

Master d'Astronomie et d'Astrophysique M2-R

PROJET INFORMATIQUE 2013–2014

Jacques Le Bourlot, Didier Pelat

5 juin 2014

1 Objectif du projet

Le projet doit permettre aux étudiants de programmer quelques unes des méthodes classiques dont le but est de générer une suite de nombres pseudo-aléatoires censée provenir d'une loi de probabilité donnée. On mettra plus particulièrement l'accent sur les simulations de Monte-Carlo par chaîne de Markov ou méthodes MCMC.

La loi de probabilité simulée rend souvent compte d'une loi physique, comme en physique statistique, ou permet de réaliser une intégration ou encore une optimisation (recherche d'un maximum global). Cette simulation peut également permettre de générer des bruits instrumentaux afin de valider, par exemple, des algorithmes de détection.

2 Simuler des variables aléatoires

2.1 Rappel sur les variables aléatoires

Une variable aléatoire X est caractérisée par sa fonction de répartition F . Cette fonction est égale, par définition, à la probabilité pour que X ne dépasse pas un seuil x . On a

$$F(x) \stackrel{\text{déf}}{=} \text{Pr}(X \leq x). \quad (1)$$

On caractérise également X par sa densité de probabilité f , si elle existe. On a

$$f(x) = \frac{dF}{dx}, \quad F(x) = \int_{-\infty}^x f(u) du. \quad (2)$$

La forme explicite de F ou f détermine la loi suivie par X , par exemple : la loi uniforme, la loi exponentielle ou la loi normale. Pour ces trois cas on a respectivement

$$F(x) = \begin{cases} 0, & \text{si } x \leq 0 \\ x, & \text{si } 0 \leq x \leq 1 \\ 1, & \text{si } x \geq 1 \end{cases}, \quad f(x) = \begin{cases} 1, & \text{si } x \in [0, 1] \\ 0, & \text{si } x \notin [0, 1] \end{cases}, \quad (3)$$

$$F(x) = 1 - e^{-\lambda x}, \quad f(x) = \lambda e^{-\lambda x}, \quad (4)$$

$$F(x) = \frac{1}{2} \left(1 + \text{erf} \left(\frac{x - \mu}{\sqrt{2}\sigma} \right) \right), \quad f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2}, \quad (5)$$

Les quantités λ , μ et σ sont les paramètres des lois. La loi uniforme entre 0 et 1 n'a pas de paramètres, c'est un cas particulier de la loi uniforme entre a et b . Pour simplifier

l'écriture de la densité il est pratique d'utiliser l'indicatrice $\mathbf{1}_A(x)$ qui est égale à 1 si $x \in A$ et à 0 dans le cas contraire. L'ensemble A est le *support* de la loi. La densité de la loi uniforme entre 0 et 1 s'écrit alors $f(x) = \mathbf{1}_{[0,1]}(x)$. En toute rigueur, on devrait écrire pour la loi exponentielle $f(x) = \lambda e^{-\lambda x} \mathbf{1}_{[0,\infty[}(x)$. On a introduit pour la fonction de répartition de la loi normale, la fonction erf qui est, par définition, égale à

$$\operatorname{erf}(x) \stackrel{\text{déf}}{=} \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \quad (6)$$

Elle s'obtient en Fortran, version 2008, par un appel à la fonction erf. Par exemple :

```
Real :: n, n_sigma_error
n = 3.0
n_sigma_error = erf(n/Sqrt(2.0))
```

dont le résultat est l'intégrale $\int_{-3}^3 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$.

Pour indiquer qu'une variable aléatoire suit une loi donnée on écrit, par exemple pour chacun des trois cas évoqués ci-dessus

$$X \sim \mathcal{U}(0, 1), \quad X \sim \mathcal{Exp}(\lambda), \quad X \sim \mathcal{N}(\mu, \sigma^2).$$

Deux variables possédant la même fonction de répartition, sauf peut-être en leurs points de discontinuités, sont dites égales *en loi*. On note $X \stackrel{\text{Loi}}{=} Y$, mais souvent on indique simplement $X = Y$ étant sous-entendu qu'il s'agit d'une égalité en loi entre variables aléatoires.

2.2 Trois méthodes

Il existe trois méthodes classiques pour générer des nombres issus d'une loi donnée.

1. La méthode par inversion.
2. La méthode du rejet de von Neumann.
3. La méthode de Monte-Carlo par chaînes de Markov dite MCMC.

Toutes ces méthodes nécessitent l'existence d'un générateur des nombres aléatoires suivant la loi uniforme entre 0 et 1. Nous n'étudierons pas les diverses méthodes permettant d'atteindre ce but, nous nous contenterons de l'examen des procédures mises à disposition par les divers langages informatiques, ici le Fortran 90, censées délivrer des nombres suivant la loi uniforme $\mathcal{U}(0, 1)$.

2.3 Nombres pseudo-aléatoires uniformes

Le langage Fortran 90 et ses versions ultérieures intègre dans sa norme le calcul de nombres pseudo-aléatoires suivant la loi uniforme entre 0 et 1. Ce calcul se fait par l'intermédiaire de l'instruction

```
Call Random_number(harvest=tableau)
```

Le mot-clef harvest est une étiquette permettant d'identifier la variable tableau, sa présence est facultative. La variable tableau est destinée à recevoir les valeurs calculées par le programme Call Random_number.

Les générateurs de nombres aléatoires doivent pouvoir produire : soit toujours la même série de nombres (en phase de test, par exemple); soit une série différente à chaque lancement du programme (en phase d'exploitation). En Fortran 90 cette possibilité est contrôlée par un appel au sous-programme Random_seed. Si l'appel se fait sans arguments :

Call Random_seed

la série est différente d'une exécution du programme à une autre et ce dernier choisit lui-même les paramètres de la simulation. Cependant cette norme n'est pas toujours correctement implémentée et il y a lieu de vérifier que le système utilisé y est fidèle. Si l'on désire contrôler soi-même la séquence de nombres au hasard, il faut fournir un tableau : seed de graines de départ dont la taille dépend de l'implémentation particulière du compilateur fortran. On obtient ce contrôle grâce à deux appels successifs à Random_seed :

```
Integer :: k
Integer, Allocatable, Dimension(:) :: seed
Call Random_seed(size=k)
Allocate(seed(k))
seed = [s1, s2, ..., sk]
Call Random_seed(put=seed(1:k))
```

Si l'on désire connaître le tableau des graines utilisées par le programme, il faut alors inclure les appels suivants :

```
Call Random_seed(size=k)
Allocate(seed(k))
Call Random_seed(get=seed(1:k))
Print*,seed
```

Les étiquettes size, put ou get sont obligatoires. Une bonne pratique consiste à imprimer et à sauvegarder les graines utilisées au début et à la fin du programme. De cette façon on aura la possibilité de recommencer le même calcul (pour la phase de mise au point) ou de le poursuivre.

