

Cours 6

5.4 Echantillonnage de Gibbs

Introduit par Geman et Geman (1984) dans le cadre de la restauration d'images, cet algorithme peut être vu comme une extension de l'algorithme de Tanner et Wong décrit précédemment. Le principe repose encore sur une décomposition du problème générale (simuler suivant une certaine loi) en une série de problèmes élémentaires (simuler suivant des lois conditionnelles). Considérons la densité $f(x, y_1, \dots, y_p)$. On s'intéresse à la loi marginale :

$$f(x) = \int \cdots \int f(x, y_1, \dots, y_p) dy_1 \cdots dy_p.$$

En particulier, on souhaite obtenir son espérance mathématique et sa variance. On se place ici dans le cas où les intégrations impliquées dans le calcul de la marginale sont compliquées et difficiles à effectuer même numériquement. On supposera cependant que les densités conditionnelles sont disponibles.

L'échantillonneur de Gibbs va nous permettre de générer x suivant $f(x)$ sans utiliser directement son expression supposée difficile à manipuler, mais en utilisant les densités conditionnelles.

Ainsi fabriquant un échantillon (x_1, \dots, x_m) assez grand on pourra approximer la moyenne, la variance et autre caractéristique en utilisant une loi des grands nombres :

$$\lim_{m \rightarrow +\infty} \frac{1}{m} \sum_{i=1}^m g(x_i) = E[g(x)].$$

Principe de l'algorithme : Considérons le cas élémentaire : $f(x, y)$. On suppose $f(x | y)$ et $f(y | x)$ disponibles. On peut alors générer ce qu'on appellera une séquence de Gibbs de la manière suivante : partant d'une valeur x_0 , on génère y_0 avec $\pi(\cdot | x_0)$, puis x_1 avec $\pi(\cdot | y_0)$, puis y_1 avec $\pi(\cdot | x_1)$ et ainsi de suite.

Après M itérations de ce schéma, il vient une séquence : $(x_0, y_0, x_1, y_1, \dots, x_M, y_M)$.

Pour M assez grand, x_M est une réalisation de X .

Exemple : Considérons une loi jointe de la forme :

$$f(x, y) \propto C_n^x y^{x+\alpha-1} (1-y)^{n-x+\beta-1}, \quad x = 0, \dots, n, \quad 0 \leq y \leq 1.$$

On remarque que l'on peut proposer une loi binomiale de paramètres (n, y) pour $f(x | y)$ et une loi bêta de paramètres $(x + \alpha, n - x + \beta)$ pour $f(y | x)$.

On peut donc appliquer un algorithme de Gibbs pour obtenir des réalisations de $f(x)$, la loi marginale de x en simulant alternativement une réalisation x^* d'une binomiale de paramètres (n, y^*) où y^* est la valeur courante de y obtenu à l'étape précédente, puis une nouvelle réalisation

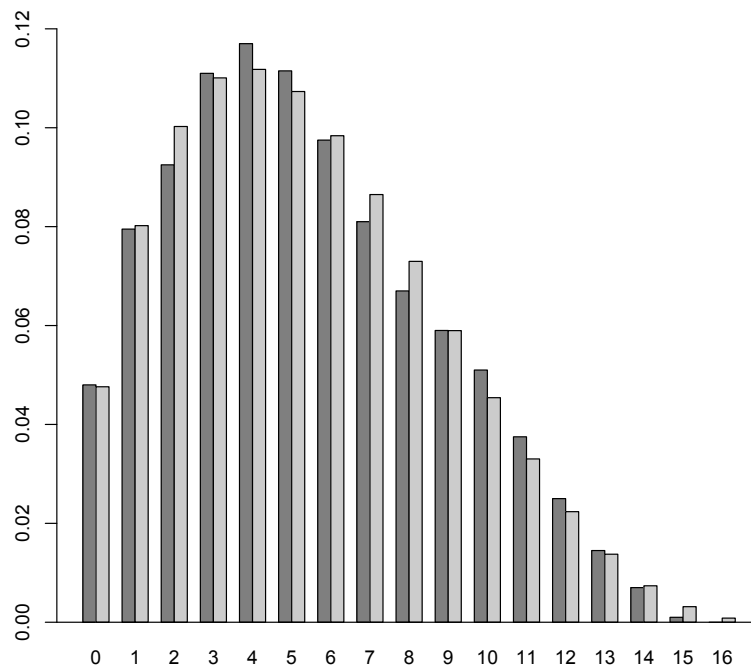
x d'une bêta de paramètre $(x^* + \alpha, n - x^*)$ (Voir TD 4, Exercice 4).

Il se trouve qu'ici la loi marginale est accessible. En effet,

$$f(x) \propto \int_0^1 f(x, y) dy \propto \int_0^1 C_n^x y^{x+\alpha-1} (1-y)^{n-x+\beta-1} dy \propto C_n^x \frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(\alpha+\beta+n)}.$$

La figure (1) donne une représentation de l'histogramme d'un échantillon de taille 2000 simulé par l'algorithme de Gibbs et de la "vraie" densité $f(x)$ qui est une bêta-binomiale pour $n = 16$, $\alpha = 2$ et $\beta = 4$.

FIGURE 1 – Représentations de l'exacte bêta-binomiale et de l'histogramme d'un échantillon de taille 2000 obtenu par l'algorithme de Gibbs (séquences de 10 itérations) pour $n = 16$, $\alpha = 2$ et $\beta = 4$.



□

Dans le cadre bayésien, l'algorithme de Gibbs va permettre d'obtenir une réalisation du paramètre $\theta = (\theta_1, \dots, \theta_m)$ suivant la loi a posteriori $\pi(\theta | x)$ dès que l'on est capable d'exprimer les lois conditionnelles : $\pi(\theta_i | \theta_j ; x)$, $j \neq i$.

