

Compte rendu stage Simulation parfaite

Antonin LAURENT

6 mai 2019

Table des matières

1	Chapitre 1	3
1.1	Définitions et théorèmes	3
1.2	Modèles étudiés	7
1.2.1	Champs aléatoires de Markov	7
1.2.2	Permutations	8
1.3	Chaînes de Markov et simulation approchée	9
1.4	Création de chaînes de Markov	10
1.5	Échantillonnage de Gibbs / Gibbs Sampler	11
1.6	Metropolis-Hastings	13
1.7	Variables aléatoires auxiliaires / Auxiliary random variables .	13
2	Chapitre 3	14

1 Chapitre 1

1.1 Définitions et théorèmes

Définition : Fonction calculable / Computable function

Une fonction est dite calculable (computable en anglais) si il existe un algorithme capable de retourner le résultat de la fonction.

Définition : Algorithme probabiliste / Randomized algorithm

Soient \mathcal{I} un ensemble d'indices tel que pour tout $I \in \mathcal{I}$, il existe une distribution π_I sur un espace d'états Ω_I . On se donne une suite de variables aléatoires X_1, X_2, \dots où les $X_I \in S_I$.

Un algorithme probabiliste est une famille de temps d'arrêts $\{T_I\}_{I \in \mathcal{I}}$ et de fonctions calculables $\{f_{I,t}\}_{I \in \mathcal{I}, t \in \{1,2,\dots\}}$. La sortie de l'algorithme est $f_{I,T_I}(X_1, \dots, X_{T_I})$.

Définition : Algorithme de simulation parfaite / Perfect simulation algorithm

Un algorithme de simulation parfaite est un algorithme probabiliste dont la sortie est une variable aléatoire qui provient d'une distribution cible. Les algorithmes de simulation parfaite sont une sous-classe des algorithmes de simulation exacte, algorithmes qui tirent d'une distribution ciblée. Cependant les algorithmes dont le temps d'arrêt T_I est déterministe (algorithmes tournant selon un nombre fini de choix aléatoires) sont généralement considérés comme algorithmes de simulation exacte mais pas comme algorithmes de simulation parfaite.

Théorème fondamental de la simulation parfaite

On suppose que pour U_1, U_2, \dots iid tel que $U_i \sim \text{Unif}([0,1])$, il existe des fonctions calculables b, g et f telles que la fonction b aie pour image $\{0,1\}$ et $\mathbb{P}(b(U) = 1) < 1$.

Pour une variable aléatoire X qui vérifie :

$$X \sim b(U)g(U) + (1 - b(U))f(X, U), \quad (1)$$

soit $T = \inf\{t : b(U_t) = 1\}$. On a alors que :

$$Y = f(\dots f(f(g(U_T), U_{T-1}), U_{T-2}), \dots, U_1)$$

a la même distribution que X et on a $\mathbb{E}[T] = \frac{1}{\mathbb{P}(b(U)=1)}$.

Preuve :

Soient X_0, X_1, \dots des tirages indépendants chacun distribués selon X , et U_1, U_2, \dots iid tel que $U_i \sim \text{Unif}([0,1])$. Pour X_t , fixons $X_{t,t} = X_t$ et récursivement, on pose :

$$X_{t,i} = b(U_{i+1})g(U_{i+1}) + (1 - b(U_{i+1}))f(X_{t,i+1}, U_{i+1}),$$

pour $i \in \{0, \dots, t-1\}$.