On suppose, à partir de maintenant, que l'on dispose d'une variable aléatoire U suivant la loi uniforme entre 0 et 1. La variable U est un être mathématique mais pour nous, d'un point de vue pratique, c'est un programme.

2.4 La méthode par inversion

2.4.1 Cas des lois continues

Un résultat classique de la théorie du changement de variables aléatoires, dit que si F désigne la fonction de répartition d'une variable aléatoire continue X , alors la variable aléatoire $F(X)$ est uniforme entre 0 et 1.

Démonstration. On connaît la fonction de répartition F de la variable X qui, par définition, est égale à :

$$F(x) \stackrel{\text{déf}}{=} \Pr(X \leq x).$$

On désire connaître G la fonction de répartition de la variable $U = F(X)$.

La variable U étant définie en tant que fonction de répartition, elle est comprise entre 0 et 1. Pour $0 \leq u \leq 1$, on a :

$$\begin{aligned} G(u) &\stackrel{\text{déf}}{=} \Pr(U \leq u), \\ &= \Pr(F(X) \leq u), \\ &= \Pr(X \leq F^{-1}(u)), \\ &= F(F^{-1}(u)) = u. \end{aligned}$$

Finalement :

$$G(u) = \begin{cases} 0, & \text{si } u \leq 0; \\ u, & \text{si } 0 \leq u \leq 1; \\ 1, & \text{si } u \geq 1. \end{cases}$$

La variable U suit une loi uniforme entre 0 et 1 (sa densité est constante). \square

On écrira $F(X) = U$, en donnant à l'égalité entre variables aléatoires le sens de l'égalité en loi.

Il est alors possible d'utiliser le changement de variables afin de générer des nombres X suivant une loi caractérisée par F . Il suffit d'écrire : $X = F^{-1}(U)$, à condition que la fonction F soit bijective, ce qui est le cas pour les variables continues. D'un point de vue technique, cette procédure est praticable si F est facilement inversible numériquement.

Exemple de la loi exponentielle. À partir de sa fonction de répartition $F(x) = 1 - e^{-\lambda x}$, on en déduit :

$$\begin{aligned} U &= 1 - e^{-\lambda X} \\ e^{-\lambda X} &= 1 - U \\ -\lambda X &= \ln(1 - U) \\ X &= -\frac{1}{\lambda} \ln(1 - U) \\ X &= -\frac{1}{\lambda} \ln(U). \end{aligned}$$

La dernière égalité ne doit pas surprendre car les variables $1 - U$ et U possèdent la même fonction de répartition et sont, par conséquent, égales en loi. Ainsi, pour générer une variable aléatoire suivant la loi exponentielle de paramètre λ , il suffit d'avoir à disposition un générateur de nombres suivant la loi exponentielle de paramètre $\lambda' = 1$ et de diviser les résultats par λ .

2.4.2 Cas des lois discrètes

Considérons le cas d'une variable aléatoire discrète X prenant les valeurs X_k avec la probabilité p_k : $\Pr(X = X_k) = p_k$ pour $k \in \mathbb{N}$. Il est facile de simuler la variable X à l'aide d'une variable U uniforme entre 0 et 1, en remarquant que d'après la définition de la fonction de répartition de la loi uniforme

$$\Pr(F(X_{k-1}) < U \leq F(X_k)) = F(X_k) - F(X_{k-1}) = p_k.$$

La méthode consiste alors à générer une variable U uniforme entre 0 et 1, repérer l'intervalle (X_{k-1}, X_k) où $F(X_{k-1}) < U \leq F(X_k)$ et d'affecter à X la borne supérieure de (X_{k-1}, X_k) . La valeur X_k est donc la plus petite valeur x telle que $F(x) \geq U$. Si on introduit l'inverse généralisée $F^{(-1)}$ de la façon suivante

$$F^{(-1)}(u) = \inf\{x | F(x) \geq u\}, \quad (7)$$

on peut alors étendre la méthode par inversion au cas discret et écrire

$$X_k = F^{(-1)}(U). \quad (8)$$

Comme l'inverse généralisée coïncide avec l'inverse si F est continue, la validité de cette définition s'étend à toutes les variables aléatoires quelles soient discrètes ou continues.

Exemple : la loi géométrique. La loi géométrique indique le rang du premier succès dans une suite d'épreuves indépendantes de Bernoulli de paramètre p (jeu de pile ou face). Ce paramètre est la probabilité de succès d'une épreuve. Si la variable X suit une loi géométrique, on a $\Pr(X = k) = p(1 - p)^{k-1}$, $k = 1, 2, \dots$. Sa fonction de répartition est $F(x) = 1 - (1 - p)^{\lfloor x \rfloor}$. Afin de simuler X , on écrit

$$\begin{aligned} U &= 1 - (1 - p)^{\lfloor X \rfloor}, \\ (1 - p)^{\lfloor X \rfloor} &= 1 - U = U, \\ \lfloor X \rfloor \ln(1 - p) &= \ln(U), \\ \lfloor X \rfloor &= \frac{\ln(U)}{\ln(1 - p)}. \end{aligned}$$

Le point délicat est d'inverser, au sens généralisé, la fonction $\lfloor x \rfloor = y$. D'après la définition (7) de l'inverse généralisé, x est la plus petite valeur telle que $\lfloor x \rfloor \geq y$, mais $\lfloor x \rfloor$ étant un entier, x est alors le plus petit entier supérieur ou égal à y , autrement dit $\lceil y \rceil$. On obtient finalement

$$X = \left\lceil \frac{\ln(U)}{\ln(1 - p)} \right\rceil.$$

2.5 La méthode du rejet de von Neumann

La méthode du rejet, ou méthode par réjection (voir [9]), consiste à simuler une variable aléatoire X de densité connue f à partir d'une autre variable Y de densité g dont la simulation, par exemple par la méthode d'inversion, est facile à mettre en œuvre. La densité g doit pouvoir, après multiplication par une constante M , majorer f pour tout x de son support

$$\forall x, g(x) \neq 0, \frac{f(x)}{g(x)} \leq M. \quad (9)$$

La densité f est la loi *cible* et g est la loi *enveloppe*. La procédure est la suivante

Méthode de von Neumann

1. On génère Y suivant la loi g , facile à simuler. Soit y la valeur trouvée.
2. On génère u d'après la variable U qui suit la loi uniforme entre 0 et 1. Les variables Y et U doivent être indépendantes.
3. Si $u \leq \frac{1}{M} \frac{f(y)}{g(y)}$, alors $x = y$, sinon on retourne à l'étape 1.

Afin de prouver que la variable X générée de cette façon, suit effectivement la loi de densité f , il faut calculer la probabilité conditionnelle $\Pr(Y \leq x | U \leq \frac{1}{M} \frac{f(Y)}{g(Y)})$ donnant la probabilité de succès du test à l'étape 3 de la procédure, et montrer qu'elle est égale à $\Pr\{X \leq x\}$.

