Échantillonneur de Gibbs

L'échantilloneur de Gibbs est un exemple d'algorithme de type Metropolis-Hastings dans le cas où nous cherchons à simuler une loi π sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ et où pour tout $x \in \mathbb{R}^d$, x peut être écrit $x = (x_1, \ldots, x_d)$ de telle sorte à ce que pour tout $1 \leq j \leq d$, nous sachions simuler $\pi(\cdot|x_{-j})$ où $x_{-j} = (x_\ell)_{\ell \neq j}$. Une itération de l'algorithme consiste alors simplement à simuler alternativement chaque x_j, $1 \leq j \leq d$, conditionnellement aux autres.

Warm-up

Soit (X, Y) un couple de variables de loi gaussienne centrée de matrice de covariance

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} ,$$

où $\rho \in (0, 1)$. Écrire un échantillonneur de Gibbs permettant de simuler approximativement la loi de (X, Y).

Nous savons bien sûr simuler exactement la loi de (X, Y), il s'agit d'un exemple illustratif pour comprendre l'échantilloneur de Gibbs. Introduisons le vecteur $Z = X - \rho Y$. Alors,

$$\mathrm{Cov}(Z, Y) = \mathrm{Cov}(X, Y) - \rho \mathrm{Cov}(Y, Y) = 0,.$$

Ainsi, puisque le vecteur (Y, Z) est gaussien, Y et Z sont indépendantes. On écrit ensuite

$$\mathbb{E}[X|Y] = \mathbb{E}[Z + \rho Y|Y] = \mathbb{E}[Z] + \rho Y = \rho Y .$$

et

$$\mathbb{V}[X|Y] = \mathbb{E}[(X - \mathbb{E}[X|Y])^2|Y] = \mathbb{E}[(X - \rho Y)^2|Y] = \mathbb{E}[Z^2|Y] = \mathbb{V}[Z] = 1 - \rho^2 .$$

La loi conditionnelle de X sachant Y, notée $\pi_{X|Y}(\cdot|Y)$, est donc une gaussienne de moyenne ρY et de variance $1 - \rho^2$. On obtient un résultat similaire pour la loi de Y sachant X, notée $\pi_{Y|X}(\cdot|X)$, par symétrie. Ainsi une itération d'un échantillonneur de Gibbs lorsque l'état courant est (X_k, Y_k) serait :

- *Simuler $X_{k+1} \sim \pi_{X|Y}(\cdot|Y_k)$.*
- *Simuler $Y_{k+1} \sim \pi_{Y|X}(\cdot|X_{k+1})$.*

Échantillonneur pour un mélange gaussien

Soit $K \geq 2$ et $n \geq 1$. Notons S_K le K-simplexe i.e. l'ensemble des K-uplets de réels positifs de somme égale à 1. On considère le modèle suivant.

- Le vecteur $p = (p_1, \ldots, p_K) \in \mathsf{S}_K$ suit une loi de densité proportionnelle à $p \mapsto \prod_{k=1}^K p_k^{\gamma_k - 1}$ où pour tout $1 \leq k \leq K$, $\gamma_k > 0$. Il s'agit d'une loi de Dirichlet permettant d'obtenir des échantillons sur S_K.
- Le vecteur $s^2 = (s_1^2, \ldots, s_K^2)$ contient des variables mutuellement indépendantes, et telles que pour tout $1 \leq k \leq K$, s_k^2 a une loi inverse-gamma de paramètres, $\lambda_k/2$ et $\beta_k/2$, i.e. de densité proportionnelle à $u \mapsto u^{-\lambda_k/2-1} \exp(-\beta_k/(2u))$ sur $\mathbb{R}_+^*$.
- Conditionnellement à (p, s^2), le vecteur $m = (m_1, \ldots, m_K)$ est constitué de variables indépendantes et pour tout $1 \leq k \leq K$, la loi conditionnelle de m_k est gaussienne de moyenne α_k et de variance s_k^2/λ_k.

