Supplementary Materials for "Deviance Information Criteria for Mixtures of Distributions" by Chanmin Kim*

* Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118

Appendix A

In Equation (13),

$$\begin{aligned} E_{\theta|y_{(i)}}[p(y_i|\boldsymbol{\theta})] &= \int p(y_i|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{y}_{(i)})d\boldsymbol{\theta} \\ &= \int p(y_i|\boldsymbol{\theta})\frac{p(\mathbf{y}_{(i)}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{p(\mathbf{y}_{(i)})}d\boldsymbol{\theta} \\ &= \int \frac{p(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{p(\mathbf{y}_{(i)})}d\boldsymbol{\theta} \\ &= \frac{p(\mathbf{y})}{p(\mathbf{y}_{(i)})}, \end{aligned} \tag{1}$$

and

$$\begin{aligned} &V_{\theta|y_{(i)}}[p(y_i|\boldsymbol{\theta})] \\ &= \int \left\{p(y_i|\boldsymbol{\theta})^2 - 2p(y_i|\boldsymbol{\theta})E_{\theta|y_{(i)}}[p(y_i|\boldsymbol{\theta})] + E_{\theta|y_{(i)}}[p(y_i|\boldsymbol{\theta})]^2\right\}\frac{p(\mathbf{y}_{(i)}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{p(\mathbf{y}_{(i)})}d\boldsymbol{\theta} \\ &= \int \left\{p(y_i|\boldsymbol{\theta})^2 - 2p(y_i|\boldsymbol{\theta})\frac{p(\mathbf{y})}{p(\mathbf{y}_{(i)})} + \frac{p(\mathbf{y})^2}{p(\mathbf{y}_{(i)})^2}\right\}\frac{p(\mathbf{y}_{(i)}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{p(\mathbf{y}_{(i)})}d\boldsymbol{\theta} \\ &= \frac{1}{p(\mathbf{y}_{(i)})}\int p(y_i|\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} - 2\frac{p(\mathbf{y})}{p(\mathbf{y}_{(i)})^2}\int p(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} + \frac{p(\mathbf{y})^2}{p(\mathbf{y}_{(i)})^2}\times\frac{p(\mathbf{y}_{(i)})}{p(\mathbf{y}_{(i)})} \\ &= \frac{p(\mathbf{y})}{p(\mathbf{y}_{(i)})}p(y_i|\mathbf{y}) - 2\frac{p(\mathbf{y})^2}{p(\mathbf{y}_{(i)})^2} + \frac{p(\mathbf{y})^2}{p(\mathbf{y}_{(i)})^2}, \end{aligned} \tag{2}$$

where Equations (1) and (2) follow from the assumption that y_i's are conditionally independent of each other given $\boldsymbol{\theta}$. Then it is easy to calculate that

$$\frac{V_{\theta|y_{(i)}}[p(y_i|\boldsymbol{\theta})]}{E_{\theta|y_{(i)}}[p(y_i|\boldsymbol{\theta})]^2} = \frac{p(y_i|\mathbf{y}) - p(y_i|\mathbf{y}_{(i)})}{p(y_i|\mathbf{y}_{(i)})}.$$

Appendix B

The weight of importance sampling, b_h^i, can be specified as

$$\begin{aligned} b_h^i &= \frac{\pi(\boldsymbol{\theta}_h \mid \mathbf{y}_{(i)})}{\pi(\boldsymbol{\theta}_h \mid \mathbf{y})} \\ &= \frac{p(\mathbf{y}_{(i)}|\boldsymbol{\theta}_h)\,\pi(\boldsymbol{\theta}_h)}{f(\mathbf{y}_{(i)})} \bigg/ \frac{p(\mathbf{y}|\boldsymbol{\theta}_h)\,\pi(\boldsymbol{\theta}_h)}{f(\mathbf{y})} \\ &= \frac{f(\mathbf{y})\,p(\mathbf{y}_{(i)}|\boldsymbol{\theta}_h)}{f(\mathbf{y}_{(i)})\,p(\mathbf{y}|\boldsymbol{\theta}_h)} \\ &= \frac{f(y_i|\mathbf{y}_{(i)})}{p(y_i|\boldsymbol{\theta}_h, \mathbf{y}_{(i)})}, \end{aligned} \tag{3}$$

where

$$\begin{aligned} f(y_i|\mathbf{y}_{(i)}) &= 1 \bigg/ \int \frac{p(\mathbf{y}_{(i)}, \boldsymbol{\theta})}{p(\mathbf{y}, \boldsymbol{\theta})} \pi(\boldsymbol{\theta} \mid \mathbf{y}) d\boldsymbol{\theta} \\ &= 1 \bigg/ \int \frac{1}{p(y_i|\boldsymbol{\theta}, \mathbf{y}_{(i)})} \pi(\boldsymbol{\theta} \mid \mathbf{y}) d\boldsymbol{\theta} \\ &\approx 1 \bigg/ \frac{1}{J} \sum_{j=1}^{J} \frac{1}{p(y_i \mid \boldsymbol{\theta}_j, \mathbf{y}_{(i)})} . \end{aligned}$$

Under the conditional independence of y_i given $\boldsymbol{\theta}$, $p(y_i \mid \boldsymbol{\theta}_h, \mathbf{y}_{(i)})$ simplifies to $p(y_i \mid \boldsymbol{\theta}_h)$. Then Equation (3) is rewritten as

$$b_h^i \approx \frac{1}{\left\{ \frac{1}{J} \sum_{j=1}^{J} \frac{1}{p(y_i \mid \boldsymbol{\theta}_j)} \right\} p(y_i \mid \boldsymbol{\theta}_h)}.$$

Appendix C : BUGS code for the Bayesian nonparametric model in the Galaxy dataset example.

```
for( i in 1 : N ) {
    S[i] ~ dcat(pi[])
    y[i] ~ dnorm(mu[S[i]], tau[S[i]])
}
# Precision Parameter
alpha ~ dunif(0.1,5)
# Constructive DPP
p[1] <- r[1]
for (j in 2 : C) {
    p[j] <- r[j] * (1 - r[j - 1]) * p[j -1 ] / r[j - 1]
}
p.sum <- sum(p[])
for (j in 1:C){
    mu[j] ~ dnorm(mu0, tau0)
    tau[j] ~ dgamma(1,1)
    sigma[j] <- 1/tau[j]
    r[j] ~ dbeta(1, alpha)
    # scaling to ensure sum to 1
    pi[j] <- p[j] / p.sum
}
# hierarchical prior on theta[i] or preset parameters
mu0 ~ dnorm(0,0.001)
tau0 ~dgamma(0.1, 0.1)
```