Démonstration. Le couple de variables aléatoires indépendantes (Y, U) possède la loi conjointe $f_{Y,U}$ donnée par l'expression $f_{Y,U}(y, u) = g(y)\mathbf{1}_{[0,1]}(u)$. Calculons la fonction de répartition de Y sujette à la condition pour que X soit identifié à Y . C'est une probabilité conditionnelle qui se calcule à l'aide de la formule $\Pr(A|B) = \Pr(A, B) / \Pr(B)$. Il vient

$$\Pr\left(Y \leq x \mid U \leq \frac{1}{M} \frac{f(Y)}{g(Y)}\right) = \frac{\Pr\left(Y \leq x, U \leq \frac{1}{M} \frac{f(Y)}{g(Y)}\right)}{\Pr\left(U \leq \frac{1}{M} \frac{f(Y)}{g(Y)}\right)}. \quad (10)$$

Le dénominateur est la probabilité de succès de la méthode, c'est-à-dire la probabilité de produire le nombre X à l'étape 3. En notant que $\frac{1}{M} \frac{f(y)}{g(y)} \leq 1$, il vient

$$\begin{aligned} \Pr\left(U \leq \frac{1}{M} \frac{f(Y)}{g(Y)}\right) &= \int_{-\infty}^{+\infty} g(y) \, dy \int_0^{\frac{1}{M} \frac{f(y)}{g(y)}} du, \\ &= \int_{-\infty}^{+\infty} g(y) \frac{1}{M} \frac{f(y)}{g(y)} \, dy, \\ &= \frac{1}{M} \int_{-\infty}^{+\infty} f(y) \, dy = \frac{1}{M}. \end{aligned}$$

Évaluons à présent le numérateur de (10)

$$\begin{aligned} \Pr\left(Y \leq x, U \leq \frac{1}{M} \frac{f(Y)}{g(Y)}\right) &= \int_{-\infty}^x g(y) \, dy \int_0^{\frac{1}{M} \frac{f(y)}{g(y)}} du, \\ &= \int_{-\infty}^x g(y) \frac{1}{M} \frac{f(y)}{g(y)} \, dy, \\ &= \frac{1}{M} \int_{-\infty}^x f(y) \, dy. \end{aligned}$$

De sorte que (10) s'écrit à présent

$$\Pr\left(Y \leq x \mid U \leq \frac{1}{M} \frac{f(Y)}{g(Y)}\right) = \frac{\frac{1}{M} \int_{-\infty}^x f(y) \, dy}{\frac{1}{M}} = \Pr(X \leq x).$$

La loi suivie par X est donc identique à la loi conditionnelle de Y . □

La quantité $1/M$ mesurant la probabilité de succès de la méthode il faut, afin d'obtenir l'efficacité maximum, que la constante M soit la plus petite possible tout en respectant la condition (9). On a

$$\frac{f(x)}{g(x)} \leq \max_x \frac{f(x)}{g(x)} \leq M.$$

On peut alors choisir $M_{\text{opt}} = \max_x f(x)/g(x)$ et obtenir ainsi l'efficacité maximum de la méthode, soit M_{opt}^{-1} .

3 Chaînes de Markov

3.1 Définition

Une chaîne de Markov est une suite ordonnée de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ où les X_n prennent leurs valeurs dans un espace des états \mathcal{E} . Nous ne considérerons ici que le cas où l'espace \mathcal{E} est de dimension finie : $\text{card}(\mathcal{E}) = m \in \mathbb{N}$. On note x_k un état particulier parmi les m possibles et bien souvent on y fait référence seulement par son indice k . L'indice n représente souvent le temps. L'évolution des variables X_n avec n est caractéristique des chaînes de Markov.

3.2 Évolution des variables de la chaîne

La première variable X_0 se trouve dans un état x_k avec la probabilité $\pi_0(x_k)$, c'est-à-dire : $\Pr(X_0 = x_k) = \pi_0(x_k)$. On adoptera la convention qui consiste à faire référence à l'adresse de x_k plutôt qu'à sa valeur et on notera $\Pr(X_0 = k) = \pi_0(k)$.

La propriété spécifique des chaînes de Markov réside en ce que le passage de la variable actuelle X_n à la variable suivante X_{n+1} ne dépend pas des valeurs des variables précédentes (X_{n-1}, \dots, X_0) . On a

$$\Pr(X_{n+1} = j | X_n = i, X_{n-1} = i_1, \dots, X_0 = i_n) = \Pr(X_{n+1} = j | X_n = i). \quad (11)$$

La probabilité $\Pr(X_{n+1} = j | X_n = i)$ est appelée *probabilité de transition*. On supposera en outre que cette probabilité ne dépend pas de n (chaîne de Markov homogène). On posera

$$\Pr(X_{n+1} = j | X_n = i) = p_{ij}. \quad (12)$$

Après la première transition, la variable X_1 prend ses valeurs dans \mathcal{E} avec la probabilité π_1 . En vertu de la formule des probabilités totales on a

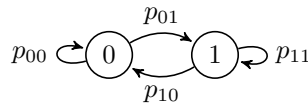
$$\pi_1(j) = \sum_{i=1}^m \pi_0(i) p_{ij}. \quad (13)$$

Et d'une manière générale, par récurrence

$$\pi_{n+1}(j) = \sum_{i=1}^m \pi_n(i) p_{ij}. \quad (14)$$

3.3 Graphe d'une chaîne de Markov

Il est pratique de visualiser une chaîne de Markov à l'aide d'un graphe où les états de la chaîne sont représentés par des nœuds et les probabilités de transition par des flèches reliant un nœud à un autre. À titre d'exemple voici le graphe d'une chaîne de Markov à deux états notés 0 et 1.



Une réalisation de cette chaîne de Markov est formée d'une suite, en général quelconque, de 0 et de 1. Cette suite peut être interprétée comme représentant l'indicatrice d'un sous-ensemble de \mathbb{N} . Il existe alors une infinité non dénombrable ($\aleph_1 = 2^{\aleph_0}$) de chaînes de Markov à deux états (et *a fortiori* à $m \geq 2$ états).

3.4 État stationnaire

On s'intéresse aux chaînes qui au cours de leurs évolutions possèdent une probabilité π_n qui tend vers une loi invariante par l'action des p_{ij} . Cet état π , s'il existe, est dit *état stationnaire*. L'état stationnaire est obtenu lorsque pour chaque nœud le bilan global est nul, c'est-à-dire que les probabilités sortantes sont égales aux probabilités rentrantes. Ecrivons ces bilans pour la chaîne à deux états illustrée ci-dessus. Pour l'état 0 on a :

$$\pi(0)p_{00} + \pi(1)p_{10} = \pi(0)p_{00} + \pi(1)p_{10}, \quad (15)$$

et pour l'état 1 :

$$\pi(1)p_{11} + \pi(1)p_{10} = \pi(0)p_{01} + \pi(1)p_{11} . \quad (16)$$

Si $(p_{01}, p_{10}) = (0, 0)$, les conditions initiales $\pi_0(0)$ et $\pi_0(1)$ restent invariantes avec n , il vient alors $\pi(0) = \pi_0(0)$, $\pi(1) = \pi_0(1)$.