On a alors d'après la relation (1) : $X_{t,0} \sim X$. On montre ce résultat pour les premiers indices t , les suivants sont prouvés de la même manière.

t=0

$$X_{0,0} = X_0$$

comme on l'a posé précédemment, d'où

$$X_{0,0} \sim X.$$

t=1

$$X_{1,0} = b(U_1)g(U_1) + (1 - b(U_1))f(X_{1,1}, U_1)$$

Or,

$$X_{1,1} = X_1 \sim X.$$

En remplaçant dans l'équation précédente, on obtient :

$$X_{1,0} = b(U_1)g(U_1) + (1 - b(U_1))f(X_1, U_1)$$

D'après la relation (1), on obtient alors $X_{1,0} = X_1 \sim X$.

t=2

$$\begin{aligned} X_{2,0} &= b(U_1)g(U_1) + (1 - b(U_1))f(X_{2,1}, U_1) \\ X_{2,1} &= b(U_2)g(U_2) + (1 - b(U_2))f(X_{2,2}, U_2) \\ &= b(U_2)g(U_2) + (1 - b(U_2))f(X_2, U_2) \end{aligned}$$

À nouveau, par la relation (1), on a : $X_{2,1} = X_2$ On remplace donc par cette valeur dans $X_{2,0}$ et on obtient :

$$X_{2,0} = b(U_1)g(U_1) + (1 - b(U_1))f(X_2, U_1)$$

D'après la relation (1), on obtient alors $X_{2,0} = X_2 \sim X$.

On a donc $X_{0,0}, X_{1,0}, X_{2,0}, \dots$ de même distribution que X mais pas forcément indépendants. On considère à présent la variable Y énoncée dans le théorème. On a alors la relation suivante : $X_{t,0} = Y$ si $t \geq T$. On illustre ce résultat avec comme exemple $T = 2$ et $t = 2, 3$:

On a donc :

$$Y = f(g(U_2), U_1)$$

Montrons que $X_{2,0} = X_{3,0} = Y$. Étant donné que $T = 2$, on a $b(U_1) = 0$ et $b(U_2) = 1$, d'où :

$$\begin{aligned} X_{2,0} &= f(X_{2,1}, U_1) \\ X_{2,1} &= g(U_2) \end{aligned}$$

on remplace dans la première équation et on obtient

$$X_{2,0} = f(g(U_2), U_1)$$

D'où le résultat pour t=2. Voyons pour t=3.

De même, puisque $T = 2$, on a $b(U_1) = 0$ et $b(U_2) = 1$, d'où :

$$\begin{aligned} X_{3,0} &= f(X_{3,1}, U_1) \\ X_{3,1} &= g(U_2) \end{aligned}$$

on remplace dans la première équation et on obtient

$$X_{3,0} = f(g(U_2), U_1)$$

D'où le résultat.

Ensuite, puisque les U_i sont indépendants, on a que : $\mathbb{P}(T > t) = (1 - \mathbb{P}(b(U) = 1))^t$ et puisque, par hypothèse, $\mathbb{P}(b(U) = 1) > 0$, on a la relation qui tend vers 0 pour t tendant vers l'infini. Il ne reste plus qu'à montrer $Y \sim X$. Pour tout t , pour tout ensemble C :

$$\mathbb{P}(Y \in C) = \mathbb{P}(Y \in C, t \geq T) + \mathbb{P}(Y \in C, t < T)$$

puisque $X_{t,0} = Y$ si $t \geq T$, on a

$$\begin{aligned} &= \mathbb{P}(X_{t,0} \in C, t \geq T) + \mathbb{P}(Y \in C, t < T) \\ &= \mathbb{P}(X_{t,0} \in C) - \mathbb{P}(X_{t,0} \in C, t < T) + \mathbb{P}(Y \in C, t < T) \end{aligned}$$

puisque $X_{t,0} \sim X$, on a

$$= \mathbb{P}(X \in C) - \mathbb{P}(X_{t,0} \in C, t < T) + \mathbb{P}(Y \in C, t < T)$$

Les deux derniers termes sont bornés par $\mathbb{P}(T > t) = (1 - \mathbb{P}(b(U) = 1))^t$, et puisque l'équation est vraie pour n'importe quel t , on obtient : $\mathbb{P}(Y \in C) = \mathbb{P}(X \in C)$ pour tout ensemble C , donc $Y \sim X$. Le fait que $\mathbb{E}[T] = \frac{1}{\mathbb{P}(b(U)=1)}$ provient du fait que T suit une loi géométrique de paramètre $\mathbb{P}(b(U) = 1)$

Définition : Algorithme arretable / Interruptible algorithm

Dans le théorème fondamental de la simulation parfaite, si X et T sont indépendants, on dit que l'algorithme est arretable, sinon il est non-arretable.

