CHAÎNES DE MARKOV ET APPLICATIONS
2007–08
VERSION RÉVISÉE

Anthony Phan

Département de Mathématiques
Boulevard Marie et Pierre Curie,
Téléport 2,
BP 30179, F-86962 Chasseneuil–Futuroscope cedex.

Introduction aux chaînes de Markov à espace d'état fini ou dénombrable, propriétés asymptotiques, applications.

Quelques références bibliographiques

- Dominique Foata et Aimé Fuchs. *Calcul des Probabilités, cours, exercices et problèmes corrigés*, Dunod, seconde édition (1998). Un bon livre pour débuter en probabilités.
- Dominique Foata et Aimé Fuchs. *Processus stochastiques, cours, exercices corrigés*, Dunod (2002). Ce livre est la suite du précédent et est plus dirigé vers des applications telles que l'actuariat.
- Walter Rudin. *Analyse réelle et complexe*, Masson (1991), une bonne référence pour la théorie de la mesure (point de vue des espaces topologiques).
- Philippe Barbe et Michel Ledoux. *Probabilité*, Belin (1998). Plus technique que l'ouvrage de D. Foata et A. Fuchs, ce livre fournit de nombreuses démonstrations rigoureuses à des énoncés qu'on est souvent amené à admettre pour cause de temps.
- J.R. Norris. *Markov Chains*, Cambridge Series in Statistical and Probabilistic Mathematics (1997–99). Un livre complet et assez clair qui présente un petit appendice sur la théorie des probabilités.

Nous recommandons particulièrement les deux derniers ouvrages. Cependant bien d'autres livres existent sur le sujet qui sont tout aussi bien.

Chapitre premier

CHAINES DE MARKOV À TEMPS DISCRET

1. Chaînes de Markov, définitions et premières propriétés

Soit E un ensemble fini ou dénombrable, muni de sa tribu discrète. Toute mesure λ sur E équivaut à la donnée d'une suite $(\lambda\{x\})_{x\in E}$ de nombres réels positifs ou nuls, c'est une mesure de probabilité si et seulement si $\sum_{x\in E}\lambda\{x\}=1$. Abusivement, nous noterons $\lambda=(\lambda\{x\})_{x\in E}$. La suite λ pourra être identifiée avec la matrice ligne, éventuellement infinie, correspondante.

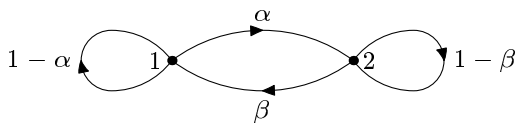
Soit $(\Omega,\mathcal{A},\mathbb{P})$ un espace probabilisé. Nous rappelons que pour qu'une application $X:\Omega\to E$ soit une variable aléatoire il est nécessaire et suffisant que pour tout $x\in E$, $\{X=x\}\in\mathcal{A}$; la loi de X est alors la mesure de probabilité sur E définie par la suite $\lambda\{x\}=\mathbb{P}\{X=x\}$, $x\in E$.

Une matrice $P=(P(x,y))_{x,y\in E}$ est une *matrice stochastique* (ou *de transition*) si chacune de ses lignes est une mesure de probabilité sur E, autrement dit si ses coefficients sont des réels positifs et dans chaque ligne la somme des coefficients vaut 1. Chaque coefficient $P(x,y)$ s'interprète comme la probabilité en partant de x d'aller en y.

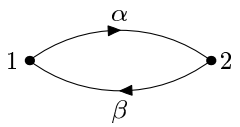
On peut décrire toute matrice stochastique P par un diagramme dont les sommets représentent les points de E et les flêches joignant ces sommets représentent les coefficients correspondants de la matrice. On omet toujours de représenter une flêche correspondant à un coefficient nul, et souvent aussi celles qui correspondent à un coefficient diagonal puisque le fait que $\sum_{y\in E}P(x,y)=1$ permet de l'identifier lorsqu'on connaît les autres coefficients de la ligne correspondante.

Exemples. — *a)* Soit $E=\{1,2\}$. Toute matrice stochastique P de taille 2×2 s'écrit

$$P=\begin{pmatrix}1-\alpha & \alpha\\ \beta & 1-\beta\end{pmatrix}$$



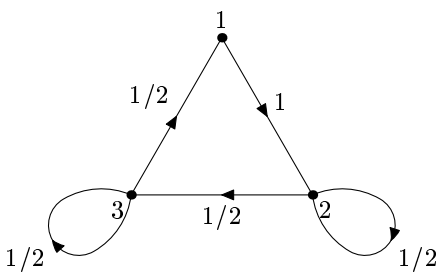
ou



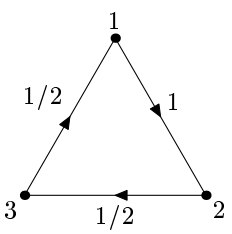
avec α, $\beta\in[0,1]$. Le premier diagramme montre les coefficients diagonaux, pas le second.

b) Pour $E=\{1,2,3\}$, considérons un exemple particulier :

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



ou



Définition 1. — Une suite $(X_n)_{n\geqslant 0}$ de variables aléatoires à valeurs dans E est une *chaîne de Markov* si pour tout $n\geqslant 0$, X_{n+1} et $(X_0,\dots,X_n)$ sont indépendants conditionnellement à X_n, c'est-à-dire si pour tous $x_0,\dots,x_{n+1}\in E$

$$\mathbb{P}\{X_{n+1}=x_{n+1}\,|\,X_0=x_0,\dots,X_n=x_n\}=\mathbb{P}\{X_{n+1}=x_{n+1}\,|\,X_n=x_n\}$$

lorsque ces probabilités conditionnelles ont un sens.

Si $(X_n)_{n\geqslant 0}$ est une chaîne de Markov, nous noterons

$$P^{(n,n+1)}(x,y) = \mathbb{P}\{X_{n+1} = y \,|\, X_n = x\}, \qquad x, y \in E,$$

et $P^{(n,n+1)}$ la matrice stochastique associée. Il se peut que certaines de ses lignes ne soient pas définies (pour les lignes d'indice $x \in E$ tel que $\mathbb{P}\{X_n = x\} = 0$). Nous admettrons que ça n'a aucune importance et ignorerons cette petite ambiguïté par la suite.

Définition 2. — Une chaîne de Markov $(X_n)_{n\geqslant 0}$ à valeurs dans E est *homogène* si les probabilités conditionnelles ci-dessus ne dépendent pas de n, plus précisément s'il existe une matrice stochastique $P = (P(x,y))_{x,y\in E}$ telle que

$$P(x,y) = \mathbb{P}\{X_{n+1} = y \,|\, X_n = x\},$$

pour tout $n \geqslant 0$ et x, $y \in E$ pour lesquels cette probabilité conditionnelle est définie.

La loi λ de X_0 est appelée *loi initiale* de la chaîne, et la matrice stochastique P est appelée *matrice de transition* de la chaîne. On abrègera en employant l'expression « $(X_n)_{n\geqslant 0}$ est *une chaîne de Markov* (λ, P) ».

Remarque. — Lorsque on considère une suite finie de variables aléatoires $(X_n)_{0\leqslant n\leqslant N}$, on emploie le terme de *chaîne finie*. La définition ci-dessus s'adapte immédiatement à ce contexte.

Exercice 1. — Soit $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires à valeurs dans E. Pour tout $n \geqslant 0$, soient $\mathcal{F}_n = \sigma\{X_0, \ldots, X_n\}$ et $\mathcal{G}_n = \sigma\{X_n, X_{n+1}, \ldots\}$ les tribus respectivement du passé et du futur à l'instant n.

Montrer que $(X_n)_{n\geqslant 0}$ est une chaîne de Markov si et seulement si pour tout $n \geqslant 0$, $x \in E$, les tribus $\mathcal{F}_n$ et $\mathcal{G}_n$ sont indépendantes conditionnellement à $\{X_n = x\}$.

Théorème 1. — *Une suite finie de variables aléatoires $(X_n)_{0\leqslant n\leqslant N}$ à valeurs dans E est une chaîne de Markov (λ, P) si et seulement si pour tous $x_0, \ldots, x_N \in E$*

$$\mathbb{P}\{X_0 = x_0, \ldots, X_N = x_N\} = \lambda\{x_0\} \times P(x_0, x_1) \times \cdots \times P(x_{N-1}, x_N). \tag{1.1}$$

Remarques. — *a*) Cet énoncé présente l'avantage de ne plus faire intervenir de probabilités conditionnelles.

b) Si une suite de variables aléatoires $(X_n)_{n\geqslant 0}$ est telle que pour tout $N \geqslant 0$, $(X_n)_{0\leqslant n\leqslant N}$ est une chaîne de Markov (λ, P), alors il est clair que $(X_n)_{n\geqslant 0}$ est une chaîne de Markov (λ, P).

Démonstration. — Supposons que $(X_n)_{0\leqslant n\leqslant N}$ est une chaîne de Markov (λ, P). Alors, en utilisant l'identité $\mathbb{P}(A \cap B) = \mathbb{P}(B \,|\, A) \times \mathbb{P}(A)$ de manière répétée, on obtient

$$\begin{aligned}
\mathbb{P}\{X_0 = x_0, \ldots, X_N = x_N\} &= \mathbb{P}\{X_N = x_N \,|\, X_0 = x_0, \ldots, X_{N-1} = x_{N-1}\} \\
&\quad \times \mathbb{P}\{X_0 = x_0, \ldots, X_{N-1} = x_{N-1}\} \\
(\ldots) &= \mathbb{P}\{X_N = x_N \,|\, X_0 = x_0, \ldots, X_{N-1} = x_{N-1}\} \\
&\quad \times \mathbb{P}\{X_{N-1} = x_{N-1} \,|\, X_0 = x_0, \ldots, X_{N-2} = x_{N-2}\} \\
&\quad \cdots \times \mathbb{P}\{X_1 = x_1 \,|\, X_0 = x_0\} \times \mathbb{P}\{X_0 = x_0\} \\
(\text{Propriété de Markov}) &= \mathbb{P}\{X_N = x_N \,|\, X_{N-1} = x_{N-1}\} \\
&\quad \times \mathbb{P}\{X_{N-1} = x_{N-1} \,|\, X_{N-2} = x_{N-2}\} \\
&\quad \cdots \times \mathbb{P}\{X_1 = x_1 \,|\, X_0 = x_0\} \times \mathbb{P}\{X_0 = x_0\} \\
&= P(x_{N-1}, x_N) \times P(x_{N-2}, x_{N-1}) \times \cdots \times P(x_0, x_1) \times \lambda\{x_0\}.
\end{aligned}$$

donc (1.1) est vérifiée.

Réciproquement, supposons que l'égalité (1.1) soit vérifiée. Puisque $\sum_{y\in E} P(x,y) = 1$, on constate que (1.1) est aussi vérifiée avec $N' = N-1$, et ainsi en descendant, on constate que

$$\mathbb{P}\{X_0 = x_0, \dots, X_n = x_n\} = \lambda\{x_0\} \times P(x_0, x_1) \times \cdots \times P(x_{n-1}, x_n)$$

pour tous $0 \leqslant n \leqslant N$ et $x_0, \dots, x_n \in E$. Pour $n = 0$, on obtient que la loi de X_0 est λ : $\mathbb{P}\{X_0 = x\} = \lambda\{x\}$ pour tout $x \in E$. Pour $0 \leqslant n \leqslant N-1$, lorsque c'est défini, on a

$$\begin{aligned}\mathbb{P}\{X_{n+1} = x_{n+1} \,|\, X_0 = x_0, \dots, X_n = x_n\} &= \frac{\mathbb{P}\{X_0 = x_0, \dots, X_{n+1} = x_{n+1}\}}{\mathbb{P}\{X_0 = x_0, \dots, X_n = x_n\}} \\ &= \frac{\lambda\{x_0\} \times P(x_0, x_1) \times \cdots \times P(x_n, x_{n+1})}{\lambda\{x_0\} \times P(x_0, x_1) \times \cdots \times P(x_{n-1}, x_n)} = P(x_n, x_{n+1}).\end{aligned}$$

Donc, $(X_n)_{0\leqslant n\leqslant N}$ est une chaîne de Markov (λ, P). □

Exemples. — a) Les marches aléatoires dans $\mathbb{Z}$. — Soient $X_0, \xi_1, \dots, \xi_n, \dots$ des variables aléatoires indépendantes à valeurs dans $\mathbb{Z}$. Pour tout $n \geqslant 1$, on pose $X_n = X_0 + \xi_1 + \cdots + \xi_n$. La suite de variables aléatoires $(X_n)_{n\geqslant 0}$ est une chaîne de Markov à valeurs dans $\mathbb{Z}$ non nécessairement homogène.

C'est une chaîne de Markov homogène dès que les $(\xi_n)_{n\geqslant 1}$ ont même loi $\nu = (\nu\{x\})_{x\in\mathbb{Z}}$. Sa matrice de transition P a pour coefficients $P(x,y) = \nu\{y-x\}$, $x, y \in \mathbb{Z}$.

Vérifions cela. Soient $x_0, \dots, x_{n+1} \in \mathbb{Z}$. Sous réserve que la probabilité conditionnelle suivante soit bien définie, on a

$$\begin{aligned}\mathbb{P}\{X_{n+1} = x_{n+1} \,|\, X_0 = x_0, \dots, X_n = x_n\} &= \mathbb{P}\{X_n + \xi_{n+1} = x_{n+1} \,|\, X_0 = x_0, \dots, X_n = x_n\} \\ &= \mathbb{P}\{x_n + \xi_{n+1} = x_{n+1} \,|\, X_0 = x_0, \dots, X_n = x_n\} \\ &= \mathbb{P}\{\xi_{n+1} = x_{n+1} - x_n \,|\, X_0 = x_0, \dots, X_n = x_n\} \\ &= \mathbb{P}\{\xi_{n+1} = x_{n+1} - x_n\},\end{aligned}$$

la dernière égalité résultant de l'indépendance de ξ_{k+1} et $(X_0, \dots, X_n)$. Comme, de même, par indépendance de ξ_{k+1} et de X_n, on a

$$\mathbb{P}\{X_{n+1} = x_{n+1} \,|\, X_n = x_n\} = \mathbb{P}\{\xi_{n+1} = x_{n+1} - x_n \,|\, X_n = x_n\} = \mathbb{P}\{\xi_{n+1} = x_{n+1} - x_n\},$$

on en déduit que

$$\mathbb{P}\{X_{n+1} = x_{n+1} \,|\, X_0 = x_0, \dots, X_n = x_n\} = \mathbb{P}\{X_{n+1} = x_{n+1} \,|\, X_n = x_n\}.$$

La chaîne est donc bien de Markov. Si de plus les $(\xi_n)_{n\geqslant 1}$ ont même loi ν, on a

$$\mathbb{P}\{X_{n+1} = y \,|\, X_n = x\} = \mathbb{P}\{\xi_{n+1} = y - x\} = \nu\{y - x\},$$

ce qui prouve ce qui était annoncé.

Tout ceci se généralise à toute situation multidimensionnelle discrète avec les mêmes arguments.

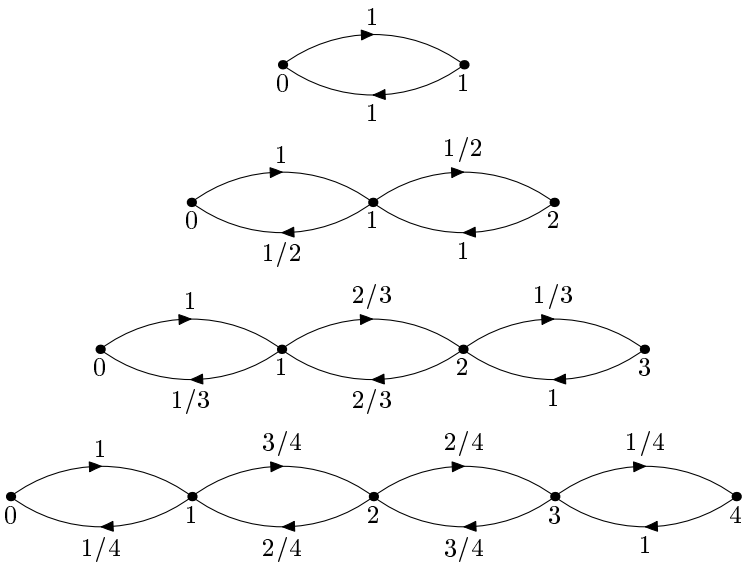
b) L'urne d'Ehrenfest. — Nous considérons deux urnes A et B et N boules numérotées de 1 à N réparties dans les deux urnes. On tire uniformément un nombre entre 1 et N. La boule portant ce numéro est alors déplacée dans l'autre urne (si elle est dans A, elle va dans B, si elle est dans B, elle va dans A). Et on continue...

Soit X_n le nombre de boules dans l'urne A à l'étape n, qui est une variable aléatoire à valeurs dans $E = \{0, 1, \dots, N\}$. En admettant que $(X_n)_{n\geqslant 0}$ est bien une chaîne de Markov

homogène, sa matrice de transition s'écrit alors

$$P = \begin{pmatrix} 0 & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ \frac{1}{N} & 0 & \frac{N-1}{N} & \ddots & & & \vdots \\ 0 & \frac{2}{N} & 0 & \frac{N-2}{N} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \frac{N-2}{N} & 0 & \frac{2}{N} & 0 \\ \vdots & & & \ddots & \frac{N-1}{N} & 0 & \frac{1}{N} \\ 0 & \cdots & \cdots & \cdots & 0 & 1 & 0 \end{pmatrix}.$$

Et voici quelques diagrammes pour $N = 1, \ldots, 4$:



Proposition 1. — (i) *Soient $U = (U_1, \ldots, U_n)$ un n-uplet de variables aléatoires indépendantes à valeurs dans un espace mesurable $(F, \mathcal{F})$, et $f_1, \ldots, f_n : F \times E \to E$ des applications mesurables. Soit X_0 une variable aléatoire à valeurs dans E indépendante de U. En posant*

$$X_k = f_k(U_k, X_{k-1}), \qquad 1 \leqslant k \leqslant n,$$

alors $(X_k)_{0 \leqslant k \leqslant n}$ est une chaîne de Markov finie.

(ii) *Réciproquement, on peut réaliser toute chaîne de Markov de cette manière : étant donnée la loi d'une chaîne de Markov $(Y_0, \ldots, Y_n)$, c'est-à-dire la donnée de la mesure de probabilité initiale λ, et des matrices de transition $P^{(0,1)}, \ldots, P^{(n-1,n)}$, il est possible de construire de la manière précédente une chaîne de Markov ayant même loi.*

Démonstration. — (i) Pour tout $k \in \{0, \ldots, n-1\}$, X_k est une fonction de $(X_0, U_1, \ldots, U_k)$ et est donc indépendante de U_{k+1}. On a alors

$$\begin{aligned} \mathbb{P}\{X_{k+1} = x_{k+1} \,|\, X_0 = x_0, \ldots, X_k = x_k\} &= \mathbb{P}\{f_k(U_{k+1}, X_k) = x_{k+1} \,|\, X_0 = x_0, \ldots, X_k = x_k\} \\ &= \mathbb{P}\{f_k(U_{k+1}, x_k) = x_{k+1} \,|\, X_0 = x_0, \ldots, X_k = x_k\} \\ \text{(indépendance de } U_{k+1} \text{ et } (X_0, \ldots, X_k)) &= \mathbb{P}\{f_k(U_{k+1}, x_k) = x_{k+1}\} \\ \text{(indépendance de } U_{k+1} \text{ et } X_k) &= \mathbb{P}\{f_k(U_{k+1}, x_k) = x_{k+1} \,|\, X_k = x_k\} \\ &= \mathbb{P}\{f_k(U_{k+1}, X_k) = x_{k+1} \,|\, X_k = x_k\} \\ &= \mathbb{P}\{X_{k+1} = x_{k+1} \,|\, X_k = x_k\}. \end{aligned}$$

D'où la propriété de Markov.

(ii) Réciproquement, on se donne X_0 et $(U_1, \dots, U_n)$ indépendantes telles que X_0 est à valeurs dans E et de loi λ et chaque U_k est à valeurs dans $F = E^E$ muni de $\mathcal{F} = \mathcal{P}(E)^{\otimes E}$ sa tribu produit (tout point $u \in F$ est une application de E dans E) et dont la loi satisfait la condition

$$\mathbb{P}\{U_k(x) = y\} = P^{(k-1,k)}(x, y), \qquad x, y \in E$$

(il suffit pour cela que la loi de U_k soit la loi produit $\bigotimes_{x\in E} P^{(k-1,k)}(x, .)$). On définit ensuite $f : F \times E \to E$ par $f(u, x) = u(x)$ qui est bien entendu mesurable pour la tribu produit. La suite $(X_0,\ X_k = f(U_k, X_{k-1}), 1 \leqslant k \leqslant n)$ est markovienne d'après le point (i), et, pour x, $y \in E$ et $1 \leqslant k \leqslant n$, on a

$$\begin{aligned}\mathbb{P}\{X_k = y \,|\, X_{k-1} = x\} &= \mathbb{P}\{f(U_k, X_{k-1}) = y \,|\, X_{k-1} = x\} \\ &= \mathbb{P}\{f(U_k, x) = y \,|\, X_{k-1} = x\} = \mathbb{P}\{f(U_k, x) = y\} = \mathbb{P}\{U_k(x) = y\} = P^{(k-1,k)}(x, y)\end{aligned}$$

où on se sera servi du fait que U_k et X_{k-1} sont indépendantes. □

Remarques. — *a*) Dans la construction précédente nous n'avons pas explicité sur quel espace probabilisé vivaient X_0 et $(U_1, \dots, U_n)$. Il suffit de prendre $\Omega = E \otimes F \otimes \cdots \otimes F$ muni de sa tribu produit et de la mesure de probabilité

$$\lambda \otimes \left(\bigotimes_{x\in E} P^{(0,1)}(x, .)\right) \otimes \cdots \otimes \left(\bigotimes_{x\in E} P^{(n-1,n)}(x, .)\right),$$

les variables aléatoires étant alors les projections canoniques correspondant à ce produit.

b) Lorsque E est fini, E^E est lui aussi fini et sa tribu produit est sa tribu discrète. En revanche ce n'est plus le cas si E est infini dénombrable, E^E a alors la puissance du continu (le même cardinal que $\mathbb{R}$), mais cela reste suffisamment raisonnable pour même dans ce cas effectuer des simulations réalistes sur ordinateur.

c) Le point (ii) montre qu'avec la donnée de λ, $P^{(0,1)}, \dots, P^{(n-1,n)}$, on peut construire une chaîne de Markov ayant cette loi initiale et ces matrices de transition. C'est donc un *théorème d'existence.*

Le théorème suivant traduit la propriété de Markov comme une propriété d'absence de mémoire, c'est-à-dire que si on se décale dans le temps, l'évolution ultérieure ne tiendra compte que de l'état à cet instant là sans tenir compte des états antérieurs.

Théorème 2. — *Soit $(X_n)_{n\geqslant 0}$ une chaîne de Markov (λ, P). Alors pour tout $x \in E$ et $m \geqslant 0$, conditionnellement à $\{X_m = x\}$, $(X_{n+m})_{n\geqslant 0}$ est une chaîne de Markov $(\delta_{\{x\}}, P)$ qui est indépendante de $(X_0, \dots, X_m)$.*

Démonstration. — Il s'agit de montrer que pour tout événement de la forme

$$A = \{X_0 = x_0, \dots, X_m = x_m\}$$

on a

$$\begin{aligned}&\mathbb{P}\big(\{X_m = x_m, \dots, X_{m+n} = x_{m+n}\} \cap A \;\big|\; X_m = x\big) \\ &\qquad = \delta(x, x_m) \times P(x_m, x_{m+1}) \times \cdots \times P(x_{m+n-1}, x_{m+n}) \times \mathbb{P}(A \,|\, X_m = x),\end{aligned}$$

où $\delta(x, x_m)$ vaut 1 si $x = x_m$, 0 sinon (symbole de Kronecker). En multipliant le premier membre par $\mathbb{P}\{X_m = x\}$, on obtient clairement

$$\lambda\{x_0\} \times P(x_0, x_1) \times \cdots \times P(x_{m+n-1}, x_{m+n}) \times \delta(x, x_m).$$

En faisant de même pour le second membre, on obtient le même résultat. D'où l'égalité annoncée. □

Loi de X_n en fonction de λ et de P. — Nous rappelons qu'une mesure de probabilité λ sur E est identifiée à un vecteur ligne indexé par E, et qu'une matrice stochastique P est ainsi une matrice indexée par $E \times E$. Ainsi le produit matriciel λP est une matrice ligne et on vérifie sans peine qu'elle correspond à une mesure de probabilité sur E :

$$(\lambda P)\{y\} = \sum_{x\in E} \lambda\{x\} \times P(x,y) \geqslant 0$$

et

$$\sum_{y\in E} (\lambda P)\{y\} = \sum_{y\in E}\sum_{x\in E} \lambda\{x\} \times P(x,y) = \sum_{x\in E} \lambda\{x\} \sum_{y\in E} P(x,y) = \sum_{x\in E} \lambda\{x\} = 1.$$

De même, il est facile de vérifier que le produit de deux matrices stochastiques est une matrice stochastique : si Q est une matrice stochastique indexée par $E \times E$,

$$(PQ)(x,y) = \sum_{z\in E} P(x,z)Q(z,y) \geqslant 0$$

et

$$\sum_{y\in E} (PQ)(x,y) = \sum_{y\in E}\sum_{z\in E} P(x,z)Q(z,y) = \sum_{z\in E} P(x,z) \sum_{y\in E} Q(z,y) = \sum_{z\in E} P(x,z) = 1.$$

Ainsi P^n est une matrice stochastique pour tout $n \geqslant 0$, où on aura convenu que $P^0 = \mathrm{Id}$. Nous omettrons certaines parenthèses lorsqu'il n'y aura pas d'ambiguïté. Ainsi on notera par exemple $\lambda P\{x\}$ au lieu de $(\lambda P)\{x\}$ et $PQ(x,y)$ au lieu de $(PQ)(x,y)$.

Lorsque $\lambda\{x\} = \mathbb{P}\{X_0 = x\} > 0$, on note $\mathbb{P}^x = \mathbb{P}\{\,.\,|\,X_0 = x\}$. Par la propriété de Markov, $(X_n)_{n\geqslant 0}$ est une chaîne de Markov $(\delta_{\{x\}}, P)$ sous la mesure de probabilité $\mathbb{P}^x$, notamment sa loi ne dépend plus de λ.

Théorème 3. — *Soit $(X_n)_{n\geqslant 0}$ est une chaîne de Markov (λ, P). Pour tous m, $n \geqslant 0$, on a*

(i) $\mathbb{P}\{X_n = y\} = (\lambda P^n)\{y\}$;

(ii) $\mathbb{P}^x\{X_n = y\} = \mathbb{P}\{X_{m+n} = y \,|\, X_m = x\} = P^n(x,y)$.

Démonstration. — (i) L'événement $\{X_n = y\}$ se décompose en la réunion disjointe d'événements $\{X_0 = x_0, \dots, X_{m-1} = x_{m-1}, X_n = y\}$, $x_0, \dots, x_{m-1} \in E$, et on connaît la probabilité de chacun d'entre eux. On a donc

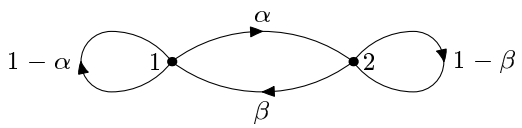
$$\begin{aligned}
\mathbb{P}\{X_n = y\} &= \mathbb{P}\Bigg(\bigcup_{x_0,\dots,x_{m-1}\in E} \{X_0 = x_0, \dots, X_{m-1} = x_{m-1}, X_n = y\}\Bigg) \\
&= \sum_{x_0,\dots,x_{m-1}\in E} \mathbb{P}\{X_0 = x_0, \dots, X_{m-1} = x_{m-1}, X_n = y\} \\
&= \sum_{x_0,\dots,x_{m-1}\in E} \lambda\{x_0\} \times P(x_0,x_1) \times \cdots \times P(x_{m-2},x_{m-1}) \times P(x_{m-1},y) \\
&= \lambda P^m\{y\},
\end{aligned}$$

puisqu'on reconnaît dans la dernière expression le coefficient d'indice y de la matrice ligne λP^m.

(ii) Par la propriété de Markov, conditionnellement à $\{X_m = x\}$, la suite de variables aléatoires $(X_{m+n})_{n\geqslant 0}$ est une chaîne de Markov $(\delta_{\{x\}}, P)$. On obtient alors le résultat en prenant $\lambda = \delta_{\{x\}}$ dans (i). □

Exemple. — Une chaîne de Markov homogène à 2 états a une matrice et un diagramme de transition de la forme

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$



avec α, $\beta \in [0,1]$. Comme $P^{n+1} = P^n \times P$, $n \geqslant 0$, on a

$$P^{n+1}(1,1) = P^n(1,1) \times (1-\alpha) + P^n(1,2) \times \beta\,;$$

puisque $P^n(1,1) + P^n(1,2) = 1$, en éliminant $P^n(1,2)$, on obtient

$$P^{n+1}(1,1) = P^n(1,1) \times (1-\alpha-\beta) + \beta.$$

En se souvenant (ou le redémontrant), qu'une suite récurrente réelle $(u_n)_{n\geqslant 0}$ telle que $u_{n+1} = au_n + b$, $n \geqslant 0$, vérifie pour tout $n \geqslant 0$

$$u_n = a^n u_0 + a^{n-1}b + \cdots + ab + b = a^n u_0 + \frac{1-a^n}{1-a} \times b = \frac{b}{1-a} + \frac{(1-a)u_0 - b}{1-a} \times a^n,$$

lorsque $a \neq 1$. On obtient ici, puisque $P^0(1,1) = 1$, lorsque $\alpha + \beta > 0$,

$$P^n(1,1) = \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n, \qquad \text{pour tout } n \geqslant 0,$$

(si $\alpha+\beta = 0$, $P^n = \mathrm{Id}$ pour tout $n \geqslant 0$). On en déduit immédiatement $P^n(1,2) = 1 - P^n(1,1)$. Les coefficients $P^n(2,1)$ et $P^n(2,2)$ peuvent être calculés de la même façon.

Exercice 2. — On considère une chaîne de Markov à 3 états $\{1,2,3\}$ de matrice de transition

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

Trouver une formule générale pour la suite $(P^n)(1,1))_{n\geqslant 0}$ (*indication :* on pourra se servir des valeurs propres de P).

Exercice 3. — Soient $(X_n)_{n\geqslant 0}$ est une chaîne de Markov (λ, P) et k entier strictement positif. Montrer que si on pose $Y_n = X_{kn}$, pour tout $n \geqslant 0$, alors la suite de variables aléatoires $(Y_n)_{n\geqslant 0}$ est une chaîne de Markov (λ, P^k).

