Model Specification for Double Metropolis Algorithm

Naihui Zhou

March 12, 2018

Y_i is the gene expression data for gene i, $i = 1, 2, \ldots, n$
The entire human genome can be represented by an undirected graph with n vertices.
An edge exists if two genes are spatially close to each other.
This information is obtained from the HiC contact map by (Jin et al. 2013).

$Y_i \sim \text{Poisson}(\lambda_i)$
Let $w_i = log(\lambda_i)$, then the **data model** is

$$f(y_i|w_i) = \frac{1}{y_i!} exp(-e^{w_i} + w_i y_i) \tag{1}$$

The **mixing distribution** is a Markov Random Field,

$$w_i|w(N_i) \sim N(\mu_i, \tau^2) \tag{2}$$

where

$$\mu_i = \alpha + \eta \sum_{j \in N_i} (w_j - \alpha) \tag{3}$$

.
Priors are assumed to be independent, $\pi(\alpha, \eta, \tau^2) = \pi(\alpha)\pi(\eta)\pi(\tau^2)$.

The **posterior distribution** for w_i is then

$$\begin{aligned} p(w_i|\alpha, \eta, \tau^2, \boldsymbol{w}) &\propto f(y_i|w_i) g(\boldsymbol{w}|\alpha, \eta, \tau^2) && (4) \\ &\propto f(y_i|w_i) g(w_i|w(N_i), \alpha, \eta, \tau^2) h(w_{-i}|\alpha, \eta, \tau^2) && (5) \\ &\propto f(y_i|w_i) g(w_i|w(N_i), \alpha, \eta, \tau^2) && (6) \end{aligned}$$

For a conclique C, denote $\boldsymbol{w_C} = \{w_i : i \in C\}$.
The **posterior distribution** for $\boldsymbol{w_C}$ is

$$\begin{aligned} p(\boldsymbol{w_C}|\alpha, \eta, \tau^2, \boldsymbol{w}) &\propto [\prod_{i \in C} f(y_i|w_i)] g(\boldsymbol{w_C}|\boldsymbol{w_{-C}}, \alpha, \eta, \tau^2) && (7) \\ &\propto \prod_{i \in C} [f(y_i|w_i) g(w_i|w(N_i), \alpha, \eta, \tau^2)] && (8) \end{aligned}$$

The **posterior distribution** of α is

$$\begin{aligned} g_1(\alpha|\eta, \tau^2, \boldsymbol{y}, \boldsymbol{w}) &\propto \pi(\alpha) g(\boldsymbol{w}|\alpha, \eta, \tau^2) && (9) \\ &= \pi(\alpha) \frac{exp(Q(\boldsymbol{w}|\alpha, \eta, \tau^2))}{\int exp(Q(\boldsymbol{w}|\alpha, \eta, \tau^2)) d\boldsymbol{w}} && (10) \\ &= \pi(\alpha) C(\alpha) exp(Q(\boldsymbol{w}|\alpha, \eta, \tau^2)) && (11) \end{aligned}$$

where

$$Q(\boldsymbol{w}|\alpha, \eta, \tau^2) = \sum_{1 \le i \le j} H_i(w_i|\alpha, \eta, \tau^2) + \sum \sum_{1 \le i < j \le n} H_{i,j}(w_i, w_j|\alpha, \eta, \tau^2) \tag{12}$$

and

$$C(\alpha) = 1/\int exp(Q(\boldsymbol{w}|\alpha, \eta, \tau^2))d\boldsymbol{w} \tag{13}$$

can be seen as an unknown constant containing α.

The function Q is called a negpotential function, and can be reduced to the form in (12) according to (Kaiser and Cressie 2000) and the Clifford and Hammersley theorem.

In summary, the posterior distribution (4) can be fully specified, but the posterior distribution (7) as well as posterior distributions of η and τ^2 contains an intractable normalizing constant that is dependent on the variable of interest. Therefore we resolve to the double metropolis algorithm(Liang 2010) to simulate from these distributions.

References

Jin, Fulai, Yan Li, Jesse R Dixon, Siddarth Selvaraj, Zhen Ye, Ah Young Lee, Chia-An Yen, Anthony D Schmitt, Celso A Espinoza, and Bing Ren. 2013. "A High-Resolution Map of the Three-Dimensional Chromatin Interactome in Human Cells." *Nature* 503 (7475). Nature Publishing Group: 290.

Kaiser, Mark S, and Noel Cressie. 2000. "The Construction of Multivariate Distributions from Markov Random Fields." *Journal of Multivariate Analysis* 73 (2). Elsevier: 199–220.

Liang, Faming. 2010. "A Double Metropolis–Hastings Sampler for Spatial Models with Intractable Normalizing Constants." *Journal of Statistical Computation and Simulation* 80 (9). Taylor & Francis: 1007–22.