MRF Model Specification

Naihui Zhou

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.
$Y_i \sim \text{Poisson}(\lambda_i)$
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 univariate approach

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

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

can be seen as an unknown constant containing α

The multivariate approach

$$
\boldsymbol{w} \sim mvN(\boldsymbol{\mu},\boldsymbol{\Sigma}) \tag{14}
$$

where $\boldsymbol{\Sigma} = (I-C)^{-}M$, $C=\{c_{ij}\}$, $c_{ij} = \begin{cases}\eta, & j\in N_i\\ 0, & o.w.\end{cases}$, $c_{ij}=c_{ji}$, $M = diag(\tau^2)$