Chaînes de Markov

Arthur Charpentier

École Nationale de la Statistique et d'Analyse de l'Information

- notes de cours à usage exclusif des étudiants de l'ENSAI -

- ne pas diffuser, ne pas citer -

1 Quelques motivations

1.1 La sentinelle de la tour carrée

Lors d'une ronde, une méthode simple pour "*tromper l'ennemi*" est d'avancer ou reculer de manière aléatoire, en tirant à pile ou face. Comme précédement, si X_n désigne la position du garde à l'instant n, l'évolution peut se modéliser à l'aide d'une matrice de transition,

$$P = \begin{array}{c} \\ N \\ E \\ S \\ O \end{array} \begin{array}{c} \begin{array}{cccc} N & E & S & O \end{array} \\ \begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix} \end{array}$$

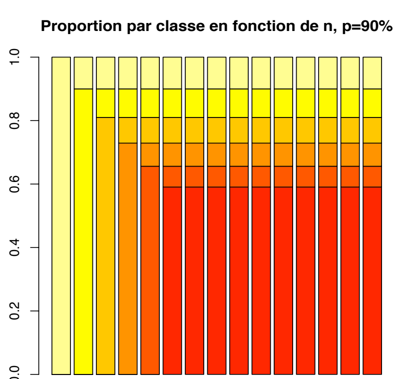
1.2 Le système bonus-malus en assurance auto

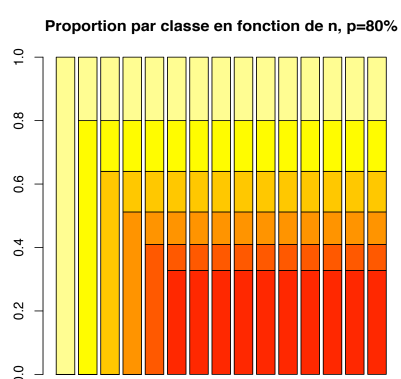
A Hong Kong, il existe 6 classes de tarification, de 1 (fort bonus) à 6 fort malus

- si un assuré n'a pas eu de sinistre, il passe de i à $\max\{1, i-1\}$,
- si l'assuré a eu au moins un sinistre, il passe de i à 6.

Si p est la probabilité de ne pas avoir de sinistre,

$$P = \begin{pmatrix} p & 0 & 0 & 0 & 0 & 1-p \\ p & 0 & 0 & 0 & 0 & 1-p \\ 0 & p & 0 & 0 & 0 & 1-p \\ 0 & 0 & p & 0 & 0 & 1-p \\ 0 & 0 & 0 & p & 0 & 1-p \\ 0 & 0 & 0 & 0 & p & 1-p \end{pmatrix}$$

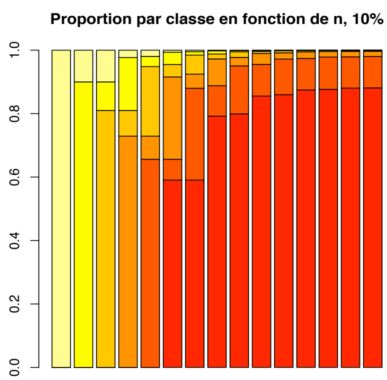


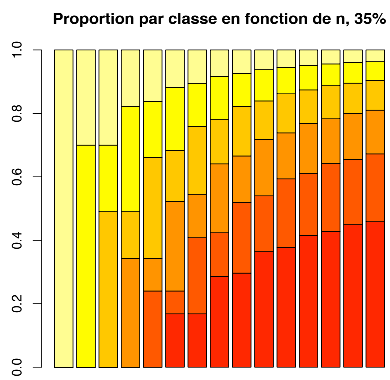


- si un assuré n'a pas eu de sinistre, il passe de i à $\max\{1, i-1\}$,
- si l'assuré a eu k sinistres, il passe de i à $\min\{6, i+k\}$.

Si le nombre de sinistres suit une loi de Poisson $p_k = \mathbb{P}(N = k) = e^{-\lambda}\lambda^k/k!$, $k \in \mathbb{N}$, et $\overline{p}_k = p_0 + ... + p_k$,

$$P = \begin{pmatrix} p_0 & p_1 & p_2 & p_3 & p_4 & 1-\overline{p}_4 \\ p_0 & 0 & p_1 & p_2 & p_3 & 1-\overline{p}_3 \\ 0 & p_0 & 0 & p_1 & p_2 & 1-\overline{p}_2 \\ 0 & 0 & p_0 & 0 & p_1 & 1-\overline{p}_1 \\ 0 & 0 & 0 & p_0 & 0 & 1-p_0 \\ 0 & 0 & 0 & 0 & p_0 & 1-p_0 \end{pmatrix}$$





1.3 La ruine d'un joueur

Deux joueurs jouent à pile ou face, chaque fois que X gagne, il touche 1 de Y, et réciproquement. Ils partent respectivement d'un capital X_0 et Y_0, et le jeu s'arrête lorsqu'un joueur n'a plus d'argent pour payer.

La fortune d'un joueur prend les valeurs $\{0, 1, 2, ..., X_0 + Y_0\}$. Si le joueur X possède une fortune $X_n = k$ à la date n, à la date $n+1$

- sa fortune devient $k-1$ avec probabilité p et$k+1$ avec probabilité $1-p$ si $0 < k < X_0 + Y_0$,
- sa fortune reste en 0 avec probabilité 1 si $k = 0$,
- sa fortune reste en $X_0 + Y_0$ avec probabilité 1 si $k = X_0 + Y_0$.

1.4 Le modèle stepping stone

On considère un tableau $n \times n$, où chaque cellule a une couleur choisie parmi k. A chaque date, on choisit une cellule au hasard, qui prend la couleur d'un de ses voisins immédiat, au hasard. Les figures suivantes montrent l'évolution de ce système pour $k = 2$ couleurs, et $n = 40$, puis $k = 8$ couleurs, et $n = 80$. On peut montrer qu'il s'agit d'une chaîne de Markov absorbante au sens où des régions absorbantes vont se former,

Ce type de modèle apparaît naturellement en génétique[1].

[1]Une animation de cet algorithme est téléchargeable sur

`http://www.crest.fr/pageperso/charpent/MC-ST-0001.avi`

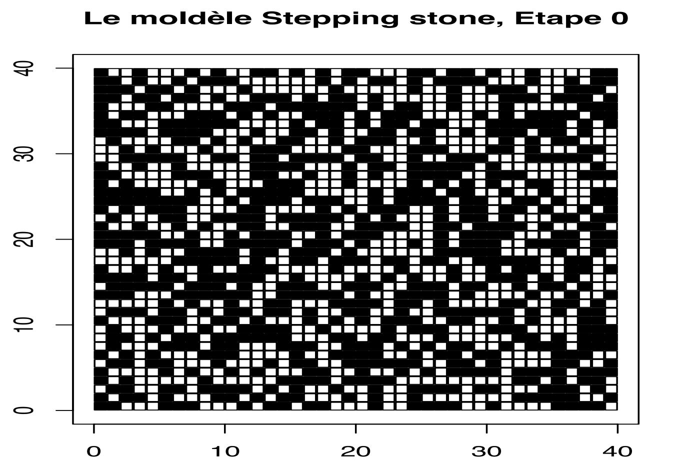


Figure 1: Le modèle *stepping stone*, $k = 2$ et $n = 40$.

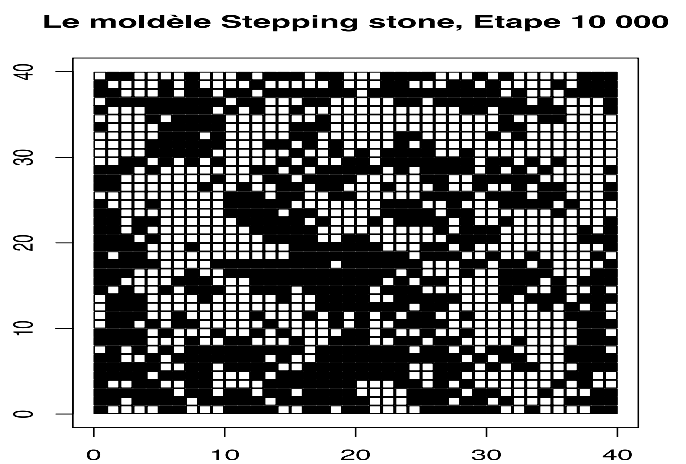


Figure 2: Le modèle *stepping stone*, $k = 2$ et $n = 40$.

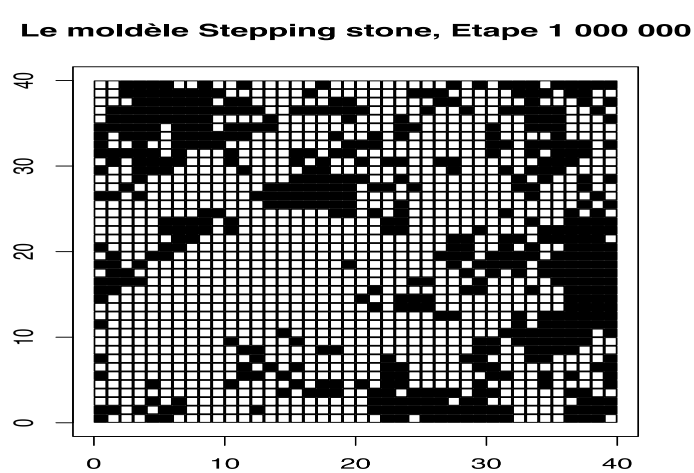


Figure 3: Le modèle *stepping stone*, $k = 2$ et $n = 40$.

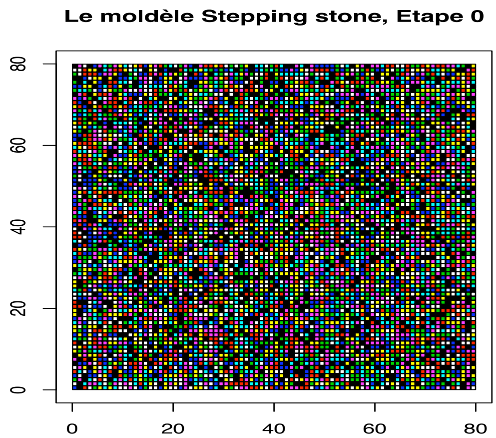


Figure 4: Le modèle *stepping stone*, $k = 8$ et $n = 80$.

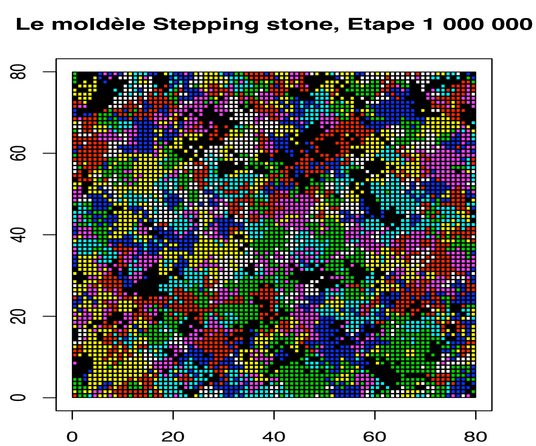


Figure 5: Le modèle *stepping stone*, $k = 8$ et $n = 80$.

1.5 Une application économique

La plupart des entreprises qui émettent des obligations sont notées par des agences de notations (Moody's ou Standard & Poors). Les notes sont de la forme AAA, AA, A, BBB, BB, B, CCC, du plus sûr au plus risqué. De matrice de transition sur un an existent:

	AAA	AA	A	BBB	BB	B	CCC	défaut
AAA	$90,81\%$	$8,33\%$	$0,68\%$	$0,06\%$	$0,12\%$	$0,00\%$	$0,00\%$	$0,00\%$
AA	$0,70\%$	$90,65\%$	$7,79\%$	$0,64\%$	$0,06\%$	$0,14\%$	$0,02\%$	$0,00\%$
A	$0,09\%$	$2,27\%$	$91,05\%$	$5,52\%$	$0,74\%$	$0,26\%$	$0,01\%$	$0,06\%$
BBB	$0,02\%$	$0,33\%$	$5,95\%$	$86,93\%$	$5,30\%$	$1,17\%$	$0,12\%$	$0,18\%$
BB	$0,02\%$	$0,14\%$	$0,67\%$	$7,73\%$	$80,53\%$	$8,84\%$	$1,00\%$	$1,06\%$
B	$0,00\%$	$0,11\%$	$0,24\%$	$0,43\%$	$6,48\%$	$83,46\%$	$4,08\%$	$5,20\%$
CCC	$0,22\%$	$0,00\%$	$0,22\%$	$1,30\%$	$2,38\%$	$5,00\%$	$64,85\%$	$19,79\%$
défaut	$0,00\%$	$0,00\%$	$0,00\%$	$0,00\%$	$0,00\%$	$0,00\%$	$0,00\%$	100%

1.6 Applications lexicographiques

Cette application a été proposée dans la modélisation initiale d'Andrei Andreevich Markov, en 1913. On considère une suite de 20 000 caractères pris dans *Eugène Onegin* d'Alexandre Pouchkine, et on distingue entre les voyelles, et les consonnes.

Heedless of the proud world's enjoyment, I prize the attention of my friends, and only wish that my employment could have been turned to worthier ends ...

En russe, il avait obtenue la matrice de transition suivante,

$$P = \begin{pmatrix} 12,8\% & 87,2\% \\ 66,3\% & 33,7\% \end{pmatrix},$$

i.e. la probabilité qu'une voyelle soit suivie d'une consonne est de $87,2\%$. La loi limite associé cette matrice est $\pi = (43,2\% \;\; 56,8\%)$, ce qui correspond aux fréquences des voyelles et des consonnes respectivement.

Notons qu'en français les fréquences sont $\pi = (45,6\% \;\; 54,4\%)$, pour l'italien $\pi = (47,4\% \;\; 52,6\%)$, et pour l'allemand $\pi = (38,5\% \;\; 61,5\%)$.

1.7 Diffusion d'un gaz et urne

Les chaînes de Markov avaient été introduites *avant* les travaux de Markov:

En 1889, Galton a introduit des chaînes de Markov pour étudier le problème de la disparition de noms de famille.

En 1907, Ehrenfest a introduit des chaînes de Markov pour étudier la diffusion d'un gaz.

1.8 Chaînes de Markov et jeux de cartes

En 1912, Poincarré a introduit des chaînes de Markov pour étudier le problème du battage de cartes. Ce problème a été étudié par la suite par Borel & Cheron (1940), Doob (1954) ou Thorpe (1972). Le nombre de méthodes pour battre les cartes est presque infini,

1 On coupe en deux, et on intervertit les deux tas (figure 1.8),

2 On prend une carte au hasard, on la met au dessus du tas, et on recommence (figure **??**).

Toutes ces méthodes conduisent plus ou moins vite (parfois jamais) à une répartition uniformément aléatoire des cartes.

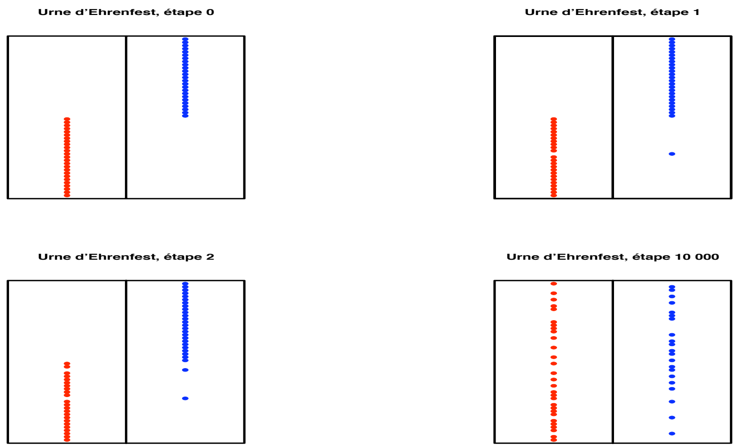


Figure 6: L'urne d'Ehrenfest.

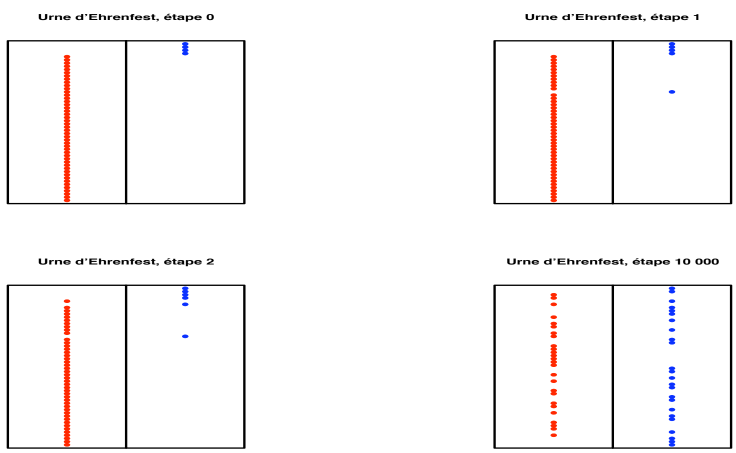


Figure 7: L'urne d'Ehrenfest.

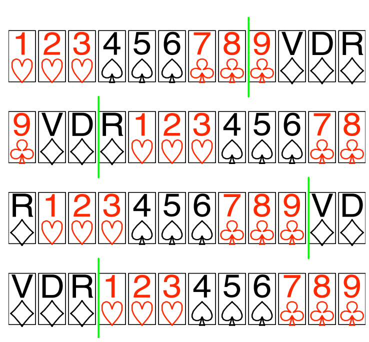


Figure 8: Le battage de cartes.

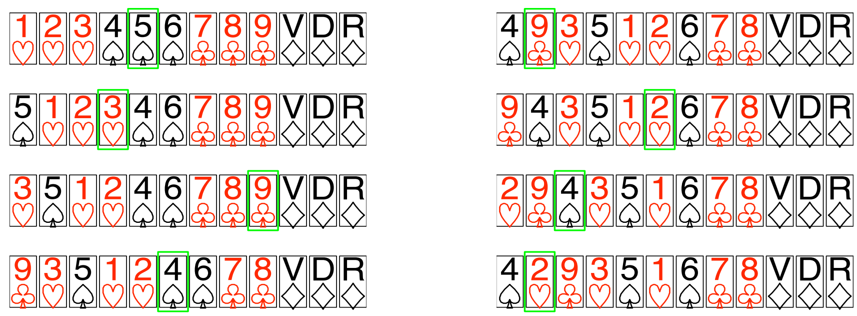


Figure 9: Le battage de cartes.

1.9 Chaînes de Markov et ADN

L'ADN est une succession de nucléotique, l'adémine, la cytosine, la guanine et la thymine. Par exemple gactgaactctgag...

Rappelons qu'il y a en réalité deux brins complémentaires, le a s'associant toujours au t, et le c avec le g. On note cg le dinucléotide c suivi de g. Ces dinucléotide ont naturellement tendance à disparaître (par *méthylation* c mute facitement en t). Les séquences cg est alors plus rare que ce que l'on pourrait espérer. Dans certaines régions toutefois, appelée *ilôts*, on trouve beaucoup de cg.