Définition : Algorithme de simulation parfaite à lecture unique / Read once perfect simulation algorithm

Un algorithme de simulation parfaite qui utilise $X \sim b(U)g(U) + (1 - b(U))f(X, U)$ est un algorithme à lecture unique si $f(X, u) = f(X, u')$ pour tout u, u' . Sinon, c'est un algorithme à lecture double.

En général, un algorithme arretable est préférable à un algorithme non-arretable, et un algorithme à lecture unique est préférable à un algorithme à lecture double.

1.2 Modèles étudiés

1.2.1 Champs aléatoires de Markov

On considère un graphe $G = (V, E)$. V est l'ensemble des nœuds et E est l'ensemble des arêtes. On notera par la suite Δ le degré du graphe (égal au degré du nœud ayant un nombre de voisins maximal). On notera par Ω l'ensemble des labels des nœuds.

Définition : Ensemble séparant / Separating set

On dit qu'un sous-ensemble de nœuds S sépare les nœuds i et j si tout chemin du graphe menant i à j passe par S .

Définition : Champ aléatoire de Markov / Markov random field

Pour un graphe $G = (V, E)$ et un ensemble de labels Ω , on dit que la distribution de X sur Ω est un champ aléatoire de Markov si, pour tous les nœuds i et j , pour tous les ensembles S séparant i de j , on a : $[X(i)|X(j), X(S)] \sim [X(i)|X(S)]$. Un état $x \in \Omega$ est appelé une configuration.

Définition : Clique

Une clique est un sous-ensemble de nœuds du graphe tel que chaque paire de nœuds soit connectée par une arête.

Théorème de Hammersley Clifford

Pour un graphe fini $G = (V, E)$, la distribution π est un champ aléatoire de Markov si elle a pour densité f_X et qu'il existe des fonctions ϕ_C pour toute clique C telles que f_X puisse s'écrire :

$$f_X(x) = \frac{1}{Z} \prod_{C \in \text{cliques}(G)} \phi_C(x)$$

Définition : Auto-modèle / Auto-model

On dit qu'un champ aléatoire de Markov est un auto-modèle, si il existe des fonctions f_i et g_{ij} telles que la densité de $X \sim \pi$ peut-être écrite sous la forme :

$$f_X(x) = \frac{1}{Z} \left[\prod_{i \in V} f_i(X(i)) \right] \left[\prod_{\{i,j\} \in E} g_{\{i,j\}}(X(i), X(j)) \right]$$

Dans la suite, on définit des exemples bien connus d'auto-modèles que l'on étudiera par la suite.

Exemple 1 : Modèle d'Ising

Le modèle d'Ising est un auto-modèle de paramètres μ et β , où $\Omega = \{-1, 1\}^V$, et $f(c) = \exp(\mu c)$, $g(c_1, c_2) = \exp(\beta \mathbf{1}(c_1 = c_2))$. Dans la littérature, on appelle μ magnétisation et β la température inverse. Lorsque $\beta > 0$, on dit que le modèle est ferromagnétique. Lorsque $\mu = 0$, on dit que le modèle est sans champ magnétique.

Exemple 2 : Modèle Hard-core gas

Le modèle hard-core gas est un auto-modèle défini sur $\{0, 1\}^V$ et de paramètre $\lambda > 0$ où $f(c) = \lambda^c$ et $g(c_1, c_2) = 1 - c_1 c_2$. Lorsqu'un nœud ν a pour label 1, on dit que le nœud est occupé, sinon, il est inoccupé.

Exemple 3 : Modèle de Strauss

Le modèle de Strauss est un auto-modèle sur $\{0, 1\}^V$ de paramètres $\lambda > 0$ et $\gamma \in [0, 1]$ où $f(c) = \lambda^c$ et $g(c_1, c_2) = 1 + (\gamma - 1)c_1 c_2$. Lorsqu'un nœud ν a pour label 1, on dit que le nœud est occupé, sinon, il est inoccupé.

