[1_A 1_{A_n}\right]}{\mathbb{P}(A_n)}.$$

On en déduit que cette probabilité conditionnelle est obtenue par "moyennisation" de la v.a. 1_A sur les événements engendrant $\mathcal{F}$. L'étape suivante est donc de remplacer cette indicatrice par une v.a. intégrable, comme on le fait dans le cours d'Analyse fonctionnelle pour construire l'intégrale de Lebesgue (passage des indicatrices aux fonctions étagées puis aux fonctions mesurables positives puis enfin aux fonctions intégrables).

Définition 0.2.2. *Soit X une v.a.r. intégrable et $\mathcal{F}$ une sous-tribu de $\mathcal{A}$ engendrée par une partition $(A_n)_{n\in\mathbb{N}_*}$ de Ω. De même que précédemment, notons $\mathcal{N} := \{n \in \mathbb{N}_* : \mathbb{P}(A_n) > 0\}$. On appelle espérance conditionnelle de X sachant la tribu $\mathcal{F}$ la v.a.r.*

$$\mathbb{E}\left[X \mid \mathcal{F}\right](\omega) := \sum_{n\in\mathcal{N}} \mathbb{E}\left[X \mid A_n\right] 1_{A_n}(\omega), \quad \omega\in\Omega,$$

où $\mathbb{E}\left[X \mid A_n\right]$ désigne le ratio $\mathbb{E}\left[X\, 1_{A_n}\right]/\mathbb{P}(A_n)$.

Remarquons plusieurs propriétés. Tout d'abord, l'espérance conditionnelle est clairement $\mathcal{F}$-mesurable, linéaire et est positive si X l'est. De plus, un exercice classique est de montrer que si X est de carré intégrable alors par l'inégalité de Cauchy-Schwarz l'espérance conditionnelle est aussi de carré integrable. De plus elle vaut en moyenne l'espérance de X. En effet, en permutant espérance et somme (ce qui est permis car en réalité la somme ne comporte qu'un seul terme à cause de la présence de l'indicatrice), on a

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{F}\right]\right] &= \sum_{n\in\mathcal{N}} \mathbb{E}\left[X \mid A_n\right] \mathbb{E}\left[1_{A_n}\right] \\
&= \sum_{n\in\mathcal{N}} \mathbb{E}\left[X\, 1_{A_n}\right] \\
&= \mathbb{E}\left[X \sum_{n\in\mathcal{N}} 1_{A_n}\right] \\
&= \mathbb{E}\left[X\right],
\end{aligned}$$

la famille $(A_n)_{n\in\mathbb{N}_*}$ formant une partition de Ω.
Par ailleurs, si X s'écrit $X = \sum_{i=1}^{p} \alpha_i \, 1_{A_i}$ avec $p \in \mathbb{N}_* \cup \{+\infty\}$ et où les α_i sont des constantes, alors X est mesurable par rapport à la tribu $\mathcal{F}$ et

$$\begin{aligned} \mathbb{E}\,[X \mid \mathcal{F}] &= \sum_{n\in\mathcal{N}} \sum_{i=1}^{p} \alpha_i \, \mathbb{E}\,[1_{A_i} \mid A_n] \; 1_{A_n} \\ &= \sum_{i=1}^{p} \alpha_i \, 1_{A_i} \\ &= X, \end{aligned}$$

en ayant utilisé encore une fois le fait que $(A_n)_{n\in\mathbb{N}_*}$ est une partition de Ω.
Si les tribus $\sigma(X)$ et $\mathcal{F}$ sont indépendantes, c'est-à-dire que tout événement $\sigma(X)$-mesurable est indépendant de tout événement $\mathcal{F}$-mesurable, alors on démontre facilement que

$$\mathbb{E}\,[X \mid \mathcal{F}] = \mathbb{E}\,[X]\,.$$

Enfin, terminons par la propriété clé satisfaite par l'espérance conditionnelle, qui est à la base de la prochaine définition et qu'on laissera en exercice: pour toute v.a.r. $\mathcal{F}$-mesurable et bornée Z on a

$$\mathbb{E}\,[X\,Z] \;=\; \mathbb{E}\,[\mathbb{E}\,[X \mid \mathcal{F}]\; Z]\,.$$

À présent, on va généraliser l'espérance conditionnelle à une tribu arbitraire, non nécessairement engendrée par une partition. En particulier, on va pouvoir conditionner par une v.a.r. générale. La définition que nous donnons ci-dessous est celle dans $L^2(\Omega, \mathcal{A}, \mathscr{P})$, que l'on notera $L^2(\mathcal{A})$ (on utilisera ce même raccourci de notation pour tous les espaces L^p) car elle nous permet d'exploiter la structure hilbertienne de $L^2(\mathcal{A})$. Pour une extension de cette définition à $L^1(\mathcal{A})$ (on rappelle que $L^2(\mathcal{A}) \subset L^1(\mathcal{A})$ par l'inégalité de Cauchy-Schwarz), la v.a.r. Z doit être supposée bornée afin que les deux espérances ci-dessous aient un sens.

Définition et théorème 0.2.3. *Soit X une v.a.r. dans $L^2(\mathcal{A})$ et soit $\mathcal{F}$ une sous-tribu quelconque de $\mathcal{A}$. Alors il existe une unique v.a.r. Y dans $L^2(\mathcal{F})$, i.e. dans $L^2(\mathcal{A})$ et $\mathcal{F}$-mesurable, telle que pour toute v.a.r. $Z \in L^2(\mathcal{F})$,*

$$\mathbb{E}\,[X\,Z] \;=\; \mathbb{E}\,[Y\,Z]\,.$$

On note $Y = \mathbb{E}\,[X \mid \mathcal{F}]$: c'est l'espérance conditionnelle de X sachant la tribu $\mathcal{F}$.

Démonstration. Posons $F = L^2(\mathcal{F})$, qui est un espace de Hilbert donc complet (i.e. toute suite de Cauchy converge pour la norme associée). De plus, F étant inclus dans $L^2(\mathcal{A})$, il est fermé (tout sous-espace complet d'un espace métrique, non nécessairement complet, est fermé). Ainsi, par le théorème du supplémentaire orthogonal d'un fermé dans un espace de Hilbert, on a $L^2(\mathcal{A}) = F \oplus F^\perp$, c'est-à-dire que l'espace $L^2(\mathcal{A})$ se décompose comme la somme directe de F et de son orthogonal $F^\perp$ défini par

$$F^\perp := \{U \in L^2(\mathcal{A}) : \mathbb{E}[U\,Z] = 0 \text{ pour tout } Z \in F\}.$$

En d'autres termes, toute v.a.r. de $L^2(\mathcal{A})$ se décompose de manière unique comme la somme de 2 v.a.r., l'une dans F et l'autre dans $F^\perp$. Lorsque l'on applique ce résultat à X on obtient l'existence et l'unicité d'une v.a.r. $Y \in F$ telle que $X = Y + (X - Y)$ où $X - Y$ est dans $F^\perp$. Ainsi, on en déduit que pour tout $Z \in F$,

$$\mathbb{E}[(X - Y)\, Z] = 0,$$

c'est-à-dire le résultat désiré. ∎

Cette définition généralise le cas de l'espérance conditionnelle par rapport à une tribu engendrée par une partition. Cependant, nous ne l'utiliserons pas en pratique car nous aurons d'autres outils à notre disposition pour calculer des espérances conditionnelles.

Comme l'espérance classique, l'espérance conditionnelle satisfait les propriétés de linéarité, de positivité, de convergences monotone et dominée, mais aussi les inégalités de Jensen et de Cauchy-Schwarz, de Hölder, etc...

Proposition 0.2.4. *Soit X une v.a.r. intégrable. Alors l'espérance conditionnelle vérifie les propriétés suivantes:*

(i) $\mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{F}\right]\right] = \mathbb{E}\left[X\right]$.

(ii) *Si les tribus $\sigma(X)$ et $\mathcal{F}$ sont indépendantes, alors $\mathbb{E}\left[X \mid \mathcal{F}\right] = \mathbb{E}\left[X\right]$.*

(iii) *Si X est $\mathcal{F}$-mesurable, alors $\mathbb{E}\left[X \mid \mathcal{F}\right] = X$.*

(iv) *Si $\mathcal{G} \subset \mathcal{F}$ alors $\mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{F}\right] \mid \mathcal{G}\right] = \mathbb{E}\left[X \mid \mathcal{G}\right]$.*

(v) *Si X est de carré intégrable, alors $\mathbb{E}\left[X \mid \mathcal{F}\right]$ est le projeté orthogonal de X sur le sous-espace fermé $L^2(\mathcal{F})$.*

Démonstration. (i): prendre $Z = 1$ qui est bien $\mathcal{F}$-mesurable et bornée.
(ii): Supposons les tribus $\sigma(X)$ et $\mathcal{F}$ indépendantes. Alors pour toute v.a.r. $\mathcal{F}$-mesurable et bornée Z,

$$\mathbb{E}\left[X\, Z\right] \;=\; \mathbb{E}\left[X\right]\, \mathbb{E}\left[Z\right].$$

Or $\mathbb{E}\left[X\right]$ étant constante, elle est bien $\mathcal{F}$-mesurable et donc on obtient le résultat par unicité de l'espérance conditionnelle.
(iii): Même raisonnement que précédemment.
(iv): Il suffit de l'écrire (un peu pénible tout de même).
(v): La v.a.r. $Y := \mathbb{E}[X \mid \mathcal{F}]$ est le projeté orthogonal de X sur le fermé $F := L^2(\mathcal{F})$ si et seulement si

$$Y \in F \quad \text{and} \quad \|X - Y\|_{L^2(\mathcal{A})} \;=\; \inf_{U \in F} \|X - U\|_{L^2(\mathcal{A})}.$$

On sait déjà que $Y \in F$. Ainsi, soit $Z \in F$ et posons $U := X - Y \in F^\perp$ (comme on l'a vu ci-dessus) et $V := Y - Z \in F$ comme différence de 2 éléments de F. Alors $\mathbb{E}\left[UV\right] = 0$ et l'on obtient:

$$\mathbb{E}\left[(X - Z)^2\right] \;=\; \mathbb{E}\left[(U + V)^2\right]$$

$$
\begin{aligned}
&= \mathbb{E}\left[U^2\right] + \mathbb{E}\left[V^2\right] + 2\,\mathbb{E}\left[UV\right] \\
&= \mathbb{E}\left[U^2\right] + \mathbb{E}\left[V^2\right],
\end{aligned}
$$

quantité qui est minimale pour le choix de $V = 0$, i.e. $Z = Y$. ■

Mentionnons que l'espérance conditionnelle par rapport à la tribu $\sigma(X)$ se note classiquement $\mathbb{E}[\,\cdot \mid X]$. Si l'on considère une v.a.r. Y intégrable, alors vu que $\mathbb{E}[Y \mid X]$ est $\sigma(X)$-mesurable, le lemme de Doob indique qu'il existe une fonction borélienne $h : \mathbb{R} \to \mathbb{R}$ telle que

$$\mathbb{E}[Y \mid X] = h(X).$$

Dans ce cas, on notera dans la suite du cours $\mathbb{E}[Y \mid X = x]$ la quantité $h(x)$, pour tout $x \in \mathbb{R}$. Attention, vous remarquerez que cette notation est quelque peu abusive car si X est continue alors $\mathbb{P}(X = x) = 0$ pour tout $x \in \mathbb{R}$ et l'on sait qu'il est interdit de conditionner par un événement de probabilité nulle.