Si $(p_{01}, p_{10}) \neq (0, 0)$ et $(p_{01}, p_{10}) \neq (1, 1)$, les équations (15) et (16) sont identiques et la solution générale est $\pi(0) = \lambda p_{10}$ et $\pi(1) = \lambda p_{01}$, où λ est un réel. La constante λ est trouvée en imposant la condition de normalisation $\pi(0) + \pi(1) = 1$. Il vient

$$\pi(0) = \frac{p_{10}}{p_{01} + p_{10}}, \quad \pi(1) = \frac{p_{01}}{p_{01} + p_{10}} . \quad (17)$$

Il existe alors un état stationnaire $\pi = \lim_{n \rightarrow \infty} \pi_n$. La convergence de π_n vers π est géométrique. On a

$$\begin{aligned} \pi_n(0) &= \pi(0) + (\pi_0(0) - \pi(0))(1 - p_{01} - p_{10})^n, \\ \pi_n(1) &= \pi(1) + (\pi_0(1) - \pi(1))(1 - p_{01} - p_{10})^n. \end{aligned}$$

Si $(p_{01}, p_{10}) = (1, 1)$, la solution (17) précédente donne $\pi(0) = \pi(1) = 0.5$, elle n'est acceptable que si, au départ, $\pi_0(0) = \pi_0(1) = 0.5$. Sinon la chaîne oscille entre les états 0 et 1, et on a $\pi_n(0) = \pi_0(0)$ si n est pair et $\pi_n(0) = \pi_0(1)$ si n est impair et un comportement équivalent pour $\pi_n(1)$. Il n'existe pas d'état stationnaire.

Une chaîne de Markov quelconque ne possède pas nécessairement d'état stationnaire, il faut des conditions particulières pour cela. Nous ne nous intéresserons qu'à l'une d'entre elles appelée *bilan détaillé* parce que les méthodes MCMC ne font appel qu'à cette dernière. Ce n'est par ailleurs qu'une condition suffisante de stationnarité.

3.5 Le bilan détaillé

Reprenons les équations de bilan global (15) et (16), par exemple pour l'équation (15), on peut écrire

$$\begin{aligned} \pi(0)p_{00} + \pi(0)p_{01} &= \pi(0)p_{00} + \pi(1)p_{10}, \\ \pi(0)(p_{00} + p_{01}) &= \pi(0)p_{00} + \pi(1)p_{10}. \end{aligned}$$

Mais $p_{00} + p_{01} = 1$ car on a dressé la liste exhaustive de toutes les probabilités sortantes à partir de l'état 0. Il vient

$$\pi(0) = \pi(0)p_{00} + \pi(1)p_{10} . \quad (18)$$

Cette équation est un cas particulier de l'équation (14) pour l'état stationnaire π . Pour une chaîne quelconque, le bilan global peut s'exprimer par l'équation aux valeurs propres

$$\pi(j) = \sum_{i=1}^m \pi(i)p_{ij} . \quad (19)$$

Mais reconsidérons l'équation (15), en la simplifiant par $\pi(0)p_{00}$, elle peut alors s'écrire

$$\pi(0)p_{01} = \pi(1)p_{10} . \quad (20)$$

Cette équation est celle du *bilan détaillé* entre les états 0 et 1. Elle indique que la probabilité de passer de l'état 0 à l'état 1 est égale à celle de passer de l'état 1 à l'état 0. Dans le cas d'une chaîne à deux états, le bilan global implique le bilan détaillé, mais

ceci n'est pas assuré pour une chaîne à plus de deux états. L'équation générale du bilan détaillé (detailed balance en anglais) est

$$\forall i, j \in \mathcal{E}, \quad \pi(i)p_{ij} = \pi(j)p_{ji}. \quad (21)$$

Dans la pratique on ne s'intéresse pas aux cas triviaux où p_{ij} et p_{ji} sont nuls, où lorsque $i = j$.

Le bilan détaillé implique le bilan global, ce que l'on démontre en sommant l'équation (21) sur les indices i . Il vient

$$\begin{aligned} \sum_i \pi(i)p_{ij} &= \sum_i \pi(j)p_{ji}, \\ \sum_i \pi(i)p_{ij} &= \pi(j) \sum_i p_{ji}, \\ \sum_i \pi(i)p_{ij} &= \pi(j). \end{aligned}$$

On retrouve ainsi l'équation du bilan global. La réciproque n'est pas vraie, pour une chaîne à 3 états (0, 1 et 2 par exemple); lorsqu'une solution stationnaire existe, le bilan global n'implique le bilan détaillé que si $p_{01}p_{12}p_{20} = p_{02}p_{21}p_{10}$.

Une chaîne de Markov possédant un état stationnaire satisfaisant le bilan détaillé est dite *réversible*. En physique statistique, un système satisfaisant le bilan détaillé entre tous ses niveaux d'énergies est dit à l'équilibre thermodynamique local.

4 Les méthodes MCMC

4.1 Principe

La simulation d'une loi de densité f par la méthode MCMC, consiste à construire une chaîne de Markov réversible possédant la loi f comme état stationnaire. Si π désigne l'état stationnaire de la chaîne, on vise alors à construire une chaîne telle que $\pi = f$. Lorsque l'état stationnaire est atteint, les valeurs successives de la chaîne simulent la loi f . La loi f est dite loi *cible*. En réalité, la densité que l'on cherche à simuler peut n'être connue qu'à une constante multiplicative λ près $f = \lambda g$.

4.2 Mise en œuvre

Il n'est en général pas possible de construire directement la chaîne de loi invariante égale à π , on passe pour cela par l'intermédiaire d'une autre chaîne possédant les probabilités de transitions q_{ij} . On pose

$$q_{ij} = q(y_j|x_i), \quad q_{ji} = q(x_i|y_j). \quad (22)$$

La valeur x_i est l'état actuel de la chaîne et y_j est un état futur potentiel qui va ou non être accepté en fonction du résultat d'un certain test. La fonction q est la densité d'une loi (éventuellement conditionnelle) à peu près quelconque appelée loi de *proposition*. Cette loi q doit être facilement simulable.

