TP 4 - Computational Statistics: Improve the Metropolis-Hastings algorithm

Emilio Picard

1 Exercice 3

1.1 Question 1

Calculons la densité a-posteriori de $(X, \mu, \sigma^2, \tau^2)$, soit $\mathbb{P}(X, \mu, \sigma^2, \tau^2 | Y)$:

$$\begin{aligned}\mathbb{P}(X, \mu, \sigma^2, \tau^2 \mid Y) &\propto \mathbb{P}(Y \mid X, \mu, \sigma^2, \tau^2)\mathbb{P}(X, \mu, \sigma^2, \tau^2) &(1)\\ &\propto \mathbb{P}(Y \mid X, \mu, \sigma^2, \tau^2)\mathbb{P}(X \mid \mu, \sigma^2, \tau^2)\mathbb{P}(\mu, \sigma^2, \tau^2) &(2)\end{aligned}$$

Calculons les 3 quantités ci-dessus:
(i) On sait que, pour tout $i, j \in \{1 \dots, N\} \times \{1, k_i\}$, $Y_{i,j} \mid (X_i, \mu, \sigma^2, \tau^2) \sim \mathcal{N}(X_i, \tau^2)$, donc on obtient:

$$\mathbb{P}(Y \mid X, \mu, \sigma^2, \tau^2) \propto \left(\frac{1}{\tau^2}\right)^{\frac{k}{2}} \exp\left(-\frac{\sum_{i=1}^N \sum_{j=1}^{k_i} (y_{i,j} - X_i)^2}{2\tau^2}\right).$$

(ii) De plus, pour tout $i \in \{1, \dots, N\}$, $X_i \sim \mathcal{N}(\mu, \sigma^2)$ donc de la même manière on obtient:

$$\mathbb{P}(X \mid \mu, \sigma^2, \tau^2) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{N}{2}} \exp\left(-\frac{\sum_{i=1}^N (X_i - \mu)^2}{2\sigma^2}\right).$$

(iii) Il nous reste l'expression de la loi a priori, que l'énoncé nous donne. Finalement, on a:

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

D'où:

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

1.2 Question 2

Calculons l'évolution des paramètres pour les itérations de l'algorithme:
Pour σ^2: $\mathbb{P}(\sigma^2 \mid Y, X, \mu, \tau^2) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{N}{2}+\alpha+1} \exp\left(-\frac{\sum_i^N (X_i-\mu)^2}{2\sigma^2} - \frac{\beta}{\sigma^2}\right)$
On reconnaît une loi Inverse Gamma
$\boxed{\mathbb{P}(\sigma^2 \mid Y, X, \mu, \tau^2) \propto \Gamma^{-1}\left(\frac{N}{2}+\alpha, \frac{\sum_i^N (X_i-\mu)^2}{2}+\beta\right).}$
Pour τ^2: $\mathbb{P}(\tau^2 \mid Y, X, \mu, \sigma^2) \propto \left(\frac{1}{\tau^2}\right)^{\frac{k}{2}+\gamma+1} \exp\left(-\frac{\sum_i^N \sum_{j=1}^{k_i} (y_{i,j}-X_i)^2}{2\tau^2} - \frac{\beta}{\tau^2}\right)$
On reconnaît également un loi Inverse Gamma:
$\boxed{\mathbb{P}(\tau^2 \mid Y, X, \mu, \sigma^2) \propto \Gamma^{-1}\left(\frac{k}{2}+\gamma, \frac{\sum_i^N \sum_{j=1}^{k_i} (y_{i,j}-X_i)^2}{2}+\beta\right).}$