Pour terminer ce chapitre, regardons en pratique comment calculer une espérance conditionnelle du type $\mathbb{E}[Y \mid X = x]$ lorsque l'on a des informations sur la loi jointe du couple (X, Y).

Cas discret: Tout d'abord donnons-nous deux v.a. discrètes X et Y, chacune à valeurs dans des espaces au plus dénombrables E et F respectivement. On suppose de plus que Y est intégrable et que $\mathbb{P}(X = x) > 0$ pour au moins un $x \in E$. Alors l'espérance conditionnelle $\mathbb{E}[Y \mid X = x]$ est bien définie et vaut

$$
\begin{aligned}
\mathbb{E}[Y \mid X = x] &= \sum_{y \in F} y\, \mathbb{P}(Y = y \mid X = x) \\
&= \sum_{y \in F} y\, \frac{\mathbb{P}(X = x; Y = y)}{\mathbb{P}(X = x)}.
\end{aligned}
$$

Pour obtenir l'expression explicite de l'espérance conditionnelle $\mathbb{E}[Y \mid X]$, il reste à remplacer x par la v.a. X dans la formule précédente, la tribu $\sigma(X)$ étant engendrée par la partition $\{X = x\}_{x \in E}$.

Cas continu: à présent, si X et Y sont deux v.a.r. de densités f_X et f_Y respectivement, alors en supposant de plus que Y est intégrable et que f_X vérifie $f_X(x) > 0$ pour au moins un $x \in \mathbb{R}$, on a

$$
\begin{aligned}
\mathbb{E}[Y \mid X = x] &= \int_{\mathbb{R}} y\, f_Y^{X=x}(y)\, dy \\
&= \int_{\mathbb{R}} y\, \frac{f_{X,Y}(x, y)}{f_X(x)}\, dy,
\end{aligned}
$$

où $f_Y^{X=x}$ désigne la densité conditionnelle de Y sachant $X = x$ et $f_{X,Y}$ est la densité jointe du couple (X, Y).

La aussi, pour en déduire l'expression explicite de l'espérance conditionnelle $\mathbb{E}[Y \mid X]$, on remplace x par la v.a. X dans le résultat obtenu et le tour est joué.

Chapitre 1

Vecteurs gaussiens

1.1 Définition et premières propriétés des vecteurs gaussiens

Si X est un vecteur aléatoire en dimension d (i.e. une v.a. à valeurs dans $\mathbb{R}^d$ vue comme un vecteur colonne) tel que chacune de ses coordonnées X_i, $i \in \{1, \dots, d\}$, soit dans $L^2(\mathcal{A})$, alors on définit dans la suite son vecteur espérance et sa matrice de covariance par

$$\mathbb{E}[X] = m := \begin{pmatrix} \mathbb{E}[X_1] \\ \cdot \\ \cdot \\ \cdot \\ \mathbb{E}[X_d] \end{pmatrix} \qquad \text{et} \qquad \mathrm{Var}(X) = \Gamma := \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,d} \\ \sigma_{2,1} & \sigma_{2,2} & \cdots & \sigma_{2,d} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \sigma_{d,1} & \sigma_{d,2} & \cdots & \sigma_{d,d} \end{pmatrix},$$

où $\sigma_{i,j}$ désigne la covariance entre les variables X_i et X_j:

$$\sigma_{i,j} \;=\; \mathrm{Cov}(X_i, X_j) \;:=\; \mathbb{E}[(X_i - m_i)\,(X_j - m_j)] \;=\; \mathbb{E}[X_i\, X_j] - m_i\, m_j,$$

où $m_i := \mathbb{E}[X_i]$ pour tout $i \in \{1, \dots, d\}$. En particulier, cette covariance est bien définie par l'inégalité de Cauchy-Schwarz. De plus, on remarque que la matrice réelle Γ est symétrique et qu'elle peut s'écrire comme

$$\Gamma \;=\; \mathbb{E}\left[(X - m)\,(X - m)^T\right] \;=\; \mathbb{E}\left[X\, X^T\right] - m\, m^T,$$

où le symbole T désigne la transposition et l'espérance d'une matrice est définie comme la matrice des espérances de chacun de ses éléments. On en déduit qu'elle est semi-définie positive au sens où pour tout $x \in \mathbb{R}^d$,

$$x^T \Gamma x \;=\; \mathbb{E}\left[\underbrace{x^T\,(X - m)}_{\in\mathbb{R}}\,\underbrace{(X - m)^T\, x}_{\in\mathbb{R}}\right] \;=\; \mathbb{E}\left[\left(x^T\,(X - m)\right)^2\right] \;\geq\; 0.$$

Dans la suite de ce chapitre, on supposera que les vecteurs gaussiens considérés sont non dégénérés, c'est-à-dire que $\det \Gamma \neq 0$ (elle est donc définie positive, i.e. $x^T \Gamma x > 0$ pour tout $x \neq 0$ dans $\mathbb{R}^d$).

Définition 1.1.1. *Soit X un vecteur aléatoire en dimension d de vecteur espérance m et de matrice de covariance Γ. Il est dit gaussien si la densité jointe est donnée par*

$$f_X(x) = \frac{1}{(2\pi)^{d/2}\sqrt{\det \Gamma}} \exp\left(-\frac{1}{2}(x-m)^T \Gamma^{-1}(x-m)\right), \quad x \in \mathbb{R}^d.$$

On note alors $X \sim \mathcal{N}_d(m, \Gamma)$ (et $\mathcal{N}(m, \Gamma)$ si $d = 1$).

Ainsi, comme dans le cas unidimensionnel, la donnée du vecteur espérance et de la matrice de covariance caractérise la loi d'un vecteur gaussien. La fonction caractéristique étant vue comme la transformée de Fourier de la densité, que l'on sait injective dans $L^1(\mathcal{A})$, on a la caractérisation suivante de la loi d'un vecteur gaussien.

Proposition 1.1.2. *Un vecteur aléatoire X est gaussien de vecteur espérance m et de matrice de covariance Γ si et seulement si sa fonction caractéristique est donnée par*

$$\phi_X(\theta) := \mathbb{E}[e^{i\theta^T X}] = \exp\left(i\theta^T m - \frac{1}{2}\,\theta^T \Gamma \theta\right), \quad \theta \in \mathbb{R}^d.$$

À présent, posons-nous la question suivante: un vecteur gaussien a-t-il toutes ses composantes unidimensionnelles gaussiennes ? Et réciproquement, suffit-il d'avoir toutes les coordonnées gaussiennes pour que le vecteur associé soit gaussien ? La proposition suivante permet de répondre à la question.

Proposition 1.1.3. *Un vecteur aléatoire X est gaussien si et seulement si $\theta^T X$ est une v.a. gaussienne unidimensionnelle pour tout $\theta \in \mathbb{R}^d$ différent du vecteur nul, i.e. toute combinaison linéaire non nulle de ses coordonnées est gaussienne.*

Démonstration. Notons respectivement m et Γ le vecteur espérance et la matrice de covariance du vecteur X. Supposons le gaussien. Alors pour tout $\theta \in \mathbb{R}^d$ différent du vecteur nul, la fonction caractéristique de la v.a.r. $\theta^T X$ s'écrit pour tout $u \in \mathbb{R}$:

$$\begin{aligned} \phi_{\theta^T X}(u) &= \mathbb{E}\left[e^{iu\,\theta^T X}\right] \\ &= \phi_X(u\theta) \\ &= \exp\left(iu\theta^T m - \frac{u^2}{2}\,\theta^T \Gamma \theta\right). \end{aligned}$$

Ainsi, on en déduit que pour tout $\theta \in \mathbb{R}^d$ non nul, la v.a.r. $\theta^T X$ suit une loi gaussienne d'espérance $\theta^T m$ et de variance $\theta^T \Gamma \theta$, qui est strictement positive car Γ est définie positive. Réciproquement, si $\theta^T X$ suit une loi gaussienne pour tout $\theta \in \mathbb{R}^d$ non nul, alors

$$\left\{ \begin{array}{lllllll} \mathbb{E}\left[\theta^T X\right] & = & \sum_{i=1}^d \theta_i\, \mathbb{E}[X_i] & = & \theta^T m; & & \\ \mathrm{Var}\left(\theta^T X\right) & = & \mathrm{Var}\left(\sum_{i=1}^d \theta_i\, X_i\right) & = & \sum_{i,j=1}^d \theta_i \theta_j\, \mathrm{Cov}(X_i, X_j) & = & \theta^T \Gamma \theta. \end{array} \right.$$

Alors on a pour tout $u \in \mathbb{R}$

$$\phi_{\theta^T X}(u) \;=\; \exp\left(iu\theta^T m - \frac{u^2}{2}\,\theta^T \Gamma \theta\right),$$

et en choisissant $u = 1$ on obtient que pour tout $\theta \in \mathbb{R}^d$ non nul,

$$\phi_X(\theta) \;=\; \exp\left(i\theta^T m - \frac{1}{2}\,\theta^T \Gamma \theta\right),$$

c'est-à-dire que $X \sim \mathcal{N}_d(m, \Gamma)$. ■

Ainsi, chacune des coordonnées d'un vecteur gaussien suit forcément une loi gaussienne. En revanche, la réciproque est fausse: il se peut que 2 variables aléatoires réelles Y et Z soient gaussiennes sans que le vecteur $(Y, Z)^T$ soit un vecteur gaussien. En effet, considérons une v.a.r. $Y \sim \mathcal{N}(0, 1)$, indépendante d'une variable ε dite de Rademacher, de loi donnée par

$$\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = \frac{1}{2}.$$

On peut montrerque $Z = \varepsilon Y$ est une v.a.r. gaussienne alors que le vecteur $(Y, Z)^T$ ne l'est pas. Ainsi, demander à ce qu'un vecteur aléatoire soit gaussien requiert plus que le caractère gaussien des coordonnées.