Soit X_1 le premier nucléotide d'une séquence d'ADN, de loi de probabilité $\boldsymbol{\lambda} = (\lambda_a, \lambda_c, \lambda_g, \lambda_t)$, où

$$\lambda_a = \mathbb{P}(X_1 = \text{a}), \lambda_c = \mathbb{P}(X_1 = \text{c}), \lambda_g = \mathbb{P}(X_1 = \text{g}) \text{ et } \lambda_t = \mathbb{P}(X_1 = \text{t}),$$

avec $\lambda_a + \lambda_c + \lambda_g + \lambda_t = 1$ et $\lambda_a, \lambda_c, \lambda_g, \lambda_t \geq 0$.

Il peut paraître légitime de penser que la suite des nucléotides $(X_n)_{n\in\mathbb{N}}$ n'est pas indépendante. Un modèle Markovien devrait pouvoir faire l'affaire.

Dans la région d'ilôts cg, la matrice de transition (observée empiriquement)

$$\text{pour le vecteur } X = \begin{pmatrix} a \\ c \\ g \\ t \end{pmatrix} \text{ est } P = \begin{pmatrix} 18,0\% & 27,4\% & 42,6\% & 12,0\% \\ 17,1\% & 36,8\% & 27,4\% & 18,8\% \\ 16,1\% & 33,9\% & 37,5\% & 12,5\% \\ 7,9\% & 35,5\% & 38,4\% & 18,6\% \end{pmatrix}$$

et dans la region où région l'on n'est pas en présence d'ilôts cg,

$$P = \begin{pmatrix} 30,0\% & 20,5\% & 28,5\% & 21,0\% \\ 32,2\% & 29,8\% & 7,8\% & 30,2\% \\ 24,8\% & 24,6\% & 29,8\% & 20,8\% \\ 17,7\% & 23,9\% & 29,2\% & 29,2\% \end{pmatrix}$$

1.10 Application au jeu de Monopoly

Le jeu de Monopoly peut être vu comme une marche aléatoire sur 40 cases, avec pour chaque case, des répercussions sur le déplacement (aller en prison) et/ou la fortune (payer car un adversaire a construit un hotel, ou recevoir 20 000 francs).

La position d'un joueur est une chaîne de Markov à 40 états (dont un de probabilité nulle "*allez en prison*").

Il est ainsi possible de montrer que l'avenue Henri Martin est la plus fréquentée (3,2%) alors que la rue Lecourbe et l'avenue des Champs Elysées sont les moins fréquentée (2,1%), et que la zone orange l'emporte sur la zone rouge.

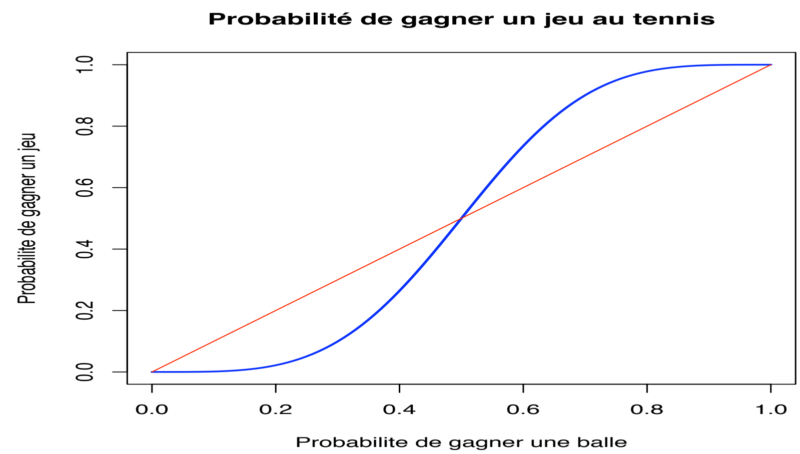
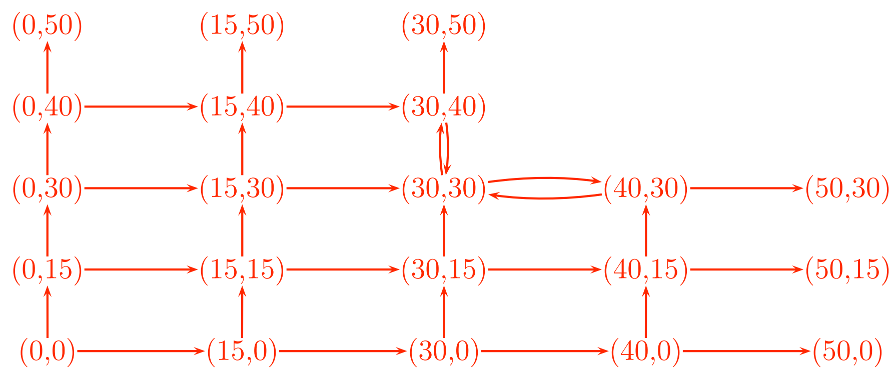


Figure 10: Probabilité de gagner un jeu au tennis.

1.11 Modélisation d'une partie de tennis

Le jeu de tennis. Considérons l'évolution du score au cours d'un jeu de tennis. Un joueur gagne le jeu s'il atteint 50.



1.12 Les processus de file d'attente

On considère une file d'attente. A chaque date n arrive un nouveau client avec probabilité p et pas de client avec probabilité $1-p$. Un client dans la file qui se fait servir quitte la file avec probabilité q, ou attend encore avec probabilité $1-q$. On note X_n le nombre de clients présents dans la file à la date n. $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov.

A l'aide de cette modélisation (et en généralisant), on peut regarder s'il est optimal - lorsqu'il y a plusieurs guichets - de créer une file unique, ou que les clients choisissent leur file en arrivant (en choisissant *a priori* la file la moins longue).

1.13 Considération plus générales, introduction aux *processus*

Un processus $(X_t)_{t\in\mathcal{T}}$ est une collection de variables aléatoires à valeurs dans $\mathcal{X}$. Pour tout $\omega\in\Omega$, la suite $(X_t(\omega))$ sera appelée une trajectoire.

Parmi les processus *usuels*, on distinguera

- le cas où $\mathcal{T}$ est discret ($\mathbb{N}$ ou $\mathbb{Z}$), et $\mathcal{X}$ est dénombrable: chaîne de Markov,
- le cas où $\mathcal{T}$ est discret ($\mathbb{N}$ ou $\mathbb{Z}$), et $\mathcal{X}=\mathbb{R}$: séries temporelles,
- le cas où $\mathcal{T}$ est continu ($\mathbb{R}$ ou $\mathbb{R}^+$), et le cas où $\mathcal{X}$ est dénombrable: processus de Poisson,

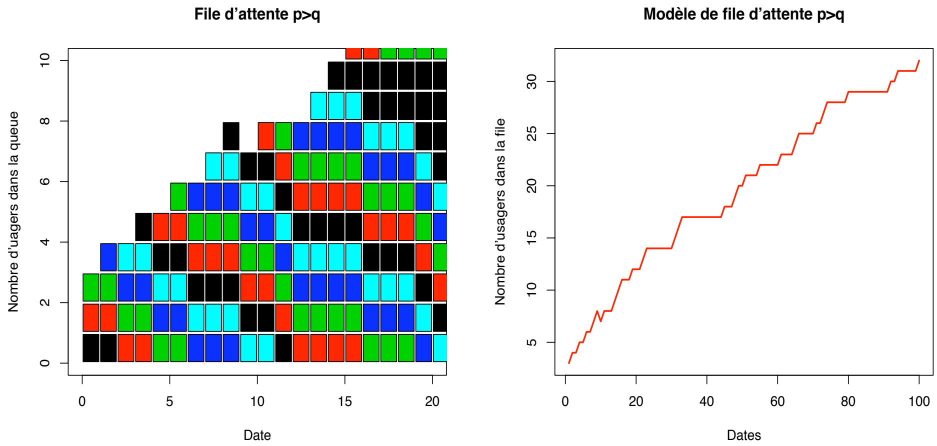


Figure 11: Evolution de la file d'attente, $p > q$.

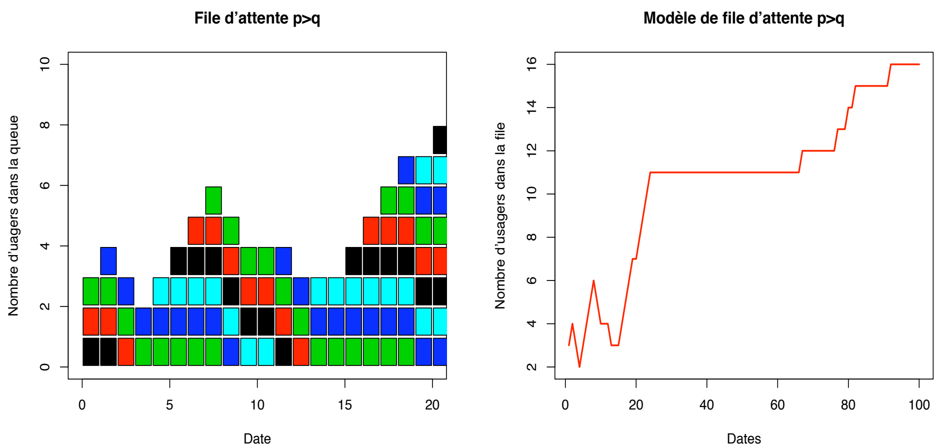


Figure 12: Evolution de la file d'attente, $p > q$.

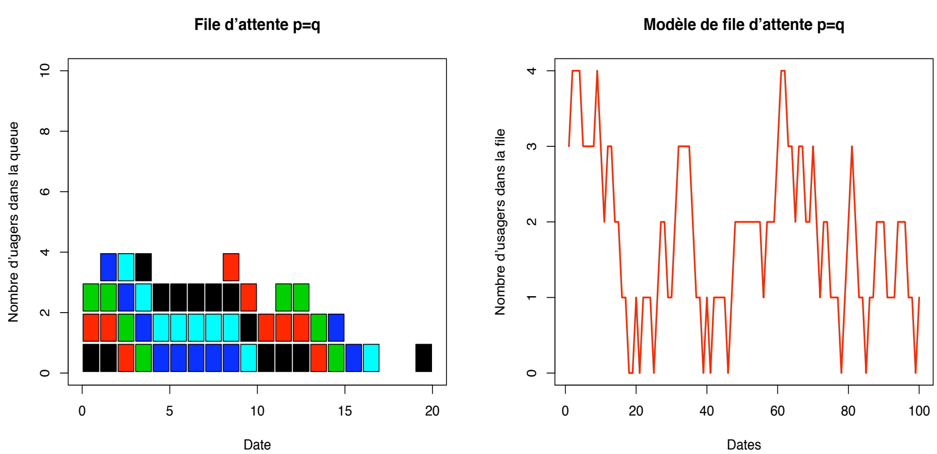


Figure 13: Evolution de la file d'attente, $p = q$.

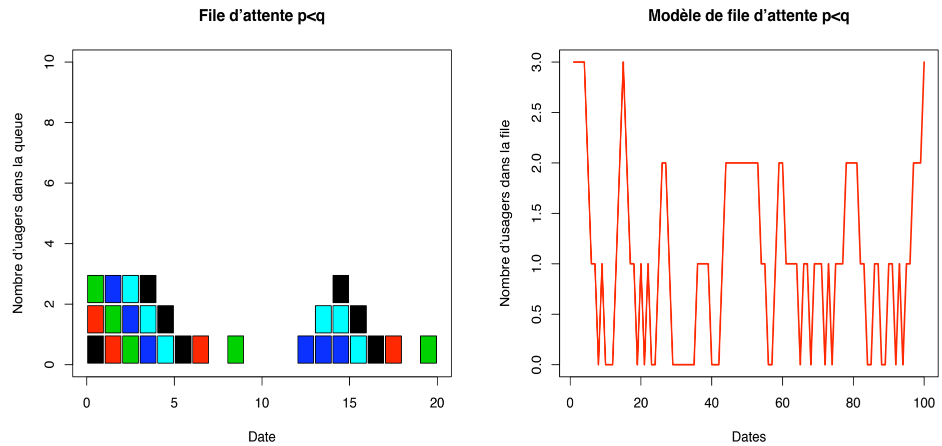
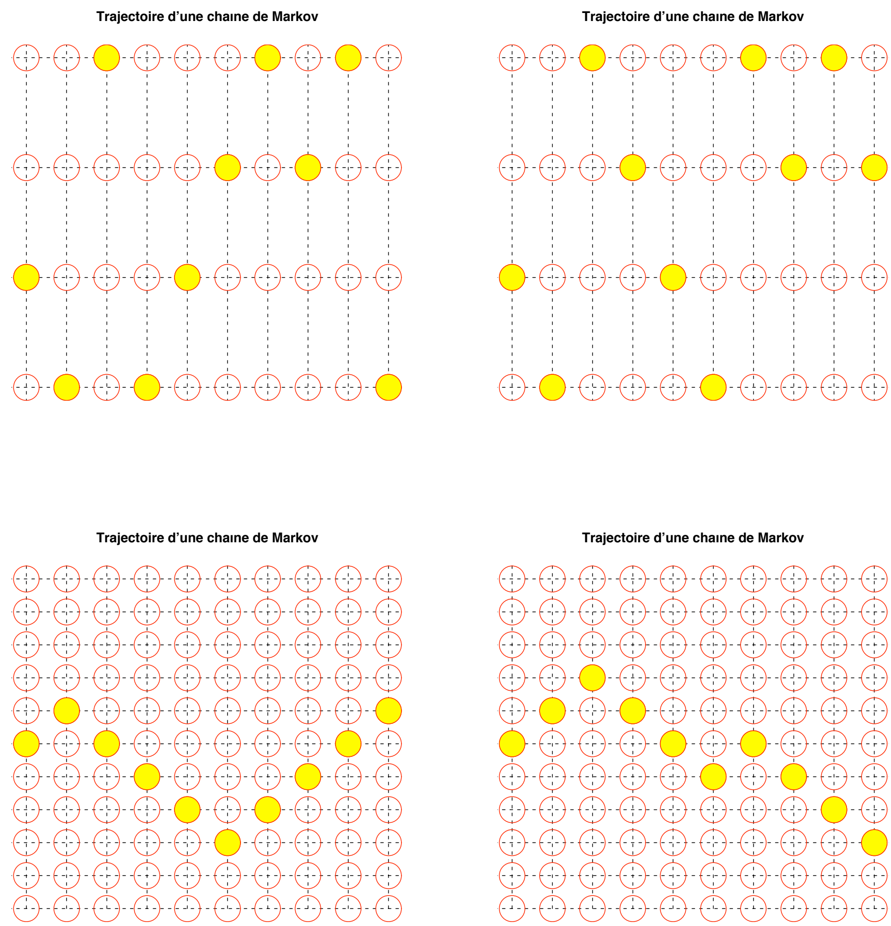
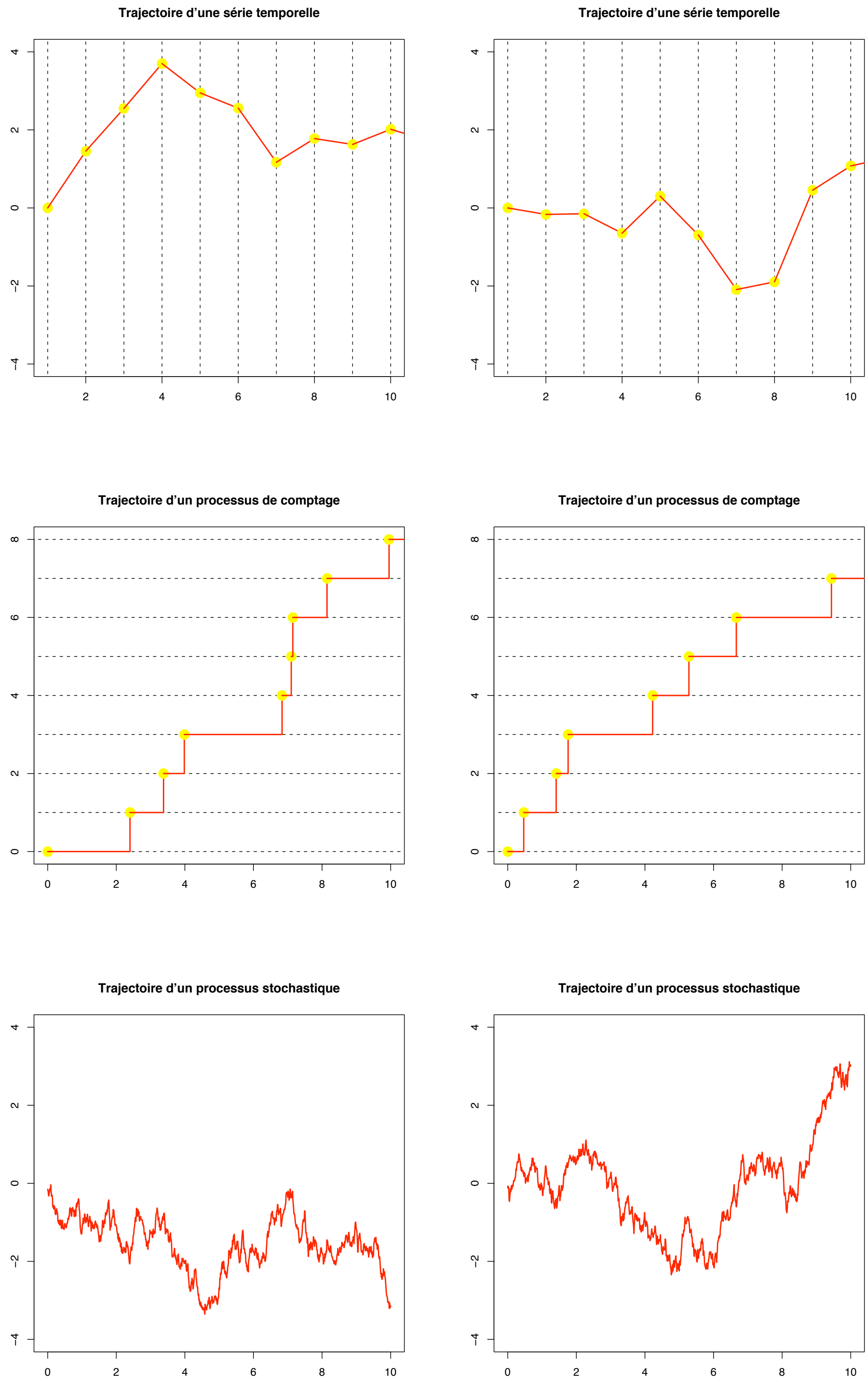


Figure 14: Evolution de la file d'attente, $p < q$.

- le cas où $\mathcal{T}$ est continu et $\mathcal{X} = \mathbb{R}$: processus stochastiques.





2 Quelques petits rappels de probabilité

L'étude des chaînes de Markov va reposer sur l'étude de la loi de X_{n+1} sachant $X_n = x_n$.

2.1 La loi conditionnelle

Rappelons la formule de Bayes : $\mathbb{P}(A|B) = \dfrac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$.

Aussi, $\mathbb{P}(X_{n+1} = x_{n+1}|X_n = x_n) = \dfrac{\mathbb{P}(X_n = x_n, X_{n+1} = x_{n+1})}{\mathbb{P}(X_n = x_n)}$.