Exercice 4. — *Marches aléatoires simples dans* $\mathbb{Z}$. — On considère pour espace d'états $E = \mathbb{Z}$ et $p \in [0,1]$ un réel. En posant $q = 1-p$, on définit une matrice de transition P par

$$P(x,y) = \begin{cases} p & \text{si } y = x+1 \\ q & \text{si } y = x-1 \\ 0 & \text{sinon.} \end{cases}$$

Montrer que

$$P^n(x,y) = \begin{cases} \dbinom{n}{(n+y-x)/2} p^{(n+y-x)/2} q^{(n-y-x)/2} & \text{si } n+y-x \text{ est pair} \\ 0 & \text{sinon,} \end{cases}$$

où

$$\binom{n}{k} = \frac{n\,!}{k\,!\,(n-k)\,!}$$

pour $k \in \{0, \ldots, n\}$ et 0 sinon.

2. Classes d'équivalence d'une chaîne de Markov

Nous considérons une chaîne de Markov $(X_n)_{n\geqslant 0}$ d'espace d'états E et de matrice de transition P. Nous supposerons pouvoir envisager toute loi initiale, en particulier les mesures de probabilité $\delta_{\{x\}}$ et donc les mesures de probabilité (conditionnelles) $\mathbb{P}^x$ pour chacunes desquelles $(X_n)_{n\geqslant 0}$ est une chaîne de Markov $(\delta_{\{x\}}, P)$.

La vérité est que nous nous donnons deux espaces mesurables $(\Omega, \mathcal{A})$, $(E, \mathcal{E})$, une suite d'applications mesurables $(X_n : (\Omega, \mathcal{A}) \to (E, \mathcal{E}))_{n\geqslant 0}$ et une famille de mesures de probabilité $(\mathbb{P}^x)_{x\in E}$ tels que pour tout $x \in E$, vu sous $\mathbb{P}^x$, $(X_n)_{n\geqslant 0}$ est une chaîne de Markov de loi initiale $\delta_{\{x\}}$ et de matrice de transition P. Dans ce cadre, la réalisation d'une chaîne de Markov de loi initiale π s'obtient en regardant le schéma initial de ce cadre sous la mesure initiale $\mathbb{P}^\pi$ qui est égale à la combinaison correspondantes des lois initiales $\mathbb{P}^x$.

Définition 3. — Soient x et $y \in E$ deux états.

(i) On dit que x *mène* à y, et on note $x \rightsquigarrow y$, si et seulement si $\mathbb{P}^x\{\exists\, n \geqslant 0 : X_n = y\} > 0$.

(ii) On dit que x et y *communiquent*, et on note $x \leftrightsquigarrow y$, si et seulement si $x \rightsquigarrow y$ et $y \rightsquigarrow x$.

Théorème 4. — *Pour deux états x et $y \in E$, les assertions suivantes sont équivalentes :*

(i) $x \rightsquigarrow y$;

(ii) *il existe une suite finie d'états $x = x_0, x_1, \ldots, x_{n-1}, x_n = y \in E$ tels que $P(x_0, x_1) \times \cdots \times P(x_{n-1}, x_n) > 0$;*

(iii) *il existe $n \geqslant 0$ tel que $P^n(x, y) > 0$.*

Démonstration. — Supposons (i) vérifié. Alors on a

$$\begin{aligned}
0 < \mathbb{P}^x\{\exists\, n \geqslant 0 : X_n = y\} &= \mathbb{P}^x\left(\bigcup_{n\geqslant 0}\{X_n = y\}\right) \leqslant \sum_{n\geqslant 0}\mathbb{P}^x\{X_n = y\} \\
&= \sum_{n\geqslant 0}\mathbb{P}^x\left(\bigcup_{x_1,\ldots,x_{n-1}\in E}\{X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = y\}\right) \\
&= \sum_{n\geqslant 0}\sum_{x_1,\ldots,x_{n-1}\in E}\mathbb{P}^x\{X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = y\} \\
&= \sum_{n\geqslant 0}\sum_{x_1,\ldots,x_{n-1}\in E} P(x, x_1) \times \cdots \times P(x_{n-1}, y).
\end{aligned}$$

L'un au moins des produits figurant dans cette somme multiple est strictement positif, le point (ii) en résulte.

Supposons le point (ii) vérifié. Puisque, par produit matriciel,

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

si l'un des termes de la somme est strictement positif, il en est de même de $P^n(x, y)$. Donc le point (iii) est satisfait.

Supposons le point (iii) vérifié pour un certain $n \geqslant 0$. On a alors

$$\mathbb{P}^x\{\exists\, m \geqslant 0 : X_m = y\} = \mathbb{P}^x\left(\bigcup_{m\geqslant 0}\{X_m = y\}\right) \geqslant \mathbb{P}^x\{X_n = y\} = P^n(x, y) > 0.$$

Donc le point (i) est alors satisfait. □

En considérant $n = 0$, on a clairement $x \rightsquigarrow x$ pour tout $x \in E$, c'est-à-dire que la relation « mener à » est réflexive. De plus avec le point (ii) du théorème précédent, il est clair que cette relation est transitive : si $x \rightsquigarrow y$ et $y \rightsquigarrow z$, soient une suite finie d'états $x = x_0, x_1, \ldots, x_{n-1}, x_n = y, x_{n+1}, \ldots, x_{m+n-1}, x_{m+n} = z \in E$ tels que

$$P(x_0, x_1) \times \cdots \times P(x_{n-1}, x_n) > 0 \qquad \text{et} \qquad P(x_n, x_{n+1}) \times \cdots \times P(x_{m+n-1}, x_{m+n}) > 0,$$

on a alors

$$P(x_0, x_1) \times \cdots \times P(x_{n-1}, x_n) \times P(x_n, x_{n+1}) \times \cdots \times P(x_{m+n-1}, x_{m+n}) > 0$$

et donc $x \rightsquigarrow z$.

La relation « communiquer » est elle aussi réflexive et transitive, comme elle est par construction symétrique, c'est une relation d'équivalence.

Définition 4. — (i) Les classes d'équivalence pour la relation « communiquer » sur E sont appelées *classes communiquantes*.

(ii) Une classe communiquante $C \subset E$ est *fermée* si et seulement si

$$x \in C \text{ et } x \rightsquigarrow y \quad \Longrightarrow \quad y \in C,$$

autrement dit, si pour tout $y \in E \backslash C$, $\mathbb{P}^x\{\exists\, n \geqslant 0 : X_n = y\} = 0$: la chaîne issue de x, une fois rentrée dans C, ne peut plus en sortir.

(iii) Un état $x \in E$ est *absorbant* si le singleton $\{x\}$ est une classe communiquante fermée, c'est-à-dire si $P(x, x) = 1$ et évidemment $P(x, y) = 0$ pour tout $y \neq x$.

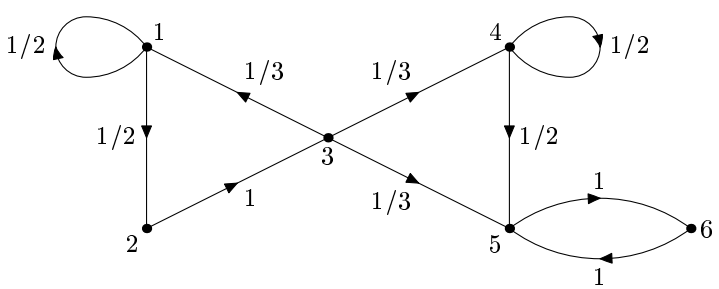
(iv) Une chaîne de Markov, ou une matrice de transition, est *irréductible* si et seulement si il n'existe qu'une seule classe communiquante, l'espace d'états E tout entier.

Exercice 5. — Dans le cas d'une chaîne de Markov à deux états, discuter en fonction de α et β des relations entre les deux points, des classes communiquantes, des classes fermées. Faire de même dans le cas de l'urne d'Ehrenfest pour $N \in \mathbb{N}$ quelconque.

Exemple. — Sur l'espace d'états $E = \{1, 2, 3, 4, 5, 6\}$, considérons la matrice stochastique

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

Elle correspond au diagramme de transition suivant :

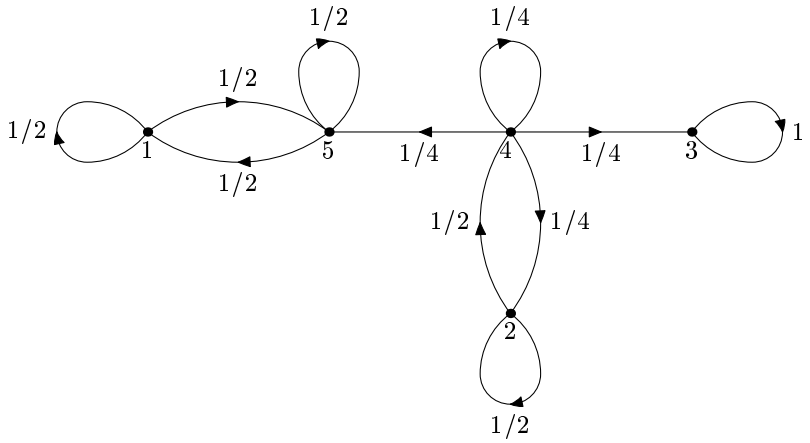


Les classes comuniquantes sont alors $\{1, 2, 3\}$, $\{4\}$ et $\{5, 6\}$, et seule cette dernière classe est fermée.

Exercice 6. — Identifier les classes communiquantes, en précisant lesquelles sont fermées, associées à

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

Réponse. — Le diagramme de transition associé à la matrice de transition précédente est



Les classes communiquantes sont $\{1,5\}$, $\{3\}$, $\{2,4\}$, et seules les deux premières sont fermées.

Exercice 7. — Montrer que toute matrice de transition sur un espace d'état fini a au moins une classe fermée. Trouver un exemple de matrice de transition sans classe fermée.

3. Temps d'atteinte et probabilités d'absorption

Définition 5. — Soit A une partie de E. Le *temps d'atteinte* par $(X_n)_{n\geqslant 0}$ de A est la variable aléatoire $H^A : (\Omega, \mathcal{A}, \mathbb{P}) \to \bar{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$ définie par

$$H^A(\omega) = \inf\{n \geqslant 0 : X_n(\omega) \in A\}, \qquad \omega \in \Omega,$$

où on convient que $\inf \varnothing = \infty$. La probabilité partant de $x \in E$ que $(X_n)_{n\geqslant 0}$ atteigne A est

$$h_x^A = \mathbb{P}^x\{H^A < \infty\}.$$

Lorsque A est une classe fermée, les probabilités h_x^A sont appelées probabilités d'absorption. Le *temps moyen* partant de $x \in E$ pris par $(X_n)_{n\geqslant 0}$ pour atteindre A est

$$k_x^A = \mathbb{E}^x\big[H^A\big] = \sum_{n\geqslant 0} n \times \mathbb{P}^x\{H^A = n\} + \infty \times \mathbb{P}\{H^A = \infty\}.$$

Théorème 5. — *Le vecteur des probabilités d'atteinte $h^A = (h_x^A)_{x\in E}$ est la solution positive minimale du système*

$$h_x^A = \begin{cases} 1 & \text{si } x \in A \\ \displaystyle\sum_{y\in E} P(x,y) \times h_y^A & \text{si } x \notin A. \end{cases} \tag{1.2}$$

La minimalité signifiant que si $h = (h_x)_{x\in E} \in \mathbb{R}_+^E$ est une autre solution positive de (1.2)*, alors $h_x \geqslant h_x^A$ pour tout $x \in E$.*

Démonstration. — Montrons tout d'abord que h^A vérifie le système (1.2). Si $X_0 = x \in A$, alors $H^A = 0$ et donc $h_x^A = 1$. Si $X_0 = x \notin A$, on a alors $H^A \geqslant 1$ et ainsi

$$\begin{aligned} h_x^A &= \mathbb{P}^x\{H^A < \infty\} = \sum_{y\in E} \mathbb{P}^x\{H^A < \infty, X_1 = y\} \\ &= \sum_{y\in E} \mathbb{P}^x\{X_1 = y\} \times \mathbb{P}^x\{H^A < \infty \,|\, X_1 = y\} \\ &= \sum_{y\in E} \mathbb{P}^x\{X_1 = y\} \times \mathbb{P}^y\{H^A < \infty\} = \sum_{y\in E} P(x,y) \times h_y^A \end{aligned}$$

l'avant dernière égalité résultant de l'application de la propriété de Markov.

Supposons que $h = (h_x)_{x\in E}$ soit une solution positive de (1.2). On a donc $h_x = 1$ pour tout $x \in A$ et si $x \notin A$

$$h_x = \sum_{y\in E} P(x,y) \times h_y = \sum_{y\in A} P(x,y) \times h_y + \sum_{y\notin A} P(x,y) \times h_y = \sum_{y\in A} P(x,y) + \sum_{y\notin A} P(x,y) \times h_y.$$

En itérant,

$$\begin{aligned} h_x &= \sum_{y\in A} P(x,y) + \sum_{y\notin A} P(x,y)\Bigg(\sum_{z\in A} P(y,z) + \sum_{z\notin A} P(y,z) \times h_z\Bigg) \\ &= \sum_{y\in A} P(x,y) + \sum_{y\notin A, z\in A} P(x,y) \times P(y,z) + \sum_{y\notin A, z\notin A} P(x,y) \times P(y,z) \times h_z \\ &= \mathbb{P}^x\{X_1 \in A\} + \mathbb{P}^x\{X_1 \notin A, X_2 \in A\} + \sum_{y\notin A, z\notin A} P(x,y) \times P(y,z) \times h_z. \end{aligned}$$

En itérant encore, on obtient ainsi pour tout $n \geqslant 1$,

$$\begin{aligned} h_x = \mathbb{P}^x\{X_1 \in A\} + \cdots + \mathbb{P}^x\{X_1 \notin A, \ldots, X_{n-1} \notin A, X_n \in A\} \\ + \sum_{y_1\notin A, \ldots, y_n\notin A} P(x,y_1) \times \cdots \times P(y_{n-1}, y_n) \times h_{y_n}. \end{aligned}$$

Puisque tous les h_x sont positifs, on a alors

$$h_x \geqslant \mathbb{P}^x\{X_1 \in A\} + \cdots + \mathbb{P}^x\{X_1 \notin A, \ldots, X_{n-1} \notin A, X_n \in A\} = \mathbb{P}^x\{H^A \leqslant n\}$$

pour tout $n \in \mathbb{N}$. Puisque la suite de ces probabilités tend en croissant vers $h_x^A = \mathbb{P}^x\{H^A < \infty\}$, on a en passant à la limite sur l'inégalité ci-dessus $h_x \geqslant h_x^A$. □

Exercice 8 (ruine du joueur). — On considère une chaîne de Markov $(X_n)_{n\geqslant 0}$ d'espace d'état $E = \mathbb{N}$ et de matrice de transition P vérifiant

$$P(0,0) = 1, \qquad \text{et} \qquad P(x,x+1) = p, \quad P(x,x-1) = 1-p = q, \quad \text{pour tout } x \in \{1,2,\ldots\},$$

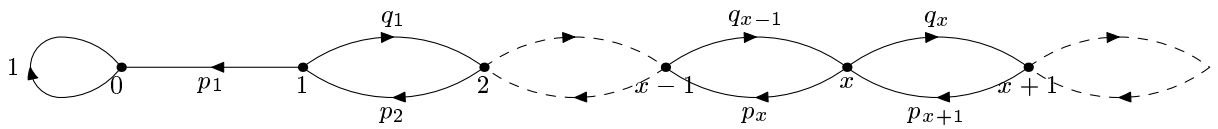
les autres coefficients étant nuls, $p \in [0,1]$ étant un réel donné. La variable X_n représente la fortune d'un joueur à l'instant n jouant contre un adversaire infiniment riche, la probabilité $h_x = h_x^0$ est alors la probabilité de perdre avec une fortune initiale $x \in \mathbb{N}$.

Montrer que si $p \leqslant q$, $h_x = 1$ pour tout $x \in \mathbb{N}$ (distinguer les cas $p < q$ et $p = q = 1/2$), et si $p > q$, $h_x = (q/p)^x$. On établira une relation de récurrence entre les h_x.

Exercice 9 (naissance et mort). — On considère $E = \mathbb{N}$, $(p_x)_{x\geqslant 1}$ une suite de réels dans $[0,1]$ et la matrice de transition P définie par

$$P(0,0) = 1, \qquad \text{et} \qquad P(x,x-1) = p_x, \quad P(x,x+1) = 1-p_x = q_x, \quad \text{pour tout } x \geqslant 1,$$

tous les autres coefficients étant nuls. Chaque variable X_n représente la taille de la population après n changements, si la population est de taille x, p_x est la probabilité qu'il y ait une mort, $q_x = 1 - p_x$ est celle qu'il y ait une naissance. La probabilité d'extinction de la population est alors $h_x = h_x^0$ lorsque la taille de la population est initialement x.



En posant $u_x = h_{x-1} - h_x$, montrer que $p_x \times u_{x+1} = q_x \times u_x$, puis que

$$u_{x+1} = \frac{q_x}{p_x}\, u_x = \frac{q_x \times \cdots \times q_1}{p_x \times \cdots \times p_1}\, u_x = \gamma_x \times u_1.$$

En constatant que $u_1 + \cdots + u_x = h_0 - h_x$, en déduire que

$$h_x = 1 - u_1(\gamma_0 + \cdots + \gamma_{x-1})$$

où $\gamma_0 = 1$. Il reste à déterminer u_1.

Si $\sum_{x=0}^{\infty} \gamma_x = \infty$, montrer que $u_1 = 0$ et qu'ainsi $h_x = 1$ pour tout x.

Si $\sum_{x=0}^{\infty} \gamma_x < \infty$, montrer que $u_1 = 1/\sum_{x=0}^{\infty} \gamma_x$ est le choix qui donne les γ_x et que

$$h_x = \frac{\sum_{y=x}^{\infty} \gamma_y}{\sum_{y=0}^{\infty} \gamma_y}.$$

Ainsi, lorsque $x \geqslant 1$, on a $h_x < 1$, donc la population survit avec une probabilité strictement positive.

Nous rappelons qu'on note $k_x^A = \mathbb{E}^x[H^A]$ le temps moyen d'atteinte de A en partant de x.

THÉORÈME 6. — *Le vecteur des temps moyens d'atteinte $k^A = (k_x^A)_{x\in E}$ est la solution positive minimale du système*

$$k_x^A = \begin{cases} 0 & \text{si } x \in A \\ 1 + \displaystyle\sum_{y\in E} P(x,y) \times k_y^A & \text{si } x \notin A. \end{cases} \tag{1.3}$$

La minimalité signifiant que si $k = (k_x)_{x\in E} \in \mathbb{R}_+^E$ est une autre solution positive de (1.3)*, alors $k_x \geqslant k_x^A$ pour tout $x \in E$.*

Remarque. — Si $A \in \mathcal{A}$ est un événement de probabilité strictement positive et $H : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ est une variable aléatoire, on pose

$$\mathbb{E}[H \,|\, A] = \mathbb{E}[H \times \mathbb{1}_A]/\,\mathbb{P}(A)$$

qui est appelée espérance de H sachant A. On constate aisément que c'est l'espérance de H sous la mesure de probabilité $\mathbb{P}(\,.\,|\,A)$. Si $(A_n)_n$ est un système complet d'événements et si $\mathbb{E}[H]$ a un sens, on peut montrer que

$$\mathbb{E}[H] = \sum_n \mathbb{E}[H \times \mathbb{1}_{A_n}] = \sum_n \mathbb{P}(A_n) \times \mathbb{E}[H \,|\, A_n]$$

et on convient que cette dernière somme a toujours un sens même si pour certains n, $\mathbb{P}(A_n) = 0$, auquel cas $\mathbb{P}(A_n) \times \mathbb{E}[H \,|\, A_n] = \mathbb{E}[H \times \mathbb{1}_{A_n}] = 0$.

Démonstration. — Montrons tout d'abord que k^A vérifie le système (1.3). Si $X_0 = x \in A$, alors $H^A = 0$ et donc $k_x^A = 0$. Si $X_0 = x \notin A$, on a alors $H^A \geqslant 1$ et ainsi

$$\begin{aligned} k_x^A &= \mathbb{E}^x\big[H^A\big] = \sum_{y\in E} \mathbb{E}^x\big[H^A \times \mathbb{1}_{\{X_1=y\}}\big] = \sum_{y\in E} \mathbb{P}^x\{X_1 = y\} \times \mathbb{E}^x\big[H^A \;\big|\; X_1 = y\big] \\ &= \sum_{y\in E} \mathbb{P}^x\{X_1 = y\} \times \mathbb{E}^y\big[1 + H^A\big] = \sum_{y\in E} \mathbb{P}^x\{X_1 = y\} \times \big(1 + \mathbb{E}^y\big[H^A\big]\big) \\ &= \sum_{y\in E} \mathbb{P}^x\{X_1 = y\} + \sum_{y\in E} \mathbb{P}^x\{X_1 = y\} \times \mathbb{E}^y\big[H^A\big] = 1 + \sum_{y\in E} P(x,y) \times k_y^A \end{aligned}$$

le fait que $\mathbb{E}^x[H^A \mid X_1 = y] = \mathbb{E}^y[1 + H^A]$ résultant de l'application de la propriété de Markov puisque $x \notin A$.

Supposons que $k = (k_x)_{x\in E}$ soit une solution positive de (1.3). On a donc $k_x = 0$ pour tout $x \in A$ et si $x \notin A$

$$k_x = 1 + \sum_{y\in E} P(x,y) \times k_y = 1 + \sum_{y\in A} P(x,y) \times k_y + \sum_{y\notin A} P(x,y) \times k_y = 1 + \sum_{y\notin A} P(x,y) \times k_y.$$

En itérant,

$$\begin{aligned} k_x &= 1 + \sum_{y\notin A} P(x,y)\left(1 + \sum_{z\notin A} P(y,z) \times k_z\right) \\ &= 1 + \sum_{y\notin A} P(x,y) + \sum_{y\notin A, z\notin A} P(x,y) \times P(y,z) \times k_z \\ &= \mathbb{P}^x\big\{H^A \geqslant 1\big\} + \mathbb{P}^x\big\{H^A \geqslant 2\big\} + \sum_{y\notin A, z\notin A} P(x,y) \times P(y,z) \times k_z. \end{aligned}$$

En itérant encore, on obtient ainsi pour tout $n \geqslant 1$,

$$k_x = \mathbb{P}^x\big\{H^A \geqslant 1\big\} + \cdots + \mathbb{P}^x\big\{H^A \geqslant n\big\} + \sum_{y_1\notin A, \ldots, y_n\notin A} P(x,y_1) \times \cdots \times P(y_{n-1},y_n) \times k_{y_n}.$$

Puisque tous les k_x sont positifs, on a alors

$$k_x \geqslant \mathbb{P}^x\big\{H^A \geqslant 1\big\} + \cdots + \mathbb{P}^x\big\{H^A \geqslant n\big\}$$

et ainsi

$$k_x \geqslant \sum_{n\geqslant 1} \mathbb{P}^x\big\{H^A \geqslant n\big\} = \mathbb{E}^x\big[H^A\big] = k_x^A,$$

d'où la conclusion. □

Exercice 10. — Démontrer le résultat utilisé à la fin de la démonstration précédente : si $H : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{N} \cup \{\infty\}$ est une variable aléatoire, on a

$$\mathbb{E}[H] = \sum_{n\geqslant 1} \mathbb{P}\{H \geqslant n\} = \sum_{n\geqslant 0} \mathbb{P}\{H > n\}.$$

On pourra distinguer les cas $\mathbb{P}\{H = \infty\} > 0$ et $\mathbb{P}\{H = \infty\} = 0$.

Exercice 11 (ruine du joueur, suite). — Montrer que si $p < q$ on a $k_x = x/(1-2p)$, et si $p \geqslant q$ on a $k_x = \infty$ (distinguer les cas $p > q$ et $p = q = 1/2$). On établira une relation de récurrence entre les k_x.

4. **La propriété de Markov forte**

La propriété de Markov, aussi appelée propriété de Markov simple, affirme que conditionnellement à $\{X_m = x\}$, $(X_{m+n})_{n\geqslant 0}$ se comporte de la même manière que $(X_n)_{n\geqslant 0}$ conditionnellement à $\{X_0 = x\}$. C'est une propriété d'invariance par translation selon temps constant m. Dans le cadre des chaînes de markov à temps discret et à espace d'états au plus dénombrable, cette propriété s'étend sans difficulté ou hypothèse supplémentaire à des temps aléatoires : les temps d'arrêt.

Sauf mention supplémentaire, on considère dans cette section une suite quelconque de variables aléatoires $X = (X_n)_{n\geqslant 0}$ définies sur um même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et prenant leurs valeurs dans des espaces mesurables quelconques. Pour $n \geqslant 0$, on note $\mathcal{F}_n^X = \sigma\{X_0, \ldots, X_n\}$ la tribu sur Ω engendrée par $X_1, \ldots, X_n$ qui est une sous-tribu de $\mathcal{A}$. La suite $(\mathcal{F}_n^X)_{n\geqslant 0}$ est une suite croissante de tribus, c'est une *filtration* qui est dite être la *filtration naturelle* de $X = (X_n)_{n\geqslant 0}$.

Définition 6. — Une variable aléatoire $T : (\Omega, \mathcal{A}, \mathbb{P}) \to \bar{\mathbb{N}}$ est un *temps d'arrêt* de la suite $X = (X_n)_{n\geqslant 0}$ si pour tout $n \geqslant 0$, l'événement $\{T \leqslant n\}$ appartient à la tribu $\mathcal{F}_n^X$, c'est-à-dire qu'il ne dépend que de $X_1, \ldots, X_n$.

Remarques. — *a*) En écrivant $\{T \leqslant n\} = \{T = 0\} \cup \ldots \cup \{T = n\}$, on constate que la condition précédente équivaut à demander que $\{T = n\} \in \mathcal{F}_n^X$ pour tout $n \geqslant 0$, ce qui est particulier à la situation où le temps est discret.

b) Si T est un temps d'arrêt, $\{T = \infty\}$ est évidemment mesurable puisqu'on suppose toujours que T est une variable aléatoire. On peut préciser que cet événement appartient à la tribu $\mathcal{F}_\infty^X = \bigvee_{n\geqslant 0} \mathcal{F}_n^X$ qui est évidemment incluse dans $\mathcal{A}$.

c) La définition des temps d'arrêt est liée à la filtration, ce qui suggère des généralisations en particulier quand le temps n'est plus $\mathbb{N}$ mais $\mathbb{R}_+$ ou tout ensemble ordonné. Par ailleurs, dans les énoncés qui suivent nous ne préciserons pas nécessairement la filtration ou, moins généralement, la suite de variables aléatoires, qui est, ou sont, sous-entendue.

Théorème 7. — (i) *Toute variable aléatoire $T : (\Omega, \mathcal{A}, \mathbb{P}) \to \bar{\mathbb{N}}$ constante est un temps d'arrêt.*

(ii) *Si $(T_k)_k$ est une famille finie ou dénombrable de temps d'arrêt, alors*

$$\inf_k T_k \qquad \textit{et} \qquad \sup_k T_k$$

sont des temps d'arrêt. En particulier, si S et T sont deux temps d'arrêt, alors $S \wedge T$ et $S \vee T$ sont des temps d'arrêt.

(iii) *Si T est un temps d'arrêt et $m \geqslant 0$, alors $T + m$ est un temps d'arrêt.*

Démonstration. — (i) Si T est constant égal à $m \in \bar{N}$, on a, pour tout $n \geqslant 0$, $\{T \leqslant n\} = \{m \leqslant n\} \in \{\varnothing, \Omega\} \subset \mathcal{F}_n^X$.

(ii) Pour $n \in \mathbb{N}$, on a

$$\{\inf_k T_k \leqslant n\} = \bigcup_k \{T_k \leqslant n\} \in \mathcal{F}_n^X \qquad \text{et} \qquad \{\sup_k T_k \leqslant n\} = \bigcap_k \{T_k \leqslant n\} \in \mathcal{F}_n^X.$$

(iii) Pour $n \in \mathbb{N}$, on a $\{T + m \leqslant n\} = \{T \leqslant n - m\}$. Ainsi, si $m > n$, $\{T + m \leqslant n\} = \varnothing \in \mathcal{F}_n^X$, et si $m \leqslant n$, $\{T + m \leqslant n\} \in \mathcal{F}_{n-m}^X \subset \mathcal{F}_n^X$. □

Exemple et contre-exemples. — *a*) Le temps d'atteinte H^A de $A \subset E$ par une suite de variables aléatoires $(X_n)_{n\geqslant 0}$ à valeurs dans E est un temps d'arrêt puisque, pour tout $n \geqslant 0$, $\{H^A \leqslant n\} = \bigcup_{k=0}^{n} \{X_k \in A\} \in \mathcal{F}_n^X$.

b) Le dernier temps de retour $L^A = \sup\{n > 0 : X_n \in A\}$ n'est généralement pas un temps d'arrêt car

$$\{L^A \leqslant n\} = \left(\bigcup_{k=1}^{n} \{X_k \in A\}\right) \cap \left(\bigcap_{k\geqslant n+1} \{X_z \notin A\}\right)$$

et donc dépend *a priori* de $\{X_{n+1}, X_{n+2}, \ldots\}$.

c) Si T est un temps d'arrêt et $m > 0$, en général $T - m$ ou $(T - m) \vee 0$ ne sont pas des temps d'arrêt.

Exercice 12. — Si S et T sont deux temps d'arrêt, montrer que $S + T$ est un temps d'arrêt. (*Avertissement :* cette propriété repose sur le fait que l'ensemble des temps (finis) est l'ensemble des entiers naturels.)

Définition 7. — Soit T un temps d'arrêt d'une suite de variables aléatoires $X = (X_n)_{n\geqslant 0}$. On appelle *tribu des événements antérieurs à* T la tribu $\mathcal{F}_T^X$ définie par

$$\begin{aligned} A \in \mathcal{F}_T^X &\iff \text{pour tout } n \geqslant 0,\ A \cap \{T \leqslant n\} \in \mathcal{F}_n^X \text{ et } A \in \mathcal{A} \\ &\iff \text{pour tout } n \geqslant 0,\ A \cap \{T = n\} \in \mathcal{F}_n^X \text{ et } A \in \mathcal{A} \end{aligned}$$

Montrons rapidement que $\mathcal{F}_T^X$ est une tribu : $\varnothing \in \mathcal{F}_T^X$ puisque $\varnothing \cap \{T \leqslant n\} = \varnothing \in \mathcal{F}_n^X$ pour tout $n \geqslant 0$; si $A \in \mathcal{F}_T^X$, alors $A^c \in \mathcal{A}$ et $A^c \cap \{T \leqslant n\} = \{T \leqslant m\} \backslash (A \cap \{T \leqslant n\}) \in \mathcal{F}_n^X$ pour tout $n \geqslant 0$, donc $A^c \in \mathcal{F}_T^X$; si $(A_k)_k$ est une famille au plus dénombrable d'éléments de $\mathcal{F}_T^X$, alors $\bigcup_k A_k \in \mathcal{A}$ et $(\bigcup_k A_k) \cap \{T \leqslant n\} = \bigcup_k (A_k \cap \{T \leqslant n\}) \in \mathcal{F}_n^X$ pour tout $n \geqslant 0$, donc $\bigcup_k A_k \in \mathcal{F}_T^X$.