Le résultat qui suit est en quelque sorte un "miracle" du cadre gaussien, au sens où il suffit de démontrer la nullité des différentes covariances afin d'obtenir l'indépendance des coordonnées.

Proposition 1.1.4 (Miracle gaussien)**.** *Soit X un vecteur gaussien. Alors ses coordonnées sont indépendantes si et seulement si sa matrice de covariance est diagonale.*

Démonstration. Les coordonnées du vecteur gaussien X sont indépendantes si et seulement si sa densité f_X est le produit des densités des coordonnées. Ceci équivaut à dire que la forme quadratique $(x - m)^T \Gamma^{-1} (x - m)$ apparaissant dans f_X s'écrit de la forme $\sum_{i=1}^d \alpha_i\,(x_i - m_i)^2$. En d'autres termes, il n'y a aucun terme croisé du type $(x_i - m_i)(x_j - m_j)$ pour $i \neq j$, c'est-à-dire que la matrice Γ^{-1} (donc Γ) est diagonale. ■

1.2 Quelques autres propriétés en vrac

Donnons dans ce bref paragraphe quelques propriétés que l'on rencontre usuellement lorsque l'on s'attaque à un problème faisant intervenir des vecteurs gaussiens.

Proposition 1.2.1 (Transformation linéaire)**.** *Soit A une matrice inversible $d \times d$ et b un vecteur dans $\mathbb{R}^d$. Considérons le vecteur gaussien $X \sim \mathcal{N}_d(m, \Gamma)$. Alors le vecteur $AX + b$ est gaussien de vecteur espérance $Am + b$ et de matrice de covariance $A\Gamma A^T$.*
En particulier, si $X \sim \mathcal{N}_d(0, \sigma^2 I_d)$ et que la matrice A est orthogonale, i.e. $AA^T = I_d$, alors les vecteurs X et AX ont même loi.

Démonstration. Très simple, il suffit de faire les calculs. On remarquera au passage que la matrice $A\Gamma A^T$ est bien semi-définie positive, et même définie positive car A est supposée inversible. ■

Proposition 1.2.2. *Un vecteur aléatoire d-dimensionnel X est gaussien de vecteur espérance m et de matrice de covariance Γ si et seulement si X peut s'écrire $X = AU + m$, où $U \sim \mathcal{N}_d(0, I_d)$ et A est une matrice carrée $d \times d$ inversible et vérifiant $AA^T = \Gamma$.*

Démonstration. La matrice Γ est réelle symétrique donc diagonalisable dans une base orthonormale. De plus, étant définie positive, toutes ses valeurs propres λ_i sont strictement positives. Notons P la matrice de passage que l'on prend orthogonale, i.e. $PP^T = I_d$. Alors la matrice carrée $d \times d$ donnée par $A := PD$, où D est la matrice diagonale formée par les $\sqrt{\lambda_i}$, est inversible et satisfait $AA^T = \Gamma$. Enfin, la proposition précédente entraîne que $U := A^{-1}(X - m)$ est un vecteur gaussien centré et de matrice de covariance l'identité. ■

Proposition 1.2.3 (Projection gaussienne)**.** *Notons $X = (X_1, \ldots, X_d)^T$ et $Y = (Y_1, \cdots, Y_{d'})^T$ et supposons que le vecteur aléatoire $(X, Y)^T$ soit gaussien dans $\mathbb{R}^{d+d'}$. Alors la loi conditionnelle de Y sachant $X = x$ est elle-aussi gaussienne et il existe des constantes $a, b_1, \ldots, b_d \in \mathbb{R}^{d'}$ telles que*

$$\mathbb{E}[Y|X] = a + \sum_{i=1}^{d} b_i X_i.$$

De plus, le vecteur aléatoire $Y - \mathbb{E}[Y|X]$ est gaussien centré et indépendant de X.

Démonstration. Nous n'allons démontrer que le cas $d = d' = 1$, le cas général en étant une adaptation immédiate. Ainsi, soit $(X, Y)^T$ un vecteur gaussien dans $\mathbb{R}^2$. On montre tout d'abord que le vecteur $(X, Y - a - bX)^T$ est gaussien dans $\mathbb{R}^2$ pour tous $a, b \in \mathbb{R}$. Soit (a_0, b_0) l'unique couple de valeurs pour lesquelles la quantité $\mathbb{E}\left[(Y - a - bX)^2\right]$ est minimale (ce couple peut être calculé explicitement). En d'autres termes, $a_0 + b_0 X$ est le projeté orthogonal de Y sur l'espace $\mathrm{Vect}\{1, X\} \subset L^2(\mathcal{A})$. La v.a.r. gaussienne $\varepsilon_0 := Y - a_0 - b_0 X$ étant par conséquent dans l'espace $\mathrm{Vect}\{1, X\}^{\perp}$, on a alors que $\mathbb{E}[\varepsilon_0] = \mathbb{E}[\varepsilon_0 X] = 0$, c'est-à-dire que ε_0 est centrée et indépendante de X, le vecteur $(X, \varepsilon_0)^T$ étant gaussien dans $\mathbb{R}^2$ d'après ce qui précède. Il en résulte enfin que

$$\mathbb{E}[Y \mid X] = \mathbb{E}\left[\varepsilon_0 + a_0 + b_0 X \mid X\right] = a_0 + b_0 X,$$

d'où le résultat. ■

Proposition 1.2.4 (Théorème Central Limite multidimensionnel)**.** *Soit $(X_n)_{n \geq 1}$ une suite de vecteurs aléatoires i.i.d. à valeurs dans $\mathbb{R}^d$, et dont les coordonnées sont de carré intégrable. Notons $m := \mathbb{E}[X_1]$ et $\Gamma := \mathrm{Var}(X_1)$. Alors on a la convergence en loi suivante:*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - m) \underset{n \to \infty}{\longrightarrow} \mathcal{N}_d(0, \Gamma).$$

Démonstration. Admise car quelque peu similaire au cas de la dimension 1. ■

Chapitre 2

Chaînes de Markov

Cette partie est consacrée à l'étude de l'objet principal de ce cours. Les chaînes de Markov, qui constituent de nos jours un inépuisable domaine de recherche, sont très souvent utilisées pour modéliser des phénomènes aléatoires évoluant au cours du temps et comportant une forme de dépendance. Dans ce qui va suivre, nous ne verrons que des chaînes de Markov à valeurs dans un ensemble unidimensionnel au plus dénombrable. Mentionnons tout de même qu'il existe une théorie générale de ces chaînes à valeurs dans des espaces plus riches comme $\mathbb{R}^d$.

2.1 Introduction aux chaînes de Markov

Dans la suite, l'ensemble E est appelé espace des états et est supposé au plus dénombrable, c'est-à-dire fini ou dénombrable (en bijection avec $\mathbb{N}$).

Définition 2.1.1. *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. à valeurs dans E. Cette suite est une chaîne de Markov (homogène) si pour tous $n \in \mathbb{N}$ et $x_0, x_1, \ldots, x_{n-1}, x, y \in E$,*

(i) $\mathbb{P}\left(X_{n+1} = y \mid X_n = x; X_{n-1} = x_{n-1}; \ldots; X_0 = x_0\right) = \mathbb{P}\left(X_{n+1} = y \mid X_n = x\right).$

(ii) la probabilité $\mathbb{P}\left(X_{n+1} = y \mid X_n = x\right)$ ne dépend pas de n.

Dans ce cas on note $P(x, y) = \mathbb{P}\left(X_{n+1} = y \mid X_n = x\right)$: c'est la probabilité de transition de l'état x à l'état y.

En d'autres termes, si X_n représente le présent, les X_k le passé lorsque $k \in \{0, \ldots, n-1\}$, et le futur quand $k \geq n+1$, on dit que “le futur est indépendant du passé, conditionnellement au présent”. Par ailleurs, la loi de X_0 est appelée loi initiale de la chaîne.

Dans la suite de ce cours, la suite $(X_n)_{n\in\mathbb{N}}$ désigne une chaîne de Markov à valeurs dans E et de matrice de transition P. Notons que l'ensemble $P = \left(P(x, y)\right)_{x,y\in E}$ définit une matrice carrée de taille $\mathrm{Card}(E)$ (éventuellement infinie si E n'est pas de cardinal fini) dont chacun des éléments est un nombre compris entre 0 et 1 et vérifiant

$$\sum_{y\in E} P(x, y) = 1, \quad x \in E,$$

c'est-à-dire que lorsque l'on se fixe une ligne de la matrice, la somme sur les colonnes vaut 1. Une matrice vérifiant de telles propriétés est dite matrice stochastique. Dans la pratique, on représente la chaîne de Markov par un graphe orienté et valué dont les noeuds sont les états de E. Un noeud x est relié à un noeud y par une arête orientée (xy) si $P(x,y)>0$, et l'arête reçoit la valeur $P(x,y)$.

Proposition 2.1.2. *Notons μ la loi initiale de la chaîne. Alors pour tous $n,k\in\mathbb{N}$ et tous $x_0,\dots,x_{n+k}\in E$,*
(i) $\mathbb{P}\left(X_n=x_n;X_{n-1}=x_{n-1};\dots;X_0=x_0\right)=\mu(x_0)P(x_0,x_1)\dots P(x_{n-1},x_n)$.
(ii) $\mathbb{P}\left(X_{n+k}=x_{n+k};\dots;X_{n+1}=x_{n+1}\mid X_n=x_n;\dots;X_0=x_0\right)=P(x_n,x_{n+1})\dots P(x_{n+k-1},x_{n+k})$.