La formule des probabilités totales : si $(B_k)_{k\in\mathcal{K}}$ forme une partition de $\mathcal{E}$, alors

$$\mathbb{P}(A) = \sum_{k\in\mathcal{K}} \mathbb{P}(A\cap B_k) = \sum_{k\in\mathcal{K}} \mathbb{P}(A|B_k)\times\mathbb{P}(B_k).$$

Aussi, $\mathbb{P}(X_{n+1}=x_{n+1}) == \sum_{x_n\in\mathcal{E}} \mathbb{P}(X_{n+1}=x_{n+1}|X_n=x_n)\times\mathbb{P}(X_n=x_n)$.

2.2 Un tout petit mot sur l'indépendance

X_n et X_{n+1} sont indépendantes si et seulement si $\mathbb{P}(X_{n+1}=x_{n+1}|X_n=x_n)$ ne dépend pas de x_n. Dans ce cas $\{X_0,X_1,...,X_n,X_{n+1}\}$ est un échantillon indépendant.

Si les X_i sont de même loi, et si les variables sont de variance finie, rappelons que l'on a une loi (faible) des grands nombres (théorème de Khintchine), qui garantie que

$$\lim_{n\to\infty}\mathbb{P}\left(\left|\frac{X_1+...+X_n}{n}-\mathbb{E}(X)\right|>\varepsilon\right)=0,$$

qui se montre à l'aide de l'inégalité de Tchebychev, et donc $\dfrac{X_1+...+X_n}{n}\xrightarrow{\mathbb{P}}\mathbb{E}(X)$.

La loi (forte) des grands nombres garantie une convergence presque sûr dès lors que les variables sont d'espérance finie,

$$\mathbb{P}\left(\lim_{n\to\infty}\frac{X_1+...+X_n}{n}(\omega)=\mathbb{E}(X)\right)=1.$$

3 Quelques définitions

Soit $\mathcal{E}$ un espace dénombrable, appelé espace d'états. Les $i\in\mathcal{E}$ sont appelés états. $\mathcal{E}$ sera isomorphe à $\{1,\ldots,k\}$, ou à $\mathbb{N}$.

Exemple 1. *Dans le cas de la sentinelle sur la tour carrée, $\{N,S,E,O\}$ sera isomorphe à $\{1,2,3,4\}$.*

Une mesure sur $\mathcal{E}$ est un vecteur $\boldsymbol{\lambda}=(\lambda_i,i\in\mathcal{E})$ tel que $0\leq\lambda_i<\infty$. On parlera de distribution de probabilité si $\boldsymbol{\lambda}'\mathbf{1}=1$.

On se donne un espace de probabilité $(\Omega,\mathcal{F},\mathbb{P})$.

Une variable aléatoire X à valeurs dans $\mathcal{E}$ est une fonction $X:\Omega\to\mathcal{E}$. On pose alors

$$\lambda_i=\mathbb{P}(X=i)=\mathbb{P}(\{\omega|X(\omega)=i\}),\ \text{et}\ \boldsymbol{\lambda}=(\lambda_i)_{i\in\mathcal{E}}.$$

Définition 2. *On considère une suite de variables aléatoires à valeurs dans $\mathcal{E}$, $(X_n)_{n\in\mathbb{N}}$. Cette suite est une chaîne de Markov si pour tout $n\geq 1$, et pour tout $i_0,i_1,\ldots,i_{n-1},i_n,i_{n+1}$ telle que $\mathbb{P}(X_0=i_0,\ldots,X_n=i_n)>0$, on ait*

$$\mathbb{P}(X_{n+1}=i_{n+1}|X_n=i_n,X_{n-1}=i_{n-1},\ldots,X_0=i_0)=\mathbb{P}(X_{n+1}=i_{n+1}|X_n=i_n). \tag{1}$$

On parlera d'absence de mémoire.

Remarque 3. *On parlera aussi de processus Markovien à l'ordre 1. Plus généralement, on peut définir un processus Markovien à l'ordre k si*

$$\begin{aligned}
& \mathbb{P}(X_{n+1}=i_{n+1}|X_n=i_n,X_{n-1}=i_{n-1},\ldots,X_0=i_0) \quad (2)\\
= \; & \mathbb{P}(X_{n+1}=i_{n+1}|X_n=i_n,\ldots,X_{n-k+1}=i_{n-k+1}). \quad (3)
\end{aligned}$$

En posant $\boldsymbol{X}_n=(X_n,\ldots,X_{n-k+1})$, à valeurs dans $\mathcal{E}^k$, notons que l'équation précédante s'écrit

$$\mathbb{P}(\boldsymbol{X}_{n+1}=\boldsymbol{i}_{n+1}|\boldsymbol{X}_n=\boldsymbol{i}_n,\boldsymbol{X}_{n-1}=\boldsymbol{i}_{n-1},\ldots,\boldsymbol{X}_{k-1}=\boldsymbol{i}_{k-1})=\mathbb{P}(\boldsymbol{X}_{n+1}=\boldsymbol{i}_{n+1}|\boldsymbol{X}_n=\boldsymbol{i}_n).$$

On retrouve un processus Markovien à l'ordre 1, sur $\boldsymbol{X}_n=(X_n,\ldots,X_{n-k+1})$.

Définition 4. *Une chaîne de Markov $(X_n)_{n\in\mathbb{N}}$ est dite homogène (dans le temps) si $\mathbb{P}(X_{n+1}=i_{n+1}|X_n=i_n)$ ne dépend pas de n.*

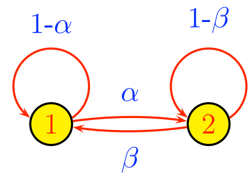
La dynamique du processus est alors entièrement caractérisée par les $p_{i,j} = \mathbb{P}(X_{n+1} = j | X_n = i)$, appelées probabilité de transition de l'état i à l'état j si la chaîne est homogène, ou plus généralement $p_{i,j}^{(n)} = \mathbb{P}(X_{n+1} = j | X_n = i)$.

Définition 5. *Une matrice $P = (p_{i,j}, i, j \in I)$ est une matrice stochastique, ou matrice de transition, si chaque ligne $\boldsymbol{p}_i = (p_{i,j}, j \in I)$ est une distribution de probabilité.*

Pour tout $i, j \in \mathcal{E}$ $p_{i,j}$ peut alors s'interpréter comme la probabilité d'aller à l'état j sachant qu'on se trouve à l'état i à l'instant d'avant.

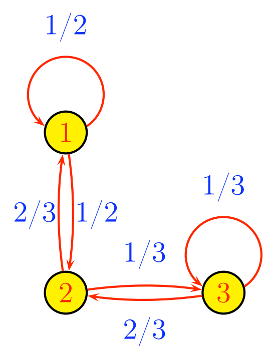
Remarque 6. *A toute matrice de transition, on peut associer un* graphe orienté*. Les sommets sont les états de la chaîne, et l'orientation est donnée par la probabilité $p_{i,j} > 0$.*

Exemple 7. *La matrice* $P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$ *est une matrice stochastique.*

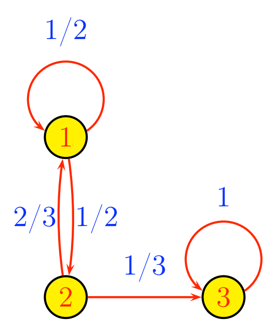


Les chaînes de Markov peuvent être interprétées comme des marches aléatoires (à probabilité non uniformes) sur un graphe.

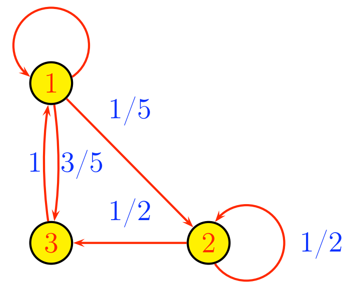
Exemple 8. *La matrice* $P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 2/3 & 0 & 1/3 \\ 0 & 2/3 & 1/3 \end{pmatrix}$ *est une matrice stochastique.*



Exemple 9. *La matrice* $P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 2/3 & 0 & 1/3 \\ 0 & 0 & 1 \end{pmatrix}$ *est une matrice stochastique.*



Exemple 10. *La matrice* $P = \begin{pmatrix} 1/5 & 1/5 & 3/5 \\ 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \end{pmatrix}$ *est une matrice stochastique.*



Définition 11. *Une suite de variable aléatoire* $(X_n)_{n\in\mathbb{N}}$ *est une chaîne de Markov de distribution initiale* $\boldsymbol{\lambda}$ *et de matrice de transition* $P = [p_{i,j}]$ *si*

- X_0 *a pour distribution initiale* $\boldsymbol{\lambda}$*, i.e.* $\mathbb{P}(X_0 = i) = \lambda_i$*,*
- *le probabilité que* $X_{n+1} = j$ *sachant que* $X_n = i$ *est* $p_{i,j}$*, i.e.* $\mathbb{P}(X_{n+1} = j | X_n = i) = p_{i,j}$*.*

Théorème 12. (Propriété de Markov) *Si* $(X_n)_{n\in\mathbb{N}}$ *est une chaîne de Markov de distribution initiale* $\boldsymbol{\lambda}$ *et de matrice de transition* P*, alors conditionnellement à* $X_{n_0} = i$*,* $(X_{n_0+n})_{n\in\mathbb{N}}$ *est une chaîne de Markov, de distribution initiale* δ_i *et de matrice de transition* P*, indépendante des variables aléatoires* $\{X_0, ..., X_{n_0}\}$*.*

Proposition 13. *si* P *est une matrice stochastique et* $\boldsymbol{\lambda}$ *une distribution de probabilité,* $(\boldsymbol{\lambda} P)_j = \sum_i \lambda_i p_{i,j}$ *est une mesure de probabilité.*

Proposition 14. *Une suite récurrente aléatoire* $X_{n+1} = f(X_n, U_{n+1})$ *avec* (U_n) *suite de v.a. i.i.d. à valeurs dans* $\mathcal{F}$ *et indépendantes de* X_0*,* $f : \mathcal{E} \times \mathcal{F} \to \mathcal{E}$ *mesurable est une chaîne de Markov homogène à valeurs dans* $\mathcal{E}$*.*

En effet

$$\mathbb{P}(X_{n+1} = x_{n+1} | X_0 = x_0, \ldots, X_n = x_n) = \mathbb{P}(f(x_n, U_1) = x_{n+1}).$$

Proposition 15. *Réciproquement, si* (X_n) *est une chaîne de Markov homogène à valeurs dans* $\mathcal{E}$*, de matrice de transition* P*, alors il existe une suite de variables aléatoires i.i.d.* (U_n) *à valeurs dans* F *et indépendantes de* X_0 *telle que* $X_{n+1} = f(X_n, U_{n+1})$ *où.*

Il suffit de prendre $f(x, y) = \inf\{u : \mathbb{P}(x,]-\infty, u]) > y\}$.

Exemple 16. $\mathcal{E} = \mathbb{N}$*,* $F = \{0, 1\}$ *et* $f(x, u) = \begin{cases} x + u & \text{si } u = 1 \\ 0 & \text{si } u = 0 \end{cases}$*.*

On notera $P^{(n)} = (p_{i,j}^{(n)})$ la matrice de probabilité, en partant de l'état i en l'instant 0, d'arriver à l'état j à l'instant n,

$$p_{i,j}^{(n)} = \mathbb{P}(X_n = j | X_0 = i), \text{ pour tout } i, j \in \mathcal{E}.$$

Proposition 17. (Propriété de Markov) *Si* $(X_n)_{n\in\mathbb{N}}$ *est une chaîne de Markov de distribution initiale* $\boldsymbol{\lambda}$ *et de matrice de transition* P*, pour tout* $n, k \in \mathbb{N}$*,*

- $\mathbb{P}(X_n = i) = (\boldsymbol{\lambda} P^{(n)})_i$
- $\mathbb{P}(X_{n+k} = j | X_n = i) = p_{i,j}^{(k)}$

Théorème 18. (Equation de Chapman-Kolmogorov (1)) *Pour tout* $n \in \mathbb{N}$*,* $P^{(n)} = P^n$*.*

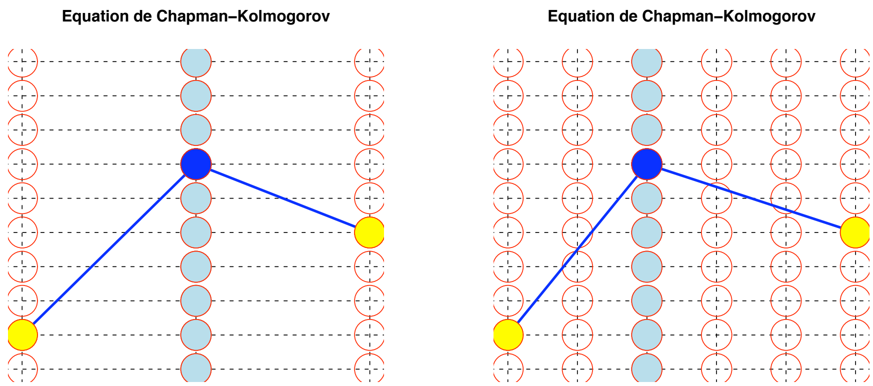
Théorème 19. (Equation de Chapman-Kolmogorov (2)) *Si* $(X_n)_{n\in\mathbb{N}}$ *est une chaîne de Markov de distribution initiale* $\boldsymbol{\lambda}$ *et de matrice de transition* P*, pour tout* $n, k \in \mathbb{N}$*, pour tout* $h = 1, ..., k-1$*,*

$$\mathbb{P}(X_{n+k} = j | X_n = i) = \sum_{u\in\mathcal{E}} \mathbb{P}(X_{n+k} = j | X_{n+h} = u) \times \mathbb{P}(X_{n+h} = u | X_n = i),$$

En effet, $P^k = P^h \times P^{k-h}$ pour tout $h = 1, ..., k-1$.

Corollaire 20. (Equation de Chapman-Kolmogorov (3)) *Si* $p_{i,j}^{(k)} = \mathbb{P}(X_{n+k} = j | X_n = i)$, *alors*

$$p_{i,j}^{(k)} = \sum_{u \in \mathcal{E}} p_{i,u}^{(h)} \times p_{u,j}^{(k-n)}.$$



Exemple 21. *Revenons à l'exemple où* $P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$. *Alors*

$$P^2 = \begin{pmatrix} (1-\alpha)^2 + \alpha\beta & 1 - [(1-\alpha)^2 + \alpha\beta] \\ 1 - [(1-\beta)^2 + \alpha\beta] & (1-\beta)^2 + \alpha\beta \end{pmatrix}$$

Plus générallement, notons que de $P^{n+1} = P^n \times P$, *on en déduit que*

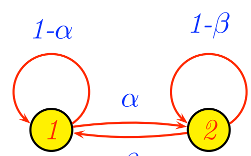
$$p_{(11)}^{n+1} = (1-\alpha)p_{11}^{(n)} + \beta p_{12}^{(n+1)} = (1-\alpha)p_{11}^{(n)} + \beta[1 - p_{11}^{(n+1)}] = (1-\alpha-\beta)p_{11}^{(n)} + \beta.$$

Cette suite définie par récurence, avec comme valeur initiale $p_{11}^{(0)} = 1$ *admet pour unique solution*

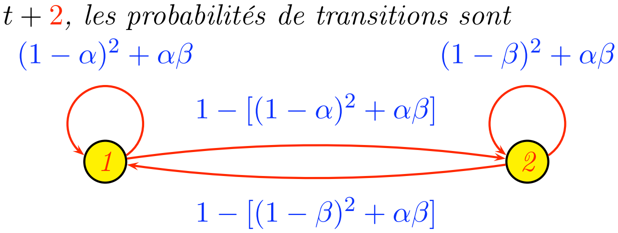
$$p_{11}^{(n)} = \frac{\beta}{\alpha+\beta} + \frac{\beta}{\alpha+\beta}[1-\alpha-\beta]^n,$$

dès lors que $\alpha + \beta > 0$.

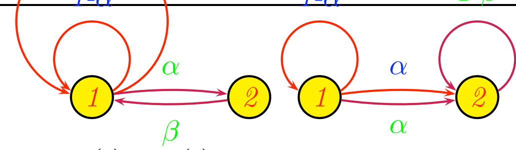
Aussi, entre les dates t *et* $t+1$, *les probabilités de transitions sont*



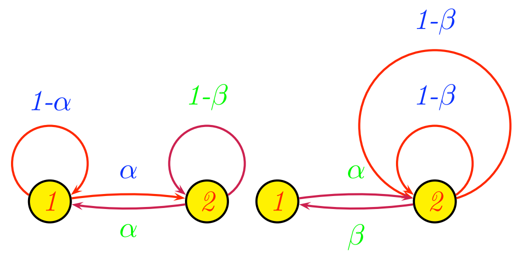
entre les dates t *et* $t+2$, *les probabilités de transitions sont*



Les probabilités de transition $p_{1,1}^{(2)}$ *et* $p_{1,2}^{(2)}$ *s'interprètent ainsi*



et les probabilités de transition $p_{2,1}^{(2)}$ *et* $p_{2,2}^{(2)}$



4 Classification des états

L'irréductibilité va nous assurer que tout point de l'espace $\mathcal{E}$ peut être atteint par la chaîne de Markov, mais aucune information n'est apportée quant au nombre de passage par cet état.

On introduira la notion de récurrence afin de formaliser cette idée.

Définition 22. *Étant donnée une chaîne de Markov* $(X_n)_{n\in\mathbb{N}}$ *de distribution initiale* $\boldsymbol{\lambda}$ *et de matrice de transition* P*, on dira que* j *est atteignable depuis* j *si* $\mathbb{P}(X_{n+k} = j | X_n = i) > 0$ *pour un* $k \in \mathbb{N}$ *On notera* $i \to j$ *. On dira que les états* i *et* j *communiquent , noté* $i \leftrightarrow j$ *si* $i \to j$ *et* $j \to i$*.*

Proposition 23. $i \to j$ *si et seulement si il existe* $n \in \mathbb{N}$ *tel que* $p_{i,j}^{(n)} > 0$*.*

Proposition 24. *Si* $i \to j$ *et* $i \to k$*, alors* $i \to k$*.*

Proposition 25. *La relation* $\leftrightarrow$ *est une classe d'équivalence.*

À partir de cette propriété on peut partitionner $\mathcal{E}$ en ensembles de valeurs qui communiquent.

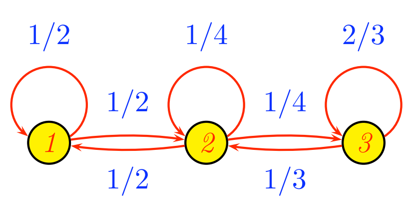
Définition 26. *Une classe* $\mathcal{C}$ *est dite fermée si* $i \in \mathcal{C}$ *et* $i \to j$ *implique* $j \in \mathcal{C}$*.*

Définition 27. *Si la chaîne de Markov ne possède qu'une unique classe, c'est à dire que tous ses éléments communiquent, la chaîne sera dite irréductible.*

On ne peut pas sortir d'une classe fermée.