1.2.2 Permutations

Un problème classique de distribution sur les permutations est lié à la recherche du permanent d'une matrice non-négative.

Exemple

Pour une matrice non-négative $w(i, j)$, notons :

$$w(\sigma) = \prod_{i=1}^n w(i, \sigma(i))$$

Tant qu'il existe au moins une permutation σ telle que $w(i, \sigma(i)) > 0$ pour tout i , cela donne une densité non normalisée sur l'ensemble des permutations. La constante de normalisation pour cette densité est alors appelée le permanent de la matrice $w(i, j)$. Si aucun σ ne vérifie cette relation, le permanent est 0.

Une autre distribution importante que l'on va considérer est la distribution uniforme sur les permutations où certains objets doivent avoir une position plus faible que les autres.

Définition : relation d'ordre partiel / partial order

Soit un ensemble P . Une relation d'ordre partiel sur P est une relation binaire \preceq telle que pour tout $a, b, c \in P$, la relation est :

1. (Réflexive) $a \preceq a$
2. (Antisymétrique) Si $a \preceq b$ et $b \preceq a$, alors $a = b$
3. (Transitive) Si $a \preceq b$ et $b \preceq c$, alors $a \preceq c$

Un ensemble disposant d'une relation d'ordre partiel est parfois appelé poset d'après l'anglais partially ordered set.

Définition : Extension linéaire d'un ordre partiel / Linear extension of a partial order

Une extension linéaire d'un ordre partiel sur $1, \dots, n$ est une permutation pour laquelle si i et j sont tels que $\sigma(i) \prec \sigma(j)$, alors $i < j$.

1.3 Chaînes de Markov et simulation approchée

Jusqu'au développement des algorithmes de simulation parfaite/exacte, la manière principale d'obtenir une réalisation d'une distribution ciblée était une méthode d'approche. Plusieurs algorithmes et méthodes ont été mis en place et réfère à l'ensemble de ces méthodes sous le nom de Chaîne de Markov Monte Carlo (Markov Chain Monte Carlo, MCMC dans la littérature).

On ne rappellera pas ici les principales définitions pour les chaînes de Markov mais d'autres seront nécessaires pour la suite. On s'intéressera notamment aux chaînes de Harris et au théorème ergodique associé.

Définition : Chaîne de Harris / Harris Chain

Une chaîne de Markov $\{X_t\}$ sur un espace d'état Ω est une chaîne de Harris si il existe des ensembles mesurables $A, B \in \Omega$ et $\epsilon > 0$ pour $x \in A$ et $y \in B$, et une mesure de probabilité ρ où $\rho(B) = 1$ tels que l'on ait :

1. Pour $T_A = \inf \{t \geq 0 : X_t \in A\}$, $(\forall z \in \Omega)(\mathbb{P}(T_A < \infty | X_0 = z) > 0)$.
2. Si $x \in A$ et $C \subseteq B$, alors $\mathbb{P}(X_1 \in C | X_0 = x) \geq \epsilon \rho(C)$

Définition : Chaîne récurrente / Recurrent chain

Soit $R = \inf \{n > 0 : X_n \in A\}$. On dit qu'une chaîne de Harris est une chaîne récurrente si pour tout $x \in A$, $\mathbb{P}(R < \infty | X_0 = x) = 1$. Une chaîne qui n'est pas récurrente est dite transiente.

Définition : Chaîne apériodique / Aperiodic chain

Une chaîne de Harris récurrente est apériodique si pour tout $x \in \Omega$, il existe n tel que pour tout $n' \geq n$, $\mathbb{P}(X_{n'} \in A | X_0 = x) > 0$

Théorème : Théorème ergodique pour les chaînes de Harris

Soit X_n une chaîne de Harris récurrente et apériodique de distribution stationnaire π . Si $\mathbb{P}(R < \infty | X_0 = x) = 1$ pour tout x , alors, pour $t \rightarrow \infty$, pour tout ensemble mesurable C et pour tout x :

$$|\mathbb{P}(X_t \in C | X_0 = x) - \pi(C)| \rightarrow 0$$

Ce théorème est le cœur des méthodes MCMC, puisqu'il "suffit" de construire une chaîne de Harris ayant pour distribution stationnaire la distribution souhaitée et de la faire tourner pendant un nombre infini de pas. Cependant, n'ayant pas un temps infini, les utilisateurs font tourner leurs algorithmes durant un grand nombre de pas et espèrent arriver dans la distribution stationnaire.