Démonstration. Démontrons seulement le (i), la deuxième égalité étant à peu près similaire. En utilisant la propriété de Markov, on a pour tout $n\in\mathbb{N}$ et tous $x_0,\dots,x_n\in E$,

$$\begin{aligned}
&\mathbb{P}\left(X_n=x_n;X_{n-1}=x_{n-1};\dots;X_0=x_0\right)\\
&= \mathbb{P}\left(X_n=x_n\mid X_{n-1}=x_{n-1};\dots;X_0=x_0\right)\mathbb{P}\left(X_{n-1}=x_{n-1};\dots;X_0=x_0\right)\\
&= P(x_{n-1},x_n)\,\mathbb{P}\left(X_{n-1}=x_{n-1};\dots;X_0=x_0\right)\\
&= P(x_{n-1},x_n)\,\mathbb{P}\left(X_{n-1}=x_{n-1}\mid X_{n-2}=x_{n-2};\dots;X_0=x_0\right)\mathbb{P}\left(X_{n-2}=x_{n-2};\dots;X_0=x_0\right)\\
&= P(x_{n-1},x_n)\,P(x_{n-2},x_{n-1})\,\mathbb{P}\left(X_{n-2}=x_{n-2};\dots;X_0=x_0\right)\\
&= \dots\\
&= P(x_{n-1},x_n)\,P(x_{n-2},x_{n-1})\dots P(x_1,x_2)\,\mathbb{P}\left(X_1=x_1;X_0=x_0\right)\\
&= P(x_{n-1},x_n)\,P(x_{n-2},x_{n-1})\dots P(x_1,x_2)\,\mathbb{P}\left(X_1=x_1\mid X_0=x_0\right)\mathbb{P}\left(X_0=x_0\right)\\
&= P(x_{n-1},x_n)\,P(x_{n-2},x_{n-1})\dots P(x_1,x_2)\,P(x_0,x_1)\,\mathbb{P}\left(X_0=x_0\right).
\end{aligned}$$

■

À présent, on définit par récurrence les puissances n-ièmes de la matrice de transition P: pour tout $n\in\mathbb{N}_*$,

$$P^n(x,y)=\sum_{z\in E}P(x,z)\,P^{n-1}(z,y)=\sum_{z\in E}P^{n-1}(x,z)\,P(z,y),\quad x,y\in E,$$

avec par convention $P^0(x,y)=1_{\{x=y\}}$. On montre facilement par récurrence que pour tout $n\in\mathbb{N}$, la matrice P^n est stochastique, c'est-à-dire que chaque élément appartient à l'intervalle $[0,1]$ et

$$\sum_{y\in E}P^n(x,y)=1,\quad x\in E.$$

L'intérêt d'introduire ces puissances successives de la matrice P réside dans le résultat suivant.

Proposition 2.1.3. *Étant donné $n\in\mathbb{N}$, la loi conditionnelle de X_n sachant X_0 est donnée par la matrice P^n, i.e.*

$$\mathbb{P}\left(X_n=y\mid X_0=x\right)=P^n(x,y),\quad x,y\in E.$$

On notera parfois dans la suite $\mathbb{P}_x$ et $\mathbb{E}_x$ respectivement la probabilité et l'espérance sachant $X_0=x$.

Démonstration. On a pour tous $x, y \in E$,

$$\begin{aligned}
\mathbb{P}_x(X_n = y) &= \sum_{x_1,\ldots,x_{n-1}\in E} \mathbb{P}_x(X_n = y; X_{n-1} = x_{n-1}; \ldots; X_1 = x_1) \\
&= \sum_{x_1,\ldots,x_{n-1}\in E} P(x, x_1)\, P(x_1, x_2) \ldots P(x_{n-1}, y) \\
&= \sum_{x_2,\ldots,x_{n-1}\in E} P^2(x, x_2) \ldots P(x_{n-1}, y) \\
&= \sum_{x_3,\ldots,x_{n-1}\in E} P^3(x, x_3) \ldots P(x_{n-1}, y) \\
&= \ldots \\
&= \sum_{x_{n-1}\in E} P^{n-1}(x, x_{n-1})\, P(x_{n-1}, y) \\
&= P^n(x, y).
\end{aligned}$$

■

Étant donnée une probabilité π sur E, on définit la quantité

$$\pi P(y) = \sum_{x\in E} \pi(x)\, P(x, y), \quad y \in E.$$

On vérifie que πP est bien une probabilité sur E, qui est reliée à la loi de la chaîne de la manière suivante.

Proposition 2.1.4. *Si X_0 suit la loi π alors pour tout $n \in \mathbb{N}$ la loi de X_n est donnée par πP^n.*

Démonstration. Pour tout $y \in E$, on a par la FPT:

$$\begin{aligned}
\mathbb{P}(X_n = y) &= \sum_{x\in E} \mathbb{P}(X_n = y \mid X_0 = x)\, \mathbb{P}(X_0 = x) \\
&= \sum_{x\in E} P^n(x, y)\, \pi(x) \\
&= \pi P^n(y),
\end{aligned}$$

ce qui achève la preuve. ■

Ainsi, si $\mathrm{card}(E)$ est fini, alors pour déterminer la loi de X_n il suffit de calculer la matrice itérée P^n, donc la diagonaliser.

2.2 Classification des chaînes de Markov

Définition 2.2.1. *On dit qu'un état x mène à un état y, et l'on note $x \rightsquigarrow y$, s'il existe $n \in \mathbb{N}$ tel que $P^n(x, y) > 0$. Autrement dit, partant de l'état x, la chaîne peut atteindre y en n étapes.*

Proposition 2.2.2. *On définit une relation binaire sur E de la manière suivante: pour tous $x, y \in E$,*

$$x \sim y \quad \Longleftrightarrow \quad x \rightsquigarrow y \quad \text{et} \quad y \rightsquigarrow x.$$

Cette relation de communication entre états est une relation d'équivalence sur E. Ainsi, l'ensemble des classes d'équivalence (mentionnées comme classes, ou classes de communication dans la suite) est une partition de l'espace des états E.

Démonstration. Pour montrer que cette relation est bien une relation d'équivalence, il faut montrer qu'elle est réflexive, symétrique et transitive. Elle est clairement symétrique, i.e. $x \sim y \Leftrightarrow y \sim x$, et réflexive car pour tout $x \in E$, on a par convention $P^0(x,x) = 1$. Montrons qu'elle est transitive. Soient $x, y, z \in E$ tels que $x \sim y$ et $y \sim z$. Par un argument de symétrie, il est suffisant d'établir que $x \rightsquigarrow z$. On sait qu'il existe $n, m \in \mathbb{N}$ tels que $P^n(x,y) > 0$ et $P^m(y,z) > 0$. D'où

$$P^{n+m}(x,z) \;=\; \sum_{u \in E} P^n(x,u)\, P^m(u,z) \;\geq\; P^n(x,y)\, P^m(y,z) \;>\; 0,$$

ce qui termine la démonstration. ∎

Notons qu'il est possible de sortir d'une classe. Par contre, une fois sorti, il est impossible d'y revenir, sinon cela signifie que l'on n'en serait pas sorti !! Le résultat suivant nous permet d'exhiber un critère simple pour montrer qu'un état mène à un autre état.

Proposition 2.2.3. *Un état x mène à un autre état y si et seulement s'il existe $n \in \mathbb{N}_*$ et $x_1, x_2, \ldots, x_{n-1} \in E$ tels que*

$$P(x,x_1)\, P(x_1,x_2)\, \ldots\, P(x_{n-1},y) > 0.$$

Démonstration. Immédiate en remarquant que pour tout $n \in \mathbb{N}$,

$$P^n(x,y) \;=\; \sum_{x_1,\ldots,x_{n-1} \in E} P(x,x_1)\, P(x_1,x_2) \ldots P(x_{n-1},y).$$

∎

Ainsi, pour montrer qu'un état x mène à y, il suffit de trouver un chemin d'arêtes sur le graphe associé, partant de x et arrivant en y.

Définition 2.2.4. *Une classe de communication $\mathcal{C}$ est dite fermée si pour tout $x \in \mathcal{C}$ tel que $x \rightsquigarrow y$, alors $y \in \mathcal{C}$. Dans le cas contraire, elle est dite ouverte.*
Un état x est dit absorbant si le singleton $\{x\}$ est une classe fermée, i.e. $P(x,x) = 1$.
La chaîne de Markov est dite irréductible si tous les états communiquent, c'est-à-dire qu'il n'y a qu'une seule classe de communication : l'espace des états E tout entier.

Nous allons maintenant introduire les propriétés de classe des chaînes de Markov, à savoir la récurrence et la transience. Au préalable, nous devons définir deux objets très importants, le nombre de visites d'un état et le temps de retour en un état.

Définition 2.2.5. *Soit $y \in E$ un état. Le nombre de visites de y par la chaîne est défini par*

$$V_y := \sum_{n\in\mathbb{N}} 1_{\{X_n=y\}} = \sum_{n\in\mathbb{N}} 1_{\{y\}}(X_n).$$

Le temps de retour en y de la chaîne est défini par

$$T_y := \inf\{n \geq 1 : X_n = y\},$$

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

Notons que l'on parle de temps de retour car n commence à 1 et non à 0. D'ailleurs on a $T_y := \inf\{n \geq 0 : X_n = y\}$ dès que $X_0 \neq y$. Cette v.a. est un temps d'arrêt relativement à la suite croissante en $n \in \mathbb{N}$ de tribus $\mathcal{F}_n := \sigma(X_0, X_1, \ldots, X_n)$, c'est-à-dire que pour tout $n \in \mathbb{N}_*$,

$$\{T_y \leq n\} = \bigcup_{k=1}^{n} \{X_k = y\} \in \mathcal{F}_n.$$

Définition 2.2.6. *Un état $y \in E$ est dit récurrent si $\mathbb{P}_y(V_y = \infty) = 1$ et transitoire si $\mathbb{P}_y(V_y = \infty) = 0$.*

La récurrence traduit la propriété suivante: “partant d'un état, la chaîne y repasse à coup sûr une infinité de fois” tandis que la transience: “partant d'un état, la chaîne n'y repasse à coup sûr qu'un nombre fini de fois”. Remarquons qu'*a priori*, il se peut que $\mathbb{P}_y(V_y = \infty) \in]0,1[$: un état peut être ni récurrent ni transitoire. On va voir dans la suite que cette situation est impossible.