- Conditionnellement à $\theta = (p, m, s^2)$, les $(Z_i, X_i)_{1\le i\le n}$ sont indépendantes et telles que :
  - pour tout $1 \le k \le K$, $Z_i \in \{1, \dots K\}$ et $Z_i = k$ avec probabilité p_k ;
  - conditionnellement à (θ, Z_i), $X_i \sim \mathcal{N}(m_{Z_i}, s^2_{Z_i})$.

La densité jointe de toutes les variables peut alors s'écrire :

$$\pi : (\theta, x, z) \mapsto \pi(p) \left\{ \prod_{k=1}^{K} \pi(s_k^2)\pi(m_k|s_k^2) \right\} \left\{ \prod_{i=1}^{n} \pi(z_i|\theta)\pi(x_i|z_i, \theta) \right\} ,$$

où $\pi(w_1|w_2)$ (resp. $\pi(w_1)$) est une notation générique pour la densité de la loi conditionnelle de la variable W_1 sachant W_2 (resp. pour la densité marginale de W_1).

1. Montrer que la loi a posteriori de θ s'écrit :

$$\pi(\theta|x) \propto \pi(\theta) \prod_{i=1}^{n} \left(\sum_{k=1}^{K} p_k \varphi_{m_k, s_k^2}(x_i) \right) ,$$

où φ_{m_k,s_k^2} est la densité gaussienne de moyenne m_k et de variance s_k^2.

Pour écrire cette loi conditionnelle, il suffit d'écrire que la densité conditionnelle est proportionnelle à la densité jointe et ensuite de ne conserver que les quantités qui dépendent de θ. On obtient alors,

$$\pi(\theta|x) \propto \pi(\theta, x) \propto \pi(\theta)\pi(x|\theta) \propto \pi(\theta) \prod_{i=1}^{n} \pi(x_i|\theta) \propto \pi(\theta) \prod_{i=1}^{n} \left(\sum_{k=1}^{K} p_k \varphi_{m_k,s_k^2}(x_i) \right) .$$

Nous ne savons pas simuler cette loi directement. Nous proposons donc de simuler la loi de (θ, Z) sachant X à l'aide d'un échantillonneur de Gibbs, i.e. en simulant alternativement la loi de θ sachant (Z, X) et la loi de Z sachant (θ, X).

2. Écrire la densité de la loi conditionnelle de Z sachant (X, θ).

En procédant comme à la question précédente, on obtient

$$\begin{aligned}\pi(z|\theta, x) \propto \pi(\theta, z, x) \propto \pi(z|\theta)\pi(x|z, \theta) &\propto \prod_{i=1}^{n} \pi(z_i|\theta)\pi(x_i|z_i\theta) ,\\ &\propto \prod_{i=1}^{n} \left(\sum_{k=1}^{K} \mathbb{1}_{z_i=k} p_k \varphi_{m_k,s_k^2}(x_i) \right) .\end{aligned}$$

Conditionnellement à (X, θ), les $(Z_1, \dots, Z_n)$ sont donc indépendants et la probabilité conditionnelle de l'événement $\{Z_i = k\}$ est proportionnelle à $p_k \varphi_{m_k,s_k^2}(x_i)$ pour $1 \le k \le K$, $1 \le i \le n$.

3. Écrire la densité de la loi conditionnelle de θ sachant (Z, X).

La densité de la loi conditionnelle de θ sachant (Z, X) s'écrit

$$\pi(\theta|z, x) \propto \pi(\theta, z, x) \propto \pi(\theta)\pi(z|\theta)\pi(x|z, \theta) .$$

Il suffit alors de calculer $\pi(p|z)$ et $\pi(m, s^2|x, z)$.

$$\begin{aligned}\pi(p|z) &\propto \pi(\theta, z, x) \propto \pi(p)\pi(z|p) ,\\ &\propto \prod_{k=1}^{K} p_k^{\gamma_k - 1} \prod_{i=1}^{n} p_{z_i} \propto \prod_{k=1}^{K} p_k^{\gamma_k + n_k - 1} ,\end{aligned}$$

où n_k est le nombre de z_i égaux à k. Ainsi, $\pi(p|z)$ est la densité de la loi de Dirichlet de paramètres $(\gamma_1 + n_1, \dots, \gamma_K + n_K)$.