On peut cependant déterminer à quel point la chaîne est proche de la loi stationnaire à l'aide du concept de couplage (voir chapitre 3 pour une définition et application étendue).

Définition : Couplage / Coupling

Supposons que $\{X_t\} \sim \nu_X$ et $\{Y_t\} \sim \nu_Y$. Un couplage de $\{X_t\}$ et $\{Y_t\}$ est un processus bivarié $\{(X'_t, Y'_t)\}$ tel que $\{X'_t\} \sim \nu_X$ et $\{Y'_t\} \sim \nu_Y$.

Théorème : Lemme de Couplage / Coupling Lemma

Soit $Y_0 \sim \pi$ et $X_0 = x_0$ tels que les deux variables évoluent de manière couplée. Alors, pour tout mesurable C :

$$|\mathbb{P}(X_t \in C | X_0 = x) - \pi(C)| \leq \mathbb{P}(X_t \neq Y_t).$$

1.4 Création de chaînes de Markov

Afin d'utiliser les méthodes MCMC, il faut créer des chaînes de Harris qui convergent vers la distribution π ciblée. Il est généralement mieux de créer des chaînes plus que stationnaire : les chaînes réversibles.

On utilisera la notation suivante pour la définition de réversibilité : $\pi(dx) = f(x)dx$. Et donc, pour tout mesurable A , $\pi(A) = \int_{x \in A} \pi(dx) = \int_{x \in A} f(x)dx$. On obtient alors la définition de réversibilité suivante :

Définition : Chaîne réversible

Une distribution π est réversible selon une chaîne de Markov $\{X_t\}$ si :
 $\pi(dx)\mathbb{P}(X_{t+1} \in dy|X_t = x) = \pi(dy)\mathbb{P}(X_{t+1} \in dx|X_t = y)$.

Lemme

Si π est réversible, alors π est stationnaire.

Preuve

Soit Ω l'espace d'état de la chaîne de Markov $\{X_t\}$ considérée et π réversible pour cette chaîne. Pour $X_t \sim \pi$, et C un ensemble mesurable, alors on a :

$$\begin{aligned}\mathbb{P}(X_{t+1} \in C) &= \mathbb{E}[\mathbf{1}(X_{t+1} \in C)] \\ &= \mathbb{E}[\mathbb{E}[\mathbf{1}(X_{t+1} \in C)|X_t]] \\ &= \int_{x \in \Omega} \mathbb{E}[\mathbf{1}(X_{t+1} \in C)|X_t = x]\pi(dx) \\ &= \int_{x \in \Omega} \mathbb{P}(X_{t+1} \in C|X_t = x)\pi(dx) \\ &= \int_{x \in \Omega} \int_{y \in C} \mathbb{P}(X_{t+1} \in dy|X_t = x)\pi(dx) \\ &= \int_{y \in C} \int_{x \in \Omega} \mathbb{P}(X_{t+1} \in dx|X_t = y)\pi(dy) \\ &= \int_{y \in C} \mathbb{P}(X_{t+1} \in \Omega|X_t = y)\pi(dy) \\ &= \int_{y \in C} \pi(dy) \\ &= \pi(C)\end{aligned}$$

D'où la stationnarité de π .

Plusieurs types de chaînes réversibles existent tels que l'échantillonnage de Gibbs (Gibbs sampler), Metropolis-Hastings, etc.
Ces chaînes sont présentées dans la suite.