L'idée est de corriger les probabilités de transition q_{ij} de façon à ce que les probabilités de transition p_{ij} qui en résultent assurent le bilan détaillé de la chaîne de Markov censée simuler f . On pose

$$p_{ij} = \alpha_{ij}q_{ij}, \quad \text{avec} \quad 0 \leq p_{ij} \leq 1. \quad (23)$$

Supposons que la chaîne soit dans l'état x_i . La variable auxiliaire Y_j est générée avec la probabilité q_{ij} et donne y_j , la quantité α_{ij} apparaît alors comme la probabilité d'accepter y_j en tant que nouvelle valeur x_j issue de la loi f . Pour qu'il en soit ainsi, il suffit que l'équation du bilan détaillé $\pi(i)p_{ij} = \pi(j)p_{ji}$ soit satisfaite. C'est-à-dire

$$f(x_i)\alpha_{ij}q_{ij} = f(x_j)\alpha_{ji}q_{ji}. \quad (24)$$

Si la valeur y_j n'est pas acceptée on pose $x_j = x_i$ et le bilan détaillé est alors automatiquement satisfait puisque ce qui part de l'état X_i y revient. Le choix de α_{ij} détermine une méthode MCMC particulière.

4.3 La méthode de Barker

Cette méthode consiste à trouver une formulation symétrique de (24) en posant

$$\alpha_{ij} = \frac{f(x_j)q_{ji}}{f(x_i)q_{ij} + f(x_j)q_{ji}}. \quad (25)$$

De cette façon le bilan détaillé est garanti, en effet

$$\begin{aligned} f(x_i)\alpha_{ij}q_{ij} &= f(x_i)f(x_j)q_{ji}q_{ij}/(f(x_i)q_{ij} + f(x_j)q_{ji}), \\ f(x_j)\alpha_{ji}q_{ji} &= f(x_j)f(x_i)q_{ij}q_{ji}/(f(x_j)q_{ji} + f(x_i)q_{ij}). \end{aligned}$$

Il faut simplement veiller à ce que α_{ij} soit une probabilité, mais ce point est assuré car les α_{ij} donnés par la formule (25) sont compris entre 0 et 1 (voir [2]).

Algorithme de Barker

0. Une valeur de départ x_0 est donnée.

1. À l'étape n on génère un nombre y issu de la loi de densité $q : Y \sim q(y|x_n)$.

2. On génère un nombre u issu d'une variable U uniforme entre 0 et 1.

Si $u < \alpha(x_n, y)$ alors $x_{n+1} = y$,

sinon $x_{n+1} = x_n$,

avec

$$\alpha(x, y) = \frac{f(y)q(x|y)}{f(x)q(y|x) + f(y)q(x|y)}.$$

3. On retourne en 1. pour générer la valeur suivante.

Après l'exécution de cet algorithme, on dispose d'une réalisation $(x_n)_{1 \leq n \leq N}$ de la chaîne $(X_n)_{1 \leq n \leq N}$, où N est le nombre de simulations effectuées. Techniquement il s'agit d'une réalisation d'une marginale fini-dimensionnelle de la chaîne $(X_n)_{n \in \mathbb{N}}$.

À titre d'exemple on a généré diverses valeurs issues d'une loi cible gamma de paramètres $\nu = 2$, $\lambda = 2$ à l'aide d'une loi de proposition exponentielle. Cette loi gamma particulière est la convolution de la loi exponentielle de densité $2e^{-2x}$ par elle-même. La loi de proposition est une loi exponentielle de paramètre conditionné aux valeurs des x_n , on a posé $q(y|x) = \frac{1}{x}e^{-\frac{y}{x}}$. L'espace des états \mathcal{E} est l'ensemble des nombres représentables en machine ; son cardinal est gigantesque, mais fini. Le résultat de cette simulation est présenté sur la figure 1 où on a comparé un histogramme des valeurs simulées avec la densité de la loi gamma multipliée par le nombre de simulations.

Loi gamma proposition exponentielle conditionnelle, par Barker

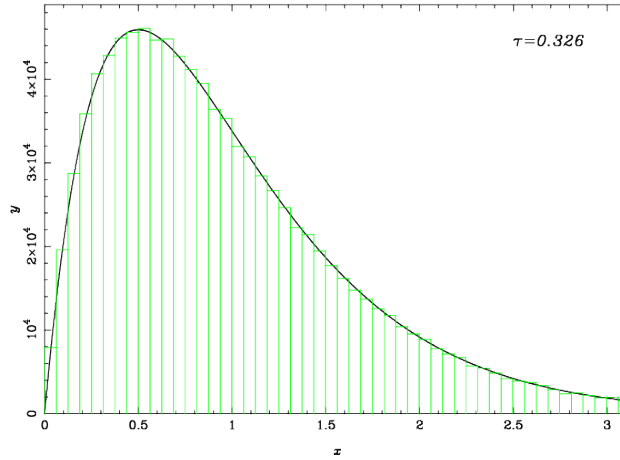


FIGURE 1 – Simulation d’un échantillon issu de la loi gamma $X \sim \mathcal{G}am(2, 2)$ à l’aide de la méthode de Barker. Il a été procédé à 10^6 simulations, le taux d’acceptation est de 32.6%.

4.4 La méthode de Metropolis

Cette méthode (voir [6]) est identique à celle de Barker à ceci près que la probabilité d’acceptation est donnée par la formule

$$\alpha_{ij} = \min \left[\frac{f(x_j)q_{ji}}{f(x_i)q_{ij}}, 1 \right]. \quad (26)$$

Il en découle l’algorithme suivant :

Algorithme de Metropolis

0. Une valeur de départ x_0 est donnée.

1. À l’étape n on génère un nombre y issu de la loi de densité $q : Y \sim q(y|x_n)$.

2. On génère un nombre u issu d’une variable U uniforme entre 0 et 1.

Si $u < \alpha(x_n, y)$ alors $x_{n+1} = y$,

sinon $x_{n+1} = x_n$,

avec

$$\alpha(x, y) = \min \left[\frac{f(y)q(x|y)}{f(x)q(y|x)}, 1 \right].$$

3. On retourne en 1. pour générer la valeur suivante.

On remarque que si $f(y)q(x|y) \geq f(x)q(y|x)$, y est systématiquement accepté, ce n’est pas le cas pour la méthode de Barker. Si $f(y)q(x|y) < f(x)q(y|x)$, y est accepté avec la probabilité $\frac{f(y)q(x|y)}{f(x)q(y|x)}$ et avec une probabilité moindre pour Barker. Le taux d’acceptation moyen de la méthode de Metropolis sera donc supérieur à celui de Barker. Il reste à montrer que la forme donnée par Metropolis de la probabilité d’acceptation α_{ij} garanti bien le bilan détaillé.

Loi gamma proposition exponentielle conditionnelle, par Metropolis-Hastings

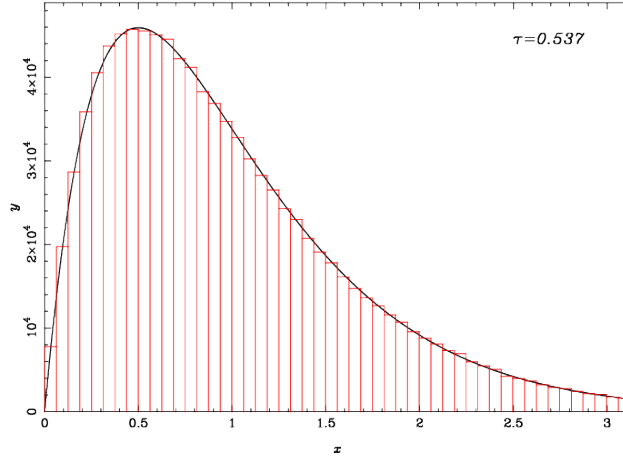


FIGURE 2 – Simulation de la loi gamma $X \sim \mathcal{G}am(2, 2)$ par la méthode de Metropolis. Il a été procédé à 10^6 simulations, le taux d'acceptation est de 53.7%.