D'autre part,

$$\begin{aligned}
\pi(m, s^2|x,z) &\propto \pi(s^2)\pi(m|s^2)\pi(z|m,s^2)\pi(x|z,m,s^2)\,,\\
&\propto \left\{\prod_{i=1}^n s_{z_i}^{-1}\exp\left(-\frac{(x_i - m_{z_i})^2}{2s_{z_i}^2}\right)\right\}\left\{\prod_{k=1}^K (s_k^2)^{-\lambda_k/2-1}\exp(-\beta_k/(2s_k^2))\right\}\\
&\qquad\qquad \times \prod_{k=1}^K s_k^{-1}\exp\left(-\frac{\lambda_k(m_k-\alpha_k)^2}{2s_k^2}\right)\,,\\
&\propto \prod_{k=1}^K (s_k^2)^{-n_k/2-\lambda_k/2-3/2}\exp(-\beta_k/(2s_k^2))\\
&\qquad\qquad \times \prod_{k=1}^K \exp\left(\frac{-1}{2s_k^2}\left(\lambda_k(m_k-\alpha_k)^2 + \sum_{i=1}^n \mathbb{1}_{z_i=k}(x_i - m_k)^2\right)\right)\,.
\end{aligned}$$

On remarque alors que

$$\begin{aligned}
&\lambda_k(m_k-\alpha_k)^2 + \sum_{i=1}^n \mathbb{1}_{z_i=k}(x_i-m_k)^2\\
&\quad = (n_k+\lambda_k)m_k^2 - 2m_k\left(\sum_{i=1}^n \mathbb{1}_{z_i=k}x_i + \lambda_k\alpha_k\right) + \lambda_k\alpha_k^2 + \sum_{i=1}^n \mathbb{1}_{z_i=k}x_i^2\,,\\
&\quad = (n_k+\lambda_k)\left(m_k - \frac{\sum_{i=1}^n \mathbb{1}_{z_i=k}x_i + \lambda_k\alpha_k}{n_k+\lambda_k}\right)^2 - \frac{\left(\sum_{i=1}^n \mathbb{1}_{z_i=k}x_i + \lambda_k\alpha_k\right)^2}{n_k+\lambda_k} + \lambda_k\alpha_k^2 + \sum_{i=1}^n \mathbb{1}_{z_i=k}x_i^2\,,\\
&\quad = (n_k+\lambda_k)\left(m_k-\tau_k\right)^2 - (n_k+\lambda_k)\tau_k^2 + \lambda_k\alpha_k^2 + \sum_{i=1}^n \mathbb{1}_{z_i=k}x_i^2\,,
\end{aligned}$$

où $\tau_k = (\sum_{i=1}^n \mathbb{1}_{z_i=k}x_i + \lambda_k\alpha_k)/(n_k+\lambda_k)$. Ainsi,

$$\begin{aligned}
\pi(m,s^2|x,z) &\propto \prod_{k=1}^K (s_k^2)^{-n_k/2-\lambda_k/2-3/2}\exp\left(-\frac{\rho_k}{2s_k^2}\right)\\
&\qquad\qquad\qquad \times \prod_{k=1}^K \exp\left(\frac{-(n_k+\lambda_k)}{2s_k^2}\left(m_k-\tau_k)^2\right)\right)\,,
\end{aligned}$$

où

$$\rho_k = \beta_k + \lambda_k\alpha_k^2 + \sum_{i=1}^n \mathbb{1}_{z_i=k}x_i^2 - (n_k+\lambda_k)\tau_k^2\,.$$

Finalement, sous $\pi(m,s^2|x,z)$, les s_k^2 sont indépendants de loi $\mathcal{IG}((n_k+\lambda_k+1)/2;\rho_k/2)$ et les m_k sont indépendants conditionnellement aux s_k^2 et de loi $\mathcal{N}(\tau_k; s_k^2/(n_k+\lambda_k))$, $1 \le k \le K$.

4. Écrire le pseudo-code de l'échantillonneur de Gibbs.

   Il suffit, à chaque itération $k \ge 0$, de simuler θ_{k+1} sachant (Z_k, X) puis de simuler Z_{k+1} sachant (θ_{k+1}, X) en utilisant les questions précédentes.