Computational statistics
TP 4 : Improve the Metropolis-Hastings algorithm

Thibault de SURREL

Exercice 1 : Adaptive Metropolis-Hastings within Gibbs sampler

Sur le notebook.

Exercice 2 : Sampling from multimodal distributions

Sur le notebook.

Exercice 3 : Bayesian analysis of a one-way random effects model

1.

On commence par calculer la loi $q(X, \mu, \sigma^2, \tau^2 \mid Y)$. D'après la formule de Bayes, on a dans un premier temps

$$q(X, \mu, \sigma^2, \tau^2 \mid Y) = \frac{q(Y \mid X, \mu, \sigma^2, \tau^2) q(X, \mu, \sigma^2, \tau^2)}{q(Y)} \propto q(Y \mid X, \mu, \sigma^2, \tau^2) q(X, \mu, \sigma^2, \tau^2)$$

car $q(Y)$ est un constante dans notre calcul. On cherche à calculer ces termes séparément :

— D'une part, pour $q(X, \mu, \sigma^2, \tau^2)$, on a déjà que

$$q(X, \mu, \sigma^2, \tau^2) = q(X \mid \mu, \sigma^2, \tau^2) \pi_{prior}(\mu, \sigma^2, \tau^2)$$

De plus, les $(X_i)_{i \in [\![1,N]\!]}$ sont indépendantes et identiquement distribuées et suivent une $\mathcal{N}(\mu, \sigma^2)$, ainsi on a

$$\begin{aligned} q(X \mid \mu, \sigma^2, \tau^2) &= \prod_{i=1}^{N} q(X_i \mid \mu, \sigma^2, \tau^2) \\ &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(X_i - \mu)^2\right] \\ &\propto (\sigma^2)^{-N/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(X_i - \mu)^2\right] \end{aligned}$$

— D'autre part, on a que les $(y_{i,j})_{(i,j) \in [\![1,N]\!] \times [\![1,k_i]\!]}$ sont indépendantes et identiquement distribuées. De plus, ici les X_i sont fixés donc, sachant X_i, comme $y_{i,j} = X_i + \varepsilon_{i,j}$ avec $\varepsilon_{i,j} \sim \mathcal{N}(0, \tau^2)$ on a que $y_{i,j} \mid X, \mu, \sigma^2, \tau^2 \sim \mathcal{N}(X_i, \tau^2)$. Ainsi,

$$q(Y \mid X, \mu, \sigma^2, \tau^2) = \prod_{i=1}^{N}\prod_{j=1}^{k_i} q(y_{i,j} \mid X, \mu, \sigma^2, \tau^2) \propto \prod_{i=1}^{N}\prod_{j=1}^{k_i} \frac{1}{\sqrt{\tau^2}} \exp\left[-\frac{1}{2\tau^2}(y_{i,j} - X_i)^2\right]$$

Ainsi, en mettant toutes les formules bout-à-bout, on a que

$$\begin{aligned} &q(X, \mu, \sigma^2, \tau^2) \\ &\propto (\sigma^2)^{-(N/2+1+\alpha)} \exp\left[-\frac{1}{\sigma^2}\left(\frac{1}{2}\sum_{i=1}^{N}(X_i - \mu)^2 + \beta\right)\right] (\tau^2)^{-(\frac{k}{2}+1+\gamma)} \exp\left[-\frac{1}{\tau^2}\left(\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{k_i}(y_{i,j} - X_i)^2 + \beta\right)\right] \end{aligned}$$

où $k = \sum_{i=1}^{N} k_i$

2.

Afin d'implémenter un échantillonneur de Gibbs, il nous faut connaître les lois des différents paramètres sachant les autres. On utilise donc la question précédente afin d'obtenir les résultats suivants :

— Pour τ^2, on a

$$q(\tau^2 \mid X, \mu, \sigma^2, Y) \propto (\tau^2)^{-(\frac{k}{2}+1+\gamma)} \exp\left[-\frac{1}{\tau^2}\left(\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{k_i}(y_{i,j} - X_i)^2 + \beta\right)\right]$$

donc, $\tau^2 \mid X, \mu, \sigma^2, Y$ suit une loi inverse-gamma de paramètres $(\frac{k}{2} + \gamma, \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{k_i}(y_{i,j} - X_i)^2 + \beta)$.

— Pour σ^2, on résonnant comme pour τ^2, on a que σ^2 suit une loi inverse-gamma de paramètres $(\frac{N}{2}+\alpha, \frac{1}{2}\sum_{i=1}^{N}(X_i - \mu)^2 + \beta)$.