Définition 28. *Un état* i *est dite absorbant si* $\{i\}$ *est une classe fermée.*

Exemple 29. *Considérons la chaîne de Markov suivante*



Cette chaîne est irréductible car tous les états communiquent. Bien que $p_{1,3} = 0$*,* $p_{1,3}^{(n)} > 0$ *pour* $n \geq 2$*.*

4.1 Pour résumer...

- Il existe trois type d'état: transients (on n'y revient pas toujours), récurrents nuls (on y revient toujours, au bout d'un temps moyen infini), ou récurents positifs (on y revient une infinité de fois, à intervalle de temps finis, en moyenne),

5 Temps d'atteinte d'une classe

Définition 30. *Étant donnée une chaîne de Markov $(X_n)_{n\in\mathbb{N}}$ de distribution initiale $\boldsymbol{\lambda}$ et de matrice de transition P, si $\mathcal{A}$ est un sous-ensemble de I, la variable aléatoire $\tau_{\mathcal{A}}$ définie par*

$$\tau_{\mathcal{A}}(\omega) = \inf\{n \geq 0, X_n(\omega) \in \mathcal{A}\}$$

est appelée temps d'atteinte de $\mathcal{A}$.

Notons que la probabilité d'atteindre un état $\mathcal{A}$ dépuis i est $p_{\mathcal{A}|i} = \mathbb{P}(\tau_{\mathcal{A}} < \infty | X_0 = i)$.

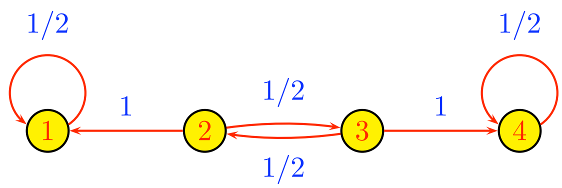
Définition 31. *Si $\mathcal{A}$ est une classe fermée de $\mathcal{E}$, $p_{\mathcal{A}|i}$ est appelée probabilité d'absorption.*

Dans ce cas, notons que le temps moyen d'atteinte de la classe $\mathcal{A}$, à partir de i est

$$e_{\mathcal{A}|i} = \mathbb{E}(\tau_{\mathcal{A}}|X_0 = i) = \sum_{n<\infty} n\mathbb{P}(\tau_{\mathcal{A}} = n) + \infty\mathbb{P}(\tau_{\mathcal{A}} = \infty).$$

On écrira ainsi $p_{\mathcal{A}|i} = \mathbb{P}(\text{atteindre } \mathcal{A}|X_0 = i)$ et $e_{\mathcal{A}|i} = \mathbb{E}(\text{temps d'atteinte de } \mathcal{A}|X_0 = i)$.

Exemple 32. *On considère la chaîne suivante*



On part de $X_0 = 2$, et on cherche à calculer la probabilité d'atteindre l'état 4 ?

On note $p_{4|i}$ la probabilité d'atteindre 4 depuis l'état i.

Notons que $p_{4|1} = 0$ et $p_{4|4} = 1$.

De plus $p_{4|2} = \frac{1}{2}p_{4|3} + \frac{1}{2}p_{4|1}$ et $p_{4|3} = \frac{1}{2}p_{4|2} + \frac{1}{2}p_{4|4}$. Aussi,

$$p_{4|2} = \frac{1}{2}p_{4|3} = \frac{1}{2}\left(\frac{1}{2}p_{4|2} + \frac{1}{2}\right).$$

Aussi à partir de $X_0 = 2$, la probabilité d'atteindre 4 est $1/3$.

Toujours à partir de $X_0 = 2$, et on cherche à calculer le temps moyen pour atteindre un état absorbant ?

On note $e_{\{1,4\}|i}$ le temps moyen d'atteindre 1 ou 4 depuis l'état i.

Notons que $e_{\{1,4\}|1} = e_{\{1,4\}|4} = 0$.

De plus $e_{\{1,4\}|2} = 1 + \frac{1}{2}e_{\{1,4\}|1} + \frac{1}{2}e_{\{1,4\}|3}$ et $e_{\{1,4\}|3} = 1 + \frac{1}{2}e_{\{1,4\}|2} + \frac{1}{2}e_{\{1,4\}|4}$. Aussi,

$$e_{\{1,4\}|2} = 1 + \frac{1}{2}e_{\{1,4\}|3} = 1 + \frac{1}{2}\left(1 + \frac{1}{2}e_{\{1,4\}|2}\right).$$

Aussi à partir de $X_0 = 2$, le temps moyen pour atteindre un état absorbant est 2.

Notons que pour le vecteur de probabilité d'atteindre $\{4\}$, on a le système suivant

$$\begin{cases} p_{4|1} = 0, \\ p_{4|2} = [p_{4|1} + p_{4|3}]/2, \\ p_{4|3} = [p_{4|2} + p_{4|4}]/2, \\ p_{4|4} = 1. \end{cases},$$

et pour le vecteur de temps moyen d'atteinte des états absorbants

$$\begin{cases} e_{\{1,4\}|1} = 0, \\ e_{\{1,4\}|2} = 1 + [e_{\{1,4\}|1} + p_{\{1,4\}|3}]/2, \\ e_{\{1,4\}|3} = 1 + [p_{\{1,4\}|2} + p_{\{1,4\}|4}]/2, \\ e_{\{1,4\}|4} = 0. \end{cases},$$

De manière plus générale, on a le résultat suivant

Proposition 33. *Le vecteur des probabilités* $\boldsymbol{p}_{\mathcal{A}} = (p_{\mathcal{A}|i})_{i\in I}$ *est la solution minimale positive du système*

$$\begin{cases} p_{\mathcal{A}|i} = 1 & si\ i \in \mathcal{A} \\ p_{\mathcal{A}|i} = \sum_{j\in I} p_{i,j} \times p_{\mathcal{A}|j} & si\ i \notin \mathcal{A} \end{cases}$$

Proposition 34. *Le vecteur des temps moyen d'atteinte de la classe* $\mathcal{A}$ $\boldsymbol{e}_{\mathcal{A}} = (e_{\mathcal{A}|i})_{i\in I}$ *est la solution minimale positive du système*

$$\begin{cases} e_{\mathcal{A}|i} = 0 & si\ i \in \mathcal{A} \\ e_{\mathcal{A}|i} = 1 + \sum_{j\notin \mathcal{A}} p_{i,j} \times e_{\mathcal{A}|j} & si\ i \notin \mathcal{A} \end{cases}$$

Définition 35. *Une variable aléatoire* T *à valeurs dans* $\{0, 1, 2, ...\} \cup \{\infty\}$ *est un temps d'arrêt pour une chaîne de Markov* $(X_n)_{n\in\mathbb{N}}$ *si l'évènement* $\{T = n\}$ *depend seulement de* $X_0, X_1, X_2, ...$

Exemple 36. *Le temps d'atteinte d'une classe* $\mathcal{A}$ *est un temps d'arrêt, avec*

$$\{T = n\} = \{X_0 \notin \mathcal{A}, X_1 \notin \mathcal{A}, X_2 \notin \mathcal{A}, ..., X_{n-1} \notin \mathcal{A}, X_n \in \mathcal{A}\}.$$

Exemple 37. *Le dernier temps de sortie d'une classe* $\mathcal{A}$, $\tau = \sup\{n \geq 0, X_n \in \mathcal{A}\}$ *n'est pas un temps d'arrêt.*

Théorème 38. (Propriété de Markov forte) *Si* $(X_n)_{n\in\mathbb{N}}$ *est une chaîne de Markov de distribution initiale* $\boldsymbol{\lambda}$ *et de matrice de transition* P*, et* T *un temps d'arrêt pour* $(X_n)_{n\in\mathbb{N}}$*, alors conditionnellement à* $\{T < \infty\}$ *et* $X_T = i$, $(X_{T+n})_{n\in\mathbb{N}}$ *est une chaîne de Markov de distribution initiale* δ_i *et de matrice de transition* P*. De plus, cette chaîne est indépendante de* $X_0, X_1, ..., X_T$.

6 Récurrence et transcience

Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov sur un ensemble dénombrable $\mathcal{E}$ de matrice de transition P. On note $(X_n^x)_{n\in\mathbb{N}}$ la chaîne partant de $X_0 = x$.

Pour $x \in E$, on introduit T_x^n la suite des instants sucessifs de retour en x définie par récurrence pour $n \geq 1$

$$T_x^1 = T_x = \inf\{k > 0 : X_k^x = x\} \quad T_x^{n+1} = \inf\{k > T_x^n : X_k^x = x\}.$$

Avec la convention $\inf\{\emptyset\} = \infty$.

Définition 39. *Soit* X_n^x *une chaîne de Markov partant de* $x \in \mathcal{E}$*. L'état* x *est dit*

1. *transient pour* P *si* $\mathbb{P}(T_x < \infty) < 1$,
2. *récurrent pour* P *si* $\mathbb{P}(T_x < \infty) = 1$.

Les états récurrents peuvent être de deux types :

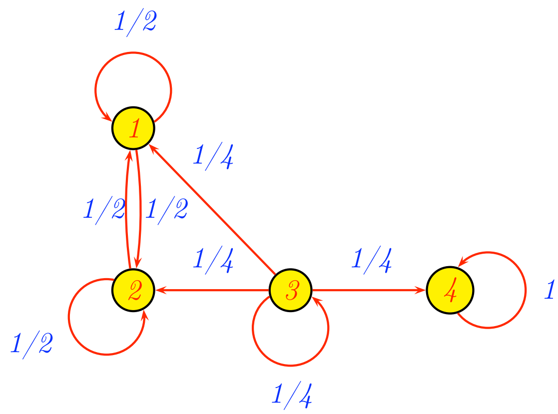
- *les états récurrents nuls si* $\mathbb{E}(T_x) = \infty$,
- *les états récurrents positifs si* $\mathbb{E}(T_x) < \infty$.

Une autre caractérisation de ces notions est la suivante,

Proposition 40. *Soit* X_n^x *une chaîne de Markov partant de* $x \in \mathcal{E}$*. L'état* x *est dit*

1. *transient pour* P *si* $\mathbb{P}(X_n = x$ *pour une infinité de valeurs de* $n | X_0 = x) = 1$,
2. *récurrent pour* P *si* $\mathbb{P}(X_n = x$ *pour une infinité de valeurs de* $n | X_0 = x) = 0$.

Exemple 41. *Considérons la chaîne de Markov suivante*



$$\text{de matrice de transition } P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Cette chaîne comporte 3 classes $\{1,2\}$, $\{3\}$ *et* $\{4\}$. *L'état* $\{4\}$ *est absorbant, les classes* $\{1,2\}$ *et* $\{4\}$ *sont récurrentes et la classe* $\{3\}$ *est transiente.*

Proposition 42. *Si $\mathcal{E}$ est un espace d'état fini, toute chaîne irréductible est récurrente.*

Les temps de passage sont reliés au nombre de visites N_x de la chaîne dans un état par la formule $\mathrm{N}_x = \sum_{k=0}^{\infty} \mathbf{1}(X_k^x = x)$ nombre de passage en x de la chaîne, $\quad N_x \geq p+1 \Leftrightarrow T_x^p < \infty$, $N_x^n = \sum_{k=0}^{n} \mathbf{1}(X_k^x = x)$ nombre de passage en x avant l'instant n, $\quad N_x^n \geq p+1 \Leftrightarrow T_x^p \leq n$.

Proposition 43. *Soit $x \in \mathcal{E}$, alors si $T_x^n < \infty$, les variables $T_x, T_x^2 - T_x^1, \ldots, T_x^{n+1} - T_x^n$ sont indépendantes et identiquement distribuées.*

Proposition 44. *Si x est récurrent, la suite (X_n^x) revient presque surement une infinité de fois à son état initial, i.e. $\mathbb{P}(N_x = \infty) = 1$.*

Proposition 45. *Si x est transient, presque surement la suite (X_n^x) visite x un nombre fini de fois. Le nombre de visite suit la loi géométrique*

$$\mathbb{P}(N_x = k) = (1-\pi)\pi^{k-1}, k \geq 1 \quad \text{avec } \pi = \mathbb{P}(T_x < \infty).$$

Si $\mathcal{C}$ est une classe de $\mathcal{E}$ dont tous les éléments communiquent, alors tous les états sont soit transients, soit récurrents.

Proposition 46. *Supposons P irréductible. Alors*

- *tous les états sont de même nature (récurrents positifs, ou récurrents nuls, ou transients).*
- *dans le cas récurrent, tous les points de $\mathcal{E}$ sont visités infiniment souvent : pour $x, y \in \mathcal{E}$*

$$\mathbb{P}(X_n^x = y \text{ pour un infinité de } n) = 1,$$

- *dans le cas transient, les sous-ensembles finis de $\mathcal{E}$ ne sont visités (qu'au plus) un nombre fini de fois : pour $A \subset \mathcal{E}$ de cardinal fini*

$$\mathbb{P}(X_n \in A \text{ pour une infinité de } n) = 0.$$

7 Premier temps d'atteinte d'un état , récurrence et transience

Définition 47. *Étant donnée une chaîne de Markov* $(X_n)_{n\in\mathbb{N}}$ *de distribution initiale* $\boldsymbol{\lambda}$ *et de matrice de transition* P*, on considère la probabilité d'atteindre un état pour la première fois un état* j*, partant de* i*, en* n *étapes, notée* $f_{i,j}^{(n)}$*, i.e.* $f_{i,j}^{(n)} = \mathbb{P}(T_{\{j\}} = n | X_0 = i)$*.*

On pose, par convention, $f_{i,j}^{(0)} = 0$.

Proposition 48. *Pour* $n \geq 0, p_{i,j}^{(n)} = \sum_{k=0}^{n} f_{i,j}^{(k)} \cdot p_{i,j}^{(n-k)}$*.*

Si le processus passe de i) j en n étapes, on s'intéresse à l'instant k où le processus à atteint j pour la première fois.

Cette dernière écriutre permet en particulier de calculer de manière récursive les $f_{i,j}^{(n)}$, en notant que $f_{i,j}^{(1)} = p_{i,j}^{(1)}$, et que

$$f_{i,j}^{(n)} = p_{i,j}^{(n)} - \sum_{k=1}^{n-1} f_{i,j}^{(k)} \cdot p_{i,j}^{(n-k)}, \text{ pour } n \geq 2.$$

On peut alors montrer le criètre de récurrence et de transience suivant

Proposition 49. *Si* $\sum_{n=0}^{\infty} p_{j,j}^{(k)} = \infty$ *l'état* j *est récurent, et si* $\sum_{n=0}^{\infty} p_{j,j}^{(k)} < \infty$ *l'état* j *est transient.*

Une démonstration peut se faire à l'aide des fonction génératrice, en posant $P_{i,j}(z) = \sum_{n\geq 0} p_{i,j}^{(n)} z^n$ et $F_{i,j}(z) = \sum_{n\geq 0} f_{i,j}^{(n)} z^n$, en notant que $P_{i,j}(z) = \delta_{i,j} + F_{i,j}(z) \cdot P_{j,j}(z)$ et du Lemme d'Abel,

Lemme 50. *Si une série de terme général* u_n *converge, et* $u = \sum_{n\geq 0} u_n$*, alors* $\lim_{z\to 1} \sum_{n\geq 0} u_n z^z = u$*. Réciproquement, si* $u_n \geq 0$*, et si* $\lim_{z\to 1} \sum_{n\geq 0} u_n z^z = u (\leq \infty)$ *alors* $u = \sum_{n\geq 0} u_n$*.*

Ce résultat permet de montrer que la récurrence est une propriété de classe,

Proposition 51. *Si* $i \leftrightarrow j$*, et si* i *est récurrent, alors* j *sera également récurrent.*

Proposition 52. *Si une chaîne de Markov a un nombre d'état fini, il existe au moins un état récurrent.*

8 Marche aléatoire et ruine du joueur

Considérons dans un premier temps la marche aléatoire[2] dans $\mathbb{Z}$.

[2] Une animation de cet algorithme est téléchargeable sur

`http://www.crest.fr/pageperso/charpent/MA-0001.avi`

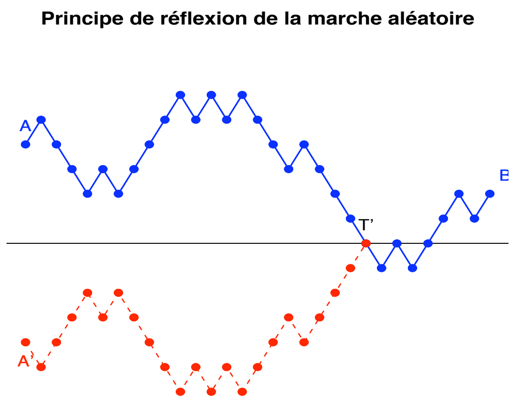


Figure 15: Principe de réflexion.

8.1 Approche fréquentiste

On considère ici la marche aléatoire, $X_n = X_{n-1} + \varepsilon_n$ pour $n \geq 1$, où ε_n est une suite i.i.d. de variables valant $+1$ avec probabilité p et -1 avec probabilité $1-p$. x_n est la richesse d'un joueur jouant à pile ou face, gagnant 1 euro dès que *pile* sort (probabilité p), et perdant 1 dès que *face* sort, au bout de n tirages.

Plaçons nous dans le cas $p = 1/2$.

Pour déterminer les probabilités de transition, on peut compter les trajectoires. Soit A l'origine ($A = (0,0)$), et un B un point atteignable en n tirages, $B = (n, x_n)$. Notons que $-n \leq x_n \leq +n$. Notons de plus que x_n est forcément de la même parité que n (si n est impair, x_n est forcément impair, et réciproquement). Le nombre de trajectoires passant de A à B est alors $N_{n,x_n} = \binom{\frac{n+x_n}{2}}{n}$. Comme le nombre total de trajectoires de longueur n est 2^n, la probabilité d'atteindre le point B est alors

$$p_{n,x_n} = \frac{1}{2^n}\binom{\frac{n+x_n}{2}}{n}.$$

Le retour à l'origine correspond à la position $x_n = 0$, qui est nécessairement possible seulement pour n pair, i.e. $n = 2k$. Aussi,

$$p_{2k,0} = \frac{1}{2^{2k}}\binom{k}{2k}.$$

On peut monter que pour n grand, à l'aide de la formule de Stirling, $p_{2k,0} \sim 1/\sqrt{\pi n}$.

Avant d'attaquer le problème du premier retour à l'origine, rappelons le principe de réflexion. Considérons désormais deux points quelconques $A = (a, x_a)$ et $B = (b, x_b)$, dont les ordonnées sont supposées strictement positives. Le nombre de trajectoires allant de $A = (a, x_a)$ à $B = (b, x_b)$ est égal au nombre de trajectoires allant de $A' = (a, -x_a)$ à $B = (b, x_b)$. C'est ce qui s'appelle principe de réflexion. Ce résultat se montre simplement par un principe de bijection, en considérant toutes les trajectoires possibles, et en introduisant le point T correspondant au premier retour en 0 (cf Figure **??**)

On peut alors utiliser ce résultat pour calculer la probabilité de ne jamais revenir en 0, entre la date 0 et la date n.