Définition 2.2.7. *Étant donné $n \in \mathbb{N}_*$ et un état $y \in E$, on définit par récurrence le temps du n-ième retour en y de la chaîne par*

$$T_y^{(n)} := \inf\{k \geq T_y^{(n-1)} + 1 : X_k = y\}, \quad \textit{avec} \quad T_y^{(0)} := 0 \quad \textit{et} \quad T_y^{(1)} := T_y.$$

Le prochain résultat est quelque peu technique mais crucial dans ce qui va suivre. Il utilise fortement la propriété de Markov.

Lemme 2.2.8. *Étant donné un état $y \in E$, on a pour tout $n \in \mathbb{N}_*$:*

$$\mathbb{P}_y\left(T_y^{(n)} = T_y^{(n-1)} + m \mid T_y^{(n-1)} < \infty\right) = \mathbb{P}_y\left(T_y = m\right), \quad m \in \mathbb{N}_*.$$

Démonstration. On a pour tout $m \in \mathbb{N}_*$:

$$\begin{aligned}
\mathbb{P}_y\left(T_y^{(n)} = T_y^{(n-1)} + m \mid T_y^{(n-1)} < \infty\right) &= \sum_{r=1}^{\infty} \mathbb{P}_y\left(T_y^{(n)} = r + m; T_y^{(n-1)} = r \mid T_y^{(n-1)} < \infty\right) \\
&= \sum_{r=1}^{\infty} \mathbb{P}_y\left(T_y^{(n)} = r + m \mid T_y^{(n-1)} = r\right) \\
&\quad \times \mathbb{P}_y\left(T_y^{(n-1)} = r \mid T_y^{(n-1)} < \infty\right)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{r=1}^{\infty} \mathbb{P}_y\left(T_y = m\right) \mathbb{P}_y\left(T_y^{(n-1)} = r \mid T_y^{(n-1)} < \infty\right) \\
&= \mathbb{P}_y\left(T_y = m\right) \mathbb{P}_y\left(T_y^{(n-1)} < \infty \mid T_y^{(n-1)} < \infty\right) \\
&= \mathbb{P}_y\left(T_y = m\right).
\end{aligned}
$$

En effet, comme l'événement $\{T_y^{(n-1)} = r\}$ appartient à la tribu $\sigma(X_1, \ldots, X_r)$, la propriété de Markov entraîne

$$
\begin{aligned}
\mathbb{P}_y\left(T_y^{(n)} = r+m \mid T_y^{(n-1)} = r\right) &= \mathbb{P}\left(X_{r+1} \neq y; \ldots; X_{r+m-1} \neq y; X_{r+m} = y \mid X_r = y\right) \\
&= \sum_{x_1 \neq y, \ldots, x_{m-1} \neq y} P(y, x_1)\, P(x_1, x_2) \,\ldots\, P(x_{m-2}, x_{m-1})\, P(x_{m-1}, y) \\
&= \mathbb{P}\left(X_1 \neq y; \ldots; X_{m-1} \neq y; X_m = y \mid X_0 = y\right) \\
&= \mathbb{P}_y\left(T_y = m\right).
\end{aligned}
$$

La démonstration du lemme est établie. ∎

Ainsi, lorsque l'on a $\mathbb{P}_y(T_y^{(n-1)} < \infty) = 1$, ce qui est vérifié dans le cas récurrent car l'on a pour tout $k \in \mathbb{N}$,

$$
1 = \mathbb{P}_y\left(V_y > k\right) = \mathbb{P}_y\left(T_y^{(k)} < \infty\right),
$$

les v.a. $T_y^{(n)} - T_y^{(n-1)}$, $n \in \mathbb{N}_*$, ont même loi que T_y. Avec un peu plus d'effort on peut montrer que les v.a. $(T_y^{(n)} - T_y^{(n-1)})_{n \in \mathbb{N}_*}$ sont indépendantes, ce qui va nous servir dans la démonstration du théorème ergodique qui viendra plus tard.
Le lemme 2.2.8 va nous permettre de démontrer le résultat suivant, qui relie le nombre de visites d'un état y par la chaîne avec le temps de retour en ce même état. On note dans la suite $f_y := \mathbb{P}_y(T_y < \infty)$.

Lemme 2.2.9. *Étant donné un état $y \in E$, on a alors pour tout $n \in \mathbb{N}$:*

$$
\mathbb{P}_y\left(V_y > n\right) = f_y^n.
$$

Démonstration. On démontre ce résultat par récurrence sur $n \in \mathbb{N}$. Trivial pour $n = 0$ car $V_y \geq 1$ dès que $X_0 = y$. Ainsi, supposons la propriété vérifiée au rang n et démontrons-la au rang $n+1$. Notons que lorsque $X_0 = y$, on a pour tout $k \in \mathbb{N}$, $\{V_y > k\} = \{T_y^{(k)} < \infty\}$. Ainsi, on obtient:

$$
\begin{aligned}
\mathbb{P}_y\left(V_y > n+1\right) &= \mathbb{P}_y\left(T_y^{(n+1)} < \infty\right) \\
&= \mathbb{P}_y\left(T_y^{(n+1)} - T_y^{(n)} < \infty; T_y^{(n)} < \infty\right) \\
&= \mathbb{P}_y\left(T_y^{(n+1)} - T_y^{(n)} < \infty \mid T_y^{(n)} < \infty\right) \mathbb{P}_y\left(T_y^{(n)} < \infty\right) \\
&= \sum_{m=1}^{\infty} \mathbb{P}_y\left(T_y^{(n+1)} - T_y^{(n)} = m \mid T_y^{(n)} < \infty\right) \mathbb{P}_y\left(V_y > n\right)
\end{aligned}
$$

$$\begin{aligned}
&= \sum_{m=1}^{\infty} \mathbb{P}_y\left(T_y = m\right) f_y^n \\
&= \mathbb{P}_y\left(T_y < \infty\right) f_y^n \\
&= f_y^{n+1},
\end{aligned}$$

où le lemme 2.2.8 et l'hypothèse de récurrence sont utilisés pour passer des lignes 4 à 5. Finalement, la propriété est bien vérifiée au rang $n+1$, donc en toute généralité. ■

À présent, nous allons pouvoir énoncer un des théorèmes principaux de cette partie, qui justifie le fait qu'un état ne peut être que récurrent ou transitoire.

Théorème 2.2.10. *Étant donné un état $y \in E$, on a la dichotomie suivante:*
(i) Si $\mathbb{P}_y(T_y < \infty) = 1$ alors y est récurrent et $\sum_{n\in\mathbb{N}} P^n(y,y) = \infty$.
(ii) Si $\mathbb{P}_y(T_y < \infty) < 1$ alors y est transitoire et $\sum_{n\in\mathbb{N}} P^n(y,y) < \infty$.

Démonstration. Supposons $f_y := \mathbb{P}_y(T_y < \infty) = 1$ et montrons que y est récurrent, i.e. $\mathbb{P}_y(V_y = \infty) = 1$. On a

$$\{V_y = \infty\} = \bigcap_{n\in\mathbb{N}} \{V_y > n\},$$

où la suite d'événements $(\{V_y > n\})_{n\in\mathbb{N}}$ est décroissante pour l'inclusion. Ainsi, on obtient par le théorème de convergence monotone

$$\mathbb{P}_y\left(V_y = \infty\right) = \mathbb{P}_y\left(\bigcap_{n\in\mathbb{N}} \{V_y > n\}\right) = \lim_{n\to\infty} \mathbb{P}_y\left(V_y > n\right) = \lim_{n\to\infty} f_y^n = 1.$$

De plus on a, en permutant espérance et somme car tous les termes sont positifs (théorème de convergence monotone),

$$\begin{aligned}
\sum_{n\in\mathbb{N}} P^n(y,y) &= \sum_{n\in\mathbb{N}} \mathbb{P}_y\left(X_n = y\right) \\
&= \sum_{n\in\mathbb{N}} \mathbb{E}_y\left[1_{\{X_n = y\}}\right] \\
&= \mathbb{E}_y\left[V_y\right] \\
&= \sum_{k\in\mathbb{N}} \mathbb{P}_y\left(V_y > k\right) \\
&= \sum_{k\in\mathbb{N}} f_y^k \\
&= \sum_{k\in\mathbb{N}} 1 \\
&= \infty.
\end{aligned}$$

Pour passer de la troisième à la quatrième ligne on a utilisé le lemme de Wald démontré ci-dessous, ainsi que le lemme 2.2.9 dans la ligne suivante.
Supposons maintenant $f_y < 1$. En suivant les deux mêmes raisonnements que précédemment, on obtient que $\mathbb{P}_y\left(V_y = \infty\right) = 0$, donc y est transitoire, et que $\sum_{n\in\mathbb{N}} P^n(y,y) < \infty$. ■

Dans la démonstration précédente nous avons utilisé le résultat suivant, connu sous le nom de lemme de Wald.

Lemme 2.2.11. *Si X est une v.a. à valeurs dans $\mathbb{N}$, alors*

$$\mathbb{E}\left[X\right] = \sum_{k\in\mathbb{N}} \mathbb{P}\left(X > k\right).$$

Démonstration. Il suffit de remarquer que X peut s'écrire de manière artificielle comme

$$X = \sum_{k=0}^{X-1} 1 = \sum_{k\in\mathbb{N}} 1_{\{X>k\}},$$

puis de passer à l'espérance des deux côtés (qui sera éventuellement infinie si X n'est pas intégrable) et d'utiliser le théorème de convergence monotone pour inverser espérance et somme. ∎

Comme nous allons le voir, la récurrence et la transience sont des propriétés de classe.