— Pour μ, on a d'après la question précédente :

$$q(\mu \mid X, Y, \sigma^2, \tau^2) \propto \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{N}(X_i - \mu)^2\right)\right]$$

On peut alors développer l'argument de l'exponentielle et obtenir, comme les X_i sont des constantes dans ce calcul :

$$q(\mu \mid X, Y, \sigma^2, \tau^2) \propto \exp\left[-\frac{1}{2\sigma^2}\left(N\mu^2 - 2\mu\sum_{i=1}^{N}X_i\right)\right] \propto \exp\left[-\frac{N}{2\sigma^2}\left(\mu - \frac{1}{N}\sum_{i=1}^{N}X_i\right)^2\right]$$

Ainsi, $\mu \mid X, Y, \sigma^2, \tau^2$ suit une $\mathcal{N}\left(\frac{1}{N}\sum_{i=1}^{N}X_i, \frac{\sigma^2}{N}\right)$.

— Enfin, pour X, cela demande un peu plus de travail. On sait que

$$\begin{aligned} q(X \mid Y, \mu, \sigma^2, \tau^2) &\propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(X_i - \mu)^2 - \frac{1}{2\tau^2}\sum_{i=1}^{N}\sum_{j=1}^{k_i}(y_{i,j} - X_i)^2\right] \\ &\propto \prod_{i=1}^{N}\underbrace{\exp\left[-\frac{1}{2\sigma^2}(X_i - \mu)^2 - \frac{1}{2\tau^2}\sum_{j=1}^{k_i}(y_{i,j} - X_i)^2\right]}_{=q(X_i \mid Y, \mu, \sigma^2, \tau^2)} \end{aligned}$$

On a alors,

$$\begin{aligned} q(X_i \mid Y, \mu, \sigma^2, \tau^2) &\propto \exp\left[-\frac{1}{2\sigma^2}(X_i^2 - 2X_i\mu + \mu^2) - \frac{1}{2\tau^2}\sum_{j=1}^{k_i}(y_{i,j}^2 - 2y_{i,j}X_i + X_i^2)\right] \\ &\propto \exp\left[-\left(\frac{1}{2\sigma^2} + \frac{k_i}{2\tau^2}\right)X_i^2 + 2X_i\left(\frac{\mu}{2\sigma^2} + \sum_{j=1}^{k_i}\frac{y_{i,j}}{2\tau^2}\right)\right] \\ &\propto \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2} + \frac{k_i}{\tau^2}\right)\left(X_i - \left(\frac{1}{\sigma^2} + \frac{k_i}{\tau^2}\right)^{-1}\left(\frac{\mu}{\sigma^2} + \sum_{j=1}^{k_i}\frac{y_{i,j}}{\tau^2}\right)\right)^2\right] \end{aligned}$$

Ainsi, on a que $X_i \mid Y, \mu, \sigma^2, \tau^2$ suit une $\mathcal{N}\left(\left(\frac{1}{\sigma^2}+\frac{k_i}{\tau^2}\right)^{-1}\left(\frac{\mu}{\sigma^2}+\sum_{j=1}^{k_i}\frac{y_{i,j}}{\tau^2}\right), \left(\frac{1}{\sigma^2}+\frac{k_i}{\tau^2}\right)^{-1}\right)$ qui se réécrit $\mathcal{N}\left(\frac{\tau^2\mu+\sigma^2\sum_{j=1}^{k_i} y_{i,j}}{k_i\sigma^2+\tau^2}, \frac{\sigma^2\tau^2}{k_i\sigma^2+\tau^2}\right)$.

3.

Pour cette question, on a besoin de connaître la loi de $X, \mu \mid \sigma^2, \tau^2, Y$. D'après les questions précédentes, on a déjà que

$$q(X,\mu \mid \sigma^2,\tau^2,Y) \propto \prod_{i=1}^{N} \underbrace{\exp\left[-\frac{1}{2\sigma^2}(X_i-\mu)^2 - \frac{1}{2\tau^2}\sum_{j=1}^{k_i}(y_{i,j}-X_i)^2\right]}_{=q(X_i,\mu|Y,\sigma^2,\tau^2)} \tag{1}$$

Ainsi, on a, pour $i \in [\![1, N]\!]$,

$$q(X_i,\mu \mid Y,\sigma^2,\tau^2) \propto \exp\left[-\frac{1}{2}\left(\left(\frac{1}{\sigma^2}+\frac{k_i}{\tau^2}\right)X_i^2 - \frac{2}{\sigma^2}X_i\mu - 2X_i\frac{\sum_{j=1}^{k_i} y_{i,j}}{\tau^2} + \frac{1}{\sigma^2}\mu\right)\right]$$