Pour cela calculons le nombre de trajectoires qui sont toujours *au dessus de la l'axe horizontal* entre $A = (1,1)$ et $B = (n, x_n)$, noté N^+_{n,x_n}. Le nombre total de trajectoire est N_{n-1,x_n-1} (compte tenu du fait que l'on part de $(1,1)$ et non pas de $(0,0)$). On peut alors écrire que $N_{n-1,x_n-1} = N^+_{n,x_n} + N^-_{n,x_n}$ où N^-_{n,x_n} est le nombre de trajectoires allant de A à B qui touche l'axe horizontal. En utilisant le principe de réflexion, notons que N^-_{n,x_n} correspond au nombre de trajectoires allant de $A' = (1,-1)$ à B, soit (par un changement d'origine), entre $(0,0)$ et

$(n-1, x_n+1)$ $N^-_{n,x_n} = N_{n-1,x_n+1}$. Aussi,

$$N^+_{n,x_n} = N_{n-1,x_n-1} - N_{n-1,x_n+1} = \binom{\frac{n+x_n}{2}-1}{n-1} - \binom{\frac{n+x_n}{2}}{n-1} = N_{n,x_n} \times \frac{x_n}{n},$$

Pour détermininer le nombre N^+_n de trajectoires qui sont toujours *au dessus de la l'axe horizontal* à partir de $A = (1,1)$, i.e. $\{X_1 > 0, ..., X_n > 0\}$, notons que, pour n pair,

$$N^+_n = 4N^+_{n-2} - \binom{\frac{n}{2}-1}{n-2} - \binom{\frac{n}{2}}{n-2}$$

(par récurence, et en comptant les chemins). Aussi, $N^+_n = \dfrac{1}{2}\dbinom{\frac{n}{2}}{n}$.

La probabilité pour qu'on n'ait aucun retour à l'origine entre 0 et la date $n = 2k$ est alors

$$\pi_{2k} = \mathbb{P}(X_0 = 0, X_1 \neq 0, X_2 \neq 0, ..., X_{2k-1} \neq 0, X_{2k} = 0).$$

Pour ce calcul, montrons quelques résultats intermédiaires. Rappelons déjà que $p_{2k,0} = \mathbb{P}(X_{2k} = 0|X_0 = 0) = p_{2k,0} = \dfrac{1}{2^{2k}}\dbinom{k}{2k}$. Montrons que $p_{2k,0} = \mathbb{P}(X_1 \neq 0, X_2 \neq 0, ..., X_{2k-1} \neq 0, X_{2k} \neq 0)$. Si la trajectoire ne coupe jamais l'axe horizontal, ce que soit on est toujours positif, soit toujours négatif, i.e.

$$\mathbb{P}(X_1 \neq 0, X_2 \neq 0, ..., X_{2k-1} \neq 0, X_{2k} \neq 0) = 2\mathbb{P}(X_1 > 0, X_2 > 0, ..., X_{2k-1} > 0, X_{2k} > 0).$$

Or d'après la question précédante,

$$\mathbb{P}(X_1 > 0, X_2 > 0, ..., X_{2k-1} > 0, X_{2k} > 0) = \frac{1}{2^{2k}} N^+_{2k} = \frac{1}{2^{2k}}\binom{k}{2k} = p_{2k,0},$$

aussi

$$\mathbb{P}(X_{2k} = 0|X_0 = 0) = p_{2k,0} = \frac{1}{2^{2k}}\binom{k}{2k}$$

$=\mathrm{P}(\mathrm{X}_1 \neq 0, X_2 \neq 0, ..., X_{2k-1} \neq 0, X_{2k} \neq 0)$.

Soit enfin τ la variable aléatoire du premier retour en 0.

$$\mathbb{P}(\tau = 2k) = \mathbb{P}(X_1 \neq 0, X_2 \neq 0, ..., X_{2k-1} \neq 0, X_{2k} = 0),$$

qui peut s'écrire

$$\mathbb{P}(\tau = 2k) = \mathbb{P}(X_1 \neq 0, X_2 \neq 0, ..., X_{2k-1} \neq 0) - \mathbb{P}(X_1 \neq 0, X_2 \neq 0, ..., X_{2k-1} \neq 0, X_{2k} \neq 0),$$

c'est à dire, en utilisant les questions précédantes,

$$\mathbb{P}(\tau = 2k) = p_{2k-2,0} - p_{2k,0} = \frac{p_{2k,0}}{2k-1} = \frac{1}{2k-1}\frac{1}{2^{2k}}\binom{k}{2k}.$$

Il est possible de montrer que cette fonction est effectivement une loi de probabilité (sur $\mathbb{N}^*$), dont la loi et la fonction de répartition sont représentées sur le graphique **??**

Parmi les autres résultats *classiques* sur la marche aléatoire, il y a la loi de l'arcsinus, liée à l'étude du dernier passage en 0 (puisque nous avons vu que la marche aléatoire avait tendance à toujours revenir en 0, on peut légitimement se poser la question).

Si $\pi_{2k,2n}$ désigne la probabilité que jusqu'à l'instant $2n$ - *inclus* - le dernier passage ait eu lieu à la date $2k$ ($\leq 2n$), alors

$$\pi_{2k,2n} = p_{2k,0} \times p_{2n-2k,0}, \text{ pour } k = 0, 1, ..., n.$$

Ce résultat s'obtient en motant que

$$\pi_{2k,2n} = \mathbb{P}(\{\cdot, X_{2k=0}\} \cap \{X_{2k+1} \neq 0, ..., X_{2n} \neq 0\}),$$

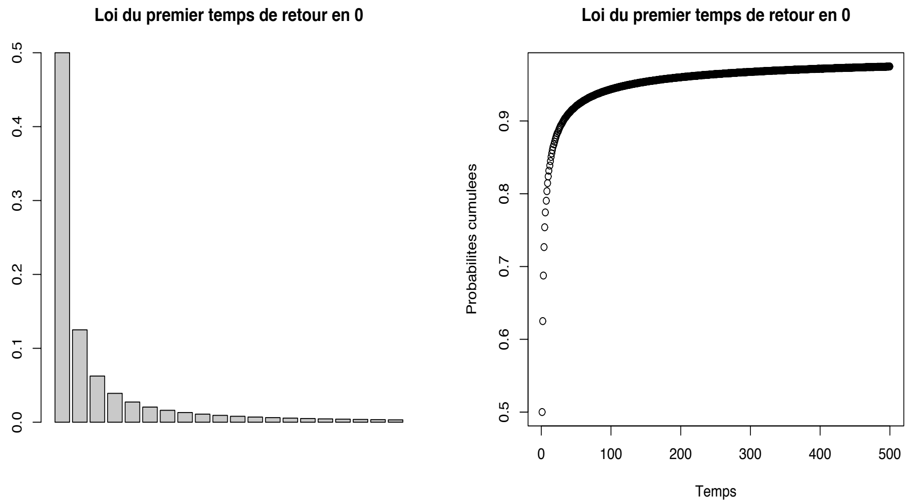


Figure 16: Loi du temps avant le premier retour en 0.

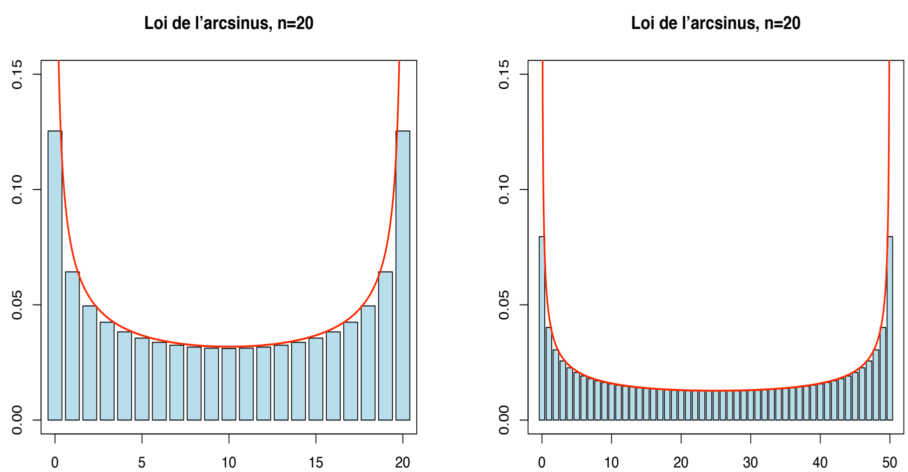


Figure 17: Loi de l'arc sinus, $n = 20$ et $n = 50$.

où ces deux évènements sont indépendants. Les probabilités associées ont été calculées auparavant. Cette loi $\pi_{2\cdot,2n}$ est appelée loi discrète de l'arcsinus, donnée par

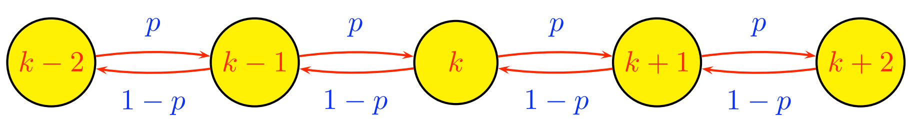
$$\pi_{2k,2n} = \frac{1}{2^{2n}}\binom{k}{2k}\binom{n-k}{2n-2k} \sim \frac{1}{\pi}\frac{1}{\sqrt{k(n-k)}} \text{ lorsque } n \to \infty.$$

Cette loi $1/\pi\sqrt{t(1-t)}$ est la densité de la loi continue de l'arc sinus.

On parle de loi de l'arcsinus car la fonction de répartition associée est de la forme $(2/\sqrt{\pi})\arcsin(\sqrt{\cdot})$. L'interprétation est que pour un jeu de pile ou face durant depuis n périodes, la probabilité pour que le dernier retour en 0 ait eu lieu dans le premier quart du jeu est $(2/\sqrt{\pi})\arcsin(\sqrt{1/4})$ soit 1/3. La probabilité pour que le dernier retour en 0 ait eu lieu dans la première moitié du jeu est $(2/\sqrt{\pi})\arcsin(\sqrt{1/2})$ soit 1/2.

8.2 Approche par les chaînes de Markov

On considère la chaîne de Markov suivante



avec comme état initial X_0. On note X_n la fortune du joueur à la date n. Les probabilités de transition sont ici

- $p_{0,0} = 1$: l'état de ruine (0) est aborbant,
- $pi, i-1 = 1-p$: on perd 1 si face sort (probabilité $1-p$),
- $p_{i,i+1} = p$: on gagne 1 si pile sort (probabilité p).

On note h_i la probabilité d'atteindre 0 sachant qu'on est à l'état i, $h_i = \mathbb{P}(\text{atteindre } 0|X_n = i)$, alors h_i est solution de

$$\begin{cases} h_0 = 1, \\ h_i = ph_{i+1} + (1-p)h_i \text{ pour } i \geq 1. \end{cases}$$

Aussi, pour $p \neq 1/2$, la forme générale de la solution est

$$h_i = A + B\left(\frac{1-p}{p}\right)^i.$$

Comme $h_i \in [0,1]$, si $p < 1-p$, alors $B = 0$ et donc $h_i = 1$ pour tout i: la ruine est certaine.

Si $p > 1-p$, alors $h_i = \left(\frac{1-p}{p}\right)^i + A\left(1 - \left(\frac{1-p}{p}\right)^i\right)$. Comme $h_i \in [0,1]$, on en déduit que $A = 0$ et donc $h_i = \left(\frac{1-p}{p}\right)^i$.

Enfin, si $p = 1/2$, $h_i = A + Bi$, et donc $h_i = 1$, comme $h_i \in [0,1]$.

Notons N_j la variable aléatoire $N_j = \inf\{n \geq 0, X_n = j\}$, correspond au premier temps d'atteinte de la fortune $j \in \mathbb{N}$.

Pour la marche aléatoire dans $\mathbb{Z}$, commençant en $X_0 = 0$.

$$p_{0,0}^{(2n)} = \mathbb{P}(X_{2n} = 0 X_0 = 0) = \binom{2n}{n}p^n(1-p)^n,$$

Rappelons que d'après la formule de Stirling, $n! \sim \sqrt{2\pi n}(n/e)^n$ lorsque $n \to \infty$. On peut alors montrer que

$$p_{0,0}^{(2n)} = \frac{2n}{!}(n!)^2[p(1-p)]^n \sim \frac{[4p(1-p)]^n}{\sqrt{2\pi n/2}} \text{ lorsque } n \to \infty.$$

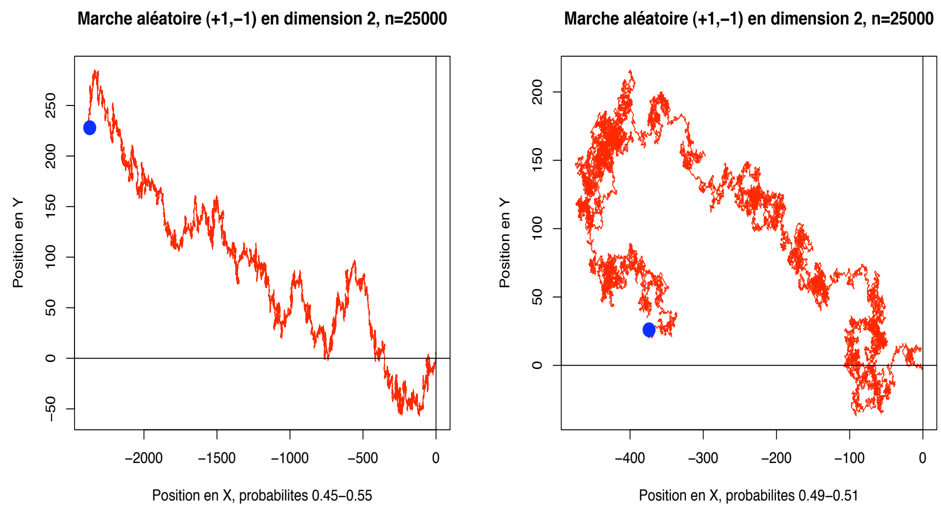


Figure 18: Marche aléatoire non uniforme dans $\mathbb{Z}^2$.

Dans le cas symmétrique où $p = 1-p = 1/2$, il existe n_0 tel que pour $n > n_0$, $p_{0,0}^{(2n)} \geq \dfrac{1}{2\sqrt{\pi n}}$. Aussi

$$\sum_{n=n_0}^{\infty} p_{0,0}^{(2n)} \geq \frac{1}{2\sqrt{\pi}} \sum_{n=n_0}^{\infty} \frac{1}{\sqrt{n}} = \infty,$$

on en déduit que la marche aléatoire est récurrente, si $p = 1/2$.

Dans le cas asymétrique où $p \neq 1-p$, $4p(1-p) < 1$, on peut montrer que

$$\sum_{n=n_0}^{\infty} p_{0,0}^{(2n)} \leq \frac{1}{2\sqrt{\pi}} \sum_{n=n_0}^{\infty} [4p(1-p)]^n < \infty,$$

on en déduit que la marche aléatoire est transient, si $p \neq 1/2$.

Pour la marche aléatoire[3] dans $\mathbb{Z}^2$, commençant en $X_0 = (0,0)$, où $p = 1/4$,

$$p_{(0,0),(0,0)}^{(2n)} = \mathbb{P}(X_{2n} = (0,0) X_0 = (0,0)) = \left(\binom{2n}{n} \left(\frac{1}{2}\right)^{2n} \right)^2 \sim \frac{2}{\pi n} \text{ lorsque } n \to \infty,$$

d'après la formule de Stirling. Aussi

$$\sum_{n=n_0}^{\infty} p_{(0,0),(0,0)}^{(2n)} \geq \frac{1}{2\pi} \sum_{n=n_0}^{\infty} \frac{1}{n} = \infty,$$

on en déduit que la marche aléatoire est récurrente, si $p = 1/4$.

La Figure **??** présente des simulations de trajectoires dans le cas où $p \neq 1/4$ (selon l'axe verticale, les probabilités sont toujours $50\% - 50\%$, mais elles passent à $49\% - 51\%$ puis $45\% - 55\%$ selon l'axe horizontal, i.e. dans ce dernier cas, les probabilités des 4 états sont $(25\%, 25\%, 22.5\%, 27.5\%)$.

Pour la marche aléatoire dans $\mathbb{Z}^3$, commençant en $X_0 = (0,0,0)$, où $p = 1/6$,

$$p_{(0,0,0),(0,0,0)}^{(2n)} = \mathbb{P}(X_{2n} = (0,0,0) X_0 = (0,0,0)) = \sum_{i,j,k \geq 0, i+j+k=n} \frac{(2n)!}{(i!j!k!)^2} \left(\frac{1}{6}\right)^{2n}$$

[3]Une animation de cet algorithme est téléchargeable sur

`http://www.crest.fr/pageperso/charpent/MA-d2-0001.avi`

En utilisant la formule de Stirling, on peut montrer que

$$p^{(2n)}_{(0,0,0),(0,0,0)} \leq \binom{2n}{n}\left(\frac{1}{2}\right)^{2n}\frac{(2n)!}{((n/3)!)^3}\left(\frac{1}{3}\right)^{n},$$

pour tout i, j, k, aussi

$$p^{(2n)}_{(0,0,0),(0,0,0)} \leq \frac{1}{2(\pi)^{3/2}}\left(\frac{6}{n}\right)^{3/2}, \text{ lorsque } n \to \infty,$$

on en déduit que la marche aléatoire est transiante, si $p = 1/6$.

9 Distributions limites ou distributions invariantes

9.1 Les distributions invariantes (ou stationnaires)

Définition 53. *Une mesure λ est dite invariante pour P si $\lambda P = \lambda$.*

On parlera aussi de loi stationnaire.

Il n'existe pas forcément de loi stationnaire pour une chaîne de Markov de matrice de transition P (par exemple la chaîne qui fait passer de i à $i+1$ avec probabilité 1). De même il peut exister une infinité de lois stationnaires. Dans le cas de la ruine du joueur, i.e. la marche aléatoire bornée inférieurement par 0, et supérieurement par s, la fortune totale des deux joueurs, $\mathcal{E} = \{0, 1, ..., s\}$, alors toute loi $\pi = (p, 0, ..., 0, 1-p)$ est une loi stationnaire, pour tout $p \in]0, 1[$.

Théorème 54. *Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov de matrice de transition P et de mesure initiale $\boldsymbol{\lambda}$ invariante pour P, alors $(X_{h+n})_{n\in\mathbb{N}}$ une chaîne de Markov de matrice de transition P et de mesure initiale $\boldsymbol{\lambda}$.*

Proof. Ceci se montre triviallement en invoquant la propriété de Markov, et en notant que, pour tout $i \in \mathcal{E}$,

$$\mathbb{P}(X_h = i) = (\boldsymbol{\lambda} P^h)_i = ([\boldsymbol{\lambda} P]P^{h-1})_i = ([\boldsymbol{\lambda}]P^{h-1})_i = ... = \boldsymbol{\lambda})_i.$$

□

9.1.1 Cas où $\mathcal{E}$ est fini

Plusieurs résultats peuvent être obtenus dans le cas fini. En effet, dans ce cas, on peut utiliser le résultat suivant

Proposition 55. *Si $\mathcal{E}$ est un espace d'état fini, alors elle admet au moins un état récurrent.*

Proof. Si $\mathcal{E} = \{1, ..., k\}$, alors $\displaystyle\sum_{i,j} p^{(n)}_{i,j} = k$ pour tout $n \in \mathbb{N}$, aussi, $\displaystyle\sum_n \left(\sum_{i,j} p^{(n)}_{i,j}\right) = \sum_{i,j}\left(\sum_n p^{(n)}_{i,j}\right) = \infty$. S'il n'existait pas d'état récurrent, tous seraient transients. Or pour un état transient j, et pour tout état i, $\displaystyle\sum_{n\geq 0} p^{(n)}_{i,j}$ est finie. Ce qui est impossible: au moins un des état est récurrent. □

Il est possible d'avoir tous les états transients dans le cas d'une chaîne sur $\mathcal{E}$ dénombrable, mais non fini.