Définition 8. — Soient $X = (X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E}))_{n\geqslant 0}$, une suite de variables aléatoires à valeurs dans un espace mesurable $(E, \mathcal{E})$ et $T : (\Omega, \mathcal{A}, \mathbb{P}) \to \bar{\mathbb{N}}$ une variable, ou un temps, aléatoire. Sur l'événement $\{T < \infty\}$, on définit la variable X au temps T par

$$\omega \in \{T < \infty\}, \qquad (X_T)(\omega) = X_{T(\omega)}(\omega).$$

Lorsque $\{T = \infty\}$ est non vide, ou plus généralement $\mathbb{P}\{T < \infty\} > 0$, on convient d'adjoindre à E un point supplémentaire δ, de munir $\bar{E} = E \cup \{\delta\}$ de la tribu $\bar{\mathcal{E}} = \sigma\{\mathcal{E}, \{\delta\}\} = \{B, B \cup \{\delta\} : B \in \mathcal{E}\}$, et finalement de poser $X_T = \delta$ sur $\{T = \infty\}$ (un tel point δ est souvent appelé « cimetière » dans la théorie des processus de Markov).

On vérifiera que X_T est une variable aléatoire. Cette propriété repose sur le fait que l'ensemble des temps est au plus dénombrable. Évidement si la variable T est finie, il est inutile d'avoir à considérer l'extension $(\bar{E}, \bar{\mathcal{E}})$.

Exercice 13. — Soit $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires et T un temps d'arrêt de cette suite.

(i) Montrer que T, $X_{T\wedge m}$ $(m \geqslant 0)$ et $X_T \times \mathbb{1}_{\{T<\infty\}}$ (lorsque l'espace d'état est de caractère numérique) sont $\mathcal{F}_T^X$-mesurables.

(ii) Montrer que si S est un autre temps d'arrêt de X et si $S \leqslant T$, alors on a $\mathcal{F}_S^X \subset \mathcal{F}_T^X$. En déduire que si S et T sont deux temps d'arrêt quelquonques, on a $\mathcal{F}_{S\wedge T}^X \subset \mathcal{F}_T^X \subset \mathcal{F}_{S\vee T}^X$.

Théorème 8 (propriété de Markov forte). — *Soit $X = (X_n)_{n\geqslant 0}$ une chaîne de Markov (λ, P) et soit T un temps d'arrêt de X. Alors conditionnellement à $\{T < \infty, X_T = x\}$, le procesus $(X_{T+n})_{n\geqslant 0}$ est une chaîne de Markov $(\delta_{\{x\}}, P)$ indépendante de $\mathcal{F}_T^X$.*

Démonstration. — Soit $A \in \mathcal{F}_T^X$. On a donc pour tout $m \geqslant 0$, $A \cap \{T = m\} \in \mathcal{F}_m^X$. D'après la propriété de Markov au temps m, on a

$$\begin{aligned}\mathbb{P}(\{X_T = y_0, X_{T+1} = y_1, \ldots, X_{T+n} = y_n\} \cap A \cap \{T = m, X_T = x\}) \\ = \mathbb{P}^x\{X_0 = y_0, X_1 = y_1, \ldots, X_n = y_n\} \times \mathbb{P}(A \cap \{T = m, X_T = x\}).\end{aligned}$$

En sommant sur $m = 0, 1, 2, \ldots$ et en divisant par $\mathbb{P}\{T < \infty, X_T = x\}$, on obtient

$$\begin{aligned}\mathbb{P}(\{X_T = y_0, X_{T+1} = y_1, \ldots, X_{T+n} = y_n\} \cap A \,|\{\, T = m, X_T = x\}) \\ = \mathbb{P}^x\{X_0 = y_0, X_1 = y_1, \ldots, X_n = y_n\} \times \mathbb{P}(A \,|\, \{T < \infty, X_T = x\}).\end{aligned}$$

D'où la conclusion. □

Exercice 14. — Soit F un sous-ensemble de E. On définit

$$H_0 = \inf\{n \geqslant 0 : X_n \in F\}, \qquad \text{et} \qquad H_{m+1} = \inf\{n > H_m : X_n \in F\}, \quad m \geqslant 0.$$

Supposons que pour tout $m \geqslant 0$, $\mathbb{P}\{H_m < \infty\} = 1$.

Vérifier que $(H_m)_{m\geqslant 0}$ est une suite strictement croissante de temps d'arrêt.

On pose pour tout $m \geqslant 0$, $Y_m = X_{H_m}$. Montrer que chaque Y_m est une variable aléatoire. La suite $(Y_m)_{m\geqslant 0}$ est la chaîne de Markov $(X_n)_{n\geqslant 0}$ observée au temps $(H_m)_{m\geqslant 0}$. Montrer que $(Y_m)_{m\geqslant 0}$ est une chaîne de Markov $(P_{X_{H_0}}, \bar{P})$ où $P_{X_{H_0}}$ est la loi de X_{H_0} et la matrice de transition $\bar{P}$ vérifie

$$\bar{P}(x, y) = h_x^y$$

où pour $y \in F$, le vecteur $(h_x^y)_{x\in E}$ est l'unique solution minimale positive de

$$h_x^y = P(x, y) + \sum_{z\notin F} P(x, z) \times h_z^y.$$

Exercice 15. — On suppose que la matrice de transition P de X n'a pas d'état absorbant. On définit

$$S_0 = 0, \qquad S_{m+1} = \inf\big\{n \geqslant S_m : X_n \neq X_{S_m}\big\}, \quad m \geqslant 0.$$

Montrer que $(S_m)_{m\geqslant 0}$ est une suite de temps d'arrêt, finis.

En posant $Z_m = X_{S_m}$, on définit alors une nouvelle chaîne de Markov : c'est la chaîne $(X_n)_{n\geqslant 0}$ observée uniquement lorsqu'elle « bouge ». Montrer que sa matrice de transition $\bar{P}$ vérifie

$$\bar{P}(x, x) = 0 \quad \text{pour tout } x \in E, \qquad \text{et} \qquad \bar{P}(x, y) = \frac{P(x, y)}{\sum_{z\neq x} P(x, z)}$$

pour tout $x, y \in E$ tels que $x \neq y$.

5. **Récurrence et transience**

Définition 9. — Soit $X = (X_n)_{n\geqslant 0}$ une chaîne de Markov de matrice de transition P. On dit qu'un état $x \in E$ est *récurrent* si et seulement si

$$\mathbb{P}^x\{X_n = x \text{ pour une infinité de } x\} = 1.$$

On dit qu'un état $x \in E$ est *transient*, ou *transitoire*, si et seulement si

$$\mathbb{P}^x\{X_n = x \text{ pour une infinité de } x\} = 0.$$

On va montrer que tout état est ou bien récurrent, ou bien transient. On appelle *premier temps de retour* T^x dans l'état $x \in E$ le temps d'arrêt défini par

$$T^x(\omega) = \inf\{n > 0 : X_n(\omega) = x\}, \qquad \omega \in \Omega,$$

où $\inf \varnothing = \infty$. On définit par récurrence la suite $(T_r^x)_{r\geqslant 0}$ des r-ièmes temps de retour dans l'état x par

$$T_0^x = 0, \qquad T_1^x = T^x, \qquad T_{r+1}^x = \inf\bigl\{n > T_r^x : X_n = x\bigr\}, \quad r \geqslant 0.$$

Pour $r \geqslant 1$, on appelle r-ième excursion la suite des valeurs prises par X entre T_{r-1}^x et T_r^x

$$\bigl(X_n : T_{r-1}^x \leqslant n < T_r^x\bigr),$$

les inégalités ayant été définies pour que sa longueur soit exactement

$$S_r^x = \begin{cases} T_r^x - T_{r-1}^x & \text{si } T_{r-1}^x < \infty \\ 0 & \text{sinon,} \end{cases}$$

et que les excursions associées à chaque trajectoire forment une partition de celle-ci.

Remarque. — On notera la différence des définitions du premier temps de retour en x

$$T^x = \inf\{n > 0 : X_n = x\} \qquad \text{et de} \qquad H^x = \inf\{n \geqslant 0 : X_n = x\}$$

le premier temps d'atteinte de x.

Lemme 1. — *Pour $r \geqslant 2$, conditionnellement à $\{T_{r-1}^x < \infty\}$, S_r^x est indépendante de $\mathcal{F}^X_{T_{r-1}^x}$ et*

$$\mathbb{P}\bigl\{S_r^x = n \;\big|\; T_{r-1}^x < \infty\bigr\} = \mathbb{P}^x\{T^x < \infty\}.$$

Démonstration. — Dans un premier temps, montrons que pour tout $r \geqslant 0$, T_r^x est un temps d'arrêt : pour $r = 0$, il est clair que $T_0^x = 0$ est un temps d'arrêt ; pour $r \geqslant 1$ et $n \geqslant 0$, on a

$$\bigl\{T_r^x \leqslant n\bigr\} = \bigcup_{1\leqslant k_1<\cdots<k_r\leqslant n} \{X_{k_1} = x, \ldots, X_{k_r} = x\}$$

événement qui est égal à la réunion finie — éventuellement vide si par exemple $r > n$ — et non nécessairement disjointe d'événements $\mathcal{F}_n^X$-mesurables, et qui est donc, lui aussi, $\mathcal{F}_n^X$-mesurable, ainsi T_r^x est un temps d'arrêt.

Le lemme ne consiste plus alors qu'en l'application de la propriété de Markov forte au temps T_{r-1}^x puisque S_r^x est le temps en partant de x à la date T_{r-1}^x pour retourner en x :

$$S_r^x = \inf\Bigl\{n \geqslant 1 : X_{T_{r-1}^x+n} = x\Bigr\}. \qquad \square$$

Le *nombre de visites* V^x dans l'état $x \in E$ est la variable aléatoire

$$V^x = \sum_{n=0}^{\infty} \mathbb{1}_{\{X_n=x\}}.$$

Pour $y \in E$, on a

$$\mathbb{E}^y[V^x] = \sum_{n=0}^{\infty} \mathbb{E}^y[\mathbb{1}_{\{X_n=x\}}] = \sum_{n=0}^{\infty} \mathbb{P}^y\{X_n = x\} = \sum_{n=0}^{\infty} P^n(y,x).$$

En particulier, on a

$$\mathbb{E}^x[V^x] = \sum_{n=0}^{\infty} P^n(x,x).$$

On définit la *probabilité de retour* en x par

$$f^x = \mathbb{P}^x\{T^x < \infty\}$$

qui est la probabilité qu'étant parti de x, la chaîne y retourne à un instant ultérieur.

LEMME 2. — *Pour tout $r \geqslant 0$, on a $\mathbb{P}^x\{V^x > r\} = (f^x)^r$.*

Démonstration. — On suppose $X_0 = x$. Alors $\{V^x > r\} = \{T_r^x < \infty\}$. Si $r = 0$ on a bien $\mathbb{P}^x\{V^x > 0\} = \mathbb{P}^x\{T_0^x < \infty\} = \mathbb{P}^x\{0 < \infty\} = 1 = (f^x)^0$. Par récurrence, supposons cette égalité vraie pour un $r \geqslant 0$, alors

$$\begin{aligned}\mathbb{P}^x\{V^x > r+1\} &= \mathbb{P}^x\big\{T_{r+1}^x < \infty\big\} = \mathbb{P}^x\big\{S_{r+1}^x < \infty, T_r^x < \infty\big\}\\ &= \mathbb{P}^x\big\{S_{r+1}^x < \infty \,\big|\, T_r^x < \infty\big\} \times \mathbb{P}^x\big\{T_r^x < \infty\big\}\\ &= \mathbb{P}^x\{T^x < \infty\} \times f_r^x = f^x \times (f^x)^r = (f^x)^{r+1}.\end{aligned}$$

Par récurrence, la conclusion du lemme s'ensuit. □

THÉORÈME 9. — *Soit $x \in E$. On a*

(i) *ou bien $\mathbb{P}^x\{T^x < \infty\} = 1$, et alors x est récurrent et $\sum_{n=0}^{\infty} P^n(x,x) = \infty$;*

(ii) *ou bien $\mathbb{P}^x\{T^x < \infty\} < 1$, et alors x est transient et $\sum_{n=0}^{\infty} P^n(x,x) < \infty$.*

Démonstration. — Si $\mathbb{P}^x\{T^x < \infty\} = f^x = 1$, on a d'après le lemme précédent

$$\mathbb{P}^x\{V^x = \infty\} = \lim_{r\to\infty} \mathbb{P}^x\{V^x > r\} = \lim_{r\to\infty} (f^x)^r = \lim_{r\to\infty} 1^r = 1,$$

donc x est récurrent et

$$\sum_{n=0}^{\infty} P^n(x,x) = \mathbb{E}^x[V^x] = \infty.$$

D'autre part, si $\mathbb{P}^x\{T^x < \infty\} = f^x < 1$, on a, toujours d'après ce lemme,

$$\sum_{n=0}^{\infty} P^n(x,x) = \mathbb{E}^x[V^x] = \sum_{r=0}^{\infty} \mathbb{P}^x\{V^x > r\} = \sum_{r=0}^{\infty} (f^x)^r = \frac{1}{1-f^x} < \infty.$$

En particulier, $\mathbb{P}^x\{V^x = \infty\} = 0$, et donc x est transitoire. □

THÉORÈME 10. — *Soit C une classe communiquante pour la chaîne de Markov X. Alors, ou bien tous les états de C sont récurrents, ou bien ils sont tous transients.*

Démonstration. — Soient x, $y \in C$, et soient m, $n \geqslant 0$ tels que

$$P^n(x,y) > 0 \qquad \text{et} \qquad P^m(y,x) > 0$$

(on rappelle que $x \leftrightsquigarrow y$ si et seulement si il existe m, $n \geqslant 0$ tels que $\mathbb{P}^x\{X_n = y\} > 0$ et $\mathbb{P}^y\{X_m = x\} > 0$). Ainsi, pour tout $r \geqslant 0$,

$$P^{n+r+m}(x,x) = \sum_{z,z'\in E} P^n(x,z) \times P^r(z,z') \times P^m(z',x) \geqslant P^n(x,y) \times P^r(y,y) \times P^m(y,x),$$

c'est-à-dire

$$P^r(y,y) \leqslant \frac{P^{n+r+m}(x,x)}{P^n(x,y) \times P^m(y,x)}\,.$$

En sommant sur $r \geqslant 0$,

$$\sum_{r=0}^{\infty} P^r(y,y) \leqslant \frac{1}{P^n(x,y) \times P^m(y,x)} \sum_{r=0}^{\infty} P^{n+r+m}(x,x).$$

Ainsi, si x est transient, la somme de droite est finie, donc celle de gauche l'est aussi, ce qui signifie d'après le théorème précédent que y est transient. Si y est récurrent, alors la somme de gauche est infinie, donc celle de droite l'est aussi et x est donc récurrent. Par symétrie du raisonnement entre x et y, on obtient l'équivalence de la récurrence de x et de y ainsi que celle de leur transience. □

THÉORÈME 11. — *Toute classe récurrente est fermée.*

Démonstration. — Soit C une classe qui n'est pas fermée (s'il en existe et qui est alors distincte de E). Donc il existe $x \in C$, $y \notin C$ et $m \geqslant 0$ tels que

$$\mathbb{P}^x\{X_m = y\} = P^m(x, y) > 0.$$

Puisqu'on a

$$\mathbb{P}^x(\{X_m = y\} \cap \{X_n = x \text{ pour une infinité de } n\}) = 0$$

car lorsque $X_m = y$, X ne retourne pas en x sinon x et y seraient dans la même classe communiquante. Ceci implique que

$$\mathbb{P}^x\{X_n = x \text{ pour une infinité de } n\} < 1$$

puisque $\mathbb{P}^x\{X_m = y\} = P^m(x, y) > 0$. Donc x n'est pas récurrent. Ainsi la classe C n'est pas récurrente, elle est donc transiente. Par contraposée de l'implication précédente, on obtient que toute classe récurrente est fermée. □

THÉORÈME 12. — *Toute classe finie et fermée est récurrente.*

Démonstration. — Soit C une classe finie et fermée. Supposons que X parte de C, c'est-à-dire que sa loi initiale est portée par C. Alors pour tout $\omega \in \Omega$, il existe $x(\omega) \in C$ tel que $X_n(\omega) = x(\omega)$ pour une infinité de n. On en déduit qu'il existe $x \in C$ tel que

$$\mathbb{P}\{X_n = x \text{ pour une infinité de } n\} > 0.$$

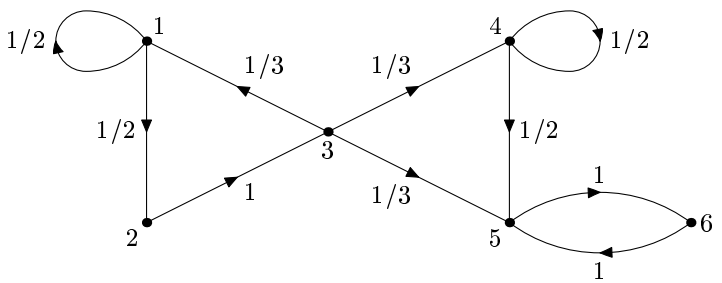
Par la propriété de Markov forte au temps T_1^x, la probabilité précédente est égale à

$$\mathbb{P}\{T_1^x < \infty\} \times \mathbb{P}\{X_n = x \text{ pour une infinité de } n\} > 0,$$

ce qui implique que x est récurrent. Donc, toute la classe C est récurrente. □

Puisqu'il est facile de déterminer les classes fermées, la récurrence ou la transience des classes finies est elle aussi facile à déterminer.

Exemple. — Dans l'exemple suivant, on constate que la seule classe récurrente est $\{5, 6\}$, les autres classes $\{1, 2, 3\}$ et $\{4\}$ sont donc transientes.



Pour ce type d'analyse, nous n'avons pas besoin de l'information précise des probabilités de transition si on convient qu'une flèche correspond toujours à une probabilité de transition strictement positive.

Remarque. — Si $C \subset E$ est une classe communiquante de cardinal infini (E est alors aussi de cardinal infini), on a seulement l'implication

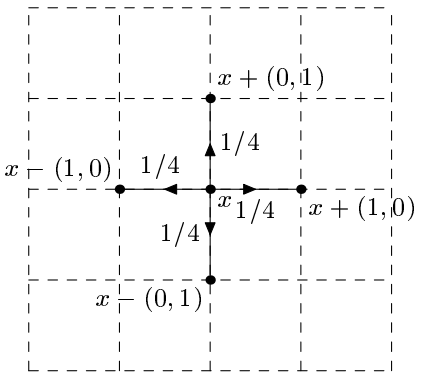
$$C \text{ est une classe non fermée} \quad \Longrightarrow \quad C \text{ est une classe transiente,}$$

ce qui n'est jamais que la contraposée de l'affirmation selon laquelle une classe récurrente est fermée. Comme on pourra le voir dans les exemples suivants, il peut exister des classes infinies qui soient à la fois fermées et transientes.

Exercice 16 (marche aléatoire simple dans $\mathbb{Z}$, suite). — Montrer que si $1 > p > q > 0$, alors l'unique classe communiquante est $\mathbb{Z}$ et elle est transiente. Montrer que si $p = q = 1/2$, l'unique classe communiquante est $\mathbb{Z}$ et qu'elle est récurrente. On pourra dans ce dernier cas utiliser la formule de Stirling

$$m\,! \approx \sqrt{2\pi m} \times \Big(\frac{m}{\mathrm{e}}\Big)^m \qquad \text{lorsque } m \to \infty.$$

Exercice 17 (marche aléatoire simple symétrique dans $\mathbb{Z}^2$). — La marche aléatoire symétrique dans $\mathbb{Z}^2$ est une chaîne de Markov telle que, pour tout n, X_n a probabilité $1/4$ de passer à l'un des 4 sites voisins



autrement dit, sa matrice de transition P est définie, pour tout $x \in \mathbb{Z}^2$, par

$$P\big(x, x+(1,0)\big) = P\big(x, x+(0,1)\big) = P\big(x, x-(1,0)\big) = P\big(x, x-(0,1)\big) = \frac{1}{4}\,,$$

les autres coefficients étant nuls.

On pose X_n^+ la projection orthogonale de X_n sur la diagonale $\{x = y\}$ et X_n^- la projection orthogonale de X_n sur l'anti-diagonale $\{x = -y\}$. Montrer que X_n^+ et X_n^- sont des marches aléatoires simples symétriques dans $\frac{1}{\sqrt{2}}\mathbb{Z}$, qui sont de plus indépendantes. En déduire que

$$P^{2n}(0,0) = \left(\binom{2n}{n} \times \Big(\frac{1}{2}\Big)^{2n}\right)^2 \approx \frac{1}{\pi n} \qquad \text{quand } n \to \infty,$$

puis que $(X_n)_{n\geqslant 0}$ est récurrente, c'est à-dire qu'il n'y a qu'une seule classe communiquante (irréductibilité), et que celle-ci est récurrente.

Exercice 18 (marche aléatoire simple symétrique dans $\mathbb{Z}^3$). — La marche aléatoire symétrique dans $\mathbb{Z}^3$ est une chaîne de Markov telle que, pour tout n, X_n a probabilité $1/6$ de passer à l'un des 6 sites voisins, autrement dit, sa matrice de transition P est définie par, pour tout $x \in \mathbb{Z}^3$,

$$\begin{aligned} P\big(x, x+(1,0,0)\big) &= P\big(x, x+(0,1,0)\big) = P\big(x, x+(0,0,1)\big) \\ &= P\big(x, x-(1,0,0)\big) = P\big(x, x-(0,1,0)\big) = P\big(x, x-(0,0,1)\big) = \frac{1}{6}\,, \end{aligned}$$

les autres coefficients étant nuls.

Montrer que

$$P^{2n}(0,0) = \binom{2n}{n}\Big(\frac{1}{2}\Big)^{2n} \times \sum_{\substack{i,y,k\geqslant 0\\ i+y+k=n}} \binom{n}{i\ \ y\ \ k}^2 \Big(\frac{1}{3}\Big)^{2n}$$

où

$$\binom{n}{i\ \ y\ \ k} = \frac{n\,!}{i\,!\ y\,!\ k\,!}$$

est un coefficient multinomial. En utilisant la formule de Stirling, montrer que

$$\sum_{n\geqslant 0} P^n(0,0) = \sum_{n\geqslant 0} P^{2n}(0,0) < \infty.$$

En déduire que $(X_n)_{n\geqslant 0}$ est transiente, c'est-à-dire qu'il n'y a qu'une seule classe communiquante (irréductibilité), et que celle-ci est transiente.

Théorème 13. — *Supposons la matrice de transition P irréductible et récurrente, c'est-à-dire qu'il n'existe qu'une seule classe communiquante et qu'elle est récurrente. Alors pour tout $y \in E$, $\mathbb{P}\{T^y < \infty\} = 1$.*

Démonstration. — D'après la formule des probabilités totales,

$$\mathbb{P}\{T^y < \infty\} = \sum_{x\in E} \mathbb{P}\{X_0 = x\} \times \mathbb{P}\{T^y < \infty \,|\, X_0 = x\} = \sum_{x\in E} \mathbb{P}\{X_0 = x\} \times \mathbb{P}^x\{T^y < \infty\}.$$

Donc il suffit de montrer que $\mathbb{P}^x\{T^y < \infty\} = 1$ pour tout $x \in E$. Soit $m \geqslant 0$. Par récurrence de la chaîne $(X_n)_{n\geqslant 0}$, on a

$$1 = \mathbb{P}^y\{X_n = y \text{ pour une infinité de } n\} \leqslant \mathbb{P}^y\{X_n = y \text{ pour un } n \geqslant m\}$$

et donc on a égalité. Par la formule des probabilités totales, puis la propriété de Markov, on a

$$\begin{aligned} 1 &= \mathbb{P}^y\{X_n = y \text{ pour un } n \geqslant m\} \\ &= \sum_{z\in E} \mathbb{P}^y\{X_m = k\} \times \mathbb{P}^y\{X_n = y \text{ pour un } n \geqslant m \,|\, X_m = k\} \\ &= \sum_{z\in E} P^m(y,z) \times \mathbb{P}^k\{T^y < \infty\}. \end{aligned}$$

Puisque $\sum_{z\in E} P^m(y,z) = 1$, on a nécessairement $\mathbb{P}^z\{T^y < \infty\} = 1$ pour tout $z \in E$ tel que $P^m(y,z) > 0$. Toujours par récurrence de la chaîne, soit $m \geqslant 0$ tel que $P^m(y,x) > 0$. D'après ce qui précède, on a $\mathbb{P}^x\{T^y < \infty\} = 1$. Finalement,

$$\mathbb{P}\{T^y < \infty\} = \sum_{x\in E} \mathbb{P}\{X_0 = x\} \times \mathbb{P}^x\{T^y < \infty\} = \sum_{x\in E} \mathbb{P}\{X_0 = x\} = 1. \qquad \square$$

Exercice 19 (décomposition du premier retour). — Soit T^y le premier temps de retour dans l'état y, et

$$f^{(n)}(x,y) = \mathbb{P}^x\{T^y = n\}.$$

Montrer que

$$P^n(x,y) = \sum_{k=1}^{n} f^{(k)}(x,y) \times P^{n-k}(y,y) \qquad \text{pour tout } n \geqslant 1.$$

En déduire que

$$P(x,y)(s) = \delta(x,y) + F(x,y)(s) \times P(y,y)(s)$$

où

$$P(x,y)(s) = \sum_{n=0}^{\infty} P^n(x,y) \times s^n \qquad \text{et} \qquad F(x,y)(s) = \sum_{n=0}^{\infty} f^{(n)}(x,y) \times s^n.$$

En déduire que $\mathbb{P}^x\{T^x < \infty\} = 1$ si et seulement si $\sum_{n=0}^{\infty} P_{i,i}^{(n)} = \infty$ et ce, sans utiliser le théorème correspondant.

6. **Mesures invariantes**

Nous rappelons qu'une mesure positive sur E est identifiée à un vecteur ligne indéxé par E de nombres positifs finis — par mesure positive, nous sous-entendons ainsi mesure positive σ-finie. Une mesure de probabilité sur E est identifiée à un tel vecteur dont la somme des coefficients est égale à 1.

Définition 10. — Une mesure positive $\lambda = (\lambda\{x\})_{x\in E}$ est *invariante* par une matrice de transition P si et seulement si

$$\lambda P = \lambda, \qquad \text{c'est-à-dire} \qquad \sum_{x\in E} \lambda\{x\} \times P(x,y) = \lambda\{y\} \quad \text{pour tout } y \in E.$$

Si λ est de plus une mesure de probabilité, on précisera que c'est une *mesure de probabilité invariante*.

Notons que si λ est une mesure invariante, tous ses multiples $\alpha\lambda$, $\alpha \geqslant 0$, sont des mesures invariantes. Si $\lambda(E) \in \{0, \infty\}$, aucun de ses multiples n'est une mesure de probabilité.

Théorème 14. — *Soit $(X_n)_{n\geqslant 0}$ une chaîne de Markov de matrice de transition P et de loi initiale π. On suppose que π est une mesure de probabilité invariante par P. Alors pour tout $m \geqslant 0$, $(X_{m+n})_{n\geqslant 0}$ est une chaîne de Markov (π, P). En particulier X_m est de loi π pour tout $m \geqslant 0$.*

Démonstration. — Par la propriété de Markov, $(X_{m+n})_{n\geqslant 0}$ est une chaîne de Markov de matrice de transition P. Puisque la loi de X_m est $\pi P^m = \pi$, on a la conclusion. □

Théorème 15. — *Supposons que l'espace d'états E est un ensemble fini et qu'il existe $x \in E$ tel que pour tout $y \in E$ la suite $\left(P^n(x,y)\right)_{n\geqslant 0}$ converge vers une limite $\pi\{y\}$. Alors la suite $\pi = (\pi\{y\})_{y\in E}$ est une mesure de probabilité invariante par P.*

Démonstration. — On a clairement $\pi\{y\} \geqslant 0$ pour tout $y \in E$. De plus

$$\sum_{y\in E} \pi\{y\} = \sum_{y\in E} \lim_{n\to\infty} P^n(x,y) = \lim_{n\to\infty} \sum_{y\in E} P^n(x,y) = \lim_{n\to\infty} 1 = 1.$$

Donc la suite π définit une mesure de probabilité sur E. De plus, on a

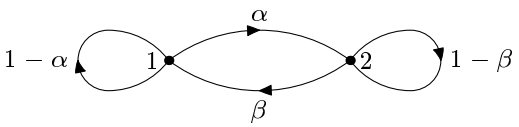
$$\begin{aligned}\pi\{y\} &= \lim_{n\to\infty} P^n(x,y) = \lim_{n\to\infty} P^{n+1}(x,y) \\ &= \lim_{n\to\infty} \sum_{z\in E} P^n(x,z)P(z,y) = \sum_{z\in E} \lim_{n\to\infty} P^n(x,z)P(z,y) = \sum_{z\in E} \pi\{z\}P(z,y) = (\pi P)\{y\},\end{aligned}$$

ce qui montre que π est invariante par P. □

Remarque. — L'hypothèse de finitude de E est importante dans la démonstration de ce théorème puisqu'elle permet d'intervertir sommes et limites sans hypothèse additionnelle sur la convergence des suites envisagées.

Exemple. — Nous considérons une chaîne de Markov à deux états. Sa matrice et son diagramme de transition s'écrivent

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$



En supposant $\alpha + \beta > 0$, lorsque n tend vers l'infini, on a

$$P^n \longrightarrow \begin{pmatrix} \dfrac{\beta}{\alpha+\beta} & \dfrac{\alpha}{\alpha+\beta} \\ \dfrac{\beta}{\alpha+\beta} & \dfrac{\alpha}{\alpha+\beta} \end{pmatrix},$$

ce qui montre que $(\beta/(\alpha+\beta), \alpha/(\alpha+\beta))$ est une mesure de probabilité invariante.

Exemple. — Si $X = (X_n)_{n\geqslant 0}$ est une marche aléatoire simple sur $\mathbb{Z}$ associée à un paramètre $p \in [0,1]$ et $q = 1-p \in [0,1]$, le calcul explicite des probabilités de transition montre que les suites $\big(P^n(x,y)\big)_{n\geqslant 0}$ tendent vers 0 pour tous x, $y \in \mathbb{Z}$. Le vecteur $(0)_{x\in\mathbb{Z}}$ est invariant par la matrice de transition de la marche mais n'est bien sûr pas une mesure de probabilité.