1.5 Échantillonnage de Gibbs / Gibbs Sampler

On présente ici l'un des plus simples échantillonneurs de Gibbs, qui agit sur un espace d'états de la forme C^V . On appelle $\nu \in V$ une dimension du problème. Pour $X_t = x$, l'échantillonneur choisit une dimension ν uniformément sur V . Soit $L(x, \nu)$ l'ensemble des états qui sont exactement

les nœuds de la configuration x sauf en ν , écrit de la manière suivante : $L(x, v) = \{y : (\forall w \in V \setminus \{v\})(y(w) = x(w))\}$. Pour $X_t = x$, l'état suivant X_{t+1} est choisi selon π à conditionné à être dans $L(x, v)$.

On présente comme exemple le modèle d'Ising précédemment vu avec $\mu = 0$. Une dimension est alors un nœud de la configuration. On choisit donc un nœud uniformément dans V , et on considère les états qui sont exactement x en tous les autres nœuds autres que ν . Pour le modèle d'Ising, la valeur en ν de x est 1 ou -1 . On note ces configurations $x_{\nu \rightarrow 1}$ et $x_{\nu \rightarrow -1}$. On choisit alors l'état suivant entre $x_{\nu \rightarrow 1}$ et $x_{\nu \rightarrow -1}$, où le choix est fait proportionnellement à π . On a donc :

$$\mathbb{P}(X_{t+1} = x_{\nu \rightarrow 1}) = \frac{\pi(\{x_{\nu \rightarrow 1}\})}{\pi(\{x_{\nu \rightarrow 1}\}) + \pi(\{x_{\nu \rightarrow -1}\})}$$

Or, pour le modèle d'Ising avec $\mu = 0$,

$$\pi(\{x\}) = \frac{1}{Z} \prod_{\{i,j\} \in E} \exp(\beta \mathbf{1}(x(i) = x(j))).$$

Après simplification, on obtient :

$$\begin{aligned} \mathbb{P}(X_{t+1} = x_{\nu \rightarrow 1}) &= \frac{\prod_{\{\nu,j\} \in E} \exp(\beta \mathbf{1}(x(j) = 1))}{\prod_{\{\nu,j\} \in E} \exp(\beta \mathbf{1}(x(j) = 1)) + \prod_{\{\nu,j\} \in E} \exp(\beta \mathbf{1}(x(j) = -1))} \\ &= \frac{\exp(\beta n_1)}{\exp(\beta n_1) + \exp(\beta n_{-1})} \end{aligned}$$

Où n_c est le nombre de voisins de ν de label c . Par exemple, si ν est adjacent à trois nœuds de label 1 et un nœud de label -1 , alors la probabilité que ν se voit labelliser 1 est $\exp(3\beta)/(\exp(3\beta) + \exp(\beta))$. Mettre ensuite en place cette méthode algorithmiquement est très simple.

Vérifions la réversibilité de la chaîne.

— Premier cas : $X_t = x_{\nu \rightarrow 1}$

— Premier sous-cas : $X_{t+1} = x_{\nu \rightarrow 1}$. Les termes de l'égalité de balance détaillée sont les mêmes, on a donc trivialement l'égalité.

— Deuxième sous-cas : $X_{t+1} = x_{\nu \rightarrow -1}$. On a alors :

$$\pi(\{x_{\nu \rightarrow 1}\})\mathbb{P}(X_{t+1} = x_{\nu \rightarrow -1} | X_t = x_{\nu \rightarrow 1}) = \exp(\beta n_1) \frac{\exp(\beta n_{-1})}{\exp(\beta n_{-1}) + \exp(\beta n_1)}$$

D'autre part :

$$\pi(\{x_{\nu \rightarrow -1}\})\mathbb{P}(X_{t+1} = x_{\nu \rightarrow 1} | X_t = x_{\nu \rightarrow -1}) = \exp(\beta n_{-1}) \frac{\exp(\beta n_1)}{\exp(\beta n_{-1}) + \exp(\beta n_1)}$$

On a donc l'égalité entre ces termes.