Proposition 56. *Si $\mathcal{E}$ est un espace d'état fini, alors $\displaystyle\lim_{n\to\infty}\frac{1}{n}\sum_{h=1}^{n} P^h = \Pi$, où Π est une matrice stochastique vérifiant $\Pi P = P\Pi = \Pi$, et $\Pi^2 = \Pi$.*

Proof. On admettra qu'une telle limite existe (algèbre linéaire sur les matrices à coefficients positifs: théorie de Perron-Frobenius, exposée pages 105-109 dans Foata & Fuchs (2004)). Notons que pour tout $i \in \mathcal{E}$,

$$\sum_j \pi_{i,j} = \sum_j \lim_{n\to\infty} \frac{1}{n} \sum_{h=1}^{n} p_{i,j}^{(h)} = \lim_{n\to\infty} \sum_j \frac{p_{i,j}^{(1)} + \dots + p_{i,j}^{(n)}}{n} = \lim_{n\to\infty} 1 = 1,$$

car $\sum_j p_{i,j}^{(h)} = 1$ pour tout i. Aussi les lignes de Π sont de somme égale à 1, Π est donc une matrice stochastique. Les dernières propriétés sont obtenues en notant que $P^{n+1} = PP^n = P^nP$. □

Théorème 57. *Si l'espace d'état $\mathcal{E}$ est fini, et s'il existe i tel que $p_{i,j}^{(n)} \to \pi_j$ pour tout j, alors $\pi = (\pi_j)$ est une distribution invariante.*

Proof. Il suffit de noter que

$$\sum_j \pi_j = \sum_j \lim_{n\to\infty} p_{i,j}^n = \lim_{n\to\infty} \sum_j p_{i,j}^n = 1,$$

où l'interversion de la limite et de la somme est valide puisque $\mathcal{E}$ est supposé fini, donc π est une mesure de probabilité. De plus

$$\pi_j = \lim_{n\to\infty} p_{i,j}^n = \lim_{n\to\infty} p_{i,j}^{n+1} = \lim_{n\to\infty} \sum_k p_{i,j}^n p_{i,j} = \sum_k \pi_k p_{k,j},$$

c'est à dire que $\pi = (\pi_j)$ est une distribution invariante. □

Rappelons que si $j \in \mathcal{E}$ est un état transient, alors pour tout i, $\lim_{n\to\infty p_{i,j}^n = 0}$ (le terme général d'une série convergent tend vers 0).

Proposition 58. *Si la chaîne est **irréductible**, et possède un nombre fini d'états, alors $\pi_j = 1/\mathbb{E}(T_j|X_0 = j)$, correspondant à l'inverse du temps de retour moyen de j à j.*

Théorème 59. *Si $\mathcal{E}$ est un espace d'état fini, alors*

- *existence: il existe au moins une mesure stationnaire,*
- *unicité: la loi stationnaire est unique si et seulement si la chaîne admet une seule classe récurrente. De plus, si $\mathcal{C}$ désigne cette classe, $\pi_j > 0$ si et seulement si $j \in \mathcal{C}$, et $\pi_j = 0$ si et seulement si $j \notin \mathcal{C}$*

Proof. On admettra l'existence, qui repose sur des résultats d'algèbre linéaire sur les matrices à coefficients positifs (théorie de Perron-Frobenius, exposée pages 105-109 dans Foata & Fuchs (2004)). L'unicité s'obtient de la manière suivante. Montrons que si la chaîne admet une seule classe récurrente, alors il existe une loi stationnaire unique, vérifiant les propriétés mentionnées. On sépare alors $\mathcal{E}$ en son unique classe récurrente $\mathcal{C}$, et $\mathcal{T}$ la réunion des classes transients. Soit $(i,j) \in \mathcal{E}$.

- si $j \in \mathcal{T}$ (et i quelconque, dans $\mathcal{E}$) alors $\lim_{n\to\infty} p_{i,j}^{(n)} = 0$.
- si $j \in \mathcal{C}$ et $i \in \mathcal{C}$ alors la restriction du processus à la classe $\mathcal{C}$ est irréductible, et d'après **??**, la limite est unique (correspondant à l'inverse d'un temps de retour),
- si $j \in \mathcal{C}$ et $i \in \mathcal{T}$, c'est un peu plus compliqué. L'idée est de passer par un état intermédiaire k, qui sera soit dans $\mathcal{C}$ soit dans $\mathcal{T}$ à une date intermidiaire avant n,

$$p_{i,j}^{(m+h)} = \sum_{k\in\mathcal{C}} p_{i,k}^{(m)} p_{k,j}^{(h)} + \sum_{k\in\mathcal{T}} p_{i,k}^{(m)} p_{k,j}^{(h)}.$$

Dans un second temps, on utilise le fait que si $p^n_{i,j} \to \pi_{i,j}$, alors par le théorème de Cesarro, $\lim_{n\to} \frac{1}{n}\sum_{h=1}^{n} p_{i,j}^{m+h} = \pi_{i,j}$, et

$$\frac{1}{n}\sum_{h=1}^{n} p_{i,j}^{(m+h)} = \sum_{k\in\mathcal{C}} p_{i,k}^{(m)} \left(\frac{1}{n}\sum_{h=1}^{n} p_{k,j}^{(h)}\right) + \sum_{k\in\mathcal{T}} p_{i,k}^{(m)} \left(\frac{1}{n}\sum_{h=1}^{n} p_{k,j}^{(h)}\right).$$

En utilisant le théorème de Cesarro sur les trois sommes,

$$\pi_{i,j} = \left[\sum_{k\in\mathcal{C}} p_{i,k}^{(m)}\right] \pi_j + \left[\sum_{k\in\mathcal{T}} p_{i,k}^{(m)}\right] \pi_{k,j}.$$

Si $m \to \infty$, $p_{i,k}^{(m)} \to 0$ pour tout $k \in \mathcal{T}$, et $\sum_{k\in\mathcal{C}} p_{i,k}^{(m)} \to 1$, et donc $\pi_{i,j} = \pi_j > 0$. Pour montrer la réciproque, raisonnons par l'absurde. Supposons qu'il existe (au moins) deux classes récurrentes $\mathcal{C}_1$ et $\mathcal{C}_2$. La matrice de transition P a alors la forme suivante

$$P = \begin{pmatrix} \star & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \star & \mathbf{0} \\ \star & \star & \star \end{pmatrix}$$

et pour tout $n \in \mathbb{N}$, P^n aura la même forme. Il est alors impossible d'avoir $\lim_{n\to\infty} \pi_{i_1,j} = \lim_{n\to\infty} \pi_{i_2,j}$ pour $i_1 \in \mathcal{C}_1$ et $i_2 \in \mathcal{C}_2$

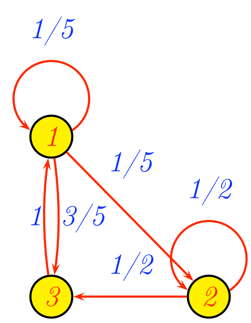
□

Exemple 60. *Considérons la chaîne de Markov de matrice de transition*

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

alors $\{2\}$ *et* $\{3,4\}$ *sont deux classes récurrentes: on n'a plus unicité de la distribution stationnaire. En effet,* $\boldsymbol{\pi} = (0, 1-2p, p, p)$ *pour tout* $p \in [0, 1/2]$.

Exemple 61. *Considérons la chaîne de l'exemple* **??**, $P = \begin{pmatrix} 1/5 & 1/5 & 3/5 \\ 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \end{pmatrix}$



Les distributions invariantes sont données par la résolution du système

$$\begin{aligned} x &= x/5 + z \\ y &= x/5 + y/2 \\ z &= 3x/5 + z/2 \end{aligned}$$

Toute solution est proportionelle au vecteur $(5, 2, 4)$. *Le vecteur de probabilité associé est*

$$\pi = \left(\frac{5}{11}, \frac{2}{11}, \frac{4}{11}\right) \sim (45.4\%, 18.2\%, 36.4\%).$$

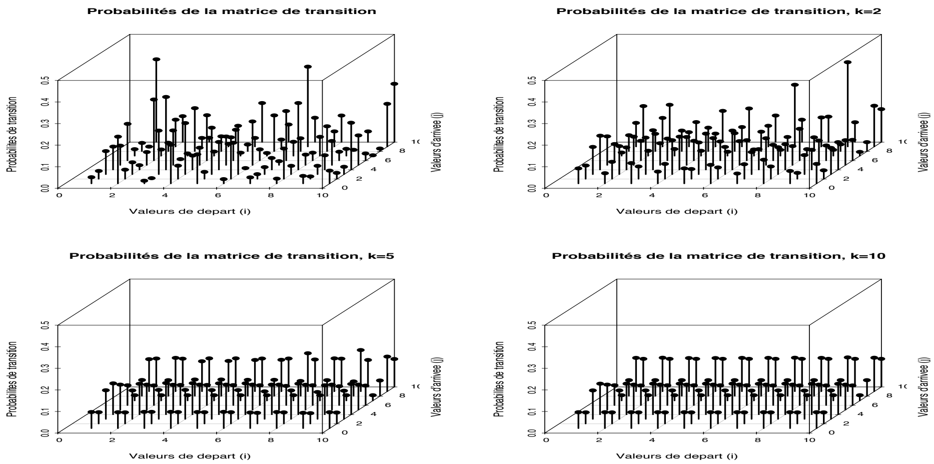


Figure 19: Calcul des puissances d'une matrice de transition.

En guise d'exemple, considérons le code suivant, qui permet de générer des matrices stochastiques aléatoires,

```
P=matrix(0,n,n)
for(i in 1:n){
s=c(0,sort(runif(n-1)),1)
p=s[2:(n+1)]-s[1:(n)]
P[i,]=p}
```

En eefet, la ligne de code `s=c(0,sort(runif(n-1)),1); p=s[2:(n+1)]-s[1:(n)]` permet de tirer au hasard un point du simplexe en dimension n, c'est à dire une mesure de probabilité sur $\{1, 2, ..., n\}$. Les figures[4] **??** suivantes montrent les probabilités des matrices P, P^2, P^5 et P^10.

9.1.2 Cas où $\mathcal{E}$ est infini

Dans le cas où $\mathcal{E}$ est dénombrable, mais non-fini, une condition suffisante n'est plus l'irréductibilité et la récurrence d'états, mais il faut ajouter la propriété de récurence positive. En particulier, la proposition **??** doit être affaiblie,

Proposition 62. *Si $\mathcal{E}$ est un espace d'état infini, alors $\displaystyle\lim_{n\to\infty} \frac{1}{n}\sum_{h=1}^{n} P^h = \Pi$, où Π est une matrice sous-stochastique (la somme des éléments par ligne est inférieure ou égale à 1) vérifiant $\Pi P = P\Pi = \Pi$.*

Il est aussi possible que $\Pi = 0$, par exemple si tous les états sont transients. En revanche

Proposition 63. *Si $\mathcal{E}$ est un espace d'état infini (dénombrable), et s'il existe au moins un état récurrent positif alors il existe au moins une mesure stationnaire.*

Si l'on suppose la chaîne irréductible et réucurrent positive, alors la probabilité stationnaire est unique.

9.2 La période

9.2.1 Périodicité d'un état $j \in \mathcal{E}$

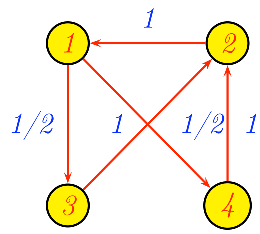
Définition 64. *Soit $j\mathcal{E}$ on appelera d période de l'état j le plus grand commun diviseur de tous les entiers n pour lesquels $p_{j,j}^{(n)} > 0$. Si $d = 1$ on dira que j est apériodique.*

[4]Des animations de cet algorithme sont téléchargeable sur

`http://www.crest.fr/pageperso/charpent/P-trans-1.avi`, `/P-trans-2.avi`, `/P-trans-3.avi`

Un état i est apériodique si $p_{i,i}^{(n)} > 0$ pour tout n suffisement grand, c'est à dire que $d = 1$.

Exemple 65. *Considérons la chaîne suivante*



Tous les états communiquent: il y a une seule classe récurrente.

Les lacets issus de 1 sont $1 \to 3 \to 2 \to 1$ *ou* $1 \to 4 \to 2 \to 1$*, de longueur* 3*, donc* $d(1) = 3$.

Les lacets issus de 2 sont $2 \to 1 \to 3 \to 2$ *ou* $2 \to 1 \to 4 \to 2$*, de longueur* 3*, donc* $d(2) = 3$.

Les lacets issus de 3 sont $3 \to 2 \to 1 \to 3$ *ou* $3 \to 2 \to 1 \to 4 \to 2 \to 1 \to 3$*, etc, de longueur* $3, 6$*, etc, donc* $d(3) = 3$.

Les lacets issus de 4 sont $4 \to 2 \to 1 \to 3$ *ou* $4 \to 2 \to 1 \to 3 \to 2 \to 1 \to 4$*, etc, de longueur* $3, 6$*, etc, donc* $d(4) = 3$.

On peut alors diviser la classe en sous-classes cycliques, $\{1\}$*,* $\{2\}$ *et* $\{3, 4\}$.

La matrice de transition peut alors s'écrire par blocs pour $\mathcal{E} = \{2, 1, 3, 4\}$*,*
$$P = \left(\begin{array}{c|c|cc} 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1/2 & 1/2 \\ \hline 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{array}\right)$$

Exemple 66. *Pour la marche aléatoire dans* $\mathbb{Z}$*, tous les états sont périodiques de période* 2.

9.2.2 Périodicité de classe

Théorème 67. *Si un état* j *est périodique de période* d *et que* $j \leftrightarrow i$*, alors* i *sera également périodique de période* d.

Cette propriété de périodicité étant une propriété de classe, on pourra parler de chaîne de markov de période d pour si P est irréductible.

Théorème 68. *Pour une chaîne irréductible, tous les états ont la même période.*

9.3 Convergence vers la loi limite

Lemme 69. *Si* P *est une matrice irréductible et possède un état apériodique* i*, alors pour tout* $j, k \in \mathcal{E}$*, il existe* n_0 *tel que* $p_{j,k}^{(n)} > 0$ *pour tout* $n \geq n_0$*. Aussi, tous les états sont apériodiques.*

Proof. Il existe r, s tels que $p_{j,i}^{(r)} > 0$ et $p_{i,k}^{(s)} > 0$ et donc
$$p_{j,k}^{(r+n+s)} \geq p_{j,i}^{(r)} \cdot p_{i,i}^{n} \cdot p_{i,k}^{(s)} > 0.$$
□

Théorème 70. *Soit* P *une matrice irréductible et apériodique, possédant une distribution invariante* π*. Étant donnée une distribution initiale* $\boldsymbol{\lambda}$*, et* $(X_n)_{n \in \mathbb{N}}$ *une chaîne de Markov* $(\boldsymbol{\lambda}, P)$*, alors*
$$\mathbb{P}(X_n = j) = \pi_j \text{ lorsque } n \to \infty \text{ pour tout } j \in \mathcal{E},$$
et en particulier $p_{i,j}^{(n)} \to \pi_j$ *lorsque* $n \to \infty$*, pour tout* $i, j \in \mathcal{E}$.

Pour démontrer ce résultat, on va utiliser un argument de couplage.

Proof. Soit $(Y_n)_{n\in\mathbb{N}}$ une chaîne de Markov $(\boldsymbol{\pi}, P)$, et pour $b \in \mathcal{E}$ on pose

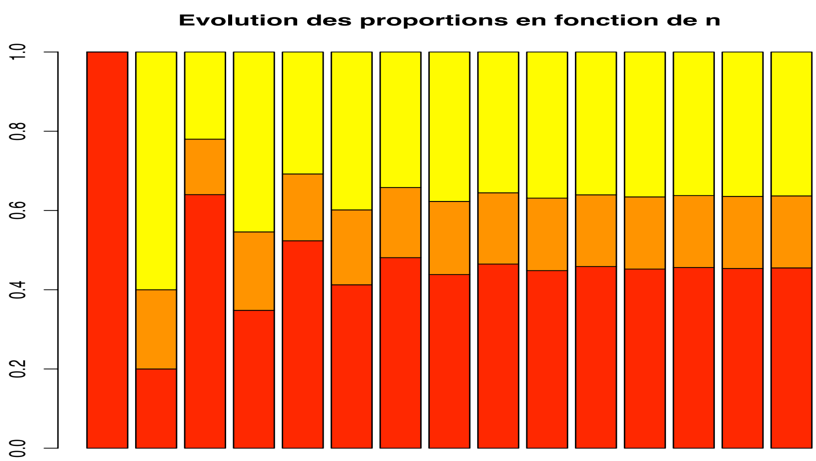
$$T_b = \inf\{n \geq 1 : X_n = Y_n = b\}.$$

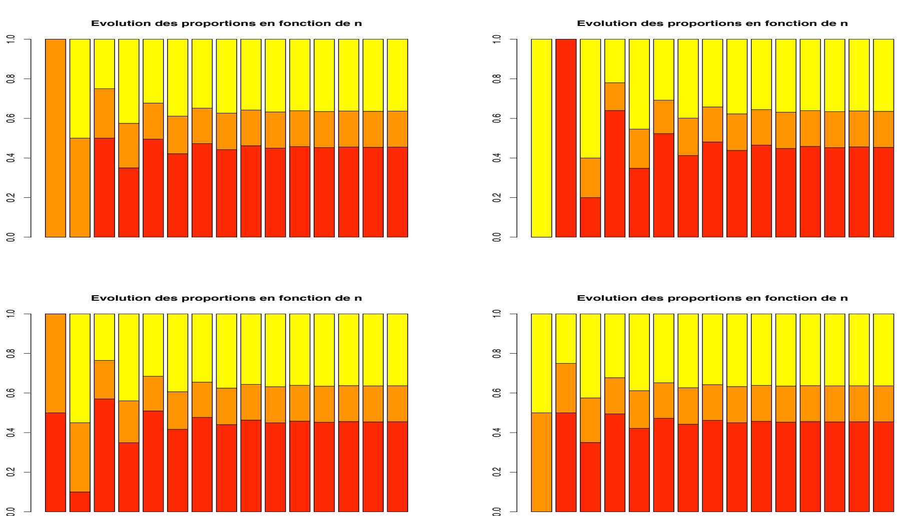
- on commence par montrer que $\mathbb{P}(T < \infty) = 1$,
- on pose $Z_n = \begin{cases} X_n & \text{si } n \leq T \\ Y_n & \text{si } n \geq T, \end{cases}$ $\square$

9.4 Interprétation de l'invariance

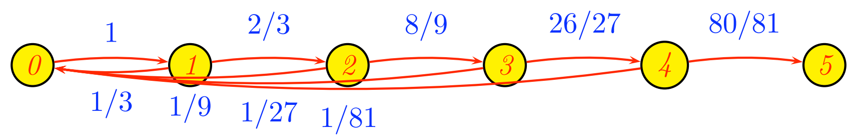
Exemple 71. *Sur l'exemple précédant, supposons* $\boldsymbol{\lambda} = (1, 0, 0)$. *Les distributions aux dates suivantes sont alors*

$$\begin{array}{ccccc}
(& 100.0\% & 00.0\% & 00.0\% &) \\
(& 20.0\% & 20.0\% & 60.0\% &) \\
(& 64.0\% & 14.0\% & 22.0\% &) \\
(& 34.8\% & 19.8\% & 45.4\% &) \\
(& 52.4\% & 16.9\% & 30.8\% &) \\
(& 41.2\% & 18.9\% & 39.9\% &) \\
(& 48.1\% & 17.7\% & 34.2\% &) \\
 & \vdots & \vdots & \vdots & \\
(& 45.4\% & 18.2\% & 36.4\% &)
\end{array}$$

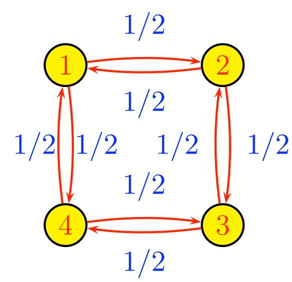




Exemple 72. *Si* $\mathcal{E}$ *est* ***infini*** *il n'existe pas forcément de loi stationnaire.*



Si la chaîne est **périodique** il n'existe pas forcément de loi stationnaire.