Si $p \neq 1/2$, la marche est asymétrique. En supposant que pour tout $n \geqslant 0$, on a $X_n = X_0 + \xi_1 + \cdots + \xi_n$, où les $(\xi_n)_{n\geqslant 0}$ sont des variables aléatoires indépendantes toutes de loi $q\delta_{\{-1\}} + p\delta_{\{1\}}$ et donc d'espérance $\mathbb{E}[\xi_1] = -q + p = 2p-1$, on a d'après la loi forte des grands nombres

$$\frac{X_n}{n} = \frac{X_0}{n} + \frac{1}{n}\sum_{k=1}^{n} \longrightarrow 0 + \mathbb{E}[\xi_1] = 2p-1 \qquad \text{presque sûrement quand } n \text{ tend vers l'infini.}$$

Puisque cette limite est différente de 0, nous obtenons un équivalent de la suite aléatoire $(X_n)_{n\geqslant 0}$:

$$X_n \sim n \times (2p-1) \qquad \text{presque sûrement quand } n \text{ tend vers l'infini.}$$

Ceci montre que, pour toute loi (de probabilité) initiale, $(X_n)_{n\geqslant 0}$ tend vers $\pm\infty$ presque sûrement, et donc qu'il y a convergence en loi vers $\delta_{\{\pm\infty\}}$ qui n'est pas une mesure de probabilité sur $\mathbb{Z}$ ($\pm\infty = +\infty$ si $p > 1/2$ et $\pm\infty = -\infty$ si $p < 1/2$).

Pour $p = 1/2$, la situation est plus délicate, nous y reviendrons plus tard.

Remarquons que dans tous les cas, la mesure uniforme $(1)_{x\in\mathbb{Z}}$ — ainsi que ses multiples — est invariante, mais ce n'est pas une mesure de probabililité.

Remarque d'algèbre linéaire. — Supposons que l'espace d'états est fini, par exemple $E = \{1,\ldots,n\}$. Rechercher un vecteur ligne $\lambda = (\lambda_1,\ldots,\lambda_n)$ tel que

$$\lambda P = (\lambda_1,\ldots,\lambda_n) \times \begin{pmatrix} p_{1,1} & \cdots & p_{1,n} \\ \vdots & \ddots & \vdots \\ p_{n,1} & \cdots & p_{n,n} \end{pmatrix} = (\lambda_1,\ldots,\lambda_n) = \lambda$$

ou encore

$$\sum_{x\in E} \lambda\{x\} \times P(x,y) = \lambda\{y\} \qquad \text{pour tout } y \in E,$$

consiste, par tranposition, à chercher chercher des solutions de ${}^tP\,{}^t\lambda = {}^t\lambda$, autrement dit à trouver λ tel que ${}^t\lambda$ soit un vecteur propre pour la valeur 1 de ${}^tP$. Or P et ${}^tP$ ont même polynôme caractéristique par invariance du déterminant par transposition :

$$\det(P - x\,\mathrm{Id}) = \det{}^t(P - x\,\mathrm{Id}) = \det({}^tP - x\,{}^t\mathrm{Id}) = \det({}^tP - x\,\mathrm{Id})$$

où x désigne l'inconnue formelle des polynômes caractéristiques. Mais P possède 1 parmi ses valeurs propres puisque

$$P \times \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

ce qui découle du fait que la somme des coefficients de chaque ligne de P vaut 1. Donc ${}^tP$ possède aussi 1 pour valeur propre, il existe alors au moins un vecteur propre non nul ${}^t\lambda$ associé. Le problème est alors de savoir si on peut trouver un tel vecteur avec des coordonnées toutes positives (au sens large). L'algèbre linéaire élémentaire ne suffit pas pour répondre à cette question car il n'y est pas tenu compte de la positivité des coefficients. Cette question peut être abordée par des méthodes analytiques, nous le ferons de manière probabiliste.

On définit le temps moyen passé en $x \in E$ entre deux visites de $z \in E$ par

$$\gamma^z\{x\} = \mathbb{E}^z\left[\sum_{n=0}^{T^z-1} \mathbb{1}_{\{X_n=x\}}\right] \qquad \text{où} \qquad T^z = \inf\{n > 0 : X_n = z\}.$$

Théorème 16. — *Soit P une matrice de transition irréductible et récurrente et $z \in E$. Alors*

(i) $\gamma^z\{z\} = 1$;

(ii) $\gamma^z = (\gamma^z\{x\})_{x\in E}$ *vérifie* $\gamma^z P = \gamma^z$;

(iii) $0 < \gamma^z\{x\} < \infty$ *pour tout* $x \in E$.

Remarques. — *a*) Nous rappelons que si E est fini et P est irréductible, alors P est récurrente. Notons qu'alors γ^z correspond à une mesure finie non nulle sur E, un de ses multiples positifs est alors une mesure de probabilité invariante et nous verrons qu'avec l'hypothèse d'irréductibilité, c'est la seule.

b) Si E est infini, P peut être irréductible sans être récurrente. De plus, γ^z ne correspond pas toujours à une mesure finie dans ce cadre.

Démonstration. — (i) Entre les instants 0 et T^z, le processus $(X_n)_{n\geqslant 0}$ ne passe en z qu'aux dates 0 et T^z. On a donc identiquement

$$\sum_{n=0}^{T^z-1} \mathbb{1}_{\{X_n=x\}} = 1$$

variable aléatoire dont l'espérance est γ^z_z qui est donc égale à 1.

(ii) Pour $n \geqslant 1$, l'événement $\{n \leqslant T^z\}$ ne dépend que de $X_0, X_1, \ldots, X_{n-1}$ puisqu'il est égal à $\{X_1 \neq z, \ldots, X_{n-1} \neq z\}$, donc d'après la propriété de Markov au temps $n-1$, on a

$$\begin{aligned}\mathbb{P}^z\{X_{n-1} = x, X_n = y, n \leqslant T^z\} &= \mathbb{P}^z\{X_{n-1} = x, n \leqslant T^z\} \times \mathbb{P}^z\{X_n = y \,|\, X_{n-1} = x, n \leqslant T^z\} \\ &= \mathbb{P}^z\{X_{n-1} = x, n \leqslant T^z\} \times \mathbb{P}^z\{X_n = y \,|\, X_{n-1} = x\} = \mathbb{P}^z\{X_{n-1} = x, n \leqslant T^z\} \times P(x,y).\end{aligned}$$

Comme P est récurrente, on a $\mathbb{P}^z\{T^z < \infty\} = 1$ donc $X_0 = X_{T^z} = z$ avec probabilité 1 sous la mesure de probabilité $\mathbb{P}^z$. Ainsi,

$$\begin{aligned}\gamma^z\{y\} &= \mathbb{E}^z\left[\sum_{n=1}^{T^z} \mathbb{1}_{\{X_n=y\}}\right] = \mathbb{E}^z\left[\sum_{n=1}^{\infty} \mathbb{1}_{\{X_n=y\}} \times \mathbb{1}_{\{n\leqslant T^z\}}\right] = \mathbb{E}^z\left[\sum_{n=1}^{\infty} \mathbb{1}_{\{X_n=y, n\leqslant T^z\}}\right] \\ &= \sum_{n=1}^{\infty} \mathbb{E}^z\left[\mathbb{1}_{\{X_n=y,n\leqslant T^z\}}\right] = \sum_{n=1}^{\infty} \mathbb{P}^z\{X_n = y, n \leqslant T^z\}\end{aligned}$$

où on notera que la première égalité s'obtient en distinguant deux cas : si $y \neq z$, on a $\mathbb{1}_{\{X_0=y\}} = \mathbb{1}_{\{X_{T^z}=y\}} = 0$, et si $y = k$, $\mathbb{1}_{\{X_0=y\}} = \mathbb{1}_{\{X_{T^z}=y\}} = 1$ avec probabilité 1 sous la mesure de probabilité $\mathbb{P}^z$, on ne change donc rien en sommant de 0 à $T^z - 1$ ou de 1 à T^z. Pour $1 \leqslant n < \infty$, on écrit l'événement $\{X_n = y, n \leqslant T^z\}$ comme la réunion disjointe des événements $\{X_{n-1} = x, X_n = y, n \leqslant T^z\}$, $x \in E$, pour obtenir

$$\gamma^z\{y\} = \sum_{n=1}^{\infty}\sum_{x\in E} \mathbb{P}^z\{X_{n-1} = x, X_n = y, n \leqslant T^z\} = \sum_{n=1}^{\infty}\sum_{x\in E} P(x,y) \times \mathbb{P}^z\{X_{n-1} = x, n \leqslant T^z\}$$

la dernière égalité résultant de la propriété de Markov. Puis finalement, par des manipulations élémentaires,

$$\begin{aligned}\gamma^z\{y\} &= \sum_{x\in E} P(x,y) \times \sum_{n=1}^{\infty} \mathbb{P}^z\{X_{n-1} = x, n \leqslant T^z\} = \sum_{x\in E} P(x,y) \times \sum_{n=1}^{\infty} \mathbb{E}^z\left[\mathbb{1}_{\{X_{n-1}=x,n\leqslant T^z\}}\right] \\ &= \sum_{x\in E} P(x,y) \times \mathbb{E}^z\left[\sum_{n=1}^{\infty} \mathbb{1}_{\{X_{n-1}=x,n\leqslant T^z\}}\right] = \sum_{x\in E} P(x,y) \times \mathbb{E}^z\left[\sum_{m=0}^{\infty} \mathbb{1}_{\{X_m=x,m\leqslant T^z-1\}}\right] \\ &= \sum_{x\in E} P(x,y) \times \mathbb{E}^z\left[\sum_{m=0}^{T^z-1} \mathbb{1}_{\{X_m=x\}}\right] = \sum_{x\in E} P(x,y) \times \gamma^z\{x\} = \sum_{x\in E} \gamma^z\{x\} \times P(x,y) \\ &= (\gamma^z P)\{y\},\end{aligned}$$

ce qui est le résultat désiré.

(iii) Puisque P est irréductible, pour tout état $x \in E$, il existe m, $n \geqslant 0$ tels que $P^m(z,x) > 0$ et $P(x,z)^{(n)} > 0$ et . Alors on a

$$\gamma^z\{x\} = \sum_{y\in E} \gamma^z\{y\}P(y,x)^{(m)} \geqslant \gamma^z\{z\}P^m(z,x) = P^m(z,x) > 0$$

et

$$\gamma^z\{x\}P(x,z)^{(n)} \leqslant \sum_{y\in E} \gamma^z\{y\}P(y,z)^{(n)} = \gamma^z_z = 1,$$

et donc finalement $0 < \gamma^z\{x\} \leqslant 1/P(x,z)^{(n)} < \infty$. □

Théorème 17. — *Soit P une matrice de transition irréductible et soit λ une mesure invariante vérifiant $\lambda\{z\} = 1$ pour un certain $z \in E$. Alors $\lambda \geqslant \gamma^z$. Si de plus P est récurrente alors $\lambda = \gamma^z$.*

Démonstration. — Pour tout $y \in E$, on a

$$\begin{aligned}
\lambda\{y\} &= \sum_{x_0 \in E} \lambda\{x_0\} \times P(x_0, y) \\
&= \sum_{x_0 \neq z} \lambda\{x_0\} \times P(x_0, y) + P(z, y) \\
&= \sum_{x_0, x_1 \neq z} \lambda\{x_1\} \times P(x_1, x_0) \times P(x_0, y) + \left(P(z, y) + \sum_{x_0 \neq z} P(z, x_0) \times P(x_0, y) \right) \\
&= \sum_{x_0, \ldots, x_n \neq z} \lambda\{x_n\} \times P(x_n, x_{n-1}) \times \cdots \times P(x_0, y) \\
&\quad + \left(P(z, y) + \sum_{x_0 \neq z} P(z, x_0) \times P(x_0, y) + \cdots + \sum_{x_0, \ldots, x_{n-1} \neq z} P(z, x_{n-1}) \times \cdots \times P(x_0, y) \right) \\
&\leqslant P(z, y) + \sum_{x_0 \neq z} P(z, x_0) \times P(x_0, y) + \cdots + \sum_{x_0, \ldots, x_{n-1} \neq z} P(z, x_{n-1}) \times \cdots \times P(x_0, y) \\
&= \mathbb{P}^z\{X_1 = y, 1 \leqslant T^z\} + \mathbb{P}^z\{X_2 = y, 2 \leqslant T^z\} + \cdots + \mathbb{P}^z\{X_n = y, n \leqslant T^z\} \\
&\longrightarrow \sum_{m=1}^{\infty} \mathbb{P}\{X_m = y, m \leqslant T^z\} = \gamma^z\{y\} \qquad \text{quand } n \text{ tend vers l'infini.}
\end{aligned}$$

Donc pour tout $y \in E$, $\lambda\{y\} \geqslant \gamma^z\{y\}$, c'est-à-dire $\lambda \geqslant \gamma^z$.

Si P est de plus récurrente, on sait d'après le théorème 17 que γ^z est invariante. Donc $\mu = \lambda - \gamma$, qui est encore une mesure positive, est elle aussi invariante. Comme P est irréductible, pour $x \in E$, il existe $n \geqslant 0$ tel que $P^n(x, z) > 0$, et

$$0 = \mu\{z\} = \sum_{y \in E} \mu\{y\} \times P^n(y, z) \geqslant \mu\{x\} \times P^n(x, z),$$

et donc $\mu\{x\} = 0$. □

Définition 11. — Soit $x \in E$ un état récurrent, c'est-à-dire tel que

$$\mathbb{P}^x\{X_n = x \text{ pour une infinité de } n\} = 1 \qquad \text{ou encore} \qquad \mathbb{P}^x\{T^x < \infty\} = 1.$$

Si le temps de retour moyen $m^x = \mathbb{E}^x[T^x]$ est fini, alors l'état x est dit *récurrent positif*. Si $m^x = \mathbb{E}^x[T^x]$ est infini, l'état x est dit *récurrent nul*.

Théorème 18. — *Soit P une matrice de transition irréductible. Les énoncés suivants sont équivalents :*

(i) *tout état est récurrent positif;*

(ii) *au moins un état est récurrent positif;*

(iii) *P a une mesure de probabilité invariante π.*

De plus, si (iii) *est réalisé, on a $m\{x\} = 1/\pi\{x\}$ pour tout $x \in E$.*

Démonstration. — (i) *implique* (ii) est évident.

(ii) *implique* (iii). Si x est un état récurrent positif, en particulier x est récurrent, donc P, qui est irréductible, est récurrente. D'après le théorème 16, $\gamma^x = (\gamma^x\{y\})_{y \in E}$ est invariant. Mais

$$\sum_{y \in E} \gamma^x\{y\} = \sum_{y \in E} \mathbb{E}^x \left[\sum_{n=0}^{T^x - 1} \mathbb{1}_{\{X_n = y\}} \right]$$

$$= \mathbb{E}^x\left[\sum_{n=0}^{T^x-1}\sum_{y\in E}\mathbb{1}_{\{X_n=y\}}\right] = \mathbb{E}^x\left[\sum_{n=0}^{T^x-1}1\right] = \mathbb{E}^x[T^x] = m\{x\} < \infty.$$

Donc, en posant $\pi\{y\} = \gamma^x\{y\}/m\{x\}$, $y \in E$, on définit une mesure de probabilité invariante sur E.

(iii) *implique* (i). Soit $z \in E$ un état. Comme π est une mesure de probabilité, il existe un état $x \in E$ tel que $\pi\{x\} > 0$. Comme P est irréductible, il existe $n \geqslant 0$ tel que $P^n(x,z) > 0$. On en déduit que

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

Soit $\lambda\{x\} = \pi\{x\}/\pi\{z\}$, $x \in E$. La mesure λ est invariante et vérifie $\lambda\{z\} = 1$. D'après le théorème précédent, on a alors $\lambda \geqslant \gamma^z$. On en déduit que

$$m\{z\} = \sum_{x\in E}\gamma^z\{x\} \leqslant \sum_{x\in E}\lambda\{x\} = \frac{1}{\pi\{z\}}\sum_{x\in E}\pi\{x\} = \frac{1}{\pi\{z\}} < \infty$$

et z est donc récurrent positif. Comme cela est vrai pour tout $z \in E$, on obtient que (i) est vérifié.

Pour le dernier point, on reprend les arguments de (iii) *implique* (i) en utilisant le fait que que P est récurrente. On a alors $\lambda = \gamma^z$ s'après le théorème 17, et l'inégalité ci-dessus est alors une égalité. □

Théorème 19. — *Soient E un ensemble fini et P une matrice de transition irréductible sur E. Alors tous les états sont récurrents positifs et P a une unique mesure de probabilité invariante.*

Démonstration. — Nous savons qu'une matrice de transition sur un ensemble fini admet au moins une classe récurrente. Puisque P est irréductible, l'unique classe E est récurrente et donc P est irréductible récurrente. D'après le théorème 16, γ^z est une mesure invariante et ses coefficients sont finis. Alors

$$m\{z\} = \sum_{x\in E}\gamma^z\{x\} < \infty$$

puisque somme finie de nombres finis. Donc z est récurrent positif pour tout $z \in E$. En posant

$$\pi\{y\} = \frac{\gamma^z\{y\}}{\sum_{x\in E}\gamma^z\{x\}}$$

on obtient une mesure de probabilité invariante. D'après le théorème précédent, celle-ci est l'unique mesure de probabilité invariante. □

Notations. — Dans la suite, nous utiliserons les notations suivantes : $\mathcal{R} \subset E$ est l'ensemble des points récurents de E, $\mathcal{T} \subset E$ est l'ensemble des points transients de E, $\mathcal{T}_\mathcal{R} \subset \mathcal{T}$ est l'ensemble des points x de $\mathcal{T}$ tel qu'il existe $y \in \mathcal{R}$ tel que $P(x,y) > 0$.

Lemme 3. — *On suppose l'espace des états E fini. Si $x \in \mathcal{T}$, alors l'une au moins des possibilités suivantes est satisfaite :*

(i) $x \in \mathcal{T}_\mathcal{R}$;

(ii) *il existe $x' \in \mathcal{T}_\mathcal{R}$ tel que $x \rightsquigarrow x'$.*

En particulier $\mathcal{R}$ est non vide.

Démonstration. — Notons $x \not\rightsquigarrow x'$ si $x \rightsquigarrow x'$ est fausse. Soit $x_0 \in \mathcal{T}$ un état transient. Montrons qu'il existe $y \in \mathcal{R}$ tel que $x_0 \rightsquigarrow y$. Par l'absurde, supposons que ce ne soit pas le cas. La classe contenant x_0 n'est pas fermée. Il existe donc x_1 n'appartenant pas à cette classe tel que $x_0 \rightsquigarrow x_1$ et bien sûr $x_1 \not\rightsquigarrow x_1$; on a de plus $x_0 \neq x_1$ et, par hypothèse, x_1 est transient puisque x_0 ne mène à aucun point récurrent. Par transitivité de la relation « mener à », x_1 vérifie la même hypothèse que x_0 : il ne mène à aucun point récurrent. On continue ainsi pour obtenir une suite

$$x_0 \rightsquigarrow x_1 \rightsquigarrow x_2 \rightsquigarrow \cdots \rightsquigarrow x_n \rightsquigarrow \cdots$$

Pour deux indices différents $m < n$, on a nécessairement $x_m \neq x_n$; nous l'avons déjà vu pour $m + 1 = n$, et si, pour $m + 1 < n$, nous avions $x_m = x_n$, alors on aurait $x_m \rightsquigarrow x_{m+1} \rightsquigarrow \cdots \rightsquigarrow x_m$ et donc $x_{m+1} \rightsquigarrow x_m$ ce qui est interdit par la construction. La suite $(x_n)_{n\geqslant 0}$ énumère ainsi une partie infinie de $\mathcal{T}$ ce qui est contradictoire avec la finitude de $\mathcal{T}$.

Soit ainsi $x \in \mathcal{T}$ et $y \in \mathcal{R}$ tels que $x \rightsquigarrow y$. Il existe alors une famille finie de points $x = x_0, x_1, \ldots, x_{n-1}, x_n = y$ telle que $P(x_0, x_1) > 0, \ldots, P(x_{n-1}, x_n) > 0$. Soit x_ℓ le premier état récurrent de cette suite. Alors pour tout $\ell \leqslant m \leqslant n$, x_m est récurrent par fermeture des classes récurrentes, et pour tout $0 \leqslant m \leqslant \ell - 1$, x_m est transient. Par construction $x' = x_{\ell-1} \in \mathcal{T}_{\mathcal{R}}$ et $x \rightsquigarrow x'$. □

Exercice 20. — Sur $\mathcal{R}$, la relation « mener à » est une relation d'équivalence.

Réponse. — Nous savons déjà que cette relation est transitive sur tout x, elle est donc transitive sur un quelconque de ses sous-ensembles, en particulier $\mathcal{R}$, et par définition elle est symétrique puisque tout point mène à lui-même en 0 pas. Pour la symétrie, si x, $y \in \mathcal{R}$ vérifient $x \rightsquigarrow y$, par fermeture de la classe communiquante de x, le point y appartient à cette même classe, et donc on a aussi $y \rightsquigarrow x$. (Vrai pour E fini ou dénombrable.) □

Lemme 4. — *Si λ est une mesure invariante telle que pour un $x \in E$, $\lambda\{x\} = 0$, alors*

$$y \rightsquigarrow x \quad \Longrightarrow \quad \lambda\{y\} = 0.$$

Démonstration. — Soit $y \in E$ tel que $y \rightsquigarrow x$. Il existe donc $y = y_0, y_1, \ldots, y_{n-1}, y_n = x$ tels que $P(y_0, y_1) \times \cdots \times P(y_{n-1}, y_n) > 0$. On a alors

$$\begin{aligned} 0 = \lambda\{x\} = (\lambda P^n)\{x\} = \sum_{z_0,\ldots,z_{n-1}} \lambda\{z_0\} \times P(z_0, z_1) \times \cdots \times P(z_{n-1}, x) \\ \geqslant \lambda\{y_0\} \times P(y_0, y_1) \times \cdots \times P(y_{n-1}, x), \end{aligned}$$

et donc $\lambda\{y\} = 0$. □

Remarque. — Ce dernier lemme est vrai même si E est dénombrable. En particulier, il permet de voir que si une mesure invariante s'annule en un point d'une classe communiquante, elle est nulle sur toute la classe ainsi que toutes celles qui peuvent mener à elles. Ce qui est un résultat assez remarquable.

Théorème 20. — *On suppose l'espace des états E fini. Soit λ une mesure invariante. Alors $\lambda(\mathcal{T}) = 0$. En particulier, pour tout état transient x, on a $\lambda\{x\} = 0$.*

Démonstration. — On a

$$\lambda(\mathcal{T}) = (\lambda P)(\mathcal{T}) = \sum_{y\in\mathcal{T}} (\lambda P)\{y\} = \sum_{y\in\mathcal{T}} \sum_{x\in E} \lambda\{x\} \times P(x, y) = \sum_{x\in E} \lambda\{x\} \sum_{y\in\mathcal{T}} P(x, y).$$

Comme pour tout $x \notin \mathcal{T}$ donc $x \in \mathcal{R}$, on a par fermeture des classes récurrentes, $P(x,y) = 0$, on a donc

$$\lambda(\mathcal{T}) = \sum_{x\in\mathcal{T}} \lambda\{x\} \sum_{y\in\mathcal{T}} P(x,y).$$

Soit $x \in \mathcal{T}$. Supposons par l'absurde que $\lambda\{x\} > 0$. D'après l'avant dernier lemme, il existe $x' \in \mathcal{T}_{\mathcal{R}}$ tel que $x \rightsquigarrow x'$ et donc, par le lemme précédent, $\lambda\{x'\} > 0$. Mais puisqu'il existe $r \in \mathcal{R}$ tel que $P(x',r) > 0$, on a

$$\sum_{y\in\mathcal{T}} P(x',y) < 1, \qquad \text{et, puisque } \lambda\{x'\} > 0, \qquad \lambda\{x'\} \sum_{y\in\mathcal{T}} P(x',y) < \lambda\{x'\}.$$

On a alors

$$\begin{aligned}\lambda(\mathcal{T}) = \sum_{x\in\mathcal{T}} \lambda\{x\} \sum_{y\in\mathcal{T}} P(x,y) &= \sum_{x\in\mathcal{T}\setminus\{x'\}} \lambda\{x\} \sum_{y\in\mathcal{T}} P(x,y) + \lambda\{x'\} \sum_{y\in\mathcal{T}} P(x',y) \\ &\leqslant \sum_{x\in\mathcal{T}\setminus\{x'\}} \lambda\{x\} \times 1 + \lambda\{x'\} \sum_{y\in\mathcal{T}} P(x',y) < \lambda(T\setminus\{x'\}) + \lambda\{x'\} = \lambda(\mathcal{T}).\end{aligned}$$

ce qui est absurde. Donc, pour tout $x \in \mathcal{T}$, $\lambda\{x\} = 0$, et ainsi $\lambda(\mathcal{T}) = 0$. □

Théorème 21. — *On suppose l'espace des états E fini. Alors il y a unicité de la mesure de probabilité invariante si et seulement si P n'a qu'une seule classe récurrente, les autres mesures invariantes étant alors multiples de celle-ci.*

Démonstration. — Si P a au moins deux classes récurrentes R_1 et R_2, désignons par P_1 et P_2 les sous-matrices qui sont les restrictions de P aux classes correspondantes. Ce sont des matrices stochastiques qui sont irréductibles et récurrentes et nous notons π_1 et π_2 leurs uniques mesures de probabilité invariantes. Alors pour tout $t \in [0,1]$, la mesure de probabilité

$$\pi_t = t \times \pi_1(\,.\cap R_1) + (1-t) \times \pi_2(\,.\cap R_2)$$

est une mesure de probabilité invariante sur E. Il n'y a donc pas unicité.

S'il n'y a qu'une seule classe récurrente $R = \mathcal{R}$, alors toute mesure de probabilité invariante est nulle sur $\mathcal{T} = R^c$ et sa restriction à R est invariante par la restriction de P à R et donc égale à l'unique mesure de probabilité invariante π_R de cette restriction. On a donc $\pi = \pi_R(\,.\cap R)$, d'où l'unicité. Pour finir, toute mesure invariante λ est finie sur E puisque E est fini, donc λ est multiple d'une mesure de probabilité invariante, et donc de π. □

Théorème 22. — *On suppose l'espace d'états E fini et on note $R_1, \ldots, R_k$ les classes récurrentes de P. Alors toute mesure invariante λ sur E s'écrit*

$$\lambda = \sum_{\ell=1}^{z} t_\ell \times \pi_\ell(\,.\cap R_\ell) \qquad \textit{avec } t_1, \ldots, t_k \in \mathbb{R}_+,$$

où, pour $1 \leqslant \ell \leqslant k$, π_ℓ est l'unique mesure de probabilité invariante sur R_ℓ par la restriction de P à R_ℓ. C'est une mesure de probabilité invariante si et seulement si $t_1 + \cdots + t_k = 1$.

Démonstration. — C'est immédiat. □

Exemples. — *a)* La marche aléatoire simple symétrique sur $\mathbb{Z}$ est irréductible, récurrente. Elle admet la mesure uniforme $(1)_{x\in\mathbb{Z}}$ comme mesure invariante, qui n'est pas finie, et nous savons que toute autre mesure invariante est multiple de celle-ci grâce au théorème 17. Cette marche n'admet donc aucune mesure de probabilité invariante.

b) De même, la marche aléatoire simple symétrique sur $\mathbb{Z}^2$ est irréductible, récurrente. Elle admet la mesure uniforme $(1)_{x\in\mathbb{Z}}$ comme mesure invariante, qui n'est pas finie. Cette marche n'admet donc aucune mesure de probabilité invariante.

c) La marche aléatoire simple symétrique sur $\mathbb{Z}^3$ est irréductible mais transiente. On constate que la mesure uniforme $(1)_{x\in\mathbb{Z}}$ en est une mesure invariante. Il est possible de montrer qu'elle n'admet pas de mesure de probabilité invariante, par exemple en la réduisant en une marche aléatoire sur $\mathbb{Z}$: notons X la marche aléatoire symétrique sur $\mathbb{Z}^3$ et Y la marche obtenue faisant un pas positif sur $\mathbb{Z}$ dès que X fait un pas positif suivant l'un des 3 axes, négatif sinon ; Y est alors une marche aléatoire simple symétrique sur $\mathbb{Z}$ et si X avait une mesure de probabilité invariante, c'est-à-dire que sous cette loi initiale les $(X_n)_{n\geqslant 0}$ auraient même loi cette loi invariante, on en déduirait qu'il en serait de même pour Y ce qui est impossible d'après le premier exemple. Dans $\mathbb{Z}^4$, on se ramènerait à la marche sur $\mathbb{Z}^2$, dans $\mathbb{Z}^n$, on ferait de manière similaire en tenant compte de la parité de n, c'est-à-dire en se ramenant ou bien à $\mathbb{Z}$ ou bien $\mathbb{Z}^2$.

d) Considérons une marche simple asymétrique non dégénérée sur $\mathbb{Z}^2$, c'est-à-dire, supposons $p \in]0,1[\backslash\{1/2\}$. Pour qu'une mesure λ soit invariante on doit avoir

$$\text{pour tout } x \in \mathbb{Z}, \qquad \lambda\{x\} = (\lambda P)\{x\} = p \times \lambda\{x-1\} + q \times \lambda\{x+1\}.$$

Grâce à des calculs antérieurs, nous connaissons les solutions finies de cette équation à l'échange de p et q près. Elles sont de la forme

$$\lambda\{x\} = A + B(p/q)^x, \qquad x \in \mathbb{Z}.$$

Puisque l'ensemble des solutions forme une famille à deux paramètres, la chaîne ne peut être récurrente d'après le théorème 17, mais nous le savions déjà.

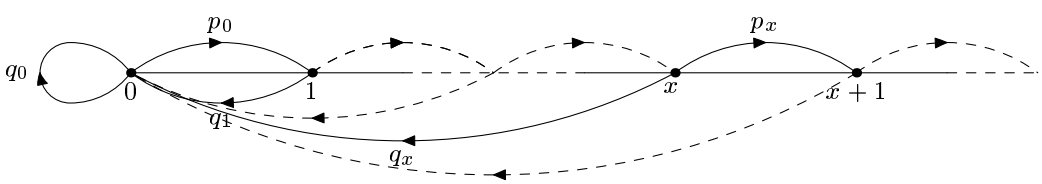
Exercice 21. — Soit F une classe fermée, en particulier si F est récurrente, associée à une matrice de transition P. Montrer que la sous-matrice de P obtenue en n'en conservant que les coefficients indexés par $F \times F$ est une matrice stochastique et qu'elle est irréductible. (Cette propriété a été utilisée dans cette section.)