Si $\alpha_{ij} = \frac{f(y)q_{ji}}{f(x)q_{ij}}$ alors $\alpha_{ji} = 1$ et il vient

$$\begin{aligned} f(x_i)p_{ij} &= f(x_i)\alpha_{ij}q_{ij} = f(x_i)\frac{f(x_j)q_{ji}}{f(x_i)q_{ij}}q_{ij} = f(x_j)q_{ji}, \\ f(x_j)p_{ji} &= f(x_j)\alpha_{ji}q_{ji} = f(x_j)q_{ji}. \end{aligned}$$

La démonstration est analogue si $\alpha_{ij} = 1$, on aurait obtenu

$$f(x_i)p_{ij} = f(x_i)q_{ij}, \quad f(x_j)p_{ji} = f(x_i)q_{ij}.$$

Le bilan détaillé est donc satisfait. La figure 2 montre une simulation de la même loi gamma que celle de la figure 1, mais par la méthode de Metropolis.

4.5 Formulation de Hastings

Hastings [5] à donné une formule pour α_{ij} qui regroupe celles de Barker et de Metropolis et peut en suggérer d'autres. Il montre que

$$\alpha_{ij} = \frac{s_{ij}}{1 + t_{ij}}, \quad (27)$$

où

$$s_{ij} = s_{ji}, \quad t_{ij} = \frac{f(x_i)q_{ij}}{f(x_j)q_{ji}}. \quad (28)$$

Toutes les valeurs s_{ij} , t_{ij} , q_{ij} étant non nulles. On retrouve la méthode de Barker en posant $s_{ij} = 1$ et celle de Metropolis en posant $s_{ij} = 1 + \min(t_{ij}, t_{ji})$. On démontre aisément, en notant que $t_{ij}t_{ji} = 1$, que la formule de Hastings satisfait le bilan détaillé. En effet

$$\begin{aligned} f(x_i)\alpha_{ij}q_{ij} &= f(x_j)\alpha_{ji}q_{ji}, \\ f(x_i)q_{ij}\frac{s_{ij}}{1 + t_{ij}} &= f(x_j)q_{ji}\frac{s_{ji}}{1 + t_{ji}}, \end{aligned}$$

Lot gamma proposition exponentielle conditionnelle, par Metropolis-Hastings

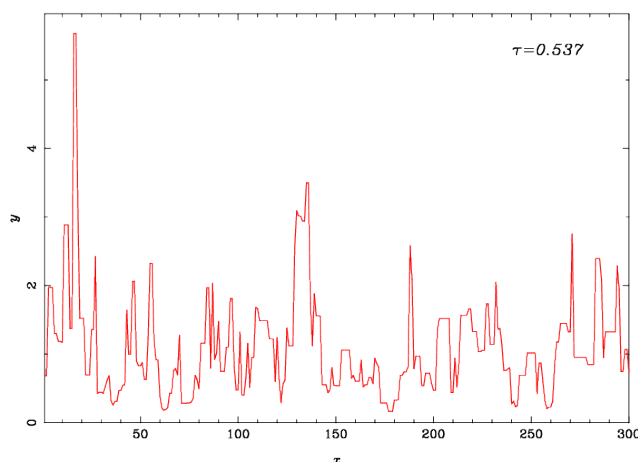


FIGURE 3 – Trois cents dernières valeurs de l'échantillon suivant la loi gamma simulé par la méthode de Metropolis. Les refus apparaissent sous forme de plateaux.

$$t_{ij} \frac{s_{ij}}{1 + t_{ij}} = \frac{s_{ij}}{1 + t_{ji}},$$

$$t_{ij} = \frac{1 + t_{ij}}{1 + t_{ji}} = t_{ij}.$$

4.6 Caractéristiques des méthodes MCMC

L'avantage principal des méthodes MCMC est qu'il est possible de les mettre en pratique sans connaître la constante de normalisation de la loi cible puisque seul le rapport $f(y)/f(x)$ intervient dans la méthode. De cette manière, on évite le calcul d'une intégrale qui peut être laborieux voir impraticable. En revanche, il faut connaître la constante de normalisation de la loi de proposition car dans le rapport $q(x|y)/q(y|x)$ la constante de normalisation du numérateur dépend de y et celle de dénominateur de x et, en général, elles ne se simplifient pas.

Le désavantage de la méthode MCMC par rapport à la méthode du rejet, par exemple, est que l'échantillon généré n'est pas formé de variables indépendantes, car si la valeur proposée est refusée, on reste sur la même valeur (voir figure 3). Pour limiter les effets de cette dépendance, il faut augmenter considérablement la taille de l'échantillon.

La loi de proposition peut être quelconque à la condition qu'elle puisse atteindre toutes les valeurs de la loi cible, éventuellement après plusieurs tirages. Un choix inapproprié de la proposition par rapport à la cible peut conduire à une convergence très lente. Un exemple classique est une cible Cauchy avec une proposition gaussienne.

4.7 L'exemple du ferromagnétisme

L'application de la méthode de Metropolis à un modèle de ferromagnétisme va démontrer toute la puissance des méthodes MCMC par rapport aux autres méthodes de Monte-Carlo. Le modèle consiste en N particules dotées d'un spin orienté vers le haut ou vers le bas. Si la variable s_i désigne l'état du spin de la particule i , on peut avoir : $s_i = +1$ ou $s_i = -1$, suivant que le spin est orienté vers le haut ou vers le bas. On

considère que cet ensemble de particules est à l'équilibre thermodynamique avec un thermostat à la température T .

On note x une configuration de spins particulière ; cette notation recouvre l'ensemble de toutes les valeurs des spins s_1, s_2, \dots, s_N . L'élément x appartient à l'ensemble des états $\mathcal{E} = \{-1, +1\}^N$. Il y a $\text{card}(\mathcal{E}) = 2^N$ configurations possibles. L'énergie E associée à une configuration de spins particulière x est donnée par l'expression

$$E(x) = -\frac{1}{2} \sum_{i,j=1}^N J_{ij} s_i s_j + h_i s_i, \quad (29)$$

où J_{ij} désigne l'interaction magnétique entre les particules i et j et h_i représente l'influence d'un éventuel champ magnétique extérieur. On a $J_{ii} = 0$ et $J_{ij} = J_{ji}$, le facteur $\frac{1}{2}$ est présent pour indiquer que, d'un point de vue énergétique, on ne compte pas deux fois la contribution de l'interaction entre deux particules distinctes. Si $J_{ij} > 0$ la contribution de l'interaction entre i et j est négative si leurs spins possèdent la même orientation et positive dans le cas contraire. La position : spins parallèles, c'est-à-dire de même orientation, est favorisée dans une situation d'équilibre où l'énergie tend à être minimale. Le ferromagnétisme correspond à $J_{ij} > 0$ et l'antiferromagnétisme à $J_{ij} < 0$, on se place dans le cas du ferromagnétisme. On demande alors l'aimantation totale $m(T) = \sum_{i=1}^N s_i$ de cet ensemble de spins dans une situation d'équilibre.