Pour des raisons de parité, $\lim_{n\to\infty} \begin{cases} \mathbb{P}(X_{2n+1} = u | X_0 = u) = 0 \\ \mathbb{P}(X_{2n} = u | X_0 = u) > 0 \end{cases}$

9.5 Pour résumer...

- Les lois stationnaires ne chargent que les évènements récurrents (positifs si $\mathcal{E}$ est seulement dénombrable),
- La loi stationnaire est unique si et seulement il n'existe qu'une classe réccurrente (positive si $\mathcal{E}$ est seulement dénombrable),
- Quand il y a une loi stationnaire, les probabilités de la chaîne convergent vers la loi stationnaire au sens de Cesarro. Il y a convergence simple si toutes les composantes récurrentes sont de période 1 (chaîne apériodique).

10 Résultats d'algèbre linéaire pour les chaînes de Markov

Lemme 73. *1 est forcément valeur propre d'une matrice stochatique.*

Proposition 74. *Soit P une matrice carrée stochastique,*

- *Si 1 est valeur propre simple de P, et si toutes les valeurs propres de P autres que 1 sont de module strictement inférieur à 1, alors la suite de matrice $(P^n)_{n\in\mathbb{N}}$ converge vers une matrice stochastique. De plus, toutes les lignes de la matrice sont identiques.*
- *Si 1 est valeur propre multiple de P, et que toutes les autres valeurs propres sont de module strictement inférieur à 1, la suite $(P^n)_{n\in\mathbb{N}}$ converge vers une matrice stochastique.*
- *Si P a au moins une valeur propre autre que 1 dont le module est égal à 1, alors la suite $(P^n)_{n\in\mathbb{N}}$ ne converge pas. Toutefois, la suite $(Q_n)_{n\in\mathbb{N}}$, définie par*

  $$Q_n = \frac{\mathbb{I} + P + \ldots + P^{n-1}}{n}$$

  converge vers une matrice stochastique (ou, dit autrement, la suite $(P^n)_{n\in\mathbb{N}}$ converge au sens de Cesaro vers une matrice stochastique).

Rappelons que si M est une matrice carrée d'ordre r, elle admet r valeurs propres. Soit s le nombre de valeurs propres distinctes, $\lambda_1, \dots, \lambda_s$, et notons $\alpha_1, \dots, \alpha_s$ leur multiplicité. Il existe P telle que

$$A = PDP^{-1} \text{ où } P = \begin{pmatrix} D_{\lambda_1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & D_{\lambda_2} & \dots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & D_{\lambda_s} \end{pmatrix},$$

où les matrices D_{λ_i} sont triangulaires supérieures, $\alpha_i \times \alpha_i$, avec λ_i sur la diagonale.

On peut écrire $D = \displaystyle\sum_{i=1}^{s}(\lambda_i J_i + u_i)$, où J_i et u_i sont de la forme

$$\begin{pmatrix} \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \star & \dots & \mathbf{0} \\ \vdots & \vdots & & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} \end{pmatrix}, \text{ où } \star = \begin{pmatrix} \lambda_i & 0 & \dots & 0 \\ 0 & \lambda_i & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \lambda_i \end{pmatrix} \text{ ou } \begin{pmatrix} 0 & * & \dots & * \\ 0 & 0 & \dots & * \\ \vdots & \vdots & & \vdots \\ 00 & \dots & 0 & \end{pmatrix}$$

Les matrices u_i et J_i vérifient des propriétés d'orthogonalité,

$$J_i J_j = J_i u_j = u_i J_j = u_i u_j = \mathbf{0} \text{ pour } i \neq j,$$

avec $J_i^2 = J_i$, $J_i u_i = u_i J_i = u_i$, et $\displaystyle\sum_{i=1}^{s} J_i = \mathbb{I}$.

Alors

$$D^n = \left(\sum_{i=1}^{s}(\lambda_i J_i + u_i)\right)^n = \sum_{i=1}^{s}(\lambda_i J_i + u_i)^n.$$

Si $\lambda_i \neq 0$, on pose $Q_i(n) = \left(J_i + \dfrac{u_i}{\lambda_i}\right)^n = \displaystyle\sum_{i=0}^{n}\binom{n}{k}\left(\frac{u_i}{\lambda_i}\right)^k$.

On a alors

$$M^n = \sum_{i=1}^{s}(\lambda_i)^n P Q_i(n) P^{-1}.$$

Si de plus on suppose que $\|\lambda\| = 1$, alors $Q_i(n) = Q_i$.

Soit Π la matrice Q_1 associée à la valeur propre 1,

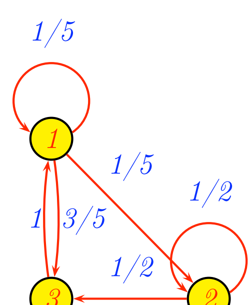
Proposition 75. *Soit P une matrice stochastique d'ordre r, alors pour tout $n \geq r$,*

$$P^n = \Pi + \sum_j (\lambda_j)^n Q_j + \sum_k (\lambda_k)^n Q_k(n),$$

où la première somme se fait sur les valeurs propres $\|\lambda_i\| = 1$ et la seconde $\|\lambda_i\| < 1$.

Notons que $\Pi^2 = \Pi$ et $Q_j^2 = Q_j$ (ce sont des projecteurs).

Exemple 76. *Considérons la matrice de transition* $P = \begin{pmatrix} 1/5 & 1/5 & 3/5 \\ 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \end{pmatrix}$,



alors $P = TDT^{-1}$, *où*

$$\begin{pmatrix} 1/5 & 1/5 & 3/5 \\ 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} -0.577 & 0.494 & 0.107 \\ -0.577 & 0.354 & -0.936 \\ -0.577 & -0.794 & 0.334 \end{pmatrix} \begin{pmatrix} 1.00 & 0 & 0 \\ 0 & -0.622 & 0 \\ 0 & 0 & 0.322 \end{pmatrix} \begin{pmatrix} -0.577 & 0.494 & 0.107 \\ -0.577 & 0.354 & -0.936 \\ -0.577 & -0.794 & 0.334 \end{pmatrix}^{-1}$$

La matrice étant diagonale, la puissance nième se calcul aisément

$$\begin{pmatrix} 1/5 & 1/5 & 3/5 \\ 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \end{pmatrix}^{n} = \begin{pmatrix} -0.577 & 0.494 & 0.107 \\ -0.577 & 0.354 & -0.936 \\ -0.577 & -0.794 & 0.334 \end{pmatrix} \begin{pmatrix} 1.00 & 0 & 0 \\ 0 & -0.622 & 0 \\ 0 & 0 & 0.322 \end{pmatrix}^{n} \begin{pmatrix} -0.577 & 0.494 & 0.107 \\ -0.577 & 0.354 & -0.936 \\ -0.577 & -0.794 & 0.334 \end{pmatrix}^{-1}$$

En particulier

$$P^5 = \begin{pmatrix} 0.413 & 0.189 & 0.398 \\ 0.421 & 0.190 & 0.388 \\ 0.524 & 0.168 & 0.308 \end{pmatrix} \text{ et } P^{10} = \begin{pmatrix} 0.458 & 0.181 & 0.360 \\ 0.457 & 0.181 & 0.361 \\ 0.448 & 0.183 & 0.369 \end{pmatrix}$$

Proposition 77. *Si la matrice de transition P est telle qu'il existe k tel que P^k n'admet que des termes stricements positifs, alors $P^n \to P = [\pi]$.*

Exemple 78. *Soit* $P = \begin{pmatrix} 3/4 & 1/4 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix}$, *alors*

$$P^2 = \begin{pmatrix} 11/16 & 3/16 & 1/8 \\ 3/8 & 5/8 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix}, P^3 = \begin{pmatrix} 39/64 & 19/64 & 3/32 \\ 19/32 & 3/32 & 5/16 \\ 3/8 & 5/8 & 0 \end{pmatrix} \text{ et } P^4 = \begin{pmatrix} 155/256 & 63/256 & 38/256 \\ 63/128 & 59/128 & 3/64 \\ 19/32 & 3/32 & 5/16 \end{pmatrix}$$

Aussi la chaîne de Markov est convergente. Or l'équation charactéristique de P est

$$detP = \begin{pmatrix} 3/4 - z & 1/4 & 0 \\ 1/2 & -z & 1/2 \\ 0 & 1 & -z \end{pmatrix} = 0$$

soit $z^3 - 3/4z^2 - 5/8z + 3/8 = 0$. Comme 1 est racine, cette équation se ramène à

$$(z-1)(z^2 + 1/4z - 3/8) = 0$$

dont les racines sont $1/2$ et $-3/4$ (de module strictemnt inférieur à 1). La chaîne de Markov est effectivement convergente. Formellement rappelons qu'il suffisait de résoudre $\pi(P - \mathbb{I}) = 0$.

Proposition 79. *Une chaîne de Markov de matrice de transition P est ergodique si et seulement si la valeur propre 1 est valeur propre simple, et est l'unique valeur propre de module 1. La loi limite de la chaîne est alors donnée par l'unique vecteur propre à gauche (ligne) de la matrice P (c'est-à-dire, les probabilités limites sont les coefficients de l'unique tel vecteur propre dont la somme des coefficients soit 1).*

11 Exemples de chaînes de Markov

11.1 Modèle de *Bernoulli-Laplace*

On considère 2 urnes, A et B, et on suppose que deux types de boules (rouges et vertes) sont introduites, en nombre k. On note X_0 le nombre de boules rouges dans la première urne.

À chaque date, on procède à un échange deux boules prises dans chacune des urnes, et on note X_n le nombre de boules rouges dans la première urne après n échanges.

La matrice de transition est donnée par

$$p_{i,i} = 2\frac{i(k-i)}{k^2}, p_{i,i-1} = \left(\frac{i}{k}\right)^2 \text{ et } p_{i,i+1} = \left(\frac{k-i}{k}\right)^2,$$

pour $0 < i, j < k$, et $p_{0,1} = 0 = p_{k,k-1} = 1$.

Cette chaîne (finie) est effectivement irréductible, puisqu'il est possible de joindre les états i et j en $|i-j|$ étapes, avec une probabilité non nulle,

$$\prod_{i=\min\{i,j\}}^{\max\{i,j\}-1} \left(\frac{k-i}{k}\right)^2.$$

Cette chaîne est apériodique car $P_{i,i} > 0$ pour tout i.

Compte tenu de la forme *quasi*-diagonale de la matrice de transition, on peut déterminer la loi invariante π, solution de $P'\pi = \pi$, ce qui revient à résoudre

$$\begin{cases} \pi_0 = p_{0,0}\pi_0 + p_{1,0}\pi_1 \\ \pi_1 = p_{0,1}\pi_0 + p_{1,1}\pi_1 + p_{2,1}\pi_2 \\ ... \\ \pi_i = p_{i-1,i}\pi_{i-1} + p_{i,i}\pi_i + p_{i+1,i}\pi_{i+1} \\ ...\pi_k = p_{k-1,k}\pi_{k-1} + p_{k,k}\pi_k \end{cases}$$

ce qui se résoud de manière récursive,

$$_1 = \frac{p_{0,1}}{p_{1,0}}\pi_0, _2 = \frac{p_{1,2}}{p_{2,1}}\frac{p_{0,1}}{p_{1,0}}\pi_0, ..., \pi_i = \frac{p_{i-1,i}}{p_{i,i-1}}\cdots\frac{p_{0,1}}{p_{1,0}}\pi_0,$$

soit, au final, $\pi_i = \left(\binom{i}{k}\right)^2 \pi_0$, soit, par normalisation

$$\pi_i = \frac{\left(\binom{i}{k}\right)^2}{\binom{i}{2k}},$$

aussi, la loi hypergéométrique $\mathcal{H}(2k, k, 1/2)$ est la loi invariante.

11.2 Urne de Ehrenfest

On considère deux urnes A et B contenant au total m boules. On tire une boule dans une des urnes choisie au hasard et on la replace dans l'autre. On cherche à décrire

$$X_n = \text{ nombre de boule dans l'urne } A \text{ après } n \text{ tirages.}$$

(X_n) est une chaîne de Markov, avec pour probabilités de transition

$$\mathbb{P}(X_{n+1} = k-1 | X_n = k) = \frac{k}{m} \text{ pour } k = 1, ..., m-1, m,$$

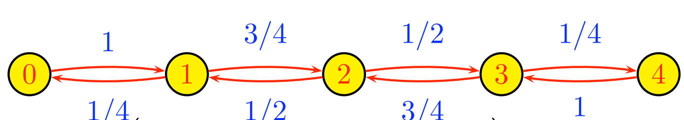
$$\mathbb{P}(X_{n+1} = k+1 | X_n = k) = \frac{m-k}{m} \text{ pour } k = 0, 1, ..., m-1,$$

$$\mathbb{P}(X_{n+1} = j | X_n = i) = 0 \text{ sinon .}$$

La matrice de transition associée est

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

Dans le cas où $n = 4$, la chaîne est alors



de matrice de transition $P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/4 & 0 & 3/4 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$

La chaîne est irréductible (donc récurente positive) et admet une probabilité invariante.

Notons que pour $i = 1, ..., m-1$,

$$\pi_i = \pi_{i-1} \times \frac{m-i-1}{m} + \pi_{i+1} \times \frac{i+1}{m},$$

avec des conditions de bords de la forme $\pi_0 = \pi_1/m$. Par récurence, $\pi_i = \pi_0 \binom{i}{m}$. Comme

$$\sum_{i=0}^{m} \pi = 1 \text{ alors } \pi_0 \sum_{i=0}^{m} \binom{i}{m} = 2^m \pi_0 = 1,$$

aussi, $\pi_i = \frac{1}{2^m}\binom{i}{m}$ pour $i = 0, ..., m$. La loi stationnaire est la loi binomiale $\mathcal{B}(m, 1/2)$.

11.3 Urne de Pólya

On considère une urne[5] avec $v + r$ boules, de deux couleurs V et R. Lorsqu'on tire une boule, on la remet dans l'urne, et on rajoute une boule de la même couleur. Il s'agit d'un *modle(simple)decontagion*.

Soient V_n et R_n le nombre de boules après n tirages. $\boldsymbol{X}_n = (V_n, R_n)$ est une chaîne de Markov dans $\mathbb{N} \times \mathbb{N}$, d'état initial $(1, 1)$, et de matrice de transition

$$\mathbb{P}((i+1, j)|(i, j) = \frac{i}{(i+j} \text{ et } \mathbb{P}((i, j+1)|(i, j) = \frac{j}{(i+j}.$$

Notons $Z_n = R_n/[V_n + R_n]$ la proportion de boules rouge.

Par récurence, on peut monter aisément que

$$\mathbb{P}\left(Z_n = \frac{k}{n+2}\right) = \frac{1}{n+1} \text{ pour } k = r, r+1, r+2, ..., r+n.$$

Soient V_n' et R_n' le nombre de boules **sorties** après n tirages, i.e. $V_n' = V_n - v$ et $R_n' = R_n - v$. Alors

$$\mathbb{P}(V_n' = i, R_n' = n-i) = \frac{n!}{i!(n-i)!} \frac{r(r+1)...(r+i-1)v(v+1)...(v+n-i-1)}{(r+v)(r+v+1)...(r+v+n-1)}$$

Notons $Z_n' = R_n'/[V_n' + R_n'] = R_n'/n$ la proportion de boules rouge **sorties**.

...

Alors

$$\mathbb{P}(Z_n' = z) = \frac{(r+nz-1)!}{(r-1)!(nz)!} \frac{(v+n(1-z)-1)!}{(v-1)!(n(1-z)!} \frac{(r+v-1)!n!}{(r+v+n-1)!}$$

En utilisant la formule de Stirling, on peut montrer que

$$\mathbb{P}(Z_n' = z) \sim \frac{(r+v-1)!}{(r-1)!(v-1)!} z^{r-1}(1-z)^{v-1}, \text{ pour } x \in [0, 1].$$

11.4 Processus de branchement et taille de population

On considère une population dont les individus se reproduisent, à chaque génération, de manière indépendante, et tels que le nombre de descendants X admet une loi de fonction génératrice g, $g(t) = \mathbb{E}(t^X)$.

On note S_n le nombre d'individus à la $n$ième génération. Notons que

$$S_n \stackrel{\mathcal{L}}{=} X_1 + ... + X_{S_{n-1}},$$

où les X_i sont i.i.d. En partant d'un seul individu, on obtient que la fonction génératrice g_n de S_n est $g_n(t) = g^n(t)$ où $g^k = g \circ g^{k-1}$.

[5]Une animation de cet algorithme est téléchargeable sur

`http://www.crest.fr/pageperso/charpent/MC-UP-0001.avi`

Si $\mathbb{P}(X=0)=0$, la chaîne est croissante, et donc la chaîne (S_n) est transiente.

Si $\mathbb{P}(X=0)\neq 0$, il est possible que la population disparaisse, et la probabilité de premier retour à 0 est $\rho_n=\mathbb{P}(S_n=0)=g_n(1)$. Notons que (ρ_n) vérifie alors une relation de récurrence $\rho_n=g(\rho_{n-1})$.

$\rho_n\to 1\in[0,1[$ si et seulement si $g'(1)>1$, c'est à dire $\mathbb{E}(X)>1$: la transience de la chaîne est associée au nombre moyen de descendant par individu.

Si $g'(1)>1$, la chaîne est transiente, et si $g'(1)\leq 1$, la chaîne est récurente.