Exercice 22. — Considérons l'urne d'Ehrenfest. Montrer que la mesure de probabilité π définie par

$$\pi\{x\} = \binom{N}{x} \times \frac{1}{2^N}, \qquad x \in \{0,1,\ldots,N\}$$

est l'unique mesure de probabilité invariante.

Exercice 23. — Nous considérons la matrice de transition sur $\mathbb{N}$ représentée par la figure suivante :



c'est-à-dire que pour tout $x \in \mathbb{N}$, $P(x,x+1) = p_x$, $P(x,0) = q_x = 1 - p_x$, $p_x \in [0,1]$, les autres coefficients étant nuls. Sous l'hypothèse de la convergence du produit infini $\prod_{i=0}^{\infty} p_x$, montrer que cette matrice de transition n'a pas de mesure invariante (autre que nulle).

Exercice 24. — Soit P une matrice de transition sur un espace d'états fini E. Montrer qu'une mesure de probabilité π est invariante si et seulement si

$$\pi \times \left(\mathrm{Id} - P + \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{pmatrix} \right) = (1, \dots, 1).$$

En déduire que si P est irréductible, alors

$$M = \mathrm{Id} - P + \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{pmatrix}$$

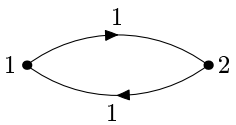
est inversible.

Remarque. — Cet exercice montre comment utiliser des méthodes matricielles pour déterminer la probabilité invariante d'une matrice de transition irréductible. Dans un cadre plus général, toujours avec E fini, il faudrait considérer chacune des sous-matrices associées à chaques classes récurrentes.

7. **Convergence vers l'équilibre**

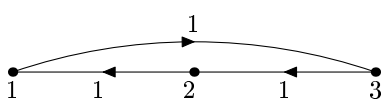
Nous avons vu que si E est fini et qu'il existe $x \in E$ tel que $\big(P^n(x,y)\big)_{n\geqslant 0}$ converge vers $\pi\{y\}$ pour tout $y \in E$, alors $\pi = (\pi\{y\})_{y\in E}$ est une mesure de probabilité invariante. Une telle convergence n'est pas toujours assurée. Déjà lorsque E est fini et possède strictement plus d'une classe récurrente, il existe plusieurs (une infinité de) mesures de probabilité invariantes ; alors une telle convergence a lieu, l'identification de la mesure limite n'est pas immédiate. De plus, lorsque P n'a qu'une seule classe récurrente ou mieux est irréductible, ce type de convergence peut n'avoir lieu pour aucun état initial comme le montrent les deux exemples qui suivent.

Exemples. — *a*) Considérons $E = \{1, 2\}$ et

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \text{  }$$

On a $P^{2n} = \mathrm{Id}$ et $P^{2n+1} = P$. Ainsi aucun des coefficients de la suite de matrices $(P^n)_{n\geqslant 0}$ ne converge. Plus encore, si la chaîne part de l'état 1 par exemple, elle occupe alternativement l'état 1 puis l'état 2. La suite des lois ne converge pas vers l'unique — P est irréductible — mesure de probabilité invariante $\frac{1}{2}\delta_{\{1\}} + \frac{1}{2}\delta_{\{2\}}$.

b) Considérons $E = \{1, 2, 3\}$ et

$$P = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \qquad \text{  }$$

Par le calcul ou en lisant sur le graphique, on a

$$P^2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \qquad P^3 = \mathrm{Id},$$

et ainsi

$$P^{3n} = \mathrm{Id}, \qquad P^{3n+1} = P, \qquad P^{3n+2} = P^2 \qquad \text{pour tout } n \geqslant 0.$$

Nous avons la même situation que précédement : quelque soit l'état initial, la suite des lois de la chaîne converge pas vers l'unique mesure invariante $\frac{1}{3}\delta_{\{1\}} + \frac{1}{3}\delta_{\{2\}} + \frac{1}{3}\delta_{\{3\}}$.

Ces deux exemples sont assez extrêmes puisque la dynamique markovienne y est déterministe. De plus, on constate dans le premier exemple que chaque état est fréquenté selon une période égale à 2, dans le second, selon une période égale à 3. Cette notion de périodicité peut être étendue à des dynamiques markoviennes non déterministe. Nous nous conterons de définir l'absence de périodicité. L'espace des états E est, sauf mention spéciale, fini ou dénombrable, et P est une matrice de transition sur E.

Définition 12. — Un état $x \in E$ est *apériodique* si il existe $N_x \geqslant 0$ tel que $P^n(x,x) > 0$ pour tout $n \geqslant N_x$, c'est-à-dire que $P^n(x,x) > 0$ pour n assez grand.

Exercice 25. — Montrer qu'un état $x \in E$ est apériodique si et seulement si

$$n_x = \mathrm{pgcd}\big\{n \geqslant 0 : P^n(x,x) > 0\big\} = 1,$$

où pgcd est l'opérateur « plus grand commun diviseur ». (Lorsque n_x est en entier supérieur ou égal à 2, l'état x est dit périodique de période n_x.)

Lemme 5. — *Supposons qu'il existe un état apériodique $x \in E$. Alors pour tous états y, z dans la classe communiquante de x, on a $P^n(y,z) > 0$ pour n assez grand. En particulier, tout état y dans la classe communiquante de x est apériodique.*

Démonstration. — Puisque y et z sont dans la classe communiquante de x, il existe r, $s \geqslant 0$ tels que $P^r(y,x) > 0$ et $P^s(x,z) > 0$. Alors

$$P^{r+n+s}(y,z) = \sum_{\ell,m \in E} P^r(y,\ell) \times P^n(\ell,m) \times P^s(m,k) \geqslant P^r(y,x) \times P^n(x,x) \times P^s(x,z) > 0$$

pour n assez grand. On a donc la conclusion. □

L'apériodicité est une propriété des classes communiquantes d'après le lemme précédent, on pourra ainsi parler de classes apériodiques. Une matrice de transition sera dite apériodique si toutes ses classes le sont.

Théorème 23 (convergence vers l'équilibre). — *Soit P une matrice de transition irréductible et apériodique, et supposons qu'il existe une mesure de probabilité invariante π. Si $(X_n)_{n\geqslant 0}$ est une chaîne de Markov (λ, P), alors*

$$\mathbb{P}\{X_n = y\} \longrightarrow \pi\{y\} \qquad \textit{quand } n \textit{ tend vers l'infini, pour tout } y \in E.$$

En particulier,

$$P^n(x,y) \longrightarrow \pi\{y\} \qquad \textit{quand } n \textit{ tend vers l'infini, pour tout } x,\, y \in E.$$

Démonstration. — On utilise un argument de couplage : on introduit une seconde chaîne de Markov $Y = (Y_n)_{n\geqslant 0}$ indépendante de $X = (X_n)_{n\geqslant 0}$ et de même matrice de transition P, puis on définit un temps d'arrêt T et considère les deux processus obtenus en recollant X avant T avec Y après T et réciproquement. Plus précisément, soit $(Y_n)_{n\geqslant 0}$ une chaîne de Markov (π, P) indépendante de $(X_n)_{n\geqslant 0}$. Fixons un état de référence $x_0 \in E$ et posons

$$T = \inf\{n \geqslant 0 : X_n = Y_n = x_0\}.$$

Première étape. — Nous montrons que $\mathbb{P}\{T < \infty\} = 1$.

La suite $(W_n)_{n\geqslant 0} = ((X_n, Y_n))_{n\geqslant 0}$ est une chaîne de Markov d'espace d'états $E \times E$, de matrice de transition $\widetilde{P}$ définie par

$$\widetilde{P}\big((x,y),(x',y')\big) = P(x,x') \times P(y,y'), \qquad (x,y),\,(x',y') \in E \times E,$$

et de loi initiale μ telle que

$$\mu\{(x,y)\} = \lambda\{x\} \times \pi\{y\}, \qquad (x,y) \in E \times E.$$

En effet, pour tous $(x_0,y_0),\dots,(x_n,y_n) \in E \times E$, on a, d'abord par indépendance, puis par application de la propriété de Markov de chaque chaîne,

$$\begin{aligned}
&\mathbb{P}\{(X_0,Y_0) = (x_0,y_0),\dots,(X_n,Y_n) = (x_n,y_n)\}\\
&\quad = \mathbb{P}\{X_0 = x_0, Y_0 = y_0, \dots, X_n = x_n, Y_n = y_n\}\\
&\quad = \mathbb{P}\big(\{X_0 = x_0,\dots,X_n = x_n\} \cap \{Y_0 = y_0,\dots,Y_n = y_n\}\big)\\
&\quad = \mathbb{P}\{X_0 = x_0,\dots,X_n = x_n\} \times \mathbb{P}\{Y_0 = y_0,\dots,Y_n = y_n\}\\
&\quad = \big(\lambda\{x_0\} \times P(x_0,x_1) \times \cdots \times P(x_{n-1},x_n)\big) \times \big(\pi\{y_0\} \times P(y_0,y_1) \times \cdots \times P(y_{n-1},y_n)\big)\\
&\quad = \big(\lambda\{x_0\} \times \pi\{y_0\}\big) \times \big(P(x_0,x_1) \times P(y_0,y_1)\big) \times \cdots \times \big(P(x_{n-1},x_n) \times P(y_{n-1},y_n)\big),
\end{aligned}$$

on a donc ce qui était annoncé.

Comme la matrice de transition P est apériodique et irréductible, pour tous états x, y, x', $y' \in E$, on a d'après le lemme précédent

$$\widetilde{P}^n\big((x,y),(x',y')\big) = P^n(x,x') \times P^n(y,y') > 0$$

pour n assez grand, et donc la matrice de transition $\widetilde{P}$ est aussi apériodique et irréductible. De plus la mesure $\tilde{\pi}$ définie par

$$\tilde{\pi}\{(x,y)\} = \pi\{x\} \times \pi\{y\}, \qquad (x,y) \in E \times E$$

est invariante par $\widetilde{P}$ puisque, pour tout $(x,y) \in E \times E$,

$$\begin{aligned}
\big(\tilde{\pi}\widetilde{P}\big)\{(x,y)\} &= \sum_{(x',y')\in E\times E} \tilde{\pi}\{(x',y')\} \times \widetilde{P}\big((x',y'),(x,y)\big)\\
&= \sum_{(x',y')\in E\times E} \pi\{x'\} \times \pi\{y'\} \times P(x',x) \times P(y,y')\\
&= \left(\sum_{x'\in E} \pi\{x'\} \times P(x',x)\right) \times \left(\sum_{y'\in E} \pi\{y'\} \times P(y',y)\right)\\
&= \pi\{x\} \times \pi\{y\} = \tilde{\pi}\{(x,y)\}.
\end{aligned}$$

Ainsi, d'après le théoreme 18, la matrice de transition $\widetilde{P}$ est récurrente positive puisqu'elle est irréductible et admet une mesure de probabilité invariante, notamment $\widetilde{P}$ est récurrente. Puisque $T = \inf\{n \geqslant 0 : (X_n,Y_n) = (x_0,x_0)\} = \inf\{n \geqslant 0 : W_n = (x_0,x_0)\}$, on a alors $\mathbb{P}\{T < \infty = 1\}$.

Deuxième étape. — Nous posons

$$Z_n = \begin{cases} X_n & \text{si } n < T\\ Y_n & \text{si } n \geqslant T, \end{cases}$$

et nous allons montrer que $(Z_n)_{n\geqslant 0}$ est une chaîne de Markov (λ, P). Il est clair que T est un temps d'arrêt de la filtration naturelle de $(W_n)_{n\geqslant 0}$. La propriété forte de Markov appliquée à la chaîne $(W_n)_{n\geqslant 0}$ au temps T fini montre que $(W_{T+n})_{n\geqslant 0}$ est une chaîne de

Markov $(\delta_{\{(b,b)\}}, \widetilde{P})$ indépendante de $\mathcal{F}_T^W$. Par symétrie, en remplaçant (X_{T+n}, Y_{T+n}) par (Y_{T+n}, X_{T+n}), on obtient encore une chaîne de Markov $\big(\delta_{\{(x_0,x_0)\}}, \widetilde{P}\big)$ indépendante de $\mathcal{F}_T^W$. En posant

$$Z'_n = \begin{cases} Y_n & \text{si } n < T \\ X_n & \text{si } n \geqslant T, \end{cases}$$

on en déduit que $(W'_n)_{n\geqslant 0} = ((Z_n, Z'_n))_{n\geqslant 0}$ est une chaîne de Markov $\big(\mu, \widetilde{P}\big)$. En particulier $(Z_n)_{n\geqslant 0}$ est une chaîne de Markov (λ, P). (...)

Troisième étape. — Pour $y \in E$, on a

$$\mathbb{P}\{Z_n = y\} = \mathbb{P}\{X_n = y, n < T\} + \mathbb{P}\{Y_n = y, n \geqslant T\}.$$

Ainsi

$$\begin{aligned} \big|\mathbb{P}\{X_n = y\} - \pi\{y\}\big| &= \big|\mathbb{P}\{X_n = y\} - \mathbb{P}\{Y_n = y\}\big| \\ &= \big|\mathbb{P}\{Z_n = y\} - \mathbb{P}\{Y_n = y\}\big| \\ &= \big|\mathbb{P}\{X_n = y, n < T\} + \mathbb{P}\{Y_n = y, n \geqslant T\} - \mathbb{P}\{Y_n = y\}\big| \\ &= \big|\mathbb{P}\{X_n = y, n < T\} - \mathbb{P}\{Y_n = y, n < T\}\big| \\ &\leqslant \mathbb{P}\{n < T\} \end{aligned}$$

Puisque $\mathbb{P}\{n < T\}$ tend vers 0 quand n tend vers l'infini, on constate que $\mathbb{P}\{X_n = y\}$ tend vers $\pi\{y\}$ quand n tend vers l'infini. □

Exercice 26 (théorème de renouvellement). — Soit $(Y_n)_{n\geqslant 1}$ une suite de variables aléatoires à valeurs dans $\mathbb{N}^* = \{1, 2, \ldots\}$ indépendantes et de même loi. On suppose que

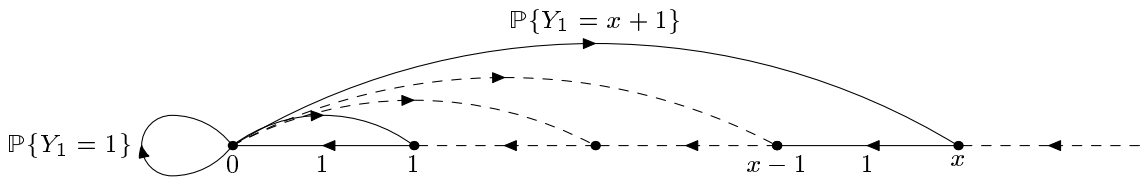
$$\operatorname{pgcd}\big\{n \geqslant 1 : \mathbb{P}\{Y_1 = n\} > 0\big\} = 1,$$

et on pose $\mu = \mathbb{E}[Y_1] \in \mathbb{R}_+ \cup \{\infty\}$.

Montrer que le processus $(X_n)_{n\geqslant 0}$ défini par $X_0 = 0$ et

$$X_n = \inf\{m \geqslant n : \exists\, k \geqslant 0,\ m = Y_1 + \cdots + Y_k\} - n, \qquad n \geqslant 1,$$

est une chaîne de Markov de matrice de transition P vérifiant $P(0, x) = \mathbb{P}\{Y_1 = x+1\}$, $x \geqslant 0$, et $P(x, x-1) = 1$, $i \geqslant 1$, les autres coefficients étant nuls.



On pourra poser $S_0 = 0$ et $S_k = Y_1 + \cdots + Y_k$ pour $k \geqslant 1$, et utiliser le fait que, si $S_{k-1} < n \leqslant S_k$, alors $X_n = S_k - n = Y_k - (n - S_{k-1})$.

Déterminer $\lim_{n\to\infty} \mathbb{P}\{X_n = 0\}$ et en déduire que

$$\mathbb{P}\{\exists\, k \geqslant 0 : n = Y_1 + \cdots + Y_k\} \longrightarrow \frac{1}{\mu} \qquad \text{quand } n \text{ tend vers l'infini.}$$

8. **Retournement du temps à une date déterministe**

La propriété (simple) de Markov consiste en l'indépendance du passé et du futur conditionnellement au présent. La symétrie de cette propriété permet immédiatement d'affirmer que si $(X_n)_{0\leqslant n\leqslant N}$ est une chaîne de Markov (dans sa filtration naturelle) alors la chaîne retournée $\big(\widehat{X}_n\big)_{0\leqslant n\leqslant N} = (X_{N-n})_{0\leqslant n\leqslant N}$ est aussi une chaîne de Markov (dans sa filtration naturelle).

Proposition 2. — *Soit $(X_n)_{n\geqslant 0}$ une chaîne de de Markov, alors $\big(\widehat{X}_n\big)_{0\leqslant n\leqslant N}$ est une chaîne de Markov. Si $\big(P^{(n,n+1)}\big)_{0\leqslant n\leqslant N-1}$ est une suite de matrice de transition de la chaîne $(X_n)_{n\geqslant 0}$, alors toute suite de matrices de transition $\big(\widehat{P}^{(n,n+1)}\big)_{0\leqslant n\leqslant N-1}$ vérifiant*

$$\mu_{N-n}\{x\}\times\big(\widehat{P}^{(n,n+1)}\big)(x,y) = \mu_{N-n-1}\{y\}\times\big(P^{(N-n-1,N-n)}\big)(y,x),\qquad x,\,y\in E,$$

où μ_n est la loi de X_n, $0\leqslant n\leqslant N$, est une suite de matrices de transition de la chaîne $(\widetilde{X}_n)_{n\geqslant 0}$.

Démonstration. — La propriété de Markov pour la chaîne retournée est clairement réalisée par indépendance du futur et du passé conditionnellement au présent. De plus, pour $x, y\in E$, on a

$$\begin{aligned}
\mu_{N-n}\{x\}\times\big(\widehat{P}^{(n,n+1)}\big)(x,y) &= \mathbb{P}\{X_{N-n}=x\}\times\mathbb{P}\big\{\widehat{X}_{n+1}=y\,\big|\,\widehat{X}_n=x\big\}\\
&= \mathbb{P}\{X_{N-n}=x\}\times\mathbb{P}\{X_{N-n-1}=y\,|\,X_{N-n}=x\}\\
&= \mathbb{P}\{X_{N-n-1}=y, X_{N-n}=x\}\\
&= \mathbb{P}\{X_{N-n-1}=y\}\times\mathbb{P}\{X_{N-n}=x\,|\,X_{N-n-1}=y\}\\
&= \mu_{N-n-1}\{y\}\times\big(P^{(N-n-1,N-n)}\big)(y,x),
\end{aligned}$$

d'où la conclusion. □

Remarques. — *a*) Une telle matrice de transition $\widehat{P}^{(n,n+1)}$, $0\leqslant N\leqslant N-1$, est définie par

$$\big(\widehat{P}^{(n,n+1)}\big)(x,y) = \frac{\mu_{N-n-1}\{y\}}{\mu_{N-n}\{x\}}\times\big(P^{(N-n-1,N-n)}\big)(y,x)$$

pour toutes les lignes x pour lesquelles le dénominateur $\mu_{N-n}\{x\}$ est strictement positif. Si, pour un $x\in E$, $\mu_{N-n}\{x\}=0$, ceci signifie que X ne passe pas par x au temps $N-n$, et donc $\widehat{X}$ n'y passe pas non plus au temps n ; on peut alors remplir la ligne correspondante de $\widehat{P}^{(n,n+1)}$ par ce qu'on veut.

b) Si la chaîne $(X_n)_{0\leqslant n\leqslant N}$ est homogène, lorsque le quotient a un sens, on a

$$\big(\widehat{P}^{(n,n+1)}\big)(x,y) = \frac{\mu_{N-n-1}\{y\}}{\mu_{N-n}\{x\}}\times P(y,x)$$

ce qui ne définit pas nécessairement une famille de matrices de transition homogène : quand on inverse le sens du temps, des évolutions qui étaient homogènes, telle que la diffusion de la poussière depuis un petit tas, deviennent assez peu naturelles ; les règles d'évolution tiennent compte des états antérieurs et ne sauraient donc être homogènes.

Théorème 24. — *Soit $(X_n)_{0\leqslant n\leqslant N}$ une chaîne de de Markov homogène de matrice de transition P. Supposons que sa loi initiale π est une mesure de probabilité invariante. Alors,*

$\big(\widehat{X}_n\big)_{0\leqslant n\leqslant N}$ *est une chaîne de de Markov homogène de loi initiale* π *et de matrice de transition* $\widehat{P}$ *vérifiant*

$$\pi\{x\}\times\hat{P}(x,y)=\pi\{y\}\times P(y,x),\qquad \textit{pour tous } x,\ y\in E.$$

Démonstration. — C'est une application immédiate de ce qui précède. □

Proposition 3. — *Soient* P *une matrice de transition admettant une mesure de probabilité invariante* π *et* $\widehat{P}$ *une matrice de transition vérifiant*

$$\pi\{x\}\times\widehat{P}(x,y)=\pi\{y\}\times P(y,x),\qquad \textit{pour tous } x,\ y\in E.$$

Alors

(i) *la mesure de probabilité* π *est invariante par* $\widehat{P}$ *;*

(ii) *si* P *est irréductible, alors* $\widehat{P}$ *est irréductible.*

Démonstration. — (i) Nous avons immédiatement, pour tout $x\in E$,

$$\big(\pi\widehat{P}\big)\{x\}=\sum_{y\in E}\pi\{y\}\times\hat{P}(y,x)=\sum_{y\in E}\pi\{x\}\times P(x,y)=\pi\{x\}\sum_{y\in E}P(x,y)=\pi\{x\},$$

ce qui montre que π est invariante par $\widehat{P}$.

(ii) Soient $x,\ y\in E$. Comme P est irréductible, la mesure de probabilité π a tous ses coefficients strictements positifs puisqu'elle est invariante. De plus, pour tous $x,\ y\in E$, il existe $y=x_n,\dots,x_0=x\in E$ tels que $P(x_n,x_{n-1})\times\cdots\times P(x_1,x_0)>0$; on a alors

$$\begin{aligned}\widehat{P}(x_0,x_1)\times\cdots\times\widehat{P}(x_{n-1},x_n)&=\frac{\pi\{x_n\}}{\pi\{x_{n-1}\}}\times P(x_n,x_{n-1})\times\cdots\times\frac{\pi\{x_1\}}{\pi\{x_0\}}\times P(x_1,x_0)\\&=\frac{\pi\{x_n\}}{\pi\{x_0\}}\times P(x_n,x_{n-1})\times\cdots\times P(x_1,x_0)>0,\end{aligned}$$

ce qui montre que $\widehat{P}$ est irréductible. □

Définition 13. — Soit P une matrice de transition. Une mesure μ sur E est *réversible* pour P si

$$\mu\{x\}\times P(x,y)=\mu\{y\}\times P(y,x),\qquad \text{pour tous } x,\ y\in E.$$

Nota bene. — Étant donnée une matrice de transition P, rien n'assure sans condition supplémentaire que P admette une mesure réversible.

Proposition 4. — *Si* μ *est une mesure réversible pour* P*, alors* μ *est invariante.*

Démonstration. — Si $x\in E$, on a

$$(\mu P)\{x\}=\sum_{y\in E}\mu\{y\}\times P(y,x)=\sum_{y\in E}\mu\{x\}\times P(x,y)=\mu\{x\}\sum_{y\in E}P(x,y)=\mu\{x\},$$

d'où l'invariance. □

Proposition 5. — *Soit* $(X_n)_{0\leqslant n\leqslant N}$ *une chaîne de Markov* (λ,P) *où* λ *est supposée être une mesure de probabilité réversible. Alors* $\widehat{P}=P$ *et la chaîne retournée* $\big(\widehat{X}_n\big)_{0\leqslant n\leqslant N}$ *a même loi que* $(X_n)_{0\leqslant n\leqslant N}$.

Démonstration. — Immédiate. Notons que la loi de ces chaînes finies est une mesure de probabilité sur E^{N+1} et pas simplement une suite de lois de probabilité sur E. □

COROLLAIRE 1. — *Si la matrice de transition P est symétrique, alors la mesure uniforme est réversible, et donc, invariante.*

Démonstration. — Immédiate. □

Exemple. — Nous considérons l'urne d'Ehrenfest. La mesure π_N sur $\{0, 1, \ldots, N\}$ définie par

$$\pi_N\{x\} = \binom{N}{x} \times \frac{1}{2^N}\,, \qquad x \in \{0, 1, \ldots, N\},$$

est réversible. En effet, si $0 \leqslant x < N$, on a

$$\pi_N\{x\} \times P(x, x+1) = \binom{N}{x} \times \frac{1}{2^N} \times \frac{N-x}{N} = \binom{N-x}{x} \times \frac{1}{2^N}$$

et

$$\pi_N\{x+1\} \times P(x+1, x) = \binom{N}{x+1} \times \frac{1}{2^N} \times \frac{x+1}{N} = \binom{N-1}{x} \times \frac{1}{2^N},$$

et toutes les autres relation envisageables donnent l'égalité de 0 avec lui-même. Comme la mesure de probabilité π_N est réversible, elle est aussi invariante.

Exemple important. — Un graphe est un ensemble au plus dénombrable de points E qualifiés de sommets, abstrait ou inclus dans un $\mathbb{R}^n$, et des arêtes reliants ces sommets, c'est-à-dire la donnée d'une relation symétrique $\mathcal{R}$ signifiant que $x \mathcal{R} y$ (ou $y \mathcal{R} x$) si et seulement si les sommets x et y sont reliés par une arête.

On peut dans certains cas convenir que tout sommet x est relié à lui-même, ce qui revient à demander que la relation $\mathcal{R}$ soit réflexive, ou au contraire exclure la réflexivité.

On définit la valence $v\{x\}$ d'un sommet $x \in E$ comme le nombre d'arêtes partant de x, c'est-à-dire $\mathrm{Card}\{y \in E : x \mathcal{R} y\}$. Nous supposerons être en présence d'un graphe tel que $1 \leqslant v\{x\} < \infty$ pour tout $x \in E$, qui est de plus connexe, c'est-à-dire qu'entre deux sommets, il existe une suite finie d'arêtes joignant ces sommets.

Nous souhaitons définir une matrice de transition P sur E qui tienne compte de manière homogène de la structure de graphe. Il y a plusieurs manière d'y parvenir, notamment si on convient, ou non, que chaque sommet est joint à lui-même.

Supposons que les sommets ne sont pas joints à eux-mêmes, et soient $p \in [0, 1]$, $q = 1 - p \in [0, 1]$. Nous définissons une matrice de transition P sur E par

$$P(x, y) = \begin{cases} q & \text{si } x = y, \\ \dfrac{p}{v\{x\}} & \text{si } x \mathcal{R} y, \text{ autrement dit si } x \text{ et } y \text{ sont joints par une arête,} \\ 0 & \text{sinon.} \end{cases}$$

On constate alors que la mesure positive $v = (v\{x\})_{x \in E}$ est réversible puisque pour tous x, $y \in E$, on a si $x \neq y$ et $x \mathcal{R} y$,

$$v\{x\} \times P(x, y) = v\{x\} \times \frac{p}{v\{x\}} = p = v\{y\} \times \frac{p}{v\{y\}} = v\{y\} \times P(y, x)$$

les autres identités pour x et y non joints par une arête étant trivialement satisfaites. Notons que l'hypothèse $1 \leqslant v\{x\} < \infty$ pour tout $x \in E$ se montre essentielle pour avoir cette propriété. La mesure positive $v = (v\{x\})_{x \in E}$ est donc invariante.

Pour obtenir une mesure de probabilité invariante à partir de v, nous devons avoir

$$V = \sum_{x \in E} v\{x\} < \infty, \qquad \text{mais comme} \qquad \mathrm{Card}\, E = \sum_{x \in E} 1 \leqslant \sum_{x \in E} v\{x\}$$

puisque tout sommet est de valence supérieure ou égale à 1, cette condition est satisfaite si et seulement si le graphe est fini, ou encore si et seulement si E est fini. Dans ce cas la mesure de probabilité $\pi = (v\{x\}/V)_{x\in E}$ est réversible et invariante.

Le cas $p = 0$ n'est guère intéressant puisqu'alors $P = \mathrm{Id}$ et donc tous les points sont absorbants. Supposons $p > 0$. L'hypothèse de connexité du graphe implique que P est irréductible.

Si E est fini, il est clair que P est récurrente, et de plus il existe $N \geqslant 0$ fini tel que tous sommets x, $y \in E$ peuvent rejoints en empruntant au plus N arêtes consécutives. Alors, en autorisant à demeurer quelques temps en un sommet intermédiaire entre x et y, ce qui correspond à $q > 0$ ou $0 < p < 1$, pour tout x, $y \in E$, x et y sont joints par au moins un chemin partiellement stationnaire de longueur N : il existe $x = x_0, x_1, \ldots, x_{N-1}, x_N = y \in E$ tels que

$$P(x, x_1) \times \cdots \times P(x_{N-1}, y) > 0.$$

Ceci montre que la matrice P^N a alors tous ses coefficients strictements positfs. On constate qu'il en est alors de même pour toute $P^{N+n} = P^N \times P^n$ par simple produit matriciel. En particulier la matrice P est apériodique.

Si $p = 1$, l'apériodicité peut ne pas être satisfaite comme le montre, par exemple, le cas du graphe à deux sommets $\{1, 2\}$. Si E est infini, la situation est plus complexe. Nous ne ferons pas de commentaire à ce sujet.

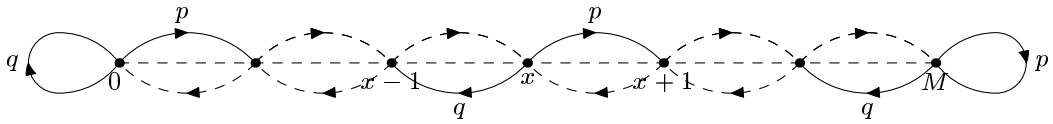
Exercice 27. — Considérons un graphe connexe fini $(E, \mathcal{R})$. Soit P définie par

$$P(x, y) = \begin{cases} \dfrac{1}{v\{x\} + 1} & \text{si } x \,\mathcal{R}\, y \text{ ou } x = y, \\ 0 & \text{sinon.} \end{cases}$$

Montrer que P est une matrice stochastique irréductible, récurrente et apériodique. Déterminer sa mesure de probabilité invariante.

Exercice 28. — Soient l'espace d'états $E = \{0, 1, \ldots, M\}$, $p = 1 - q \in]0, 1[$, et la matrice de transition P définie par

$$P(0,0) = q, \qquad P(x+1, x) = p, \qquad 0 \leqslant i \leqslant M-1\,;$$
$$P(x, x-1) = q, \quad 1 \leqslant x \leqslant M\,; \quad P(M, M) = p,$$



les autres coefficients étant nuls. Montrer que la mesure $\mu = ((p/q)^x)_{0\leqslant x\leqslant M}$ est réversible. En déduire une mesure de probabilité réversible.