La loi de Gibbs. La thermodynamique nous apprend, qu'en raison des fluctuations thermiques, un état énergétique $E(x)$ possède une probabilité $g_{E,T}(x)$ donnée par la loi de Gibbs

$$g_{E,T}(x) = \frac{1}{Z_T} \exp\left(-\frac{E(x)}{T}\right), \quad x \in \mathcal{E}, \quad Z_T = \sum_{y \in \mathcal{E}} \exp\left(-\frac{E(y)}{T}\right). \quad (30)$$

La constante de normalisation Z_T est la *fonction de partition*. (On n'a pas fait figurer la constante de Boltzmann k car toutes les quantités physiques sont supposées avoir été normalisées.) Le calcul de Z_T est pratiquement impossible à réaliser même pour un nombre de particules N relativement modeste. Pour $N = 1000$, par exemple, il faut sommer $2^{1000} \approx 10^{301}$ termes afin de calculer Z_T . Il devient alors impraticable d'utiliser la méthode d'inversion ou celle de von Neumann pour simuler le comportement du système. Ces méthodes sont incapables, par exemple, de calculer l'aimantation recherchée. En revanche, la méthode de Metropolis est extrêmement simple à mettre en œuvre. Étant donnée une proposition y suivant la loi $q(y|x)$ on accepte y avec la probabilité

$$\alpha(x, y) = \min\left[\frac{g_{E,T}(y) q(x|y)}{g_{E,T}(x) q(y|x)}, 1\right] = \min\left[\exp\left(-\frac{E(y) - E(x)}{T}\right) \frac{q(x|y)}{q(y|x)}, 1\right]. \quad (31)$$

Le choix de la loi de proposition q détermine une méthode particulière. Afin de simplifier la procédure, on pose souvent $q(x|y) = q(y|x)$. Cette prescription permet de ne pas avoir à préciser la loi de densité q . La méthode qui suit est un exemple de ce type.

Le modèle d'Ising. Dans sa version la plus simple, dite du *bain thermique*, ce modèle consiste à disposer les N particules dans un cube ou un carré suivant un maillage régulier et à ne pas considérer de champ extérieur. On suppose en outre qu'une particule n'interagit qu'avec ses quatre plus proches voisins. Pour éviter les effets de bord, on

considérera le cube ou le carré comme étant le plan fondamental d'un tore plat. L'état initial x_0 des spins est arbitraire. Pour passer de l'état x_n à l'état x_{n+1} on commence par choisir une particule au hasard de façon uniforme et on considère ses quatre plus proches voisins. Soit k l'indice de la particule centrale, les voisins sont situés dans les directions N, E, S et W, et se trouvent à une même distance de la particule k égale à la maille du réseau. (Si la particule i fait partie des plus proches de k on note $i \sim k$.) De ce fait, on ne considérera pas l'effet de la distance dans l'évaluation des éléments J_{ik} . Ils seront tous, par convention, posés égaux à un. L'état énergétique global du système ne varie qu'en fonction de l'état énergétique local de la particule et de ses quatre voisins. On peut poser $E(x) = E_k(x) + E_{-k}(x)$, où $E_k(x)$ et $E_{-k}(x)$ indiquent respectivement la somme des termes $J_{ij}s_i s_j$ où apparaît l'indice k et celle où il n'apparaît pas. Comme seul l'état du spin k est susceptible de changer, on a $E_{-k}(x) = E_{-k}(y)$. Il vient alors

$$\exp\left(-\frac{E(y) - E(x)}{T}\right) = \exp\left(-\frac{E_k(y) - E_k(x)}{T}\right),$$

où E_k désigne l'état énergétique local de la particule k donnée par l'expression

$$E_k = - \sum_{i \sim k} J_{ik} s_i s_k.$$

La proposition y consiste à changer l'état du spin de la particule centrale. Les états x et y ne diffèrent alors que par l'orientation du spin de k . On posera $q(x|y) = q(y|x)$, c'est à dire que la probabilité de changer le spin central est la même qu'il soit dans un sens ou dans un autre alors que rien ne change par ailleurs. On acceptera donc la proposition y avec la probabilité

$$\alpha(x, y) = \min\left[\exp\left(-\frac{E_k(y) - E_k(x)}{T}\right), 1\right].$$

Autrement dit, on accepte de changer le spin de k si $E_k(y) \leq E_k(x)$ et on le change avec la probabilité α dans le cas contraire. Au bout d'un certain nombre de simulations initiales, on converge vers une situation conforme à la loi de Gibbs, et ceci sans avoir eu à calculer la fonction de partition.

Algorithme d'Ising (bain thermique)

0. Une température normalisée T est donnée ainsi qu'une configuration de départ x_0 .
1. À l'étape n on choisit au hasard de façon uniforme un site k et on change son spin. Cette configuration constitue la proposition y .
2. On calcule, pour les configurations x et y , les énergies d'interactions locales $E_k(x)$ et $E_k(y)$ de k avec ses quatre plus proches voisins. On a $E_k(y) = -E_k(x)$.
3. Si $E_k(y) < E_k(x)$ alors $x_{n+1} = y$,
sinon on génère un nombre u issu d'une variable U uniforme entre 0 et 1.
Si $u < \alpha(x_n, y)$ alors $x_{n+1} = y$,
sinon $x_{n+1} = x_n$,
avec : $\alpha(x, y) = \min(\exp(2E_k(x)/T), 1)$.
4. On retourne en 1. pour générer la valeur suivante.

A Appendice

A.1 Solutions stationnaires

On donne la solution stationnaire, quand elle existe, de quelques chaînes de Markov à un nombre fini d'états.

Chaîne à deux états.

$$\pi(0) = \frac{p_{01}}{p_{01} + p_{10}}, \quad (32)$$

$$\pi(1) = \frac{p_{10}}{p_{01} + p_{10}}. \quad (33)$$

Chaîne à trois états.

$$\pi(0) \propto p_{10}p_{20} + p_{21}p_{10} + p_{12}p_{20}, \quad (34)$$

$$\pi(1) \propto p_{01}p_{21} + p_{20}p_{01} + p_{02}p_{21}, \quad (35)$$

$$\pi(2) \propto p_{10}p_{20} + p_{21}p_{10} + p_{12}p_{20}. \quad (36)$$

La constante de normalisation est donnée par la somme des membres de droite, tous positifs, des expressions précédentes.