Supposons que $g'(1)\leq 1$. Si (S_n) admet une loi invariante, de fonction caractéristique ψ, alors ψ doit vérifier

$$\psi(t)=g(t)\psi(0)+\psi(g(t))-\psi(0).$$

Exemple 80. *Si X suit une loi de Bernoulli, $\mathcal{B}(p)$, $g(t)=(1-p)+pt$, et ψ est alors solution de*

$$\psi(t)=g(t)\psi(0)+\psi(g(t))-\psi(0)=\psi((1-p)+pt)+p(t-1)\psi(0),$$

ce qui donne, de manière itérative

$$\psi(t)=\psi((1-p)+p(1-p)+...+p^{k-1}(1-p)+p^k t)+(p+p^2+...+p^k)(t-1)\psi(0),$$

pour tout $k\in\mathbb{N}$. Par passage à la limite, $k\to\infty$,

$$\psi(t)=\psi\left(\frac{1-p}{1-p}\right)+\psi\left(\frac{p}{1-p}\right)(t-1)\psi(0)=1+\frac{p}{1-p}(t-1)\psi(0),$$

i.e. $\psi(0)=1-p$ et $\psi(t)=1+p(t-1)=(1-p)+pt$: la loi de Bernoulli est invariante.

12 Théorème(s) ergodique(s)

Le premier théorème ergodique est une forme de loi de grand nombre pour des suites de variables aléatoires non indépendantes (mais markoviennes). Rappelons que de manière classique, si les X_i sont de même loi, et si les variables sont de variance finie, la loi (faible) des grands nombres (théorème de Khintchine) garantie que

$$\lim_{n\to\infty}\mathbb{P}\left(\left|\frac{X_1+...+X_n}{n}-\mathbb{E}(X)\right|>\varepsilon\right)=0,$$

qui se montre à l'aide de l'inégalité de Tchebychev, et donc $\dfrac{X_1+...+X_n}{n}\xrightarrow{\mathbb{P}}\mathbb{E}(X)$. Aussi, ce résultat peut s'écrire, pour toute fonction f intégrable,

$$\frac{1}{n}\sum_{i=1}^{n}f(X_i)\xrightarrow{\mathbb{P}}\mathbb{E}(f(X))=\int_{\mathcal{E}}f(x)d\pi(x),$$

où π désigne la loi des X_i.

La loi (forte) des grands nombres garantie une convergence presque sûr dès lors que les variables sont d'espérance finie,

$$\mathbb{P}\left(\lim_{n\to\infty}\frac{X_1+...+X_n}{n}(\omega)=\mathbb{E}(X)\right)=1.$$

Théorème 81. *Soit $(X_n)_{n\in N}$ une chaîne de Markov irréductible récurrente positive, et π son unique mesure invariante. Pour toute fonction f bornée sur $\mathcal{E}$,*

$$\frac{1}{n}\sum_{k=0}^{n-1}f(X_k)\xrightarrow{\mathbb{P}}\int_{\mathcal{E}}f(x)d\pi(x)=\sum_{x\in\mathcal{E}}\pi(x)f(x).$$

Proof. Pour tout état $x \in \mathcal{E}$, on note $N_x(n)$ le nombre de retour à l'état x avant l'instant n, $N_x(n) = \sum_{k=0}^{n} \mathbf{1}(X_k = x)$. On s'intéresse alors à toutes les excursions, entre chaque retour à l'état x, notées $\mathcal{E}_0, \mathcal{E}_1, ..., \mathcal{E}_{N_x(n)}$, chacune de longueur $L_x(k)$. Notons que $L_x(0) + ... + L_x(N_x(n) - 1) \leq n \leq L_x(0) + ... + L_x(N_x(n))$, de telle sorte que

$$\frac{L_x(0) + ... + L_x(N_x(n) - 1)}{N_x(n)} \leq \frac{n}{N_x(n)} \leq \frac{L_x(0) + ... + L_x(N_x(n))}{N_x(n)}.$$

Les excursions étant indépendantes (propriété de Markov), on peut utiliser la loi des grands nombres, et

$$\frac{L_x(0) + ... + L_x(N_x(n))}{N_x(n)} \overset{p.s.}{\to} \mathbb{E}(T_x),$$

où T_x est le temps du premier retour à l'état x, $T_x = \inf\{n \geq 1, X_n = x\}$. Aussi, $n/N_x(n)$ tend presque sûrement vers $\mathbb{E}(T_x)$, ou encore $\dfrac{N_x(n)}{x} \overset{p.s.}{\to} \dfrac{1}{\mathbb{E}(T_x)}$ (le *presque sûr* étant par rapport à $\mathbb{P}_x(\cdot) = \mathbb{P}(\cdot|X_0 = x)$, mais on peut montrer que la convergence a lieu quelle que soit la mesure initiale λ, puisque les limites sont identiques).

Pour montrer maintenant la convergence, notons que

$$\left| \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) - \sum_{x \in \mathcal{E}} \pi(x) f(x) \right| = \left| \sum_{x \in \mathcal{E}} \left(\frac{N_x(n)}{n} - \pi(x) \right) f(x) \right|.$$

On utilise le fait que f est bornée, i.e. $|f(\cdot)| \leq M$. Soit $\mathcal{F} \in \mathcal{E}$, alors

$$\begin{aligned}
\left| \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) - \sum_{x \in \mathcal{E}} \pi(x) f(x) \right| &= \left| \sum_{x \in \mathcal{E}} \left(\frac{N_x(n)}{n} - \pi(x) \right) f(x) \right| \\
&\leq M \sum_{x \in \mathcal{F}} \left| \frac{N_x(n)}{n} - \pi(x) \right| + M \sum_{x \notin \mathcal{F}} \left(\frac{N_x(n)}{n} + \pi(x) \right) \\
&\leq 2M \sum_{x \in \mathcal{F}} \left| \frac{N_x(n)}{n} - \pi(x) \right| + 2M \sum_{x \notin \mathcal{F}} pi(x).
\end{aligned}$$

La difficulté est que $\mathcal{E}$ peut être de cardinal infini, on choisit alors $\mathcal{F}$ de cardinal fini, tel que $\sum_{x \notin \mathcal{F}} pi(x)$ soit aussi petit que possible (inférieure à ε). On utilise alors la convergence étable précédement, et donc, il existe N tel que pour $n \geq N$, $\sum_{x \in \mathcal{F}} \left| \dfrac{N_x(n)}{n} - pi(x) \right| \leq \varepsilon$, ce qui garantira la convergence. □

Le second théorème ergodique est une forme de théorème centrale limite pour des suites de variables aléatoires non indépendantes (mais markoviennes). Rappelons que de manière classique, si les variables X_i sont indépendantes, de même loi (et centrées, pour simplifier l'écriture)

$$\frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} X_k \overset{\mathcal{L}}{\to} \mathcal{N}(0, \sigma^2),$$

où $\sigma^2 = \text{Var}(X) = \mathbb{E}(X) = \displaystyle\int x^2 d\pi(x)$.

On supposera (dans la démonstration, mais le théorème reste généralement vrai) que $\mathcal{E}$ est fini. On se placera également dans le cas centré, là aussi pour simplifier.

Théorème 82. *Soit $(X_n)_{n \in N}$ une chaîne de Markov irréductible, et π son unique mesure invariante sur $\mathcal{E}$ fini. Pour toute fonction f sur $\mathcal{E}$ telle que $\int_{\mathcal{E}} f(x) d\pi(x) = \sum_{x \in \mathcal{E}} \pi(x) f(x) = 0$, alors il existe σ (dépendant de f) telle que*

$$\frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} f(X_k) \overset{\mathcal{L}}{\to} \mathcal{N}(0, \sigma^2).$$

Proof. L'idée est d'écrire

$$\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1} f(X_k) = \sum_{x\in\mathcal{E}} \pi(x)f(x)\sqrt{\frac{N_x(n)}{n}}\left(\frac{\sqrt{N_x(n)}}{\pi_(x)} - \frac{n}{\sqrt{N_x(n)}}\right),$$

qui va se comportement de la même manière que

$$\sum_{x\in\mathcal{E}} \pi(x)f(x)\sqrt{\frac{N_x(n)}{n}}\frac{\mathcal{L}_x(0)+...+\mathcal{L}_x(N_x(n))}{\sqrt{N_x(n)}},$$

où $\mathcal{L}_x(k) = L_x(k) - 1/\pi(x)$. Par indépendance sur chaque excursion, on peut invoquer le théorème centrale limite qui assure que

$$\frac{\mathcal{L}_x(0)+...+\mathcal{L}_x(N_x(n))}{\sqrt{N_x(n)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

□

Notons que le calcul de la variance n'est pas forcément simple à obtenir (en pratique). Elle vaut

$$\sigma^2 = 2\sum_{x\in\mathcal{E}}\pi(x)\mu(x)f(x) - \sum_{x\in\mathcal{E}}\pi(x)f(x)^2, \text{ où } \mu(x) = \sum_{n=0}^{\infty}\mathbb{E}(f(X_n)|X_0 = x),$$

pour tout $x \in \mathcal{E}$.

13 Inversion du temps

Si $(X_n)_{n0}$ est une chaîne de Markov de matrice de transition P, conditionnellement à la valeur présente, le passé et le futur sont indépendant.

Théorème 83. *Soit $(X_n)_{0nN}$ une chaîne de Markov de matrice de transition P irréductible, de distribution invariante π, et de mesure initiale $\lambda = \pi$, et poson $Y_n = X_{N-n}$. Alors $(Y_n)_{0nN}$ est une chaîne de Markov de matrice de transition $\tilde{P}$ et de mesure initiale π où*

$$\pi_j\tilde{p}_{j,i} = \pi_i p_{i,j} \text{ pour tout } i, j,$$

et $\tilde{P}$ est irréductible, de distribution invariante π.

14 Passage au temps continu

14.1 Approche infinitésimale

Etant donnée une matrice Q , la série $\sum_{k\geq 0}\frac{Q^k}{k!}$ converge, et sera notée $\exp(Q)$.

Théorème 84. *Soit Q une matrice définie sur l'espace d'état I. Notons $P(t) = \exp(tQ)$, alors la fonction $P(\cdot)$ vérifie*

- *$P(s+t) = P(s) + P(t)$,*
- *P est l'unique solution de l'équation différencielle* forward $\frac{d}{dt}P(t) = P(t)Q$,
- *P est l'unique solution de l'équation différencielle* backward $\frac{d}{dt}P(t) = QP(t).$,
- *pour tout $k \geq 0$, $\left(\frac{d}{dt}\right)^k P(t)\Big| t = 0 = Q^k$.*

Définition 85. *Une matrice M est une Q-matrice sur I si*

- $q_{i,i} \in (-\infty, 0]$,
- $q_{i,j}$ *pour tout* $j \neq i$,
- $\sum_{j \in I} q_{i,j} = 0$.

Exemple 86. *La matrice* $\begin{pmatrix} -2 & 1 & 1 \\ -1 & 1 & 0 \\ 2 & 1 & -3 \end{pmatrix}$ *est une* Q*-matrice.*

Théorème 87. *Une matrice* Q *est une* Q*-matrice si et seulement si* $P(t) = \exp(tQ)$ *est une matrice stochastique pour tout* $t \geq 0$.

Les composantes de l'équation différentielle forward $P'(t) = P(t)Q$ sont

$$\begin{array}{lll} \mathrm{p}'_{i,i}(t) = -\lambda p_{i,i}(t) & \mathrm{p}_{i,i}(0) = 1 & \text{pour i} < \text{N} \\ \mathrm{p}'_{i,j}(t) = -\lambda p_{i,j}(t) + \lambda p_{i,j-1}(t) & \mathrm{p}_{i,j}(0) = 0 & \text{pour i<j} < \text{N} \\ \mathrm{p}'_{i,N}(t) = \lambda p_{i,N-1}(t) & \mathrm{p}_{i,N}(0) = 0 & \text{pour i} < \text{N} \end{array}$$

La première équation admet pour solution $p_{i,i}(t) = \exp(-\lambda t)$ for $i < N$, et pour $i < j < N$, $(e^{-\lambda t} p_{i,j}(t))' = e^{\lambda t} p_{i,j-1}(t)$, et par induction, $e^{-\lambda t} \dfrac{(\lambda t)^{j-i}}{(j-i)!}$. Si $i = 0$, on obtient les probabilités d'une loi de Poisson de paramètre λt.

à continuer

14.2 Processus de Poisson

Pour rappel, une variable aléatoire positive T suit une loi exponentielle de paramètre $\lambda \geq 0$ si

$$\mathbb{P}(T > t) = \exp(-\lambda t), \text{ pour tout } t \geq 0.$$

La moyenne de T est alors $\mathbb{E}(T) = 1/\lambda$.

Théorème 88. T *suit une loi exponentielle si et seulement si elle vérifie la propriété d'absence de mémoire,*

$$\mathbb{P}(T > s = t | T > s) = \mathbb{P}(T > t) \text{ pour tout } s, t \geq 0.$$

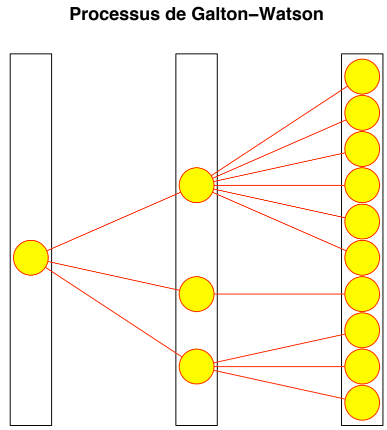
[à compléter]

15 Statistique pour les chaînes de Markov

15.1 Un petit exemple

Considérons un processus de branchement de type Galton-Watson. On part de $X_0 = 1$, et on suppose que la loi de reproduction est une loi binomiale $\mathcal{B}(N, p$, où p est inconnu. On suppose les reproductions sont indépendantes

On note X_n le nombre de descendant en n générations.



Notons que X_{n+1} sachant $X_n = i$ est la somme de i binomiales indépendantes $\mathcal{B}(N, p$, donc

$$X_{n+1}|X_n = i\ sim\mathcal{B}(Ni, p)$$

donc

$$\mathbb{P}(X_{n+1} = j|X_n = i) = \binom{Ni}{j} p^j (1-p)^{Ni-j}, = p_{i,j},\ \text{pour } j = 0, 1, ..., Ni.$$

Supposons que l'on ait observé à chaque date $x_0 = 1, x_1, ..., x_n$. La vraisemblance de cet échantillon est

$$\mathcal{L}(x_0, ..., x_n, p) = \prod_{i=0}^{n-1} p_{x_i, x_{i+1}} = \prod_{i=0}^{n-1} \binom{Nx_i}{x_{i+1}} p^{x_{i+1}} (1-p)^{Nx_i - x_{i+1}}$$

aussi $\mathcal{L}(x_0, ..., x_n, p) \left(\prod_{i=0}^{n-1} \binom{Nx_i}{x_{i+1}} \right) p^{\sum_{i=0}^{n-1} x_{i+1}} (1-p)^{N \sum_{i=0}^{n-1} x_i - x_{i+1}}$. On cherche alors

$$\widehat{p} \in \underset{\text{argmax}}{p \in (0,1)} \log \mathcal{L}(x_0, ..., x_n, p).$$

La log-vraisemblance s'écrit

$$\log \mathcal{L}(x_0, ..., x_n, p) = \text{constante} + \left(\log \frac{p}{1-p} \right) \sum_{i=0}^{n-1} x_{i+1} + N \left(\log(1-p) \right) \sum_{i=0}^{n-1} x_i.$$

Aussi, $\dfrac{\partial \log \mathcal{L}(x_0, ..., x_n, p)}{\partial \widehat{p}} = 0$ si et seulement si

$$\widehat{p} = \frac{1}{N} \left(\frac{\sum_{i=0}^{n-1} x_{i+1}}{\sum_{i=0}^{n-1} x_i} \right).$$

15.2 Généralisation

On suppose que $\mathcal{E}$ est un espace d'état fini. On cherche à estimer $P = [p_{i,j}]$, la matrice de transition.

On note $N_n^{i,j}$ le nombre de transition de i vers j observées entre les dates 0 et n,

$$N_n^{i,j} = \sum_{k=0}^{n-1} \mathbf{1}(X_k = i, X_{k+1} = j).$$

Soit $N_n^{i,\cdot}$ le nombre de passage en i,

$$N_n^{i,\cdot} = \sum_{k=0}^{n-1} \mathbf{1}(X_k = i) = \sum_{j \in \mathcal{E}} N_n^{i,j}.$$

Supposons que l'on ait observé à chaque date $x_0 = 1, x_1, ..., x_n$. La vraisemblance de cet échantillon est

$$\mathcal{L}\left(x_0, ..., x_n, P = [p_{i,j}]\right) = \lambda(x_0) \times \prod_{k=0}^{n-1} p_{x_k, x_{k+1}}.$$

On cherche alors

$$\widehat{P} = [\widehat{p}_{i,j}] \in \underset{\text{argmax}}{p_{i,j} \in (0,1)} \sum_{k=0}^{n-1} \log p_{x_k, x_{k+1}} = \sum_{i,j \in \mathcal{E}} \log p_{i,j} N_n^{i,j},$$

avec la constrainte d'avoir une matrice stochastique, i.e. $\displaystyle\sum_{j \in \mathcal{E}} p_{i,j} = 1$ pour tout i.

En posant λ_i le multiplicateur de Lagrange associé,

$$\log \mathcal{L} - \sum_{i \in \mathcal{E}} \lambda_i \left(\sum_{j \in \mathcal{E} p_{i,j} - 1} \right) = 0.$$

Pour tout $i\mathcal{E}$, on obtient que $\lambda_i = N_n^{i,\cdot}$, et donc

$$\widehat{p}_{i,j} = \frac{N_n^{i,j}}{N_n^{i,\cdot}} = \text{ estimateur empirique.}$$

D'après le théorème ergodique,

$$\frac{1}{n} N_n^{i,\cdot} = \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}(X_k = i) \overset{\mathbb{P}-as}{\to} \pi(i),$$

où λ est la mesure stationnaire associée à la chaîne P, et

$$\frac{1}{n} N_n^{i,j} = \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}(X_k = i, X_{k+1} = j) \overset{\mathbb{P}-as}{\to} \pi(i) p_{i,j},$$

aussi, pour tout $i, j \in \mathcal{E}$,

$$\widehat{p}_{i,j} \overset{\mathbb{P}-as}{\to} p_{i,j}.$$

On peut en fait montrer que cet estimateur est normalement convergent,

$$\sqrt{n}\left(\widehat{p}_{i,j} - p_{i,j}\right) \overset{\mathcal{L}}{\to} \mathcal{N}\left(0, \frac{p_{i,j}(1 - p_{i,j})}{\pi(i)}\right).$$

References

[1] Benaïm, M. & El Karoui, N. *Promenade Aléatoire*, Ed. Ecole Polytechnique.

[2] Cottrell, M., Genon-Catalot, V., Duhamel, C. & Meyren T. *Exercices de probabilits*, Ed. Cassini.

[3] Foata, D. & Fuchs, A. *Processus Stochastiques*, Ed. Dunod.

[4] Norris, J.R. *Markov Chains*, Ed. Cambridge University Press.

[5] Robert, C. *Méthodes de Monte Carlo par chaîne de Markov*, Ed. Economica.

[6] Ycart, B. *Premiers pas en statistique*, Ed. Springer.