Pour μ: $\mathbb{P}(\mu \mid Y, X, \sigma^2, \tau^2) \propto \exp\left(-\frac{\sum_i^N (X_i-\mu)^2}{2\sigma^2}\right) \propto \exp\left(-\frac{2\mu \sum_i^N X_i}{2\sigma^2} + \frac{N\mu^2}{2\sigma^2}\right) \propto \exp\left(-\frac{\left(\frac{\sum_i^N X_i}{N}-\mu\right)^2}{\frac{2\sigma^2}{N}}\right)$,

où l'on reconnaît cette fois une loi normale:
$\boxed{\mathbb{P}(\mu \mid Y, X, \sigma^2, \tau^2) \propto \mathcal{N}\left(\frac{\sum_i^N X_i}{N}, \frac{\sigma^2}{N}\right).}$.
Pour X:

$$
\begin{aligned}
\mathbb{P}(X \mid \mu, \sigma^2, \tau^2, Y) &\propto \prod_{i=1}^N \exp\left(-\frac{\sum_{j=1}^{k_i}(y_{i,j}-X_i)^2}{2\tau^2} - \frac{(X_i-\mu)^2}{2\sigma^2}\right) \\
&\propto \prod_{i=1}^N \exp\left(\frac{X_i^2 - 2\mu X_i + \mu^2}{2\sigma^2} - \frac{\sum_{j=1}^{k_i}\left(y_{i,j}^2 - 2y_{i,j}X_i + X_i^2\right)}{2\tau^2}\right) \\
&\propto \prod_{i=1}^N \exp\left(-X_i^2\left(\frac{1}{2\sigma^2}+\frac{k_i}{2\tau^2}\right) + X_i\left(\frac{2\mu}{2\sigma^2} + \frac{2\sum_{j=1}^{k_i} y_{i,j}}{2\tau^2}\right) - \left(\frac{\mu^2}{2\sigma^2} + \frac{\sum_{j=1}^{k_i} y_{i,j}^2}{2\tau^2}\right)\right) \\
&\propto \prod_{i=1}^N \exp\left(-\frac{\left(X_i - \frac{\tau^2\mu+\sigma^2\sum_{j=1}^{k_i} y_{i,j}}{\tau^2+k_i\sigma^2}\right)^2}{2\frac{\sigma^2\tau^2}{\tau^2+k_i\sigma^2}}\right).
\end{aligned}
$$

Ainsi, on reconnaît également une loi normale:
$\boxed{\forall i, \mathbb{P}(X_i \mid \mu, \sigma^2, \tau^2, Y) \propto \mathcal{N}(\frac{\tau^2\mu+\sigma^2\sum_{j=1}^{k_i} y_{i,j}}{k_i\sigma^2+\tau^2}, \frac{\sigma^2\tau^2}{k_i\sigma^2+\tau^2}).}$

1.3 Question 3

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

En utilisant les mêmes raisonnements que précédemment, on trouve que $X, \mu \mid Y, \sigma^2, \tau^2$ suit une loi normale multivariée, avec

$$\Sigma_{\text{multi}} = \left[\text{diag}\left(\frac{\sigma^2 k_i + \tau^2}{\tau^2\sigma^2}\right) \quad -\frac{1}{\sigma^2}\mathbf{1} \ \ -\frac{1}{\sigma^2}\mathbf{1}^T \quad \frac{N}{\sigma^2}\right]$$

et

$$\mu_{\text{multi}} = \Sigma^{-1}\left[\frac{\sum_{j=1}^{k_i} y_{i,j}}{\tau^2} \ \ 0\right]$$

1.4 Question 4

Dans un échantillonneur de Gibbs classique, les paramètres sont mis à jour séquentiellement en échantillonnant leurs distributions conditionnelles, tandis que les autres paramètres sont fixés. Cela rend l'algorithme plutôt simple à implémenter. Cependant, lorsque les variables sont fortement corrélées, les mises à jour peuvent entraîner un ralentissement, l'échantillonneur de Gibbs nécessitant davantage d'itérations pour bien capturer les relations entre les variables.
Par ailleurs, l'échantillonneur de Gibbs par block met à jour les paramètres corrélés ensemble dans un même bloc. Cela peut permettre une convergence plus rapide, mais au prix de distributions plus complexes à échantillonner.
Dans notre cas, cette méthode est particulièrement adaptée. Lorsque nous mettons à jour X_i et μ ensemble, nous obtenons directement la moyenne exacte de $Y_{i,j}$. Étant donné que X_i est échantillonné à partir d'une distribution gaussienne centrée sur μ, cette mise à jour conjointe de X_i et μ capture efficacement leur relation, ce qui conduit à une convergence plus rapide et précise.