- Deuxième cas : $X_t = x_{\nu \rightarrow -1}$
- Premier sous-cas : $X_{t+1} = x_{\nu \rightarrow -1}$. Les termes de l'égalité de balance détaillée sont les mêmes, on a donc trivialement l'égalité.
- Deuxième sous-cas : $X_{t+1} = x_{\nu \rightarrow 1}$. On a alors :

$$\pi(\{x_{\nu \rightarrow -1}\})\mathbb{P}(X_{t+1} = x_{\nu \rightarrow 1} | X_t = x_{\nu \rightarrow -1}) = \exp(\beta n_{-1}) \frac{\exp(\beta n_1)}{\exp(\beta n_{-1}) + \exp(\beta n_1)}$$

D'autre part :

$$\pi(\{x_{\nu \rightarrow 1}\})\mathbb{P}(X_{t+1} = x_{\nu \rightarrow -1} | X_t = x_{\nu \rightarrow 1}) = \exp(\beta n_1) \frac{\exp(\beta n_{-1})}{\exp(\beta n_{-1}) + \exp(\beta n_1)}$$

On a donc bien l'égalité entre ces termes.

Ayant donc vérifié tous les cas possibles, on obtient donc l'équation de balance détaillée qui conduit bien à la réversibilité de la chaîne.

1.6 Metropolis-Hastings

Pour la méthode de Metropolis-Hastings, pour chaque configuration x de la chaîne, on a une densité q_x . On propose ensuite un état suivant y , selon la densité q_x . L'état suivant sera donc x ou y . La probabilité avec laquelle est choisie la configuration y est donnée de telle sorte que la réversibilité s'applique. On la donne ci-dessous. Notez que, pour que cela fonctionne, il est nécessaire que si $q_x(y) > 0$, alors $q_y(x) > 0$ aussi. On a alors le changement vers y avec probabilité :

$$\min \left\{ 1, \frac{f_\pi(y)q_y(x)}{f_\pi(x)q_x(y)} \right\}$$

On notera aussi que f_π peut-être normalisée ou non.

En exemple, on considère à nouveau le modèle d'Ising avec $\mu = 0$. A chaque pas, on choisit un nœud uniformément dans V puis un label candidat pour ce nœud est choisi uniformément dans $\{-1, 1\}$. On a alors $q_x(y) = q_y(x) = 1/2$. On calcule ensuite, pour n_c le nombre de voisins de ν de label c (où c est le label proposé) et $n_{x(\nu)}$ le nombre de voisins de ν de même label que ν :

$$\frac{f_\pi(y)}{f_\pi(x)} = \frac{\exp(n_c \beta)}{\exp(n_{x(\nu)} \beta)}$$

1.7 Variables aléatoires auxiliaires / Auxiliary random variables

Dans la plupart des applications du MCMC, la densité ciblée X a une structure multiplicative. Pour ces densités, il est possible d'ajouter un vecteur Y de variables aléatoires supplémentaires telles que (X, Y) a une distribution jointe plus simple. Par exemple, on considère à nouveau le modèle

d'Ising avec $\mu = 0$. Pour chaque arête $\{i, j\}$, on crée une variable aléatoire auxiliaire $Y(\{i, j\})$ telle que sa distribution conditionnée sur X soit uniforme sur $[0, 1]$ si $X(i) \neq X(j)$ et uniforme sur $[0, \exp(\beta)]$ si on a l'égalité.

La densité jointe est alors uniforme sur :

$$X \in \{0, 1\}^V \text{ et } Y \in \{[0, \infty) : (\forall \{i, j\})(y\{i, j\}) \leq \min(\exp(\beta), 1))\}$$

2 Chapitre 3

Définition : Fonction de mise à jour / Update function

On dit que $\phi : \Omega \times [0, 1] \rightarrow \Omega$ est une fonction de mise à jour pour une chaîne de Markov $\{X_t\}$ si, pour $U \sim \text{Unif}[0, 1]$, $[X_{t+1}|X_t] \sim \phi(X_t, U)$.

La fonction ϕ est déterministe : tout l'aléatoire est contenu dans la variable U .

Toute chaîne qui peut-être simulée sur ordinateur est un exemple de fonction de mise à jour.