L'échantillonnage de Gibbs consiste à :

Partant d'un vecteur initial $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_m^{(0)})$
 A la $(p+1)^{\text{ième}}$ étape, disposant du vecteur $\theta^{(p)} = (\theta_1^{(p)}, \dots, \theta_m^{(p)})$,
 simuler
 $\theta_1^{(p+1)} = \pi(\theta_1 \mid \theta_2^{(p)}, \theta_2^{(p)}, \dots, \theta_m^{(p)}; x)$
 $\theta_2^{(p+1)} = \pi(\theta_2 \mid \theta_1^{(p+1)}, \theta_3^{(p)}, \dots, \theta_m^{(p)}; x)$
 \dots
 $\theta_m^{(p+1)} = \pi(\theta_m \mid \theta_1^{(p+1)}, \theta_2^{(p)}, \dots, \theta_{m-1}^{(p)}; x)$

Les itérations successives de cet algorithme génèrent successivement les états d'une chaîne de Markov $\{\theta^{(p)}, p > 0\}$ à valeurs $\mathbb{N}^{\otimes m}$.

La probabilité de transition de θ' vers θ a pour expression :

$$\begin{aligned} K(\theta', \theta) &= \pi(\theta_1 \mid \theta_2', \dots, \theta_m') \times \pi(\theta_2 \mid \theta_1, \theta_3', \dots, \theta_m') \\ &\times \pi(\theta_3 \mid \theta_1, \theta_2, \theta_4', \dots, \theta_m') \times \dots \times \pi(\theta_m \mid \theta_1, \dots, \theta_{m-1}). \end{aligned}$$

On montre que cette chaîne admet une mesure invariante qui est la loi a posteriori. Pour un nombre d'itérations suffisamment grand, le vecteur θ obtenu peut donc être considéré comme étant une réalisation de la loi a posteriori.

□

5.5 L'algorithme de Metropolis

Initialement développé en 1953 pour le traitement de problèmes de physique par Metropolis, Rosenbluth, Rosenbluth et Teller, cet algorithme a ensuite été beaucoup utilisé en physique statistique pour simuler des systèmes complexes. Actuellement, dans la littérature statistique, cet algorithme est présenté comme une méthode permettant de fabriquer une chaîne de Markov dont on s'est donné la loi stationnaire π . Sa mise en oeuvre présente l'avantage de ne nécessiter la définition de π qu'à une constante près.

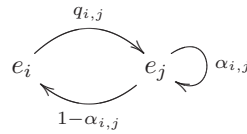
5.5.1 Cas discret

Considérons une chaîne de Markov à valeurs dans un espace d'états discrets $\{e_i\}_{i=1,2,\dots}$. Supposons cette chaîne récurrente, irréductible et de matrice de transition $Q = \{q_{i,j}\}$, symétrique de sorte qu'elle admet une loi stationnaire que l'on note π . π_i , composante de π , est la probabilité que la chaîne se trouve en l'état e_i .

Pour aller de l'état e_i vers l'état e_j , i étant différent de j , imposons le mouvement suivant à la chaîne :

1. aller de e_i vers e_j ,
2. rester en e_j avec une probabilité fixée α_{ij} ,
 sinon retourner en e_i

On a le graphe suivant :



Cette construction constitue une nouvelle chaîne de Markov dont la matrice de transition a pour terme (i, j) , $p_{i,j} = \alpha_{i,j} q_{i,j}$ si $i \neq j$. Pour que la nouvelle chaîne reste en e_i , il faut, ou bien que cette transition se produise sur la chaîne initiale, la probabilité est alors égale à q_{ii} ou bien que l'on retourne en e_i après avoir transité par un des états e_j ($j \neq i$) et la probabilité est alors égale à : $\sum_{j \neq i} q_{i,j} (1 - \alpha_{i,j})$.

On en déduit donc :

$$p_{i,i} = q_{ii} + \sum_{j \neq i} q_{ij} (1 - \alpha_{i,j}) = 1 - \sum_{j \neq i} q_{i,j} \alpha_{i,j} = 1 - \sum_{j \neq i} p_{i,j}.$$

La mesure stationnaire de la nouvelle chaîne sera π si l'on choisit comme l'ont suggéré Metropolis et al. :

$$\alpha_{i,j} = \begin{cases} 1 & \text{si } \pi_j / \pi_i > 1 \\ \pi_j / \pi_i & \text{si } \pi_j / \pi_i \leq 1 \end{cases}$$

En effet, avec ce choix, la chaîne est réversible i.e. : $\forall i, j, \pi_i p_{i,j} = \pi_j p_{j,i}$.

$$\begin{aligned} \pi_i p_{i,j} = \pi_i \alpha_{i,j} q_{ij} &= \pi_i \min\{1, \pi_j / \pi_i\} q_{i,j} \\ &= \min\{\pi_i, \pi_j\} q_{i,j} \\ &= \min\{\pi_j, \pi_i\} q_{j,i} \\ &= \pi_j \min\{1, \pi_j / \pi_i\} q_{i,j} \\ &= \pi_j \alpha_{i,j} q_{i,j} = \pi_j p_{j,i} \end{aligned}$$

On en déduit donc que π est invariante et la chaîne initiale étant récurrente et irréductible, π est unique. (Ripley 87). Si l'espace des états est fini, l'hypothèse irréductible suffit, puisque cette hypothèse entraîne la récurrence. D'autres choix sont possibles pour α_{ij} :

$$\alpha_{ij} = \frac{\pi_j}{\pi_i + \pi_j} \quad (\text{Berker 1965}).$$