Théorème 2.2.12. *Tous les éléments d'une même classe sont de même nature: ils sont tous récurrents ou tous transitoires. On dit alors de cette classe qu'elle est récurrente ou transitoire.*

Démonstration. Soit $\mathcal{C}$ une classe de communication et soient $x, y \in \mathcal{C}$ avec x supposé transitoire. Montrons que y l'est aussi. Comme x et y communiquent, il existe $m, n \in \mathbb{N}$ tels que $P^m(x, y) > 0$ et $P^n(y, x) > 0$. Alors

$$\begin{aligned} P^{m+k+n}(x, x) &= \sum_{u,v\in E} P^m(x, u)\, P^k(u, v)\, P^n(v, x) \\ &\geq P^m(x, y)\, P^k(y, y)\, P^n(y, x). \end{aligned}$$

Enfin, étant donné que x est transitoire, on a

$$\sum_{k\in\mathbb{N}} P^k(y, y) \leq \frac{1}{P^m(x, y)\, P^n(y, x)} \sum_{k\in\mathbb{N}} P^{m+k+n}(x, x) < \infty,$$

ce qui entraîne que y est aussi transitoire.

Quant à la récurrence, elle est immédiate en argumentant par l'absurde. ∎

À présent, comment peut-on établir simplement qu'une classe est récurrente ou transitoire ? La proposition suivante répond à cette question.

Proposition 2.2.13. *Soit $\mathcal{C}$ une classe de communication. Alors les propriétés suivantes sont vérifiées:*

(i) Si $\mathcal{C}$ est récurrente, alors elle est fermée.

(ii) Si $\mathcal{C}$ est fermée et finie, alors elle est récurrente.

En particulier, si l'espace d'état E est fini, alors la classe $\mathcal{C}$ est récurrente si et seulement si elle est fermée.

Démonstration. Nous n'allons démontrer que le point (i), le point (ii) utilisant la propriété de Markov forte, que nous n'énonçons pas dans ce cours. Pour ce faire, on va utiliser un raisonnement par l'absurde. Ainsi, supposons $\mathcal{C}$ récurrente et ouverte. Alors il existe un état récurrent $x \in C$, un état $y \notin C$ et $n \in \mathbb{N}$ tels que $P^n(x,y) > 0$. Lorsque l'on sort d'une classe, on ne peut y revenir donc on a

$$\begin{aligned} \mathbb{P}_x\left(\{X_n = y\} \cup \{V_x = \infty\}\right) &= \mathbb{P}_x\left(X_n = y\right) + \mathbb{P}_x\left(V_x = \infty\right) - \mathbb{P}_x\left(X_n = y; V_x = \infty\right) \\ &= P^n(x,y) + 1 - 0 \\ &> 1, \end{aligned}$$

ce qui est bien évidemment impossible: la classe $\mathcal{C}$ est donc fermée. ■

Ainsi, un état récurrent ne peut mener qu'à un état de sa classe, donc à un état récurrent. En revanche, un état transitoire peut mener à un autre état transitoire (de sa classe ou pas) comme à un état récurrent.

Pour résumer, on a les équivalences suivantes: si $x \in E$,

(i) x est un état récurrent.

(ii) $\mathbb{P}_x\left(V_x = \infty\right) = 1$.

(iii) $\mathbb{P}_x\left(T_x < \infty\right) = 1$.

(iv) $\sum_{n\in\mathbb{N}} P^n(x,x) = \infty$.

De même, les assertions suivantes sont équivalentes: si $x \in E$,

(i) x est un état transitoire.

(ii) $\mathbb{P}_x\left(V_x = \infty\right) = 0$.

(iii) $\mathbb{P}_x\left(T_x < \infty\right) < 1$.

(iv) $\sum_{n\in\mathbb{N}} P^n(x,x) < \infty$.

On peut aussi démontrer les résultats suivants:

(i) si x et y sont dans la même classe récurrente, alors on a

$$\left\{\begin{array}{lcl} \mathbb{P}_x\left(V_y = \infty\right) & = & 1; \\ \mathbb{P}_x\left(T_y < \infty\right) & = & 1; \\ \sum_{n\in\mathbb{N}} P^n(x,y) & = & \infty. \end{array}\right.$$

(ii) si x et y sont dans la même classe transitoire,

$$\left\{\begin{array}{lcl} \mathbb{P}_x\left(V_y = \infty\right) & = & 0; \\ \mathbb{P}_x\left(T_y < \infty\right) & < & 1; \\ \sum_{n\in\mathbb{N}} P^n(x,y) & < & \infty. \end{array}\right.$$

Enfin, si la chaîne est irréductible alors il n'est pas difficile de montrer que l'on peut enlever le conditionnement relatif à X_0 dans ce qui précède, c'est-à-dire que ces critères sont valables pour n'importe quelle loi initiale.

2.3 Probabilité invariante

Étant donnée une chaîne de Markov, on peut se demander si elle a vocation à se stabiliser statistiquement, c'est-à-dire existe-il une probabilité π telle que:

- si X_0 suit la loi π alors X_n la suit aussi pour tout $n \in \mathbb{N}$?
- la loi de X_n converge vers π lorsque $n \to \infty$?

Il s'avère que sous certaines hypothèses, la réponse est oui. Pour étudier ces questions, introduisons tout d'abord la notion de probabilité invariante.

Définition 2.3.1. *Soit π une mesure sur E, de masse éventuellement infinie. On dit que π est une mesure invariante pour la chaîne si l'équation suivante est vérifiée:*

$$\pi(y) = \pi P(y), \quad y \in E,$$

où l'on rappelle que la définition de πP est donnée par

$$\pi P(y) \; = \; \sum_{x \in E} \pi(x)\, P(x,y), \quad y \in E.$$

De surcroît, si π est une probabilité, on parle de probabilité invariante pour la chaîne.

Ainsi, dire que π est une probabilité invariante signifie que si X_0 a pour loi π, alors X_1 suit aussi cette loi. On en déduit de proche en proche que pour tout $n \in \mathbb{N}$, la v.a. X_n suit la loi π, i.e. que $\pi = \pi P^n$. On va voir dans la suite que sous de bonnes hypothèses, il existe une unique probabilité invariante pour la chaîne. Pour ce faire, on va considérer le nombre de visites d'un état y avant le retour en un autre état x:

$$V_y^x \; := \; \sum_{n=0}^{T_x - 1} 1_{\{X_n = y\}} \; = \; \sum_{n=0}^{\infty} 1_{\{X_n = y; T_x > n\}}.$$

On remarque que $V_x^x = 1_{\{X_0 = x\}}$ et donc que $\mathbb{E}_x \left[V_x^x \right] = \mathbb{P}_x \left(X_0 = x \right) = 1$. Étant donné un état x, notons dans la suite γ_x la mesure sur E donnée par

$$\gamma_x(y) \; := \; \mathbb{E}_x \left[V_y^x \right] \; = \; \sum_{n=0}^{\infty} \mathbb{P}_x \left(X_n = y; T_x > n \right),$$

quantité qui peut éventuellement être infinie. Commençons par plusieurs lemmes techniques.

Lemme 2.3.2. *On suppose la chaîne de Markov irréductible (une seule classe, E) et récurrente (tous les états le sont). Alors pour tout $x \in E$, la mesure γ_x est une mesure invariante pour la chaîne.*

Démonstration. On doit montrer que pour tout état $x \in E$, on a $\gamma_x = \gamma_x P$, c'est-à-dire

$$\mathbb{E}_x \left[V_y^x \right] \; = \; \sum_{z \in E} \mathbb{E}_x \left[V_z^x \right] \, P(z,y), \quad y \in E.$$

La chaîne étant supposée récurrente, on a que $T_x < \infty$ et que $X_0 = X_{T_x} = x$ lorsque $X_0 = x$, où l'on rappelle que T_x est le temps de retour de la chaîne en l'état x. En permutant espérance et somme par le théorème de convergence monotone,

$$\begin{aligned}
\mathbb{E}_x\left[V_y^x\right] &= \mathbb{E}_x\left[\sum_{n=0}^{T_x-1} 1_{\{X_n=y\}}\right] \\
&= \mathbb{E}_x\left[\sum_{n=1}^{T_x} 1_{\{X_n=y\}}\right] + \mathbb{P}_x\left(X_0=y\right) - \mathbb{P}_x\left(X_{T_x}=y\right) \\
&= \mathbb{E}_x\left[\sum_{n=1}^{\infty} 1_{\{X_n=y;T_x\geq n\}}\right] + 0 \\
&= \sum_{n=1}^{\infty} \mathbb{P}_x\left(X_n=y;T_x\geq n\right) \\
&= \sum_{n=1}^{\infty}\sum_{z\in E} \mathbb{P}_x\left(X_n=y;X_{n-1}=z;T_x\geq n\right) \\
&= \sum_{n=1}^{\infty}\sum_{z\in E} \mathbb{P}_x\left(X_n=y \mid X_{n-1}=z;T_x\geq n\right)\mathbb{P}_x\left(X_{n-1}=z;T_x\geq n\right).
\end{aligned}$$

Remarquons que l'événement $\{T_x \geq n\}$ appartient à la tribu $\sigma(X_1,\ldots,X_{n-1})$ donc par la propriété de Markov,

$$\begin{aligned}
\mathbb{E}_x\left[V_y^x\right] &= \sum_{n=1}^{\infty}\sum_{z\in E} \mathbb{P}\left(X_n=y \mid X_{n-1}=z\right)\mathbb{P}_x\left(X_{n-1}=z;T_x\geq n\right) \\
&= \sum_{n=0}^{\infty}\sum_{z\in E} P(z,y)\,\mathbb{P}_x\left(X_n=z;T_x\geq n+1\right) \\
&= \sum_{z\in E} P(z,y)\,\mathbb{E}_x\left[V_z^x\right],
\end{aligned}$$

ce qui termine la preuve. ■

La mesure γ_x étant une mesure invariante pour la chaîne, on a immédiatement que $\gamma_x = \gamma_x P^n$ pour tout $n \in \mathbb{N}$. Ceci va nous servir dans la démonstration du prochain résultat.