On sens qu'on va pouvoir mettre ça sous la forme d'une gaussienne multivariée de la variable $Z_i = (X_i \ \mu)^T$, dont la moyenne sera notée $\theta = (\theta_1 \ \theta_2)^T$ et l'inverse de la matrice de variance-covariance sera $\Sigma^{-1} = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$ (elle est symétrique car la transposée et l'inverse commutent et que la matrice de variance-covariance est symétrique). Alors, la densité de la variable Z_i s'écrit

$$q(Z_i \mid Y,\sigma^2,\tau^2) \propto \exp\left[-\frac{1}{2}(Z-\theta)^T\Sigma^{-1}(Z-\theta)\right] \tag{2}$$

En détaillant les calculs, on peux trouver que, avec nos notations, on a

$$(Z-\theta)^T\Sigma^{-1}(Z-\theta) = aX_i^2 + d\mu^2 - 2aX_i\theta_1 - 2d\mu\theta_2 + 2bX_i\mu - 2b\theta_2X_i - 2b\theta_1\mu \tag{3}$$

On peut alors identifier les différents termes de Σ^{-1} grâce à l'équation 1. On trouve

$$a = \frac{1}{\sigma^2}+\frac{k_i}{\tau^2}, \quad b = -\frac{1}{\sigma^2}, \quad d = \frac{1}{\sigma^2}$$

Alors, grâce au fait que $b = -d$, on peut réécrire l'équation 3 de la façon suivante :

$$(Z-\theta)^T\Sigma^{-1}(Z-\theta) = aX_i^2 + d\mu^2 + 2bX_i\mu - 2(a\theta_1+b\theta_2)X_i - 2d(\theta_2-\theta_1)\mu$$

Et en identifiant encore grâce à l'équation 1, comme il n'y a pas de terme avec μ tout seul, on a que

$$\begin{cases} (a\theta_1+b\theta_2) = \dfrac{\sum_{j=1}^{k_i} y_{i,j}}{\tau^2} \\ d(\theta_2-\theta_1) = 0 \end{cases} \iff \theta_1 = \theta_2 = \frac{\sum_{j=1}^{k_i} y_{i,j}}{k_i}$$

$$q(X,\mu \mid Y,\sigma^2,\tau^2) = \mathcal{N}(\theta,\Sigma)$$

avec

$$\Sigma^{-1} = \begin{pmatrix} \frac{1}{\sigma^2}+\frac{k_1}{\tau^2} & 0 & 0 & \frac{-1}{\sigma^2} \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & \frac{1}{\sigma^2}+\frac{k_N}{\tau^2} & \frac{-1}{\sigma^2} \\ \frac{-1}{\sigma^2} & \cdots & \frac{-1}{\sigma^2} & \frac{N}{\sigma^2} \end{pmatrix} \quad \text{et } \theta = \begin{pmatrix} \frac{\sum_{j=1}^{k_1} y_{1,j}}{k_1} \\ \vdots \\ \frac{\sum_{j=1}^{k_N} y_{N,j}}{k_N} \\ \frac{\sum_{i=1}^{N}\sum_{j=1}^{k_i} y_{i,j}}{\sum_{i=1}^{N} k_i} \end{pmatrix}$$

4.

L'algorithme de Gibbs permet de générer une chaîne de Markov. L'avantage est qu'on on peut initialiser notre chaîne de Markov où l'on veut et que si on laisse une période de "chauffe" suffisamment importante, cette initialisation n'aura pas d'importance sur le comportement de la chaîne. De plus, on a des garantis théoriques sur la convergence de la suite de variables aléatoires produites par l'algorithme de Gibbs vers la loi que l'on veut échantillonner avec une vitesse de convergence géométrique. Pour l'appliquer en pratique, on doit savoir facilement simuler selon les marginales de notre loi cible, sinon on perd tout l'intérêt de cette méthode. Le problème de cet algorithme est qu'on a une forte dépendance des variables générées les unes aux autres. Au vu de chaque étape de l'algorithme, les variables sont fortement corrélées les unes aux autres. C'est pourquoi on introduit l'algorithme de Gibbs par blocs.

L'algorithme de Gibbs par blocs permet d'une part de gagner en temps de calcul, car en regroupant les variables en blocs, on en a moins à générer. D'autre part, on a des garantis théoriques qui nous disent que cette version de l'algorithme permettra d'améliorer la qualité des échantillons générés.