STAT 3006: Statistical Computing
Lecture 9*

13 March

7.3 Hybrid Gibbs Sampler

In the last subsection, we have known that Gibbs sampler is sampling from full conditional functions iteratively to approximate samples from the target distribution $f(\mathbf{x})$ of our interest. In some case, for a (or some) specific conditional function $f_i(x_i|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p)$, it does not correspond to a common distribution. Directly drawing from $f_i(x_i|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p)$ is difficult. We can generate a MH sample in this step, which requires a proposal distribution $q_i(y|x_i)$. The algorithm combines the Gibbs sampler and the MH algorithm into a hybird one, thus called hybrid Gibbs sampler.

Algorithm: Hybrid Gibbs sampler to sample from $f(\mathbf{x})$.

Input: the pdf (or pmf) $f(\mathbf{x})$ and a starting point $\mathbf{x}^{(0)} = (x_1^{(0)}, \ldots, x_p^{(0)})$. $f_1, f_2, \ldots, f_p$ are full conditional functions of $f(\mathbf{x})$. Without loss of generality, $f_1, \ldots, f_{p-1}$ are common distributions that are easy to sample, but f_p is not a standard distribution to be sampled using MH algorithm.

Initialize: $t \leftarrow 0$.

Repeat

generate $x_1^{(t+1)} \sim f_1(x_1|x_2^{(t)}, x_3^{(t)}, \ldots, x_p^{(t)})$;

$\vdots$

generate $x_{p-1}^{(t+1)} \sim f_{p-1}(x_{p-1}|x_1^{(t+1)}, x_2^{(t+1)}, \ldots, x_p^{(t)})$;

(MH step) generate a proposal y from $q_p(y|x_p^{(t)})$;

calculate r, the value of $\min\left(\frac{f_p(y|x_1^{(t+1)},\ldots,x_{p-1}^{(t+1)})}{f_p(x_p^{(t)}|x_1^{(t+1)},\ldots,x_{p-1}^{(t+1)})}\frac{q(x_p^{(t)}|y)}{q(y|x_p^{(t)})}, 1\right)$.

accept $x_p^{(t+1)}$ as y with probability r;otherwise, reject y and let $x_p^{(t+1)}$ be $x_p^{(t)}$.

$t \leftarrow t+1$;

Until some criteria are met.

Output:Given a large number B, $\{\mathbf{x}^{(B)}, \mathbf{x}^{(B+1)}, \ldots\}$ are samples from the distribution f.

Remark 1. MH step only implements MH algorithm *once* in one Gibbs sampler cycle.

*If you have any question about the note, please send an email to xyluo@link.cuhk.edu.hk

7.4 One Example of Implementing Gibbs Sampler

There are n subjects. For each subject $j(j = 1, \ldots, n)$, it has a G dimensional vector $(x_{1j}, x_{2j}, \ldots, x_{Gj})$. We know n subjects can be clustered into K groups. We use C_j to represent the cluster to which subject j belongs, and π_k to represent the proportion of cluster k. Given $C_j = k$, $(x_{gj} \sim Poi(\lambda_{gk})$ for $g = 1, \ldots, G$. $\{x_{gj} : g = 1, \ldots, G, j = 1, \ldots, n\}$ are observed, and the missing data is $\mathbf{C} = \{C_j : j = 1, \ldots, n\}$.The parameters are $\pi = \{\pi_1, \ldots, \pi_K\}$ and $\Lambda = \{\lambda_{gk}\}(g = 1, \ldots, G, k = 1, \ldots, K)$.

Based on the information, we have the complete-data likelihood function

$$f(\mathbf{X}, \mathbf{C}|\pi, \Lambda) = \prod_{j=1}^{n} \prod_{k=1}^{K} \left[\pi_k \prod_{g=1}^{G} \frac{\lambda_{gk}^{x_{gj}}}{x_{gj}!} e^{-\lambda_{gk}} \right]^{I(C_j = k)}.$$

We assign the following priors on (π, Λ):

$$\begin{aligned}(\pi_1, \ldots, \pi_K) &\sim Dirichlet(\alpha_1, \alpha_2, \ldots, \alpha_K);\\ \lambda_{gk} &\sim Gamma(\alpha, \beta).\end{aligned}$$

Subsequently, the posterior distribution for $(\pi, \Lambda, \mathbf{C})$ is

$$p(\pi, \lambda, \mathbf{C}|\mathbf{X}) \propto f(\mathbf{X}, \mathbf{C}|\pi, \Lambda) p(\pi) p(\Lambda).$$

The Gibbs sampler to simulate $p(\pi, \lambda, \mathbf{C}|\mathbf{X})$ proceeds as follows.

- $p(\lambda_{gk}|-) \propto \lambda_{gk}^{\sum_{j=1}^{n} I(C_j = k) x_{gj} + \alpha - 1} e^{-\lambda_{gk}(\beta + \sum_{j=1}^{n} I(C_j = k))}$, so $\lambda_{gk} \sim Gamma(\sum_{j=1}^{n} I(C_j = k) x_{gj} + \alpha, \beta + \sum_{j=1}^{n} I(C_j = k))$.
- $p(\pi|-) \propto \prod_{k=1}^{K} \pi_k^{\sum_{j=1}^{n} I(C_j = k) + \alpha_k - 1}$, so $\pi \sim Dirichlet(\sum_{j=1}^{n} I(C_j = 1) + \alpha_1, \ldots, \sum_{j=1}^{n} I(C_j = K) + \alpha_K)$.
- $p(C_j = k|-) = \frac{\pi_k \prod_{g=1}^{G} \lambda_{gk}^{x_{gj}} e^{-\lambda_{gk}}}{\sum_{l=1}^{K} \pi_l \prod_{g=1}^{G} \lambda_{gl}^{x_{gj}} e^{-\lambda_{gl}}}$.

"$-$" means given other variables.