Lemme 2.3.3. *On suppose la chaîne irréductible et récurrente. Alors pour tout état $x \in E$, on a $0 < \gamma_x(y) < \infty$ pour tout $y \in E$.*

Démonstration. Soient $x, y \in E$ deux états. La chaîne étant irréductible, tous les états communiquent: il existe $m, n \in \mathbb{N}$ tels que $P^m(x,y) > 0$ et $P^n(y,x) > 0$. Ainsi, par le lemme 2.3.2, la mesure γ_x est invariante pour la chaîne et l'on a

$$1 = \gamma_x(x) = \sum_{z\in E} \gamma_x(z)\,P^n(z,x) \geq \gamma_x(y)\,P^n(y,x),$$

d'où $\gamma_x(y) \leq 1/P^n(y,x)$: la quantité $\gamma_x(y)$ est donc finie. Par ailleurs, on a encore par le lemme 2.3.2 que

$$\gamma_x(y) \;=\; \sum_{z\in E} \gamma_x(z)\, P^m(z,y) \;\geq\; \gamma_x(x)\, P^m(x,y) \;=\; P^m(x,y) \;>\; 0,$$

ce qui achève la démonstration. ■

Terminons par un dernier lemme technique mais ô combien important dans les démonstrations des théorèmes qui vont suivre.

Lemme 2.3.4. *On suppose la chaîne irréductible et récurrente, et qu'il existe λ une mesure invariante pour la chaîne telle que $\lambda(u) = 1$ pour un certain état $u \in E$. Alors les mesures λ et γ_u coïncident.*

Démonstration. La mesure λ étant invariante, on a pour tout état $y \in E$ et tout $n \in \mathbb{N}_*$:

$$\begin{aligned}
\lambda(y) &= \sum_{x_n \in E} \lambda(x_n)\, P(x_n,y) \\
&= \sum_{x_n \neq u} \lambda(x_n)\, P(x_n,y) + 1 \times P(u,y) \\
&= \sum_{x_n \neq u} \left(\sum_{x_{n-1}\in E} \lambda(x_{n-1})\, P(x_{n-1},x_n) \right) P(x_n,y) + P(u,y) \\
&= \sum_{x_n\neq u,\, x_{n-1}\neq u} \lambda(x_{n-1})\, P(x_{n-1},x_n)\, P(x_n,y) + \sum_{x_n\neq u} P(u,x_n)\, P(x_n,y) + P(u,y) \\
&= \sum_{x_n\neq u,\, x_{n-1}\neq u,\, x_{n-2}\neq u} \lambda(x_{n-2})\, P(x_{n-2},x_{n-1})\, P(x_{n-1},x_n)\, P(x_n,y) \\
&\quad + \sum_{x_n\neq u,\, x_{n-1}\neq u} P(u,x_{n-1})\, P(x_{n-1},x_n)\, P(x_n,y) + \sum_{x_n\neq u} P(u,x_n)\, P(x_n,y) + P(u,y) \\
&= \ldots \\
&= \sum_{x_n\neq u,\, \ldots,\, x_1\neq u} \lambda(x_1)\, P(x_1,x_2)\, \ldots\, P(x_n,y) + \sum_{x_n\neq u,\, \ldots,\, x_2\neq u} P(u,x_2)\, P(x_2,x_3)\, \ldots\, P(x_n,y) \\
&\quad + \ldots + \sum_{x_n\neq u} P(u,x_n)\, P(x_n,y) + P(u,y) \\
&= \sum_{x_n\neq u,\, \ldots,\, x_1\neq u} \lambda(x_1)\, P(x_1,x_2)\, \ldots\, P(x_n,y) + \mathbb{P}_u\left(X_n = y; T_u \geq n\right) \\
&\quad + \ldots + \mathbb{P}_u\left(X_2 = y; T_u \geq 2\right) + \mathbb{P}_u\left(X_1 = y; T_u \geq 1\right) \\
&\geq \sum_{k=1}^{n} \mathbb{P}_u\left(X_k = y; T_u \geq k\right).
\end{aligned}$$

Par le lemme 2.3.3, la quantité $\gamma_u(y)$ est bien définie. Ainsi, on peut passer à la limite dans l'expression précédente lorsque $n \to \infty$ et l'on obtient que pour tout $y \in E$,

$$\lambda(y) \;\geq\; \sum_{k=1}^{\infty} \mathbb{P}_u\left(X_k = y; T_u \geq k\right) \;=\; \gamma_u(y).$$

On en déduit que la quantité μ donnée par $\mu = \lambda - \gamma_u$ définit bien une mesure sur E. Par ailleurs, la mesure γ_u est invariante pour la chaîne par le lemme 2.3.2, donc la mesure μ est elle-aussi invariante et satisfait:

$$\mu(u) \;=\; \lambda(u) - \gamma_u(u) \;=\; 1 - 1 \;=\; 0.$$

Par irréductibilité, pour tout état $x \in E$, il existe $n \in \mathbb{N}$ tel que $P^n(x,u) > 0$. On obtient alors que

$$0 \;=\; \mu(u) \;=\; \mu P^n(u) \;=\; \sum_{x \in E} \mu(x)\, P^n(x,u) \;\geq\; \mu(x)\, P^n(x,u),$$

ce qui force à avoir $\mu(x) = 0$ pour tout $x \in E$. Autrement dit, les deux mesures λ et γ_u coïncident. ■

Revenons à T_x, le temps de retour en un état x. On remarque qu'il peut s'écrire en fonction des V_y^x:

$$T_x \;=\; \sum_{n=0}^{T_x - 1} 1 \;=\; \sum_{n=0}^{T_x - 1} \sum_{y \in E} 1_{\{X_n = y\}} \;=\; \sum_{y \in E} V_y^x,$$

d'où en passant à l'espérance, on obtient en utilisant le théorème de convergence monotone que

$$\mathbb{E}_x[T_x] < \infty \quad \Longleftrightarrow \quad \sum_{y \in E} \mathbb{E}_x[V_y^x] = \sum_{y \in E} \gamma_x(y) < \infty.$$

En particulier, d'après le lemme 2.3.3, une condition suffisante dans le cas irréductible et récurrent pour que $\mathbb{E}_x[T_x]$ soit finie pour tout état x est de supposer E de cardinal fini. Ceci nous amène à la définition suivante.

Définition 2.3.5. *Un état récurrent x est dit récurrent positif si $\mathbb{E}_x[T_x]$ est fini, et récurrent nul sinon. Si tous les états sont récurrents positifs (resp. récurrents nuls), on dit que la chaîne est récurrente positive (resp. récurrente nulle).*

Notons que dans le cas transitoire on a $\mathbb{P}_x(T_x < \infty) < 1$, ce qui entraîne immédiatement que $\mathbb{E}_x[T_x] = \infty$.

Si l'on suppose maintenant qu'un état x est récurrent positif, on définit alors la probabilité π_x sur E par:

$$\pi_x(y) \;=\; \frac{\mathbb{E}_x[V_y^x]}{\mathbb{E}_x[T_x]} \;=\; \frac{\gamma_x(y)}{\sum_{y \in E} \gamma_x(y)}, \quad y \in E.$$

Par le lemme 2.3.2, la probabilité π est invariante pour la chaîne sous l'hypothèse supplémentaire de l'irréductibilité. Cependant, en existe-il d'autres ? À présent, nous sommes en mesure de répondre à cette question en énonçant le théorème suivant.

Théorème 2.3.6. *On suppose la chaîne irréductible et récurrente. Alors les assertions suivantes sont équivalentes:*

(i) *tout état de E est récurrent positif.*

(ii) *il existe un état récurrent positif.*

(iii) *il existe une unique probabilité invariante π.*

Dans ce cas, on a $\pi = \pi_x$ pour tout $x \in E$: on prendra alors

$$\pi(y) \,=\, \pi_y(y) \,=\, \frac{1}{\mathbb{E}_y[T_y]}, \quad y \in E.$$

Démonstration. $(i) \Rightarrow (ii)$: Immédiat.

$(ii) \Rightarrow (iii)$: L'existence de π a été établie par le lemme 2.3.2. L'unicité, quant à elle, découle du lemme 2.3.4.

$(iii) \Rightarrow (i)$: Soit $y \in E$ et notons π l'unique probabilité invariante de la chaîne. On va montrer que l'état y est récurrent positif. Il est immédiat qu'il existe $x \in E$ tel que $\pi(x) > 0$ (sinon la somme $\sum_{z \in E} \pi(z)$ vaudrait 0 et non 1). De plus, par irréductibilité, il existe $n \in \mathbb{N}$ tel que $P^n(x,y) > 0$. Donc

$$\pi(y) \,=\, \pi P^n(y) \,=\, \sum_{z \in E} \pi(z)\, P^n(z,y) \,\geq\, \pi(x)\, P^n(x,y) \,>\, 0.$$

Soit λ la mesure sur E donnée pour tout $z \in E$ par $\lambda(z) = \pi(z)/\pi(y)$. Ainsi, λ est une mesure invariante pour la chaîne, avec de surcroît $\lambda(y) = 1$. En utilisant le lemme 2.3.4, les mesures λ et γ_y coïncident et il en résulte que

$$\mathbb{E}_y[T_y] \,=\, \sum_{z \in E} \gamma_y(z) \,=\, \sum_{z \in E} \lambda(z) \,=\, \sum_{z \in E} \frac{\pi(z)}{\pi(y)} \,=\, \frac{1}{\pi(y)} \,<\, \infty,$$

ce qui signifie que y est récurrent positif. Or l'état y étant arbitraire, on obtient le résultat désiré, à savoir que tout état est récurrent positif. ■

Notons que le théorème précédent admet une version "récurrence nulle", dont la démonstration est quelque peu similaire à celle que l'on vient de faire.

Théorème 2.3.7. *On suppose la chaîne irréductible et récurrente. Alors les assertions suivantes sont équivalentes:*

(i) *tout état de E est récurrent nul.*

(ii) *il existe un état récurrent nul.*

(iii) *il existe une mesure invariante λ pour la chaîne, unique à constante multiplicative près, et l'on a pour tout $x \in E$ fixé,*

$$\lambda(y) \,=\, \lambda(x)\, \gamma_x(y) \,>\, 0, \quad y \in E.$$

De plus, λ est forcément de masse infinie.

Pour terminer ce paragraphe, résumons dans le théorème qui suit les résultats importants que nous venons de voir jusqu'à présent.

Théorème 2.3.8. *On suppose la chaîne de Markov irréductible. Alors les états sont tous transitoires, tous récurrents nuls ou tous récurrents positifs. Et dans ce dernier cas seulement, la chaîne admet une unique probabilité invariante π donnée par*

$$\pi(x) = \frac{1}{\mathbb{E}_x[T_x]}, \quad x \in E.$$

Enfin, si la chaîne est irréductible sur E supposé de cardinal fini, alors tous les états sont récurrents positifs et l'existence et l'unicité de π sont assurées.