9. **Un théorème ergodique**

Une propriété ergodique affirme grossièrement l'égalité de la limite des moyennes temporelles d'une quantité mesurée sur une unique trajectoire avec la moyenne de la même quantité mesurée sur toutes les trajectoires à un instant donné.

On peut imaginer énormément de situations où il n'y a pas de propriété ergodique : mesure du revenu d'un ensemble d'individus au cours de leur vie (des inégalités peuvent perdurer), billes lancées sur un billard avec trous (une bille finira sa trajectoire dans un trou particulier, mais en moyenne sur l'ensemble des billes aucun trou n'est privilégié).

Cependant, pour le développement de la Mécanique Statistique à partir du milieu du XIXe, posséder des propriétés ergodiques sur des systèmes complexes (gaz de particules) s'est révélé

assez essentiel pour faire progresser la théorie. Dans les manuels de Thermodynamique de premier cycle, on trouve souvent l'énoncé d'un théorème ergodique dit de Birkhoff : « Pour un temps très long, un système isolé passe dans chaque micro-état un temps proportionnel à la probabilité *a priori* de cet état ». Par micro-état, il faut simplement entendre qu'on ramène un espace d'états assez grand ($\mathbb{R}^n$ par exemple) à un espace d'états au plus dénombrable ; par système isolé, il faut entendre qu'il n'y a pas de perte ou de gain de particules ou autre ; par probabilité *a priori*, ce qu'on entend est un peu plus flou, mais pour nous ça ne peut être que la mesure de probabilité invariante si elle existe et est unique.

La recherche de propriétés ergodiques pour des systèmes physiques concrets ou leurs modèles mathématiques est toujours d'actualité. On appréciera que dans le contexte qui nous occupe des réponses générales existent. Un premier exemple est la loi forte des grands nombres que nous énonçons sous une forme qui nous sera utile plus loin.

THÉORÈME 25. — *Soit $(Y_n)_{n\geqslant 0}$ une suite de variables aléatoires positives, indépendantes et de même loi de moyenne $m = \mathbb{E}[Y_1] \in [0,\infty]$. Alors,*

$$\frac{Y_1 + \cdots + Y_n}{n} \longrightarrow m \qquad \textit{quand } n \textit{ tend vers l'infini, presque sûrement.}$$

Démonstration. — Cette propriété est supposée connue lorsque $m < \infty$: c'est un cas particulier de la loi forte des grands nombres de Kolmogorov. Supposons $m = \infty$. Pour $k > 0$, posons $Y_n^{(k)} = Y_n \wedge k$. La suite de variables aléatoires obtenues vérifie les hypothèses de la loi forte des grands nombres (indépendance, même loi, espérance $m^{(k)} = \mathbb{E}[Y_1^{(k)}] < \infty$). Puisque pour tout $n \geqslant 1$, on a $Y_n^{(k)} \leqslant Y_n$, alors

$$\frac{Y_1^{(k)} + \cdots + Y_n^{(k)}}{n} \leqslant \frac{Y_1 + \cdots + Y_n}{n} \qquad \text{pour tout } n \geqslant 1.$$

le terme de gauche tendant presque sûrement vers $m^{(k)}$, on a

$$m^{(k)} \leqslant \liminf_{n\to\infty} \frac{Y_1 + \cdots + Y_n}{n} \qquad \text{presque sûrement.}$$

Or $Y_1^{(k)} = Y_1 \wedge k$ tend simplement en croissant vers Y_1, et par convergence monotone $m^{(k)} = \mathbb{E}[Y_1^{(k)}]$ tend en croissant vers $m = \mathbb{E}[Y_1] = \infty$. En faisant tendre k vers l'infini, on obtient

$$\infty = \liminf_{n\to\infty} \frac{Y_1 + \cdots + Y_n}{n} \qquad \text{presque sûrement.}$$

La limite supérieure est alors aussi égale à l'infini, donc la limite existe et est infinie presque sûrement. □

Soit $(X_n)_{n\geqslant 0}$ une chaîne de Markov de loi initiale λ et de matrice de transition p. Le nombre de visites de l'état x à l'instant n est la variable aléatoire

$$V_n^x = \sum_{k=0}^{n-1} \mathbb{1}_{\{X_k = x\}},$$

la variable aléatoire V_n^x/n est la proportion de temps passé en x avant n.

Nous en arrivons au théorème ergodique.

THÉORÈME 26. — *Soient P une matrice de transition irréductible sur E et λ une mesure de probabilité sur E. Si $(X_n)_{n\geqslant 0}$ est une chaîne de Markov (λ, P), alors*

$$\frac{V_n^x}{n} \longrightarrow \frac{1}{m_x} \qquad \textit{quand } n \textit{ tend vers l'infini, presque sûrement,}$$

où $m_x = \mathbb{E}^x[T^x]$ *est le temps moyen de retour en* x.

De plus, si la chaîne est récurrente positive, pour toute fonction bornée $f : E \to \mathbb{R}$, *on a*

$$\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \longrightarrow \bar{f} \qquad \textit{quand } n \textit{ tend vers l'infini, presque sûrement,}$$

où $\bar{f} = \sum_{x\in E} \pi\{x\} \times f(x)$ *et* π *est l'unique mesure de probabilité invariante.*

Démonstration. — Si la matrice P est transiente, le nombre total de visites $V_\infty^x = \lim_{n\to\infty}$ est fini presque sûrement, donc

$$\frac{V_n^x}{n} \leqslant \frac{V_\infty^x}{n} \longrightarrow 0 \qquad \text{quand } n \text{ tend vers l'infini, presque sûrement,}$$

ce qui est égal à $1/m_x$ puisque $m_x = \mathbb{E}^x[T^x] = \infty$ par transience.

Supposons la matrice P récurrente. Soient $x \in E$ et $T = T^x$. On sait que $\mathbb{P}\{T < \infty\} = 1$. Alors par la propriété forte de Markov, $(X_{T+n})_{n\geqslant 0}$ est une chaîne de Markov $(\delta_{\{x\}}, P)$ (et de plus indépendante de $\mathcal{F}_T^X$ la tribu des événements antérieurs à T). La proportion asymptotique du temps passé en x étant la même pour $(x_n)_{n\geqslant 0}$ et $(X_{T+n})_{n\geqslant 0}$, on peut supposer pour la suite $\lambda = \delta_{\{x\}}$.

Soit S_r^x la longueur de la r-ième excursion en dehors de x. On sait que les variables $(S_r^x)_{r\geqslant 1}$ sont indépendantes, identiquement distribuées, de moyenne m_x. D'une part, l'instant de dernière visite de x avant n est inférieur ou égal à $n-1 \leqslant n$, c'est-à-dire

$$S_1^x + \cdots + S_{V_n^x - 1}^x \leqslant n - 1 \leqslant n\,;$$

d'autre part, l'instant de première visite de x après $n-1$ est supérieur ou égal à n, c'est-à-dire

$$S_1^x + \cdots + S_{V_n^x}^x \geqslant n.$$

On en déduit que

$$\frac{S_1^x + \cdots + S_{V_n^x - 1}^x}{V_n^x} \leqslant \frac{n}{V_n^x} \leqslant \frac{S_1^x + \cdots + S_{V_n^x}^x}{V_n^x}. \qquad (*)$$

Par la loi forte des grands nombres présentée sous la forme précédente, on a :

$$\frac{S_1^x + \cdots + S_n^x}{} \longrightarrow m_x \qquad \text{quand } n \text{ tend vers l'infini, presque sûrement.}$$

Puis, comme P est récurrente,

$$V_n^x \longrightarrow \infty \qquad \text{quand } n \text{ tend vers l'infini, presque sûrement.}$$

On obtient alors par encadrement

$$\frac{n}{V_n^x} \longrightarrow m_x \qquad \text{quand } n \text{ tend vers l'infini, presque sûrement,}$$

ou encore

$$\frac{V_n^x}{n} \longrightarrow \frac{1}{m_x} \qquad \text{quand } n \text{ tend vers l'infini, presque sûrement.}$$

Supposons que P admet une mesure de probabilité invariante π. Soit $f : E \to \mathbb{R}$ une application bornée, et on supposera, par commodité, $\sup_{x\in E} |f(x)| \leqslant 1$. Alors pour tout

$F \subset E$, on a

$$\begin{aligned}
\left|\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) - \bar{f}\right| &= \left|\frac{1}{n}\sum_{k=0}^{n-1}\Big(\sum_{x\in E} f(x)\mathbb{1}_{\{X_k=x\}}\Big) f(X_k) - \sum_{x\in E} f(x)\pi\{x\}\right| \\
&= \left|\sum_{x\in E} f(x)\frac{1}{n}\sum_{k=0}^{n-1}\mathbb{1}_{\{X_k=x\}} - \sum_{x\in E} f(x)\pi\{x\}\right| \\
&= \left|\sum_{x\in E} f(x)\frac{V_n^x}{n} - \sum_{x\in E} f(x)\pi\{x\}\right| = \left|\sum_{x\in E} f(x)\left(\frac{V_n^x}{n} - \pi\{x\}\right)\right| \\
&\leqslant \left|\sum_{x\in F} f(x)\left(\frac{V_n^x}{n} - \pi\{x\}\right)\right| + \left|\sum_{x\notin F} f(x)\left(\frac{V_n^x}{n} - \pi\{x\}\right)\right| \\
&\leqslant \sum_{x\in F}|f(x)| \times \left|\frac{V_n^x}{n} - \pi\{x\}\right| + \sum_{x\notin F}|f(x)| \times \left(\frac{V_n^x}{n} + \pi\{x\}\right) \\
&\leqslant \sum_{x\in F}\left|\frac{V_n^x}{n} - \pi\{x\}\right| + \sum_{x\notin F}\left(\frac{V_n^x}{n} + \pi\{x\}\right) = \sum_{x\in F}\left|\frac{V_n^x}{n} - \pi\{x\}\right| + \sum_{x\notin F}\pi\{x\} + \sum_{x\notin F}\frac{V_n^x}{n} \\
&= \sum_{x\in F}\left|\frac{V_n^x}{n} - \pi\{x\}\right| + \sum_{x\notin F}\pi\{x\} + 1 - \sum_{x\in F}\frac{V_n^x}{n} = \sum_{x\in F}\left|\frac{V_n^x}{n} - \pi\{x\}\right| \\
&\quad + 2\sum_{x\notin F}\pi\{x\} + \sum_{x\notin F}\pi\{x\} - \sum_{x\in F}\frac{V_n^x}{n} \\
&= \sum_{x\in F}\left|\frac{V_n^x}{n} - \pi\{x\}\right| + 2\sum_{x\notin F}\pi\{x\} + \sum_{x\notin F}\left(\pi\{x\} - \frac{V_n^x}{n}\right) \leqslant 2\sum_{x\in F}\left|\frac{V_n^x}{n} - \pi\{x\}\right| + 2\sum_{x\notin F}\pi\{x\}.
\end{aligned}$$

Soit $A \subset \Omega$ l'événement sur lequel $V^x(\omega)/n \to \pi$ pour tout $x \in E$, $\omega \in A$, qui est un événement de probabilité 1. Soit $\varepsilon > 0$. On choisit $F \subset E$ fini tel que $\pi(E\backslash F) < \varepsilon/4$. Ensuite, soit N (dépendant de l'aléa $\omega \in A$) tel que pour tout $n \geqslant N$

$$\sum_{x\in F}\left|\frac{V_n^x}{n} - \pi\{x\}\right| < \frac{\varepsilon}{4} \qquad \text{sur tout } A.$$

Alors on a

$$\left|\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) - \bar{f}\right| < \varepsilon \qquad \text{sur tout } A,$$

ce qui montre la convergence sur A, c'est-à-dire la convergence presque sûre annoncée. □

Exemple. — Si P est irréductible, récurrente positive, en considérant, pour $x \in E$, $f_x : y \in E \mapsto \delta(x,y)$ (symbole de Kronecker), on obtient que

$$\frac{1}{n}\sum_{k=0}^{n-1}\mathbb{1}_{\{X_k=x\}} \longrightarrow \pi\{x\} \qquad \text{quand } n \text{ tend vers l'infini, presque sûrement,}$$

c'est-à-dire que la fréquence d'apparition de x dans la chaîne tend presque sûrement vers $\pi\{x\}$.

10. **Estimation statistique des coefficients d'une matrice de transition**

Précisons d'abord le vocabulaire des statisticiens et leur thématique. Soit $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires ; certaines lois ou mécanismes régissant cette suite sont inconnus, par exemple une quantité numérique p qui est dite *paramètre*.

Définition 14. — (i) Un *échantillon* est une suite généralement finie de variables aléatoires $(X_0, \dots, X_N)$.

(ii) Un *échantillon observé* est la suite de valeurs prises par ces variables aléatoires pour un $\omega \in \Omega$ donné, $(X_0(\omega), \dots, X_N(\omega)) = (x_0, \dots, x_n)$.

(iii) Un estimateur de p est une simple fonction (mesurable) $\Lambda_N = \ell_N(X_0, \dots, X_N)$ de l'échantillon ; un estimateur, ou plus précisément une suite d'estimateurs, est consistant si Λ_N converge en probabilité (ou presque sûrement, ce qui est plus fort) vers le paramètre p lorsque N tend vers l'infini ; il est sans biais si $\mathbb{E}[\Lambda_N] = p$.

(iv) Une estimation est un estimateur observé $\lambda_N = \ell_N(x_0, \dots, x_N)$.

Il est fréquent dans la pratique que la suite de variables aléatoires $(X_n)_{n\geqslant 0}$ soit une suite de variables aléatoires indépendantes et identiquement distribuées, auquel cas ce qui intéresse est de déterminer la loi de X_0 ou certaines de ses caractéristiques. Ceci repose le plus souvent sur la loi des grands nombres pour montrer la consistance des estimateurs.

Néanmoins, il n'est pas rare d'avoir à considérer une chaîne de Markov $(X_n)_{n\geqslant 0}$ par exemple avec des modèles biologiques liés aux chaînes d'ADN, ou encore certaines évolutions liées à la vie courante (commerce, finance) admettant une telle modélisation. Ce qui est intéressant est alors d'évaluer les coefficients de la matrice de transition associée, car, sachant qu'on est dans tel état à l'instant n, c'est elle qui permettra d'évaluer ce qu'il se passera plus tard.

Ainsi, on considère une chaîne de Markov $(X_n)_{n\geqslant 0}$ de loi initiale π et de matrice de transition P sur E. Pour un $\omega \in \Omega$, nous supposons disposer de $(X_0(\omega), \dots, X_N(\omega)) = (x_0, \dots, x_n)$ qui sont $N+1$ observations de la chaîne et nous cherchons à l'aide de celles-ci à estimer les coefficients de P. La mesure initiale π n'interviendra pas.

La fonction de log-vraisemblance d'une matrice de transition Q sur E est une fonction tenant compte des observations $(x_0, \dots, x_N)$ qui ont été faites. Elle est définie par

$$\begin{aligned}\ell(Q) &= \ln\big(Q(x_0,x_1) \times \cdots \times Q(x_{N-1},x_N)\big) \\ &= \ln Q(x_0,x_1) + \cdots + \ln Q(x_{N-1},x_N) = \sum_{x,y\in E} N(x,y) \ln Q(x,y)\end{aligned}$$

où, pour tout x, $y \in E$, $N(x,y)$ est le nombre de transitions (en un pas) observées dans $(x_0, \dots, x_N)$ de x à y. Une méthode standard et peu comprise de la Statistique pour trouver une estimation $\widehat{P}$ de P à partir des données $(x_0, \dots, x_N)$ consiste à rechercher $\widehat{P}$ parmi les matrices stochastiques qui maximisent la fonction de log-vraisemblance ℓ. Le problème ici est qu'on a affaire à une maximisation avec contrainte (matrices stochastiques). Nous allons utiliser la méthode des multiplicateurs de Lagrange. Considérons $\mathcal{O}$ l'ensemble ouvert des matrices à coefficients réels *strictement* positifs. Une matrice $Q \in \mathcal{O}$ est stochastique si elle vérifie les contraintes

$$h_x(Q) = \sum_{y\in E} Q(x,y) - 1 = 0, \qquad x \in E.$$

Notons $\mathcal{C}$ l'ensemble de telles matrices. On définit le lagrangien associé à ce problème de maximisation par

$$\mathcal{L} : \mathcal{O} \times \mathbb{R}^E \longrightarrow \mathbb{R}$$
$$(Q, \lambda) \longmapsto \ell(Q) + \sum_{x\in E} \lambda_x \times h_x(Q) = \ell(Q) + \sum_{x\in E} \lambda_x \Big(\sum_{y\in E} Q(x,y) - 1 \Big).$$

Une condition nécessaire pour que (Q, λ) soit un extremum local de cette fonctionnelle régulière est que sa différentielle s'annulle. Pour la partie en la variable λ, c'est

$$\mathrm{d}_\lambda \mathcal{L}(Q, \lambda) = \sum_{x\in E} h_x(Q)\, \mathrm{d}\lambda_x = 0,$$

donc que $h_x(Q) = 0$ pour tout $x \in E$, ce qui assure qu'alors Q est une matrice stochastique. Pour la partie en la variable Q, on a

$$\begin{aligned}
\mathrm{d}_Q \mathcal{L}_(Q, \lambda) &= \mathrm{d}_Q \Big(\ell(Q) + \sum_{x\in E} \lambda_x \times h_x(Q) \Big) \\
&= \sum_{x,y\in E} \mathrm{d}_Q \big(N(x,y) \ln Q(x,y) \big) + \sum_{x\in E} \mathrm{d}_Q \Big(\lambda_x \sum_{y\in E} Q(x,y) - 1 \Big) \\
&= \sum_{x,y\in E} N(x,y) \frac{\mathrm{d}Q(x,y)}{Q(x,y)} + \sum_{x,y\in E} \lambda_x \, \mathrm{d}Q(x,y) \\
&= \sum_{x,y\in E} \Big(\frac{N(x,y)}{Q(x,y)} - \lambda_x \Big) \mathrm{d}Q(x,y).
\end{aligned}$$

On trouve alors une unique solution, sous réserve que $\lambda_x \neq 0$,

$$\widehat{P}(x,y) = -\frac{N(x,y)}{\lambda_x}\,, \qquad x,\, y \in E.$$

Il faut ensuite préciser les λ_x pour obtenir une matrice stochastique

$$1 = \sum_{y\in E} \widehat{P}(x,y) = -\sum_{y\in E} \frac{N(x,y)}{\lambda_x} \qquad \text{donc} \qquad \lambda_x = -\sum_{y\in E} N(x,y),$$

λ_x est donc l'opposé du nombre d'apparitions de x dans $(x_0, \ldots, x_{N-1})$. La fonction ℓ étant concave sur l'ouvert convexe $\mathcal{O}$, et les contraintes étant affines, l'extremum $\widehat{P}$ ainsi trouvé est un maximum global de ℓ sur $\mathcal{C}$ d'après un théorème usuel d'optimisation. Ainsi

$$\begin{aligned}
\widehat{P}(x,y) = \frac{N(x,y)}{\sum_{z\in E} N(x,z)} &= \frac{\sum_{n=0}^{N-1} \mathbb{1}_{\{x_n = x, x_{n+1} = y\}}}{\sum_{z\in E} \sum_{n=0}^{N-1} \mathbb{1}_{\{x_n = x, x_{n+1} = z\}}} \\
&= \frac{\sum_{n=0}^{N-1} \mathbb{1}_{\{x_n = x, x_{n+1} = y\}}}{\sum_{n=0}^{N-1} \sum_{z\in E} \mathbb{1}_{\{x_n = x, x_{n+1} = z\}}} = \frac{\sum_{n=0}^{N-1} \mathbb{1}_{\{x_n = x, x_{n+1} = y\}}}{\sum_{n=0}^{N-1} \mathbb{1}_{\{x_n = x\}}}
\end{aligned}$$

qui est la proportion de transitions observées de x à y est l'estimation cherchée de $P(x,y)$, l'estimateur s'obtenant en remplaçant x_n par X_n dans la formule ci-dessus.

La méthode ci-dessus ne fonctionne pas si pour un certain x, $\lambda_x = 0$, ce qui signifie que x n'apparaît pas parmi $(x_0, \ldots, x_{N-1})$. Dans ce cas on se désintéresse de l'état x, que ce soit pour estimer un $P(x,y)$, ce qui donnerait 0 pour estimation naturelle, ou pour estimer un $P(y,x)$, puisque si $x_N \neq x$, alors x n'est pas vu, et si $x_N = x$, $x_{N-1} = y$, $\widehat{P}(y,x)$ est égal à 1 divisé par le nombre d'apparition de y parmi $(x_0, \ldots, x_{N-1})$ ce qui peut être trop peu

représentatif. On fait comme si l'état x n'existait pas. Bien sûr, si pour de plus grands N ceci cesse d'être le cas, on prend alors en compte x.

Cas de la récurrence. — Supposons la matrice P irréductible et récurrente. Le dénominateur de

$$\widehat{P}(x,y) = \frac{\sum_{n=0}^{N-1} \mathbb{1}_{\{X_n=x, X_{n+1}=y\}}}{\sum_{n=0}^{N-1} \mathbb{1}_{\{X_n=x\}}}$$

est le nombre de visites de x avant n et on sait qu'il tend vers l'infini. On peut réécrire le numérateur comme la somme des indicatrices de l'atteinte de y après un premier pas au début d'une excursion hors de x. Puisqu'on sait que les excursions sont indépendantes et de même loi, ces nouvelles indicatrices sont elles aussi indépendantes et de même loi, la loi de Bernoulli de paramètre $P(x,y)$. Par la loi forte des grands nombres, on obtient alors la convergence presque sûre de la suite d'estimateurs vers l'espérance de ces indicatrices, à savoir $P(x,y)$.

Cas de la transience. — Lorsque la matrice P est irréductible transiente, il est clair en revanche que cet estimateur est rarement consistant. En effet, un état x n'étant fréquenté qu'un nombre fini de fois, la suite des estimations est stationnaire et ne donne alors qu'une fraction pouvant s'éloigner notablement de la probabilité de transition cherchée. Il faut alors changer de méthode. On s'arrange pour faire suffisamment d'observations pour voir la chaîne quitter N fois l'état x, ce qui revient à faire partir N fois la chaîne de l'état x de manières indépendantes. En maximisant dans ce cadre

$$\sum_{x,y\in E} N(x,y) \ln Q(x,y) \qquad \text{avec} \qquad \sum_{y\in E} Q(x,y) = 1,$$

on obtient alors pour estimateur du maximum de vraisemblance

$$\widehat{P}(x,y) = \frac{N(x,y)}{N}\,.$$

Comme $N(x,y) = Y_1 + \cdots + Y_N$, où $Y_n = 1$ si la n-ième transition va de x à y, et 0 sinon, par la propriété de Markov forte, les Y_n sont des variables aléatoires indépendantes identiquement distribuées et on a par la loi forte des grands nombres

$$\widehat{P}(x,y) = \frac{N(x,y)}{N} = \frac{Y_1 + \cdots + Y_N}{N} \longrightarrow \mathbb{E}[Y_1] = P(x,y) \qquad \text{presque sûrement,}$$

d'où la consistance.

Cas de non irréductibilité. — On essaie de combiner au mieux les deux approches précédentes mais c'est très délicat car si on ne connaît pas la matrice de transition P, il faut pouvoir tester si un état est récurrent ou transient...

Chapitre II

QUELQUES APPLICATIONS

1. **Simulations de lois de variables aléatoires**

L'objectif est, étant donnée une loi de probabilité π sur un ensemble E, de définir à partir d'une variable aléatoire U uniformément distribuée sur $[0,1]$, ou d'une suite $(U_n)_{n\geqslant 0}$ de variables aléatoires indépendantes uniformément distribuées sur $[0,1]$, une variable aléatoire X à valeurs dans E de loi π.

Ce qui motive ceci est qu'on sait assez bien « simuler » une telle variable aléatoire U, ou une telle suite de variables aléatoires $(U_n)_{n\geqslant 0}$, à l'aide de procédés informatiques : l'appel d'une fonction « `random` » (`rnd`, `rand`, `srand`, `grand`, etc., en fonction du logiciel employé) retourne un nombre, généralement compris entre 0 et 1 (sinon on adapte en divisant le résultat par la plus grande valeur qu'on peut obtenir) ; des appels successifs semblent donner des résultats sans liaison apparente — indépendants — et ceux-ci se répartissent — presque — uniformément dans l'intervalle $[0,1]$.

Une méthode classique pour obtenir un générateur de nombres aléatoires est de considérer les termes successifs d'une suite récurrente linéaire

$$v_0 = \text{donnée (graine aléatoire, random seed)}, \qquad v_{n+1} = (a \times v_n + b) \bmod m,$$

où a, b et m sont des entiers. La méthode est dite « méthode des congruences linéaires ». En choisissant convenablement ces trois entiers, en particulier m grand, la suite $(u_n = v_n/m)_{n\geqslant 0}$ ressemble, en un certain sens, à une réalisation $(U_n(\omega))_{n\geqslant 0}$ de la suite $(U_n)_{n\geqslant 0}$.

Exemples. — *a)* Soit $\pi = \mathcal{B}(1,p)$ la loi de Bernoulli de paramètre $p \in [0,1]$,

$$\pi = (1-p)\delta_0 + p\delta_1.$$

La variable aléatoire

$$X = \begin{cases} 0 & \text{si } U < 1-p, \\ 1 & \text{sinon,} \end{cases}$$

est de loi π.

b) Soit $\pi = \mathcal{B}(n,p)$ la loi de Binomiale de paramètres $n \geqslant 1$ et $p \in [0,1]$,

$$\pi = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta_k.$$

En posant, pour $1 \leqslant k \leqslant n$,

$$X_k = \begin{cases} 0 & \text{si } U_k < 1-p, \\ 1 & \text{sinon,} \end{cases}$$

la variable aléatoire $X = X_1 + \cdots + X_n$ est de loi π puisque c'est la somme de n variables aléatoires indépendantes de loi de Bernoulli de paramètre p.

c) Soit π une loi de probabilité discrète sur un ensemble au plus dénombrable $E = (x_n)_{1\leqslant n<N}$ avec $N \in \mathbb{N}^* \cup \{\infty\}$. Notons $\pi_n = \pi\{x_n\}$, $1 \leqslant n < N$. En posant

$$X = x_n \qquad \text{si} \qquad \sum_{1\leqslant k\leqslant n-1} \pi_k \leqslant U < \sum_{1\leqslant k\leqslant n} \pi_k, \quad 1 \leqslant n < N,$$

la variable aléatoire X est de loi π.

d) Soit $\pi = \mathcal{U}(a,b)$ la loi uniforme sur l'intervalle $[a,b]$,

$$\pi(\mathrm{d}x) = \frac{\mathbb{1}_{[a,b]}(x)}{b-a}\,\mathrm{d}x.$$

La variable aléatoire

$$X = a + (b-a)U$$

est de loi π.

Nous rappelons que si π est une mesure de probabilité sur $\mathbb{R}$, sa fonction de répartition est l'application $F : \mathbb{R} \to [0,1]$ définie par $F(x) = \pi(]-\infty, x])$, $x \in \mathbb{R}$. C'est une fonction croissante, continue à droite et limitée à gauche, tendant respectivement vers 0 et 1 en $-\infty$ et $+\infty$.

Proposition 6. — *Soit F la fonction de répartition d'une mesure de probabilité π sur $\mathbb{R}$. On suppose que F est continue, strictement croissante sur un intervalle $]a,b[\subset \mathbb{R}$, et qu'elle vérifie*

$$\lim_{x \searrow a} F(x) = 0 \qquad \textit{et} \qquad \lim_{x \nearrow b} F(x) = 1.$$

En définissant $F^{-1} : [0,1] \to [a,b]$ comme l'application inverse de F sur $[a,b]$, la variable aléatoire $X = F^{-1}(U)$ est de loi π.

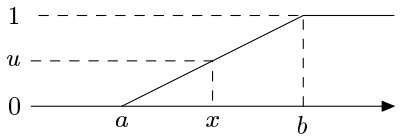
Démonstration. — Il est clair que la variable aléatoire X est à valeurs dans $[a,b]$. Notons F_X la fonction de répartition de sa loi, $F_X(x) = \mathbb{P}\{X \leqslant x\}$, $x \in \mathbb{R}$. On a immédiatement $F_X(x) = 0$ si $x < a$ et $F_X(x) = 1$ si $x \geqslant b$. Si $x \in [a,b]$,

$$F_X(x) = \mathbb{P}\{X \leqslant x\} = \mathbb{P}\{F^{-1}(U) \leqslant x\} = \mathbb{P}\{U \leqslant F(x)\} = F(x)$$

puisque U est de loi uniforme sur $[0,1]$. Donc $F_X = F$, la loi de X est π. □

Exemples. — *a)* Soit π la loi uniforme sur l'intervalle $[a,b]$, Sa fonction de répartition est donnée par

$$x \in \mathbb{R}, \qquad F(x) = \begin{cases} 0 & \text{si } x < a \\ (x-a)/(b-a) & \text{si } a \leqslant x \leqslant b \\ 1 & \text{si } x \geqslant b. \end{cases}$$



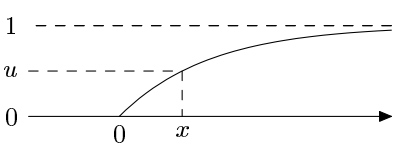
Les conditions de la proposition précédente sont statisfaites. En résolvant $u = F(x) = (x-a)/(b-a)$ pour $x \in [a,b]$, on obtient $x = a + (b-a)u$. La variable aléatoire $X = a + (b-a)U$ est donc de loi uniforme sur $[a,b]$.

b) Soit π la loi exponentielle de paramètre $\lambda > 0$,

$$\pi(\mathrm{d}x) = \mathbb{1}_{\mathbb{R}_+}(x) \times \lambda\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x.$$

Sa fonction de répartition est donnée par

$$x \in \mathbb{R}, \qquad F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - \mathrm{e}^{-\lambda x} & \text{si } x \geqslant 0. \end{cases}$$



Les conditions de la proposition précédente sont statisfaites avec $a = 0$ et $b = \infty$. En résolvant $u = F(x) = 1 - \mathrm{e}^{-\lambda x}$ pour $x \geqslant 0$, on obtient $x = -\ln(1-u)/\lambda$. La variable aléatoire

$$X = -\frac{\ln(1-U)}{\lambda}$$

est donc de loi exponentielle de paramètre λ. Comme $1-U$ est de loi uniforme sur $[0,1]$, on peut noter que la variariable aléatoire $Y = -\ln(U)/\lambda$ est aussi de loi exponentielle de paramètre λ.