Une même chaîne de Markov peut-être représentée par plusieurs fonctions de mise à jour : la fonction de mise à jour n'est pas forcément unique.

À l'aide d'une fonction de mise à jour ϕ , on peut représenter la trajectoire d'une chaîne de Markov $\{X_t\}$. En effet, soient U_0, U_1, U_2, \dots iid $\sim \text{Unif}([0, 1])$. Pour un état initial x_0 , on a : $X_1 = \phi(x_0, U_0)$ puis pour $i > 1$, $X_i = \phi(X_{i-1}, U_{i-1})$.

On notera alors la trajectoire jusqu'au temps t sous la forme :

$$\phi_t(x_0, U) = \phi(\phi(\phi(\dots(\phi(x_0, U_0), U_1), \dots, U_{t-1}))$$

Ensuite, pour n'importe quels états x_0 et y_0 dans Ω , on définit pour une fonction de mise à jour $\phi : X_t = \phi_t(x_0, U)$ et $Y_t = \phi_t(y_0, U)$ (en utilisant les mêmes valeurs de U). On appelle ce procédé un couplage. Notons qu'avec ce couplage, si il existe $t \geq 0$ tel que $X_t = Y_t$, alors on dit que les processus ont fusionné (ou se sont rejoints, etc).

Définition : Couplage / Coupling

Soit \mathcal{S} un ensemble de processus stochastiques définis sur un même ensemble \mathcal{I} et un même espace d'états Ω . Si il existe un indice $i \in \mathcal{I}$ et un état $x \in \Omega$ tels que pour tout $S \in \mathcal{S}$, on aie $S_i = x$, alors on dit que les processus stochastiques ont fusionné (ou se sont rejoints, couplés, etc).

Exemple

Soit $\Omega = \{0, 1, 2\}$. Soit une fonction de mise à jour ϕ telle que :

$$\phi(x, U) = x + \mathbf{1}(x < 2, U > \frac{1}{2}) - \mathbf{1}(x > 0, U \leq \frac{1}{2}).$$

Soient $\{X_t\}$ et $\{Y_t\}$ deux chaînes de Markov ayant ϕ comme fonction de mise à jour et telles que $X_0 = 0$ et $Y_0 = 2$. On suppose que $U_0 = 0.64, U_1 = 0.234$ et $U_2 = 0.1$. On a donc les trajectoires suivantes : $(X_0, X_1, X_2, X_3) = (0, 1, 0, 0)$ et $(Y_0, Y_1, Y_2, Y_3) = (2, 2, 1, 0)$.

Les deux chaînes ont donc fusionné au temps $t = 3$.

Théorème CFTP (Coupling From The Past)

On suppose que ϕ est une fonction de mise à jour pour une chaîne de Markov définie sur Ω , telle que pour $U = (U_0, U_{-1}, \dots, U_{-t-1}) \sim \text{Unif}([0, 1]^t)$ on ait :

- Pour $Y \sim \pi, \phi(Y, U_0) \sim \pi$
- Il existe un ensemble $A \subseteq [0, 1]^t$ tel que $\mathbb{P}(U \in A) > 0$ et $\phi_t(\Omega, A) = \{x\}$ pour un certain $x \in \Omega$

Posons alors, pour tout $x \in \Omega$,

$$Y_0 = \phi_t(x_0, U_0)\mathbf{1}(U_0 \in A) + \phi_t(\phi_t(x_0, U_{-1}), U_0)\mathbf{1}(U_{-1} \in A) + \\ \phi_t(\phi_t(\phi_t(x_0, U_{-2}), U_{-1}), U_0)\mathbf{1}(U_{-2} \in A) + \dots$$

Alors $Y_0 \sim \pi$.

Preuve

Soit $x_0 \in \Omega$. Le résultat est immédiat en utilisant le théorème fondamental de la simulation parfaite, en posant $g(U) = \phi_t(x_0, U), b(U) = \mathbf{1}(U \in A)$, et $f(X, U) = \phi_t(X, U)$.