Chaîne à quatre états.

$$\begin{aligned} \pi(0) \propto & p_{10}p_{20}p_{30} + \\ & p_{10}p_{20}p_{31} + p_{10}p_{20}p_{32} + \\ & p_{10}p_{30}p_{21} + p_{10}p_{30}p_{23} + \\ & p_{20}p_{30}p_{12} + p_{20}p_{30}p_{13} + \end{aligned} \quad (37)$$

$$p_{31}p_{21}p_{10} + p_{23}p_{31}p_{10} + p_{32}p_{21}p_{10} +$$

$$p_{32}p_{12}p_{20} + p_{31}p_{12}p_{20} + p_{13}p_{32}p_{20} +$$

$$p_{13}p_{23}p_{30} + p_{21}p_{13}p_{30} + p_{12}p_{23}p_{30},$$

$$\pi(1) \propto \text{même formule mais en changeant 0 en 1 et 1 en 0}, \quad (38)$$

$$\pi(2) \propto \text{même formule mais en changeant 0 en 2 et 2 en 0}, \quad (39)$$

$$\pi(3) \propto \text{même formule mais en changeant 0 en 3 et 3 en 0}. \quad (40)$$

La constante de normalisation s'obtient comme pour une chaîne à trois états.

Chaîne à m états. On constate que pour $m = 2, 3, 4$, le numérateur de $\pi(j)$ est égal à la somme des poids de tous les arbres du réseau orienté vers l'état racine j et passant par tous les autres états sans former de circuits fermés (spanning trees). D'après la formule de Cayley il y a m^{m-2} arbres de ce type (voir les références [1] §24, [8] et les figures 4, & 5). Le poids d'un arbre orienté est égal au produit des probabilités de transitions de ses branches prises dans le sens allant vers la racine (voir figure 5). Cette formule s'applique quel que soit le cardinal fini de l'espace des états. L'intérêt pratique de cette solution est limitée par l'explosion du nombre de termes qu'elle comporte. Pour $n = 5$, par exemple, le numérateur de $\pi(j)$ comporterait 125 termes et la constante de normalisation 625. Cependant, pour les chaînes de Markov les plus usuelles, de très nombreux termes sont nuls.

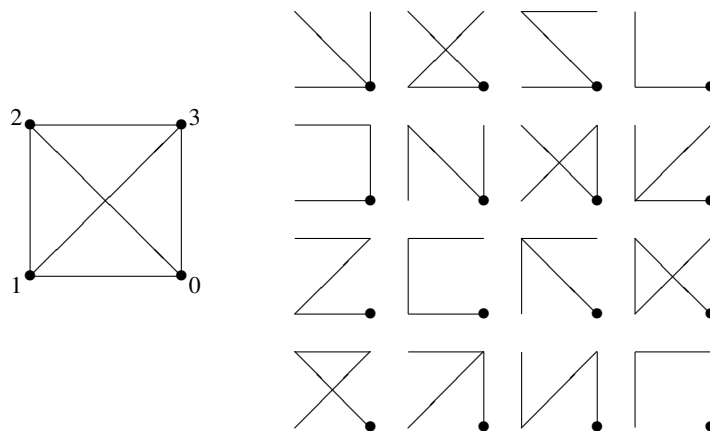


FIGURE 4 – Le graphe complet K_4 et ses 16 “spanning trees”. Ces arbres sont présentés dans l’ordre des termes de l’équation (37) conduisant à l’expression de $\pi(0)$.

Références

- [1] Aigner M., Ziegler G. M., 2001, *Proofs from The Book, Second Edition*, Springer-Verlag, Berlin. ISBN 3-540-67865-4
- [2] Barker A. A., 1965, *Monte-Carlo calculations of the radial distribution for a proton-electron plasma*, *Aus. J. Phys.*, **18**, 119–133
- [3] Fishman G. S., 1996, *Monte Carlo, Concepts, Algorithms, and Applications*, Springer-Verlag, New-York. ISBN 0-387-94527-X
- [4] Graham C., 2008, *Chaînes de Markov*, Dunod, Paris. ISBN 978-2-10-052083-1
- [5] Hastings W., 1970, *Monte Carlo sampling methods using Markov chains and their application*, *Biometrika*, **57**, 97–109
- [6] Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., Teller E., 1953, *Equation of state calculations by fast computing machines*, *J. Chem. Phys.*, **21**, 1087–1092
- [7] Robert C. P., Casella G., 2011, *Méthodes de Monte-Carlo avec R*, Springer France. ISBN 978-2-8178-0180-3
- [8] Trent H., 1954, *A note on the enumeration and listing of all possible trees in a connected linear graph*, *Proc. Nat. Acad. Sci. U.S.A.* **40**, 1004-1007
- [9] von Neumann J., 1951, *Various techniques used in connection with random digits*, *Monte Carlo Methods, Applied Mathematics Serie 12*, National Bureau of Standards, Washington D.C.

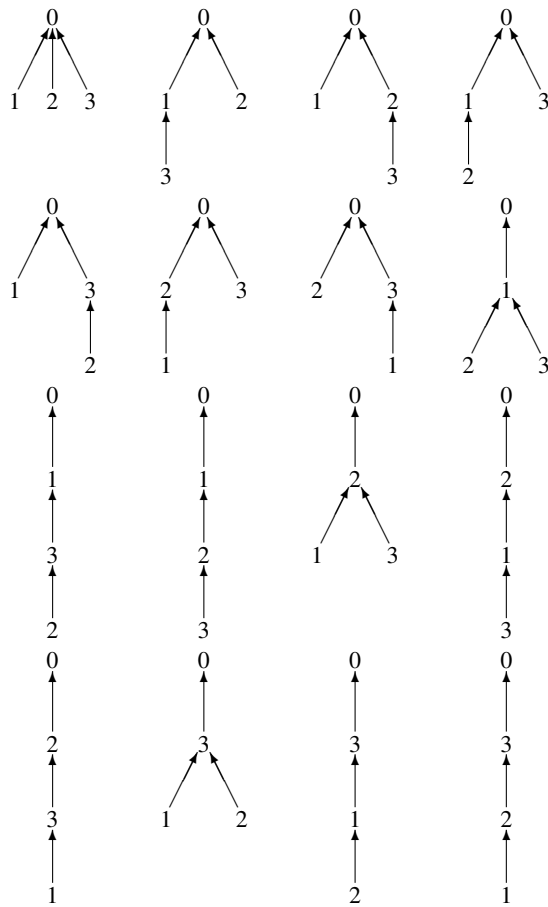


FIGURE 5 – Représentation, sous forme d’arbres orientés, de la partie droite de la figure 4. La racine est l’état 0. Le poids du premier arbre est $p_{10}p_{20}p_{30}$, celui du deuxième $p_{10}p_{20}p_{31}$ et ainsi de suite.