PROPOSITION 7. — *Soit π une loi de probabilité sur $\mathbb{R}$ et F sa fonction de répartition. En définissant l'inverse généralisé F^{-1} de F par*

$$u \in [0,1], \qquad F^{-1}(u) = \inf\{y \in \mathbb{R} : u \leqslant F(y)\} \in \bar{\mathbb{R}},$$

la variable aléatoire $X = F^{-1}(U)$ est de loi π.

Démonstration. — La preuve est laissée en exercice. Indiquons tout de même qu'il s'agit de montrer que, pour tout $x \in \mathbb{R}$, les événements $\{X \leqslant x\}$ et $\{U \leqslant F(x)\}$ sont égaux en se servant de la croissance et de la continuité à droite de F. □

Les cas où on sait inverser de manière explicite ou efficace une fonction de répartition sont rares. Il faut en général avoir recours à d'autres méthodes. Un exemple assez remarquable de cette situation est donné par l'exercice suivant. Il concerne le cas de la loi normale standard $\mathcal{N}(0,1)$ et peut aisément servir de point de départ pour des lois normales $\mathcal{N}(\mu,\sigma^2)$ sur $\mathbb{R}$, voire sur $\mathbb{R}^n$.

EXERCICE 29 (LA MÉTHODE POLAIRE). — Soient $\Theta : (\Omega,\mathcal{A},\mathbb{P}) \to [0,2\pi[$ de loi uniforme sur $[0,2\pi[$ et $R : (\Omega,\mathcal{A},\mathbb{P}) \to \mathbb{R}_+$ de loi admettant pour densité

$$r \in \mathbb{R}, \qquad p_R(r) = \mathbb{1}_{\mathbb{R}_+}(r) \times r\,\mathrm{e}^{-r^2/2},$$

indépendantes, obtenues en posant

$$R = \sqrt{-\ln U_1} \qquad \text{et} \qquad \Theta = 2\pi U_2.$$

Montrer que $X = R\cos\Theta$ et $Y = R\sin\Theta$ sont deux variables aléatoires de loi $\mathcal{N}(0,1)$ indépendantes. Établir la réciproque.

Simulation pratique de chaînes de Markov. — D'après un exemple précédent, nous savons théoriquement simuler une loi de probabilité discrète π sur un ensemble au plus dénomblable E. En pratique l'espace des états peut être grand mais est toujours fini à cause des limitations imposées par l'outil de calcul, l'ordinateur. D'ailleurs, même si on désire simuler une loi de probabilité sur un espace plus que dénombrable — $\mathbb{R}$ ou $\mathbb{R}^n$ par exemple —, cette contrainte de calcul nous ramenera toujours à un ensemble fini. Nous pouvons donc considérer pour une chaîne de Markov l'espace des états $E = \{1,\dots,N\}$ fini et P une matrice de transition sur E.

Rappelons que les appels successifs d'un générateur aléatoire fournissent une suite $(u_n)_{n\geqslant 1}$ interprétée comme une réalisation $(U_n(\omega))_{n\geqslant 1}$ d'une suite $(U_n)_{n\geqslant 1}$ de variables aléatoires indépendantes toutes de loi uniforme sur $[0,1]$. La simulation d'une chaîne de Markov $(X_n)_{n\geqslant 0}$ ne pourra consister qu'en une réalisation $(x_n)_{n\geqslant 0} = (X_n(\omega))_{n\geqslant 0}$ d'une telle suite construite à partir de $(u_n)_{n\geqslant 1}$. Supposons que la matrice P n'a pas de structure particulière, par exemple ce n'est la matrice d'une marche aléatoire simple ou de l'urne d'Ehrenfest. L'exemple de la simulation d'une loi discrète indique qu'il faudra calculer des sommes de probabilités, il est peut-être avantageux de commencer par définir la matrice S de ces sommes :

$$S(x,y) = \sum_{z\leqslant y} P(x,z)$$

Ensuite on choisit, ou tire au hasard suivant une loi initiale discrète π, la valeur de x_0. Posons $i = x_0$ et tirons $u_1 \in [0,1]$. En comparant les valeurs successives de $S(x,z)$ avec u_1, $1 \leqslant z \leqslant N$, on détermine l'unique $1 \leqslant y \leqslant N$ tel que

$$S(x,y-1) \leqslant u_1 < S(x,y)$$

où on convient que $S(x,0) = 0$ et admet que $u_1 = 1$ n'arrive jamais (sinon, on adapte). On pose alors $x_1 = y$. Puis on continue de même pour $x_2, x_3, \ldots$

Lorsque la chaîne évolue selon un schéma algorithmique plus simple que la donnée de la matrice de transition, comme pour les deux exemples donnés précédemment, on profite de cette algorithmique pour simplifier la procédure.

Exercice 30. — (i) Définir un algorithme de simulation pour une marche aléatoire simple non nécessairement symétrique sur $\mathbb{Z}$, $\mathbb{Z}^2$, ...

(ii) Définir un algorithme de simulation pour l'évolution de l'urne d'Ehrenfest.

2. Simulation de lois, l'algorithme de Hastings–Metropolis

Soit E un espace d'états au plus dénombrable et μ une mesure finie non nulle sur E. On souhaite simuler la loi de probabilité $\pi = \mu/\mu(E)$. Deux problèmes peuvent se poser. Le premier est le calcul de $\mu(E)$ qui est dans certains cas pratiques difficile, le second est qu'on veut disposer d'une méthode qui ne nécessite pas trop de calculs. Pour notre propos, on supposera $\mu\{x\} > 0$ pour tout $x \in E$, ce qui revient à ignorer les états de mesure nulle.

Exemple. — Soit E l'ensemble des permutations σ de $\{1, \ldots, 10\}$ vérifiant $\sum_{j=1}^{10} j \times \sigma(j) \geqslant 250$. Cet ensemble est fini, et non vide puisqu'il contient notamment l'ensemble des permutations laissant fixe $\{7, 8, 9, 10\}$. Nous souhaitons choisir uniformément une permutation dans E. Ici μ est la mesure uniforme sur E et $\mu(E) = \mathrm{Card}(E)$, ce cardinal semblant difficile à expliciter par des méthodes combinatoires ou autres.

Idée. — *Construire une matrice de transition P sur E, irréductible et apériodique et telle que à μ en soit une mesure réversible. En effet, on aura alors :*

(i) *La mesure de probabilité $\pi = \mu/\mu(E)$ est invariante par P.*

(ii) *Il y a convergence vers l'équilibre : si $(X_n)_{n\geqslant 0}$ est une chaîne de Markov (λ, P), alors, pour tout $x \in E$, $\mathbb{P}\{X_n = x\} \to \pi\{x\}$ quand n tend vers l'infini.*

Construction. — On considère une matrice de transition Q sur E, irréductible et apériodique telle que $Q(x,y) = 0$ implique $Q(y,x) = 0$ — construite par exemple à partir d'un graphe de sommets E.

Soit $D = \{(x,y) \in E \times E : Q(x,y) > 0\}$ et soit $\alpha : D \to]0,1]$ défini par

$$\alpha(x,y) = \min\left(\frac{Q(y,x) \times \mu\{y\}}{Q(x,y) \times \mu\{x\}}, 1\right),$$

où on notera l'importance de l'hypothèse de stricte positivité de μ pour cette définition. On définit la matrice P par

$$x \neq y \in E, \qquad P(x,y) = \begin{cases} Q(x,y) \times \alpha(x,y) & \text{si } (x,y) \in D, \\ 0 & \text{sinon,} \end{cases}$$

et

$$x \in E, \qquad P(x,x) = 1 - \sum_{y \neq x} P(x,y).$$

Ceci définit bien une matrice stochastique : les sommes sur les lignes valent 1 ; les coefficients non diagonaux sont positifs ; et pour les coefficients diagonaux on a $0 \leqslant \sum_{y\neq x} P(x,y) \leqslant \sum_{y\neq x} Q(x,y) \leqslant 1$, et donc $P(x,x) \geqslant 0$.

L'irréductibilité de P est une conséquence immédiate de celle de Q puisque si on a, pour $x_0, x_1, \ldots, x_{n-1}, x_n \in E$,

$$Q(x_0, x_1) \times \cdots \times Q(x_{n-1}, x_n) > 0 \qquad \text{alors} \qquad P(x_0, x_1) \times \cdots \times P(x_{n-1}, x_n) > 0$$

et réciproquement. L'apériodicité de P résulte aussi de celle de Q puisqu'en utilisant ce qui précède constate que $Q^n(x, y) > 0$ si et seulement si $P^n(x, y) > 0$. Nous allons voir que la mesure μ est réversible pour P. Soient $(x, y) \in D$, $x \neq y$. Si par exemple $\alpha(x, y) \leqslant 1$, alors $\alpha(y, x) = 1$ et on a

$$\begin{aligned}\mu\{x\} \times P(x, y) = \mu\{x\} \times Q(x, y) \times \alpha(x, y) &= \mu\{x\} \times Q(x, y) \times \frac{Q(y, x) \times \mu\{y\}}{Q(x, y) \times \mu\{x\}} \\ &= Q(y, x) \times \mu\{y\} = \mu\{y\} \times Q(y, x) \times \alpha(y, x) = \mu\{y\} \times P(y, x).\end{aligned}$$

Pour $(x, y) \notin D$, $x \neq y$, on a $\mu\{x\} \times P(x, y) = 0 = \mu\{y\} \times P(y, x)$. Enfin, il n'y a rien à dire pour le cas $x = y$. La matrice P ainsi construite vérifie les conditions imposées.

3. **Le recuit simulé**

La méthode du recuit provient de la métallurgie et de la verrerie : afin d'éliminer dans la structure d'un produit (métal, verre) des inhomogénéités ou fractures qui pourraient le rendre cassant, on le chauffe puis le laisse lentement refroidir afin que sa structure interne se réorganise de manière plus stable. Ceci a inspiré une méthode probabiliste de recherche de minima globaux de fonction. Elle utilise des chaînes de Markov inhomogènes, l'inhomogénéité provenant de l'abaissement progressif de la température, et donc la modification au fil du temps des règles d'évolution de la chaîne.

Soit $U : E \to \mathbb{R}_+$ une fonction positive bornée. Cette fonction sera appelée *énergie* ou *potentiel* sur l'espace des états E, espace toujours supposé au plus dénombrable muni de sa tribu discrète. On cherche à déterminer l'ensemble E_U de ses minima globaux, ensemble qui peut être éventuellement vide lorsque E est infini dénombrable.

On explore aléatoirement l'espace des états selon une matrice de transition Q sur E vérifiant

$$\begin{cases} \text{(i) il existe une mesure } \lambda \text{ sur } E \text{ réversible pour } Q\,; \\ \text{(ii) il existe } p \in \mathbb{N} \text{ tel que } Q^p \text{ ait tous ses coefficients positifs.} \end{cases} \qquad (\mathcal{P})$$

La matrice Q permettra de déterminer un nouvel état vers lequel on pourra choisir ou non de transiter.

Remarque. — La condition (i) implique que λ est une mesure invariante de Q. La condition (ii) implique que pour tout $n \geqslant 0$, Q^{p+n} a tous ses coefficients strictement positifs ; en particulier, Q est apériodique. La réciproque est vraie lorsque E est fini.

Exemple. — Lorsque E est un ensemble fini, on peut construire Q à partir d'un graphe connexe de sommets E, comme il a été vu précédemment. Lorsque E est infini, on peut, par exemple identifier E à $\mathbb{Z}$, se donner une mesure de probabilité μ sur $\mathbb{Z}$ partout strictement positive, symétrique par rapport à 0, et définir $Q(x, y) = \mu\{y - x\}$. La condition (i) est alors satisfaite avec λ la mesure uniforme sur $\mathbb{Z}$ et (ii) est vérifiée avec $p = 1$.

La température est notée $(T_n)_{n\geqslant 0}$; c'est une suite numérique, décroissante de nombres strictements positifs tendant vers 0. Pour tout $n \geqslant 0$, on note $\beta_n = 1/T_n$ l'inverse de la température. La suite $(\beta_n)_{n\geqslant 0}$ est croissante et tend vers l'infini. On définit de manière algorithmique une chaîne de Markov inhomogène $(X_n)_{n\geqslant 0}$ qui se stabilise lorsque n tend vers l'infini dans l'ensemble des minima globaux E_U de U.

Heuristique[1]. — *Supposons que pour $n \geqslant 0$, l'algorithme soit dans un état $X_n = x \in E$ d'énergie $U(x)$. Puisqu'on ignore si c'est un minimum global, on choisit un nouvel état y selon la loi $Q(x, .)$. Alors*

— *si $U(y) \leqslant U(x)$, on pose $X_{n+1} = y$;*

— *si $U(y) > U(x)$, alors on choisit*

$$X_{n+1} = \begin{cases} x & \text{avec probabilité } 1 - \mathrm{e}^{-\beta_n (U(y)-U(x))}, \\ y & \text{avec probabilité } \mathrm{e}^{-\beta_n (U(y)-U(x))}. \end{cases}$$

Remarques. — *a*) On constate que le passage d'un état x à un état y d'énergie strictement supérieure est d'autant plus difficile que la différence des énergies est grande.

b) Pour des températures proches de 0, c'est-à-dire des β_n grands, il est de plus en plus difficile de passer à un état d'énergie supérieure, il est donc plus fortement probable de passer dans un état d'énergie inférieure ou de rester dans le même état.

La difficulté est de choisir la suite de température $(T_n)_{n\geqslant 0}$ pour que $(X_n)_{n\geqslant 0}$ explore suffisamment d'états et converge vers un minimum global, autrement dit

— $(X_n)_{n\geqslant 0}$ est une chaîne de Markov (inhomogène) récurrente ;

— $(X_n)_{n\geqslant 0}$ converge en loi vers une mesure portée par l'ensemble des minima globaux de U.

Cas de la tempéture constante. — On suppose la suite $(T_n)_{n\geqslant 0}$ constante, il en est donc de même pour $(\beta_n)_{n\geqslant 0}$. Ainsi, soit $\beta > 0$. On considère la chaîne de Markov homogène $(X_n)_{n\geqslant 0}$ de matrice de transition P_β définie, pour x et $y \in E$, par

$$P_\beta(x, y) = Q(x, y)\, \mathrm{e}^{-\beta \max(U(y)-U(x),0)} = \begin{cases} Q(x, y) & \text{si } U(y) \leqslant U(x), \\ Q(x, y)\, \mathrm{e}^{-\beta(U(y)-U(x))} & \text{si } U(y) > U(x). \end{cases}$$

Cette matrice correspond à la description faite dans l'heuristique pour une température constante.

Soit

$$Z_\beta = \sum_{y\in E} \lambda\{y\}\, \mathrm{e}^{-\beta U(y)}.$$

Si $Z_\beta < \infty$, on définit la *mesure de Gibbs*, ou *mesure de Gibbs–Boltzmann*, par

$$\mu_\beta\{x\} = \frac{\lambda\{x\}}{Z_\beta}\, \mathrm{e}^{-\beta U(x)}, \qquad x \in E.$$

C'est une mesure de probabilité associée à la fonction d'énergie U et à la température $T = 1/\beta$. Elle dépend évidemment de l'espace mesuré (E, λ).

---

1. En optimisation combinatoire, Théorie des graphes et Théorie de la complexité, une heuristique est un algorithme qui fournit rapidement (en temps polynomial) une solution réalisable, pas nécessairement optimale, pour un problème d'optimisation NP-difficile. Une heuristique, où méthode approximative, est donc le contraire d'un algorithme exact qui trouve une solution optimale pour un problème donné. L'intérêt de l'heuristique étant que pour les problèmes NP-difficiles, les algorithmes exacts connus sont tous de complexité exponentielle et donc sans aucun intérêt en pratique (ni en théorie d'ailleurs). On utilise presque systématiquement une heuristique pour obtenir une première solution réalisable dans un processus de résolution exacte. (Wikipédia)

Remarque. — La constante Z_β est généralement difficile à calculer. C'est pourquoi les physiciens pour simuler une mesure de Gibbs se servent de la méthode de Hastings–Metropolis, méthode qui a d'ailleurs inspiré celle du recuit simulé.

Proposition 8. — *On suppose que les propriétés $(\mathcal{P})$ sont vérifiées, et que Z_β est fini. Alors la chaîne de Markov $(X_n)_{n\geqslant 0}$ de matrice de transition P_β est irréductible, apériodique, de loi de probabilité réversible — et donc invariante — la mesure de Gibbs μ_β.*

Démonstration. — Puisque Q est irréductible et apériodique, et que le facteur

$$\exp\bigl(-\beta\max(U(y)-U(x),0)\bigr)\qquad x,y\in E,$$

est toujours strictement positif, P_β est aussi irréductible et apériodique (voir le raisonnement semblable pour l'algorithme de Hastings–Metropolis). Il reste simplement à voir que μ_β est réversible pour P_β. Pour x, $y\in E$, on a

$$\begin{aligned}
Z_\beta\times\mu_\beta\{x\}\times P_\beta(x,y) &= \bigl(\lambda\{x\}\,\mathrm{e}^{-\beta U(x)}\bigr)\times\bigl(Q(x,y)\,\mathrm{e}^{-\beta\max(U(y)-U(x),0)}\bigr)\\
&= \bigl(\lambda\{x\}Q(x,y)\bigr)\times\bigl(\mathrm{e}^{-\beta(\max(U(y)-U(x),0)+U(x))}\bigr)\\
&= \bigl(\lambda\{y\}Q(y,x)\bigr)\times\bigl(\mathrm{e}^{-\beta(\max(U(y),U(x))}\bigr)\\
&= \bigl(\lambda\{y\}Q(y,x)\bigr)\times\bigl(Q(y,x)\,\mathrm{e}^{-\beta(\max(0,U(x)-U(y))+U(y))}\bigr)\\
&= \bigl(\lambda\{y\}\,\mathrm{e}^{-\beta U(y)}\bigr)\times\bigl(Q(y,x)\,\mathrm{e}^{-\beta\max(U(x)-U(y),0)}\bigr)\\
&= Z_\beta\times\mu_\beta\{y\}\times P_\beta(y,x)
\end{aligned}$$

du fait de la réversibilité de λ pour Q. □

Corollaire 2. — *En conservant les mêmes hypothèses qu'à la proposition précédente, pour toute loi initiale η et tout $\beta>0$, on a*

$$\lim_{n\to\infty}\left(\sup_{A\subset E}\bigl|\eta P_\beta^n(A)-\mu_\beta(A)\bigr|\right)=0.$$

Démonstration. — Par convergence vers l'équilibre, on sait que pour tout $x\in E$,

$$\eta P_\beta^n\{x\}-\mu_\beta\{x\}\longrightarrow 0\qquad\text{quand } n \text{ tend vers l'infini.}$$

Soient $\varepsilon>0$ et $F\subset E$ fini tel que $\mu_\beta(F)\geqslant 1-\varepsilon/4$, ou encore $\mu_\beta(F^c)\leqslant\varepsilon/4$. Soit $N\geqslant 0$ tel que pour tout $n\geqslant N$ et tout $x\in F$,

$$\bigl|\eta P_\beta^n\{x\}-\mu_\beta\{x\}\bigr|\leqslant\frac{\varepsilon}{4\,\mathrm{Card}(F)}\,.$$

On a alors, par l'inégalité triangulaire,

$$\begin{aligned}
\bigl|1-\eta P_\beta^n(F)\bigr| &\leqslant \bigl|1-\mu_\beta(F)\bigr|+\bigl|\mu_\beta(F)-\eta P_\beta^n(F)\bigr|\\
&\leqslant \bigl|1-\mu_\beta(F)\bigr|+\sum_{x\in F}\bigl|\mu_\beta\{x\}-\eta P_\beta^n\{x\}\bigr|\leqslant\frac{\varepsilon}{4}+\frac{\varepsilon}{4}=\frac{\varepsilon}{2}\,,
\end{aligned}$$

et ainsi $\eta P_\beta^n(F)\geqslant 1-\varepsilon/2$, ou encore $\eta P_\beta^n(F^c)\leqslant\varepsilon/2$. Soit maintenant $A\subset E$ une partie

de E. On a

$$\begin{aligned}\left|\eta P_\beta^n(A) - \mu_\beta(A)\right| &\leqslant \left|\eta P_\beta^n(A\cap F) - \mu_\beta(A\cap F)\right| + \left|\eta P_\beta^n(A\cap F^c) - \mu_\beta(A\cap F^c)\right| \\ &\leqslant \sum_{x\in A\cap F} \left|\eta P_\beta^n\{x\} - \mu_\beta\{x\}\right| + \eta P_\beta^n(A\cap F^c) + \mu_\beta(A\cap F^c) \\ &\leqslant \sum_{x\in A\cap F} \left|\eta P_\beta^n\{x\} - \mu_\beta\{x\}\right| + \eta P_\beta^n(F^c) + \mu_\beta(F^c) \\ &\leqslant \frac{\varepsilon}{4\,\mathrm{Card}\,F} \times \mathrm{Card}(A\cap F) + \frac{\varepsilon}{2} + \frac{\varepsilon}{4} \leqslant \frac{\varepsilon}{4} + \frac{\varepsilon}{2} + \frac{\varepsilon}{4} = \varepsilon.\end{aligned}$$

Ainsi, pour tout $\varepsilon > 0$, il existe $N \geqslant 0$ tel que pour tout $n \geqslant N$ et $A \subset E$, on a $\left|\eta P_\beta^n(A) - \mu_\beta(A)\right| \leqslant \varepsilon$, ce qui est la propriété annoncée. □

Remarque. — On munit généralement un ensemble E au plus dénombrable de sa tribu discrète ainsi que de sa topologie discrète. Dans ce contexte, on peut montrer que la convergence d'une suite de mesures de probabilité $(\mu_n)_{n\geqslant 0}$ vers une mesure de probabilité μ sur E équivaut à la convergence des suites $(\mu_n(A))_{n\geqslant 0}$ vers $\mu(A)$ pour tout $A \subset E$ (ou encore celle de $(\mu_n\{x\})_{n\geqslant 0}$ vers $\mu\{x\}$ pour tout $x \in E$). Le type de convergence apparaissant dans ce corollaire est donc la *convergence uniforme* sur l'ensemble des parties de E, ce qui est un résultat fort.

Avec l'hypothèse $(\mathcal{P})$, on sait que la mesure λ charge tous les points, c'est-à-dire $\lambda\{x\} > 0$ pour tout $x \in E$. Nous notons

$$m = \inf\{U(x) : x \in E\} \qquad \text{et rappelons que} \qquad E_U = \{x \in E : U(x) = m\}.$$

En multipliant le numérateur et le dénominateur par $\exp(\beta m)$, on peut alors réécrire la mesure de Gibbs sous la forme

$$\mu_\beta\{x\} = \frac{\lambda\{x\}\,\mathrm{e}^{-\beta(U(x)-m)}}{\lambda(E_U) + \sum_{y\notin E_U} \lambda\{y\}\,\mathrm{e}^{-\beta(U(y)-m)}}\,, \qquad x \in E.$$

Si on considère cette expression, le comportement de μ_β lorsque β tend vers l'infini est immédiat. Ceci est précisé dans les deux propositions qui suivent.

Proposition 9. — *On suppose $E_U \neq \varnothing$. Alors μ_β tend vers la mesure de probabilité déduite de la restriction de λ à E_U lorsque β tend vers l'infini :*

$$\lim_{\beta\to\infty} \mu_\beta\{x\} = \frac{1}{\lambda(E_U)} \times \mathbb{1}_{E_U}(x)\,\lambda\{x\}, \qquad \textit{pour tout } x \in \mathbb{E}.$$

Démonstration. — S'il est besoin d'une démonstration, celle-ci est laissée en exercice. □

Proposition 10. — *Soient $\varepsilon > 0$ et $E_U^\varepsilon = \{x \in E : U(x) \leqslant m + \varepsilon\}$. Alors*

$$\lim_{\beta\to\infty} \mu_\beta(E_U^\varepsilon) = 1.$$

Démonstration. — S'il est besoin d'une démonstration, celle-ci est laissée en exercice. □

Cas de la température variable. — On montre que le réglage optimal de la décroissance de la température au cours du temps est inverse logarithmique

$$\beta_n = \frac{1}{C}\ln(n + \mathrm{e}), \qquad n \geqslant 0,$$

où $\mathrm{e} = \mathrm{e}^1$ est la constante de Neper[2] dont le rôle est simplement de partir d'une température strictement positive. Si C est proche de 0, c'est-à-dire $(\beta_n)_{n\geqslant 0}$ croissant rapidement, la loi limite pourrait ne pas être portée par E_U et n'être portée que par des minima locaux au sens de la matrice de transition Q, et ainsi ignorer d'autres minima, en particulier des minima globaux. Si C est grand, la convergence peut être très lente, ce qui est pénalisant d'un point de vue pratique. On définit l'oscillation de U selon la matrice de transition Q par

$$\mathrm{osc}_Q(U) = \max\{U(y) - U(x) : x, y \in I \text{ tels que } Q(x,y) > 0\}.$$

THÉORÈME 27. — *On suppose la condition* $(\mathcal{P})$ *réalisée et* U *bornée. Soit* $(X_n)_{n\geqslant 0}$ *une chaîne de Markov inhomogène de loi initiale* η *définie par l'algorithme de recuit associé à la suite d'inverses de températures*

$$\beta_n = \frac{1}{C}\ln(n+\mathrm{e}), \quad n \geqslant 0, \qquad \textit{avec} \quad C > p \times \mathrm{osc}_Q(U),$$

où $p \geqslant 0$ *est tel que* Q^p *ait tous ses coefficients strictement positifs. Alors, on a*

$$\textit{pour tout } \varepsilon > 0, \qquad \lim_{n\to\infty} \mathbb{P}\{X_n \in E_U^\varepsilon\} = 1.$$

En particulier, si E *est fini,*

$$\lim_{n\to\infty} \mathbb{P}\{X_n \in E_U\} = 1.$$

Démonstration. — Admise. □

Remarque. — Il n'a été nulle part question dans ces énoncés de vitesse de convergence. Les questions de ce type sont pourtant fondamentales pour les applications numériques.

Exemple (le problème du voyageur de commerce). — Un voyageur de commerce doit visiter N villes en partant de l'une d'entre elle et en y revenant à la fin. Le problème est de trouver un ou des trajets minimisant la distance totale (le point de départ et d'arrivée n'a pas vraiment d'importance puisque le voyageur fait une boucle). Nous considérons donc N points dans un certain espace métrique de distance d (ce qui compte est de pouvoir mesurer la distance entre deux villes). L'espace des états E est l'ensemble des arrangements de ces N villes, il s'identifie donc à l'ensemble $\mathcal{S}_N$ des permutations de $\{1,\dots,N\}$ qui est fini de cardinal $N\,!$ ce qui peut être trop grand pour examiner chaque état. La fonction d'énergie s'écrit alors

$$U(\sigma) = \sum_{x=1}^{N-1} d(\sigma(x), \sigma(x+1)) + d(\sigma(N), \sigma(1)), \qquad \sigma \in \mathcal{S}_N,$$

le dernier terme étant la distance à parcourir pour retourner à son point de départ.

Une structure de graphe connexe simple de sommets $\mathcal{S}_N$ est obtenue en considérant les transpositions $\mathcal{T}_N$ de $\{1,\dots,N\}$ qui est un ensemble de cardinal $N(N-1)/2$. En effet, on sait que toute permutation est la composée d'au plus $N-1$ transpositions. En considérant $\mathcal{T}'_N = \mathcal{T}_N \cup \{\mathrm{Id}\}$ qui est de cardinal $N(N-1)/2+1$, on constate que toutes permutations σ et σ' peuvent être conjuguées par exactement $N-1$ éléments de $\mathcal{T}'_N$:

$$\sigma' = \tau_{N-1} \circ \cdots \circ \tau_1 \circ \sigma, \qquad \text{avec } \tau_1, \dots, \tau_{N-1} \in \mathcal{T}'_{N-1}.$$

2. John Napier, plus connu sous son nom francisé Neper, né à Édimbourg en 1550 et mort le 4 avril 1617 au chteau de Merchiston, est un théologien, physicien, astronome et mathématicien écossais. (Wikipédia)

Ceci suggère de définir la matrice de transition Q par

$$Q(\sigma,\sigma') = \begin{cases} \dfrac{1}{\operatorname{Card}\mathcal{T}'_N} = \dfrac{1}{N(N-1)/2+1} & \text{si } \sigma' = \tau\circ\sigma \text{ pour } \tau\in\mathcal{T}_N, \text{ ou } \sigma'=\sigma, \\ 0 & \text{sinon.} \end{cases}$$

La matrice Q a pour mesure réversible la mesure uniforme λ sur E et Q^{N-1} a tous ses coefficients strictement positifs. Il ne reste alors qu'à choisir convenablement la constante C, puis de lancer une chaîne de Markov $(X_n)_{n\geqslant 0}$ depuis un état quelconque, et enfin attendre en espérant observer assez rapidement des états d'énergie minimale.

Le choix de Q dans ce problème n'est peut-être pas le meilleur envisageable. D'une permutation donnée, on peut passer à $N(N-1)/2$ autres permutations ou conserver celle de départ, alors qu'il y a en tout $N!$ permutations. Ceci suggère que la transition à l'aide de Q est très locale et qu'il faut beaucoup de temps pour explorer l'ensemble des cas possibles, ou, au moins, un sous-ensemble significatif de celui-ci. De plus, le réglage de la constante C s'avère assez délicat. Elle est très fortement liée au choix de Q et peut s'avérer complexe à choisir.