Notons que si la chaîne est irréductible et transitoire, alors il se peut qu'il existe des mesures invariantes (il n'y a aucune raison pour que l'on ait unicité de la mesure invariante, même à multiplication près par une constante: considérer la marche aléatoire asymétrique sur $\mathbb{Z}$). Si l'on note λ une telle mesure invariante, alors on montre comme dans la démonstration du théorème 2.3.6 que $\lambda(y) > 0$ pour tout $y \in E$, mais aussi que cette mesure est forcément de masse infinie. En effet la somme $\sum_{n\in\mathbb{N}} P^n(x,y)$ est finie donc $P^n(x,y)$ tend vers 0 lorsque n tend vers l'infini. Si la quantité $\lambda(E)$ était finie, le théorème de convergence dominée appliqué lorsque $n \to \infty$ dans l'égalité

$$\lambda(y) = \sum_{x\in E} \lambda(x) \, P^n(x,y), \quad y \in E,$$

entraînerait alors que $\lambda(y) = 0$, ce qui est impossible d'après ce qui précède.

2.4 Convergence vers la probabilité invariante

Dans le cas irréductible et récurrent positif, on a vu qu'il existait une unique probabilité invariante π, c'est-à-dire une probabilité satisfaisant la propriété suivante: si X_0 suit la loi π alors pour tout $n \in \mathbb{N}$, la v.a. X_n la suit aussi. Cependant, si X_0 est distribuée selon une autre loi, peut-on espérer que la loi de X_n converge vers π ? Oui, à condition d'éviter un problème de périodicité.

Considérons la matrice de transition triviale P sur $E = \{0,1\}$ donnée par $P(0,1) = P(1,0) = 1$. Autrement dit, il n'y a pas d'aléa. En calculant les puissances successives de P, on montre que $P^{2n} = I_2$ et $P^{2n+1} = P$. Ainsi, on remarque que $P^n(x,y)$ ne peut pas converger lorsque $n \to \infty$ à cause de cette périodicité. Donnons à présent la définition de la période d'un état.

Définition 2.4.1. *On appelle période de l'état x le nombre*

$$d_x := pgcd\{n \in \mathbb{N}_* : P^n(x,x) > 0\}.$$

L'état x est dit apériodique si $d_x = 1$.

Proposition 2.4.2. *Le concept de période est une propriété de classe, c'est-à-dire que si x et y sont deux états appartenant à une même classe $\mathcal{C}$, alors ils ont même période. On parlera alors de période d'une classe.*

Démonstration. Soient $x, y \in \mathcal{C}$: il existe $m, n, \in \mathbb{N}$ tels que $P^m(x,y) > 0$ et $P^n(y,x) > 0$. Ainsi, pour tout $k \in \mathbb{N}_*$, on a

$$P^{m+k+n}(x,x) \;=\; \sum_{u,v \in E} P^m(x,u)\, P^k(u,v)\, P^n(v,x) \;\geq\; P^m(x,y)\, P^k(y,y)\, P^n(y,x),$$

d'où pour tout $k \in \mathbb{N}_*$, $P^{m+k+n}(x,x) > 0$ dès que $P^k(y,y) > 0$: on en déduit que d_x divise $m+k+n$. De plus, on a

$$P^{m+n}(x,x) \;=\; \sum_{u \in E} P^m(x,u)\, P^n(u,x) \;\geq\; P^m(x,y)\, P^n(y,x) > 0,$$

donc d_x divise $m+n$. Ainsi, d_x divise tous les entiers $k \in \mathbb{N}_*$ tels que $P^k(y,y) > 0$, c'est-à-dire le pgcd de l'ensemble $\{n \in \mathbb{N}_* : P^n(y,y) > 0\}$, donc d_y.
Enfin, par symétrie de x et de y, on a aussi que d_y divise d_x, d'où l'égalité $d_x = d_y$. ∎

On peut maintenant énoncer l'un des derniers résultats fondamentaux de ce cours. La démonstration, que l'on admettra car reposant sur un formalisme quelque peu pénible à introduire, les "chaînes de Markov produit", est basée sur une technique dite de couplage.

Théorème 2.4.3. *On suppose la chaîne irréductible et récurrente positive.*

(i) Si elle est apériodique, i.e. la seule classe E a pour période $d = 1$, alors la loi de X_n converge lorsque $n \to \infty$ vers l'unique probabilité invariante π, en oubliant sa condition initiale. Autrement dit, pour tout $x \in E$,

$$P^n(x,y) \;=\; \mathbb{P}_x(X_n = y) \underset{n\to\infty}{\longrightarrow} \pi(y), \quad y \in E.$$

(ii) Si elle est périodique de période $d > 1$, alors pour tous $x, y \in E$, il existe $r = r_{x,y} \in \{0, 1, \ldots, d-1\}$ tel que pour tout $n \in \mathbb{N}$, on ait $P^n(x,y) = 0$, sauf lorsqu'il existe $m \in \mathbb{N}$ tel que $n = md + r$, auquel cas on a

$$P^{md+r}(x,y) \underset{m\to\infty}{\longrightarrow} d\, \pi(y).$$

Nous allons terminer ce cours par l'énoncé de la Loi des Grands Nombres pour les chaînes de Markov, connue sous le nom de théorème ergodique. Dans ce qui suit, la convergence a lieu au sens de la convergence presque sûre, c'est-à-dire pour tout ω en dehors d'un certain ensemble de probabilité 0. Notons

$$V_x^n \;:=\; \sum_{k=0}^{n-1} 1_{\{X_k = x\}} \;=\; \sum_{k=0}^{n-1} 1_{\{x\}}(X_k), \quad x \in E, \quad n \in \mathbb{N}_*,$$

le nombre de visites de l'état x avant l'instant n. Si la chaîne est irréductible et transitoire alors la v.a. V_x^n est majorée par V_x, qui est une quantité finie, l'état x étant transitoire. On en déduit alors que

$$\frac{V_x^n}{n} \underset{n\to\infty}{\longrightarrow} 0 = \frac{1}{\mathbb{E}_x[T_x]},$$

c'est-à-dire que la proportion de temps passé en l'état x tend vers 0.

Théorème 2.4.4. *On suppose la chaîne irréductible et récurrente, de loi initiale quelconque. Alors pour toute mesure invariante λ, on a*

$$\frac{V_y^n}{V_x^n} \underset{n\to\infty}{\longrightarrow} \frac{\lambda(y)}{\lambda(x)} = \frac{1}{\gamma_x(y)}, \quad x, y \in E,$$

mais aussi

$$\frac{V_x^n}{n} \underset{n\to\infty}{\longrightarrow} \frac{1}{\mathbb{E}_x[T_x]}, \quad x \in E.$$

Démonstration. Nous n'allons démontrer que la seconde convergence, la première utilisant le même type d'arguments. Tout d'abord, la chaîne étant irréductible et récurrente, le temps de retour en x est fini presque sûrement et ce pour n'importe quelle loi initiale. On peut donc supposer que la chaîne part de x pour établir le résultat. De plus, la suite strictement croissante $(T_x^{(n)})_{n\in\mathbb{N}}$ des temps de retour successifs en x est bien définie et tend vers l'infini lorsque n tend vers l'infini. Comme $X_0 = x$, on a les inégalités suivantes: $T_x^{(V_x^n-1)} \leq n-1$ et $n \leq T_x^{(V_x^n)}$, d'où l'encadrement

$$\frac{T_x^{(V_x^n-1)}}{V_x^n} \leq \frac{n}{V_x^n} \leq \frac{T_x^{(V_x^n)}}{V_x^n}.$$

La chaîne étant récurrente, V_x^n tend vers l'infini lorsque n tend vers l'infini. Par ailleurs, comme on l'a vu précédemment, les v.a. $(T_x^{(n)} - T_x^{(n-1)})_{n\in\mathbb{N}_*}$ sont i.i.d. et l'on a alors par la Loi des Grands Nombres classique que

$$\frac{T_x^{(V_x^n)}}{V_x^n} = \frac{1}{V_x^n}\sum_{k=1}^{V_x^n}(T_x^{(k)} - T_x^{(k-1)}) \underset{n\to\infty}{\longrightarrow} \mathbb{E}_x[T_x].$$

Ainsi, le terme de l'encadrement de gauche tendant lui aussi vers $\mathbb{E}_x[T_x]$, on en déduit la conclusion désirée. ∎

En travaillant un peu, on peut remplacer dans le résultat précédent les indicatrices $1_{\{x\}}$ et $1_{\{y\}}$ par des fonctions f et g à valeurs réelles. On obtient ainsi un résultat rappelant celui connu pour les v.a.r. i.i.d., la mesure invariante de la chaîne remplaçant la loi commune du cas i.i.d.

Théorème 2.4.5. *On suppose la chaîne irréductible et récurrente, de loi initiale quelconque. Alors pour toute mesure invariante λ et toutes fonctions $f, g : E \to \mathbb{R}$ intégrables par rapport à λ, avec de surcroît $g > 0$, on a*

$$\lim_{n\to\infty} \frac{\sum_{k=0}^{n-1} f(X_k)}{\sum_{k=0}^{n-1} g(X_k)} = \frac{\sum_{z\in E} f(z)\,\lambda(z)}{\sum_{z\in E} g(z)\,\lambda(z)}.$$

Si la chaîne est récurrente nulle, alors pour toute fonction $f : E \to \mathbb{R}$ qui est λ-intégrable,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \;=\; 0.$$

Si la chaîne est récurrente positive, alors pour toute fonction $f : E \to \mathbb{R}$ qui est π-intégrable, où π désigne l'unique probabilité invariante, $\pi(x) = 1/\mathbb{E}_x[T_x]$, on a

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \;=\; \sum_{z\in E} f(z)\,\pi(z).$$

Bibliographie

[1] P. Baldi, L. Mazliak et P. Priouret. Martingales et chaînes de Markov. Hermann, 2000.

[2] P. Barbe et M. Ledoux. Probabilités. EDP Sciences, 2007.

[3] J.-P. Delmas. Introduction aux probabilités. Ellipses, 2000.

[4] J.R. Norris. Markov chains. Cambridge University Press, 1998.