4. **Théorie du potentiel**

Aspects classiques. — Considérons $\mathbb{R}^n$ muni de sa base canonique $(e_i)_{1\leqslant i\leqslant n}$ et de sa structure euclidienne canonique, D un domaine de $\mathbb{R}^n$, c'est-à-dire un ouvert connexe de $\mathbb{R}^n$, ∂D sa frontière (éventuellement vide si $D=\mathbb{R}^n$), ϱ une application sur D (une application de D dans $\mathbb{R}$, une mesure sur D, une distribution), et une application $f:\partial D\to\mathbb{R}$. Un *potentiel* $\phi:\bar{D}\to\mathbb{R}$ est une solution de l'équation de Poisson[3]

$$\begin{cases} -\Delta\phi(x) = \varrho(x) & \text{pour } x\in D, \\ \phi(x) = f(x) & \text{pour } x\in\partial D, \end{cases}$$

où

$$\Delta\phi(x) = \sum_{i=1}^{n} \frac{\partial^2\phi}{\partial x_i^2}(x), \qquad x\in D,$$

est le laplacien de l'application ϕ dans l'ouvert D de $\mathbb{R}^n$. Ce type d'équation apparaît de manière centrale en théorie classique de la gravitation, en électrostatique, en thermostatique, etc. Elle est étudiée lorsque D est une variété riemannienne, un graphe, ou encore toute autre structure où naturellement apparaît un laplacien.[4]

Discrétisation. — Soit $\varepsilon>0$ petit. On a approximativement

$$\frac{\partial\phi}{\partial x_i}(x) \approx \frac{\phi(x+\varepsilon e_i)-\phi(x)}{\varepsilon} \qquad \text{ou encore} \qquad \frac{\partial\phi}{\partial x_i}(x) \approx \frac{\phi(x)-\phi(x-\varepsilon e_i)}{\varepsilon}$$

et ainsi

$$\begin{aligned} \frac{\partial^2\phi}{\partial x_i^2}(x) &\approx \frac{\partial\phi/\partial x_i(x)-\partial\phi/\partial x_i(x-\varepsilon e_i)}{\varepsilon} \\ &\approx \frac{(\phi(x+\varepsilon e_i)-\phi(x))/\varepsilon-(\phi(x)-\phi(x-\varepsilon e_i)/\varepsilon)}{\varepsilon} = \frac{\phi(x+\varepsilon e_i)+\phi(x-\varepsilon e_i)-2\phi(x)}{\varepsilon^2}. \end{aligned}$$

3. Siméon Denis Poisson (21 juin 1781 à Pithiviers – 25 avril 1840 à Sceaux) est un mathématicien, géomètre et physicien français. (Wikipédia)

4. Pierre-Simon Laplace, né le 23 mars 1749 à Beaumont-en-Auge et mort le 5 mars 1827 à Paris, est un mathématicien, astronome et physicien français. Il est l'un des principaux scientifiques de la période napoléonienne. (Wikipédia)

Par conséquent,

$$\Delta\phi(x) \approx \sum_{i=1}^{n} \frac{\phi(x+\varepsilon e_i)+\phi(x-\varepsilon e_i)-2\phi(x)}{\varepsilon^2} = \frac{2n}{\varepsilon^2}\left(\sum_{i=1}^{n}\frac{1}{2n}\phi(x\pm\varepsilon e_i)-\phi(x)\right),$$

la dernière somme comportant $2n$ termes. Ainsi, sur $\varepsilon\mathbb{Z}^n$, on a

$$\Delta\phi(x) \approx \frac{2n}{\varepsilon^2}(P-\mathrm{Id})\phi(x)$$

où P est la matrice de transition de la marche aléatoire symétrique sur $\varepsilon\mathbb{Z}^n$. Une discrétisation du problème précédent s'écrit alors

$$\begin{cases} \phi(x) = P\phi(x) + c(x) & \text{pour } x \in D, \\ \phi(x) = f(x) & \text{pour } x \in \partial D, \end{cases}$$

où D et ∂D sont deux ensembles au plus dénombrables disjoints, $E = D \cup \partial D$, éventuellement $\partial D = \varnothing$, P est une matrice de transition sur E, c (égale à $\varepsilon^2/2n \times \varrho$) est la fonction de charge ou de coût, f la condition au bord.

Remarque. — Jusqu'à présent, nous avons fait agir des matrices de transition sur des mesures de probabilité en multipliant la matrice ligne correspondant à la mesure de probabilité avec la matrice carrée représentant la matrice de transition ; ainsi on obtient une nouvelle matrice ligne qui est interprétée comme une mesure de probabilité. Avec cette convention, il est naturel de voir les fonctions $\phi : E \to \mathbb{R}$ comme des matrices colonnes. En effet, si λ est une mesure sur E, l'intégrale d'une fonction $\phi = (\phi(x))_{x\in E}$ par rapport à λ s'écrit

$$\lambda(\phi) = \sum_{x\in E} \lambda\{x\} \times \phi(x) = (\ldots, \lambda\{x\}, \ldots) \times \begin{pmatrix} \vdots \\ \phi(x) \\ \vdots \end{pmatrix}.$$

Plus généralement, si $(X_n)_{n\geqslant 0}$ est une chaîne de Markov de matrice de transition P,

$$\mathbb{E}^x[\phi(X_1)] = \sum_{y\in E} P(x,y)\phi(y), \qquad \text{et ainsi} \qquad (\mathbb{E}^x[\phi(X_1)])_{x\in E} = P\phi,$$

ce qui définit une nouvelle fonction $P\phi : E \to \mathbb{R}$ qui est toujours vue comme un vecteur colonne indexé par E. Encore plus généralement, on a bien entendu

$$\mathbb{E}^x[\phi(X_n)] = \sum_{y\in E} P^n(x,y)\phi(y), \qquad \text{et ainsi} \qquad (\mathbb{E}^x[\phi(X_n)])_{x\in E} = P^n\phi.$$

Soit $(X_n)_{n\geqslant 0}$ une chaîne de Markov sur E de matrice de transition P, et T le temps d'arrêt

$$T = \inf\{n \geqslant 0 : X_n \in \partial D\} \in [0,\infty]$$

qui est le temps d'atteinte de la frontière, ou bord, ∂D.

Théorème 28. — *On suppose les fonctions $c : D \to \mathbb{R}_+$ et $f : \partial D \to \mathbb{R}_+$ à valeurs positives. En posant*

$$\phi(x) = \mathbb{E}^x\left[\sum_{0\leqslant n<T} c(X_n) + f(X_T)\mathbb{1}_{\{T<\infty\}}\right], \qquad x \in E,$$

alors :

(i) *le potentiel* $\phi : E \to \mathbb{R}_+$ *vérifie*

$$\begin{cases} \phi(x) = P\phi(x) + c(x) & \text{pour } x \in D, \\ \phi(x) = f(x) & \text{sur } x \in \partial D\,; \end{cases} \tag{$*$}$$

(ii) *si* $\psi : E \to \mathbb{R}$ *vérifie*

$$\begin{cases} \psi(x) \geqslant P\psi(x) + c(x) & \text{pour } x \in D, \\ \psi(x) \geqslant f(x) & \text{pour } x \in \partial D\,; \end{cases} \tag{$**$}$$

et $\psi(x) \geqslant 0$ *pour tout* $x \in E$*, alors* $\psi(x) \geqslant \phi(x)$ *pour tout* $x \in E$.

(iii) *Si* $\mathbb{P}^x\{T < \infty\} = 1$ *pour tout* $x \in E$*, alors* $(*)$ *a au plus une solution bornée.*

Démonstration. — (i) Il est clair que $\phi = f$ sur ∂D $(T = 0)$. Si $x \in D$, par la propriété de Markov appliquée à $T \geqslant 1$,

$$\mathbb{E}^x\left[\sum_{1\leqslant n<T} c(X_n) + f(X_T)\mathbb{1}_{\{T<\infty\}} \,\middle|\, X_1 = y\right] = \mathbb{E}^y\left[\sum_{0\leqslant n<T} c(X_n) + f(X_T)\mathbb{1}_{\{T<\infty\}} \,\middle|\, X_1 = y\right] = \phi(y).$$

Ainsi, en sortant de l'espérance définissant $\phi(x)$ le terme de la somme correspondant à $n = 0$, à savoir $c(x)$, puis en appliquant la formule des probabilités totales, on a

$$\phi(x) = c(x) + \sum_{y\in E} P(x,y)\,\mathbb{E}^y\left[\sum_{1\leqslant n<T} c(X_n) + f(X_T)\mathbb{1}_{\{T<\infty\}} \,\middle|\, X_1 = y\right] = c(x) + \sum_{y\in E} P(x,y)\phi(y),$$

et donc $(*)$ est vérifiée.

Remarque. — Si c est identiquement nulle sur D, les solutions de $(*)$ sont appelées *fonctions harmoniques* associées à l'opérateur P. Si (iii) s'applique, $\mathbb{P}^x\{T < \infty\} = 1$ pour tout $x \in E$, et on a

$$\phi(x) = \mathbb{E}^x[f(X_T)].$$

Lorsque le temps est $\mathbb{R}_+$, $(X_t)_{t\geqslant 0}$ un mouvement brownien à valeurs dans $\mathbb{R}^n$ (processus de Markov de générateur la moitié du laplacien de $\mathbb{R}^n$), le résultat analogue est le théorème de Kakutani.

(ii) Pour $n \geqslant 0$, on pose

$$\phi_n(x) = \mathbb{E}^x\left[\sum_{k=0}^{n} c(X_k)\mathbb{1}_{\{k<T\}} + f(X_T)\mathbb{1}_{\{T\leqslant n\}}\right].$$

Par convergence monotone, $(\phi_n(x))_{n\geqslant 0}$ tend en croissant vers $\phi(x)$ pour tout $x \in E$. De même qu'au point précédent,

$$\begin{cases} \phi_{n+1}(x) = c + P\phi_n(x) & \text{pour } x \in D, \\ \phi_{n+1}(x) = f(x) & \text{pour } x \in \partial D. \end{cases}$$

Supposons que ψ vérifie $(**)$ avec $\phi \geqslant 0 = \phi_0$. Alors $\psi \geqslant P\psi + c \geqslant P\phi_0 + c = \phi_1$ sur D et $\psi \geqslant f = \phi_1$ sur ∂D, donc $\psi \geqslant \phi_1$ sur E. En poursuivant de même, on obtient par récurrence que $\psi \geqslant \phi_n$ pour tout $n \geqslant 0$. Ainsi $\psi \geqslant \phi$ comme annoncé.

(iii) On va montrer que si ψ vérifie $(**)$, alors

$$\psi(x) \geqslant \phi_{n-1}(x) + \mathbb{E}^x[\psi(X_n)\mathbb{1}_{\{T\geqslant n\}}]$$

avec égalité lorsqu'il y a égalité dans (∗∗). Cela donnera une nouvelle preuve de (ii). Mais, de plus, dans le cas de l'égalité, si $|\psi(x)| \leqslant M$ et $\mathbb{P}^x\{T < \infty\} = 1$ pour tout $x \in E$, alors, quand n tend vers l'infini, on a

$$\left|\mathbb{E}^x[\psi(X_n)\mathbb{1}_{\{T \geqslant n\}}]\right| \leqslant M \times \mathbb{P}^x\{T \geqslant n\} \longrightarrow 0,$$

et donc $\psi(x) = \lim_{n\to\infty} \phi_n(x) = \phi(x)$, ce qui prouve (iii).

Pour $x \in E$, on a

$$\psi(x) \geqslant c(x) + \sum_{y\in\partial D} P(x,y)f(y) + \sum_{y\in D} P(x,y)\psi(y),$$

et ainsi, par des substitutions successives à droite de cette inégalité sur ψ,

$$\begin{aligned}
\psi(x) &\geqslant c(x) + \sum_{y\in\partial D} P(x,y)f(y) + \sum_{y\in D} P(x,y)c(y) + \cdots \\
&\quad \cdots + \sum_{y_1\in D,\dots,y_{n-1}\in D} P(x,y_1)\times\cdots P(y_{n-2},y_{n-1})c(y_{n-1}) \\
&\quad + \sum_{y_1\in D,\dots,y_{n-1}\in D, y_n\in\partial D} P(x,y_1)\times\cdots P(y_{n-1},y_n)f(y_n) \\
&\quad + \sum_{y_1\in D,\dots,y_{n-1}\in D, y_n\in D} P(x,y_1)\times\cdots P(y_{n-1},y_n)\psi(y_n) \\
&= \mathbb{E}^x\big[c(X_0)\mathbb{1}_{\{T>0\}} + f(X_1)\mathbb{1}_{\{T=1\}} + c(X_1)\mathbb{1}_{\{T>1\}} + \cdots \\
&\quad \cdots + c(X_{n-1})\mathbb{1}_{\{T>n-1\}} + f(X_n)\mathbb{1}_{\{T=n\}} + \psi(X_n)\mathbb{1}_{\{T>n\}}\big] \\
&= \phi_{n-1}(x) + \mathbb{E}^x\big[\psi(X_n)\mathbb{1}_{\{T>n\}}\big]
\end{aligned}$$

comme demandé, avec égalité lorsqu'il y a égalité dans (∗∗). □

Comme nous le verrons plus loin, le résultat du théorème suivant n'est pas sans lien avec le problème précédent.

Théorème 29. — *On suppose la fonction $c : E \to \mathbb{R}$ bornée. On pose, pour $\alpha \in [0,1[$,*

$$\phi(x) = \mathbb{E}^x\left[\sum_{n=0}^{\infty} \alpha^n c(X_n)\right].$$

Alors ϕ est l'unique solution bornée de

$$\phi(x) = \alpha \times P\phi(x) + c(x), \qquad \textit{pour tout } x \in E.$$

Démonstration. — Soit $C \geqslant 0$ tel que $|c(x)| \leqslant C$ pour tout $x \in E$. Alors,

$$|\phi(x)| \leqslant C \sum_{n=0}^{\infty} \alpha^n = \frac{C}{1-\alpha}\,.$$

Donc ϕ est aussi bornée. Par la propriété de Markov,

$$\mathbb{E}^x\left[\sum_{n=1}^{\infty} \alpha^{n-1} c(X_n) \;\middle|\; X_1 = y\right] = \mathbb{E}^y\left[\sum_{n=0}^{\infty} \alpha^n c(X_n)\right] = \phi(y).$$

Ainsi, d'une manière analogue à la preuve du théorème précédent,

$$\phi(x) = \mathbb{E}^x\left[\sum_{n=0}^{\infty} \alpha^n c(X_n)\right]$$

$$= c(x) + \alpha \sum_{y\in E} P(x,y)\, \mathbb{E}^x\left[\sum_{n=1}^{\infty} \alpha^{n-1} c(X_n) \;\middle|\; X_1 = y\right] = c(x) + \alpha \sum_{y\in E} P(x,y)\phi(y),$$

et donc $\phi = \alpha \times P\phi + c$.

D'un autre côté, soit $\psi : E \to \mathbb{R}$ bornée et vérifiant $\psi = \alpha \times P\phi + c$. On pose $M = \sup\{|\psi(x) - \phi(x)| : x \in E\} < \infty$. Comme $\psi - \phi = \alpha \times P(\psi - \phi)$, pour tout $x \in E$, on a

$$|\psi(x) - \phi(x)| \leqslant \alpha \sum_{y\in E} P(x,y) \times |\psi(y) - \phi(y)| \leqslant \alpha \times M.$$

Ainsi, $M \leqslant \alpha \times M$ et puisque $\alpha < 1$, on a $M = 0$ et donc $\psi = \phi$. □

Matrice de Green. — On suppose que $\partial D = \varnothing$. Le potentiel ϕ est alors défini par

$$\phi(x) = \mathbb{E}^x\left[\sum_{n=0}^{\infty} c(X_n)\right], \qquad x \in E,$$

où $c : E \to \mathbb{R}$ est par exemple à valeurs positives afin de ne pas rencontrer de problèmes de sommation et d'intégration. Par le théorème de Fubini, ou au moins formellement, on a

$$\phi(x) = \sum_{n=0}^{\infty} \mathbb{E}[c(X_n)] = \sum_{n=0}^{\infty} P^n c(x) = \left(\sum_{n=0}^{\infty} P^n\right) c(x) = Gc(x)$$

où G est la *matrice de Green* de la matrice de transition P :

$$G = \sum_{n=0}^{\infty} P^n.$$

Ainsi, si on sait calculer G, on sait calculer le potentiel ϕ associés à une charge c. Il est à noter que les colonnes de G sont elles-mêmes des potentiels puisqu'elles correspondent aux charges égales à 1 en un point et nulles ailleurs. De plus, si x, $y \in E$, on a

$$G(x,y) = \sum_{n=0}^{\infty} P^n(x,y) = \sum_{n=0}^{\infty} \mathbb{P}^x(X_n = y) = \sum_{n=0}^{\infty} \mathbb{E}^x[\mathbb{1}_{\{X_n=y\}}] = \mathbb{E}^x\left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X_n=y\}}\right].$$

Alors $G(x,y) = \infty$ équivaut au fait que x mène à y et y est récurrent. En particulier les coefficients diagonaux de G qui sont infinis correspondent aux états récurrents. Plus précisément, on a

$$G(x,y) = \frac{\mathbb{P}^x\{H^y < \infty\}}{1 - \mathbb{P}^y\{T^y < \infty\}},$$

où $H^y = \inf\{n \geqslant 0 : X_n = y\}$ est le temps d'atteinte de y, et $T^y = \inf\{n > 0 : X_n = y\}$ est le temps de retour en y.

Résolvantes. — Pour l'équation $\phi = \alpha \times P\phi + c$, on a de manière similaire

$$\phi(x) = \mathbb{E}^x\left[\sum_{n=0}^{\infty} \alpha^n c(X_n)\right] = \sum_{n=0}^{\infty} \alpha^n\, \mathbb{E}^x[c(X_n)] = \sum_{n=0}^{\infty} \alpha^n P^n c(x) = \left(\sum_{n=0}^{\infty} \alpha^n P^n\right) c(x) = R_\alpha c(x)$$

où la famille de matrices $(R_\alpha)_{0\leqslant\alpha<1}$ est la *résolvante* de la matrice de transition P :

$$R_\alpha = \sum_{n=0}^{\infty} \alpha^n P^n, \qquad 0 \leqslant \alpha < 1.$$

Contrairement à la matrice de Green, tous les coefficients de R_α sont finis pour $0 \leqslant \alpha < 1$. De plus, lorsque l'espace d'états est fini, l'identité géométrique permet d'établir que

$$R_\alpha = (\mathrm{Id} - \alpha P)^{-1},$$

ce qui justifie le nom de résolvante.

Exercice 31 (cas d'une frontière non vide, $\partial D \neq \varnothing$). — Soient $(X_n)_{n\geqslant 0}$ une chaîne de Markov de matrice de transition P sur un espace d'états E muni d'une partition $\{D, \partial D\}$, $T = \inf\{n \geqslant 0 : X_n \in \partial D\}$. Soit $\phi : E \to \mathbb{R}$ le potentiel associé à une charge $c : D \to \mathbb{R}$ et une condition au bord $f : \partial D \to \mathbb{R}$. On pose

$$\widetilde{X} = \begin{cases} X_n & \text{si } n \leqslant T, \\ \delta & \text{si } n > T, \end{cases}$$

où δ est un nouvel état. Montrer que $(\widetilde{X}_n)_{n\geqslant 0}$ est une chaîne de Markov et déterminer sa matrice de transition.

On définit $\tilde{c} : E \cup \{\delta\} \to \mathbb{R}$ par

$$\tilde{c}(x) = \begin{cases} c(x) & \text{si } x \in D, \\ f(x) & \text{si } x \in \partial D, \\ 0 & \text{si } x = \delta. \end{cases}$$

Vérifier que

$$\phi(x) = \mathbb{E}^x\Big[\sum_{0\leqslant n<\widetilde{T}} \tilde{c}\big(\widetilde{X}_n\big)\Big] = \mathbb{E}^x\Big[\sum_{n\geqslant 0} \tilde{c}(\widetilde{X}_n)\Big],$$

où $\widetilde{T} = T + 1 = \inf\{n \geqslant 0 : X_n = \delta\}$. Constater qu'ainsi on s'est ramené à la situation d'une frontière vide.

Exercice 32. — Soit $c : E \to \mathbb{R}_+$ une fonction positive. On partage E en la réunion disjointe de deux sous-ensembles D et ∂D, et on suppose que les équations linéaires

$$\begin{cases} \phi(x) = P\phi(x) + c(x) & \text{pour } x \in D, \\ \phi(x) = 0 & \text{pour } x \in \partial D, \end{cases}$$

ont une unique solution bornée. Montrer que la chaîne de Markov $(X_n)_{n\geqslant 0}$ de matrice de transition P est certaine d'atteindre ∂D.

Considérons une nouvelle partition $\{\widetilde{D}, \partial\widetilde{D}\}$ une nouvelle partition de E telle que $\widetilde{D} \subset D$. Montrer que les équations linéaires

$$\begin{cases} \psi(x) = P\psi(x) + c(x) & \text{pour } x \in \widetilde{D}, \\ \psi(x) = 0 & \text{pour } x \in \partial\widetilde{D}, \end{cases}$$

ont une unique solution bornée et que $\psi(x) \leqslant \phi(x)$ pour tout $x \in E$.

APPENDICE

A. **Probabilités conditionnelles, formule de Bayes**

Ce paragraphe est constitué de rappels et permet notamment de fixer les notations. Nous introduisons la notion d'indépendance conditionnelle sous une forme assez restreinte. On comprendra comment développer les définitions qui sont données lorsqu'on connaît (ce qui doit être le cas) la notion d'indépendance.

Définition 15. — Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, A, $B \in \mathcal{A}$ deux événement (sous-ensembles mesurables de Ω) tels que $\mathbb{P}(B) > 0$. On appelle *probabilité de A sachant B* le nombre

$$\mathbb{P}(A \,|\, B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}\,.$$

Remarque. — Deux événements A et $B \in \mathcal{A}$ sont indépendants si et seulement si

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B).$$

Ainsi, lorsque $\mathbb{P}(B) > 0$, A et B sont indépendants si et seulement si $\mathbb{P}(A \,|\, B) = \mathbb{P}(A)$. Notons que lorsque $\mathbb{P}(B) = 0$, alors A et B sont indépendants quelque soit $A \in \mathcal{A}$.

Proposition 11. — *L'application*

$$\begin{aligned} \mathbb{P}(\,.\,|B) : \mathcal{A} &\longrightarrow [0,1] \\ A &\longmapsto \mathbb{P}(A \,|\, B) \end{aligned}$$

est une mesure de probabilité sur $(\Omega, \mathcal{A})$.

Démonstration. — Tout d'abord, pour tout $A \in \mathcal{A}$, on a $0 \leqslant \mathbb{P}(A \cap B) \leqslant \mathbb{P}(B)$, et ainsi $\mathbb{P}(A \,|\, B) \in [0,1]$. Vérifions que c'est une mesure de probabilité. Pour commencer, la mesure de l'ensemble total Ω est 1 :

$$\mathbb{P}(\Omega \,|\, B) = \mathbb{P}(\Omega \cap B)/\,\mathbb{P}(B) = \mathbb{P}(B)/\,\mathbb{P}(B) = 1.$$

Ensuite, l'application $\mathbb{P}(\,.\,|\,B)$ est dénombrablement additive : soit $(A_n)_n$ une famille au plus dénombrable d'éléments de $\mathcal{A}$ disjoints ($A_m \cap A_n = \varnothing$ si $m \neq n$), alors, et en notant que les $(A_n \cap B)_n$ sont eux aussi disjoints,

$$\begin{aligned} \mathbb{P}\left(\bigcup_n A_n \;\middle|\; B\right) &= \frac{\mathbb{P}\big(\big(\bigcup_n A_n\big) \cap B\big)}{\mathbb{P}(B)} = \frac{\mathbb{P}\big(\bigcup_n (A_n \cap B)\big)}{\mathbb{P}(B)} \\ &= \frac{\sum_n \mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \sum_n \frac{\mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \sum_n \mathbb{P}(A_n \,|\, B) \end{aligned}$$

ce qui finit de montrer que $\mathbb{P}(\,.\,|\,B)$ est une mesure de probabilité sur $(\Omega, \mathcal{A})$. □

Définition 16. — (i) On appelle *système complet d'événements* toute partition mesurable, finie ou dénombrable $(B_n)_n \subset \mathcal{A}$ de Ω, c'est-à-dire

$$\text{pour tout } n, \quad B_n \in \mathcal{A}, \qquad \text{si } m \neq n, \quad B_m \cap B_n = \varnothing, \qquad \text{et} \qquad \bigcup_n B_n = \Omega.$$

(ii) On appelle *système complet d'événements généralisé* toute suite finie ou dénombrable $(B_n)_n \subset \mathcal{A}$ telle que

$$\text{si } m \neq n, \qquad \mathbb{P}(B_m \cap B_n) = 0 \qquad \text{et} \qquad \mathbb{P}\left(\bigcup_n B_n\right) = 1.$$

Par la suite nous ne nous servirons que de la notion de système complet d'événements. Les énoncés ultérieurs s'appliquent aussi aux systèmes complets d'événements généralisés en adaptant, facilement, les démonstrations à ce cadre plus général. On conviendra que lorsque $\mathbb{P}(A \cap B) = 0$, le produit $\mathbb{P}(A \,|\, B) \times \mathbb{P}(B)$ a toujours un sens et est nul.

Proposition 12 (formules de Bayes). — *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.*

(i) *Si A, $B \in \mathcal{A}$ vérifient $\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$, alors*

$$\mathbb{P}(B \,|\, A) = \mathbb{P}(A \,|\, B) \times \frac{\mathbb{P}(B)}{\mathbb{P}(A)}\,.$$

(ii) *Si $(B_n)_n$ est un système complet d'événements, alors pour tout $A \in \mathcal{A}$ tel que $\mathbb{P}(A) > 0$, et tout indice N tel que $\mathbb{P}(B_N) > 0$,*

$$\mathbb{P}(B_N \,|\, A) = \mathbb{P}(A \,|\, B_N) \times \frac{\mathbb{P}(B_N)}{\sum_n \mathbb{P}(A \,|\, B_n) \times \mathbb{P}(B_n)}\,.$$

Démonstration. — (i) On a

$$\mathbb{P}(B \,|\, A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \times \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \times \frac{\mathbb{P}(B)}{\mathbb{P}(A)} = \mathbb{P}(A \,|\, B) \times \frac{\mathbb{P}(B)}{\mathbb{P}(A)}\,.$$

(ii) On a, en utilisant tout d'abord l'identité précédente puis la disjonction des $(A \cap B_n)_n$,

$$\begin{aligned}
\mathbb{P}(B_N \,|\, A) &= \mathbb{P}(A \,|\, B_N) \times \frac{\mathbb{P}(B_N)}{\mathbb{P}(A)} = \mathbb{P}(B_N \,|\, A) = \mathbb{P}(A \,|\, B_N) \times \frac{\mathbb{P}(B_N)}{\mathbb{P}(A \cap \Omega)} \\
&= \mathbb{P}(A \,|\, B_N) \times \frac{\mathbb{P}(B_N)}{\mathbb{P}(A \cap (\bigcup_n B_n))} = \mathbb{P}(A \,|\, B_N) \times \frac{\mathbb{P}(B_N)}{\mathbb{P}(\bigcup_n (A \cap B_n))} \\
&= \mathbb{P}(A \,|\, B_N) \times \frac{\mathbb{P}(B_N)}{\sum_n \mathbb{P}(A \cap B_n)} = \mathbb{P}(A \,|\, B_N) \times \frac{\mathbb{P}(B_N)}{\sum_n \mathbb{P}(A \,|\, B_n) \times \mathbb{P}(B_n)}\,,
\end{aligned}$$

d'où le résultat. □

Définition 17. — (i) Soient A, B et $C \in \mathcal{A}$ avec $\mathbb{P}(C) > 0$. Les événements A et B sont indépendants conditionnellement à C (ou sachant C) si et seulement si

$$\mathbb{P}(A \cap B \,|\, C) = \mathbb{P}(A \,|\, C) \times \mathbb{P}(B \,|\, C),$$

autrement formulé, si A et B sont indépendants sous la mesure de probabilité $\mathbb{P}(\,.\,|\,C)$.

(ii) Deux variables aléatoires $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$ et $Y : (\Omega, \mathcal{A}, \mathbb{P}) \to (F, \mathcal{F})$ sont indépendantes conditionnellement à $C \in \mathcal{A}$, $\mathbb{P}(C) > 0$, si et seulement si

$$\text{pour tous } G \in \mathcal{E},\ H \in \mathcal{F}, \qquad \mathbb{P}\{X \in G,\, Y \in H \,|\, C\} = \mathbb{P}\{X \in G \,|\, C\} \times \mathbb{P}\{Y \in H \,|\, C\},$$

autrement formulé, si les variables aléatoires X et Y sont indépendantes sous la mesure de probabilité $\mathbb{P}(\,.\,|\,C)$.

(iii) Deux variables aléatoires $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$ et $Y : (\Omega, \mathcal{A}, \mathbb{P}) \to (F, \mathcal{F})$ sont indépendantes conditionnellement une variable aléatoire *discrète* $Z : (\Omega, \mathcal{A}, \mathbb{P}) \to G$, avec G fini ou dénombrable muni de sa tribu discrète, si et seulement si pour tout $z \in G$ tel

que $\mathbb{P}\{Z = z\} > 0$, les variables aléatoires X et Y sont indépendantes conditionnellement à $\{Z = z\}$.

Remarques. — *a*) Ces trois définitions rendent compte des cas d'indépendance conditionnelle qui sera principalement rencontré dans la suite. Elles sont aussi indicatives pour définir l'indépendance d'un (événement/variable aléatoire/tribu) avec un (événement/variable aléatoire/tribu) conditionnellement à un (événement/variable aléatoire), mais pas à une tribu !

b) Nous avons limité ces définitions à l'indépendance de deux objets probabilistes conditionnellement à un troisième. On peut naturellement, en s'inspirant des formulations de l'indépendance (non conditionnelle), les généraliser à des familles quelconques de tels objets probabilistes conditionnellement à un tel objet (voire plusieurs tels objets).

Exercice 33. — (i) Soient A, B et $C \in \mathcal{A}$ tels que $\mathbb{P}(B \cap C) > 0$. Montrer que A et B sont indépendants conditionnellement à C si et seulement si

$$\mathbb{P}(A \,|\, B \cap C) = \mathbb{P}(A \,|\, C).$$

On pourra noter au passage qu'ici on a nécessairement $\mathbb{P}(C) > 0$.

(ii) Soient E, F et G des ensembles au plus dénombrables munis de leurs tribus discrètes, et $X : (\Omega, \mathcal{A}, \mathbb{P}) \to E$, $Y : (\Omega, \mathcal{A}, \mathbb{P}) \to F$ et $Z : (\Omega, \mathcal{A}, \mathbb{P}) \to G$ trois variables aléatoires (discrètes, donc).

Montrer que X et Y sont indépendantes conditionnellement à Z si et seulement si, pour tout $x \in E$, $y \in F$, $z \in G$ tel que $\mathbb{P}\{Z = z\} > 0$,

$$\mathbb{P}\{X = x, Y = y \,|\, Z = z\} = \mathbb{P}\{X = x \,|\, Z = z\} \times \mathbb{P}\{Y = y \,|\, Z = z\}.$$

TABLE DES MATIÈRES

Chapitre premier. Chaines de Markov à temps discret 1
1. Chaînes de Markov, définitions et premières propriétés 1
2. Classes d'équivalence d'une chaîne de Markov 8
3. Temps d'atteinte et probabilités d'absorption 10
4. La propriété de Markov forte 14
5. Récurrence et transience 16
6. Mesures invariantes 22
7. Convergence vers l'équilibre 31
8. Retournement du temps à une date déterministe 35
9. Un théorème ergodique 38
10. Estimation statistique des coefficients d'une matrice de transition . . 42

Chapitre II. Quelques applications 45
1. Simulations de lois de variables aléatoires 45
2. Simulation de lois, l'algorithme de Hastings–Metropolis 48
3. Le recuit simulé 49
4. Théorie du potentiel 54

Appendice 60
A. Probabilités conditionnelles, formule de Bayes 60

Table des matières 63