UCLA Biostatistics 285: Homework 3

Instructor: Dr. Michele Guindani

Soumyakanti Pan, e-mail: span18@ucla.edu

June 14, 2023

1 Problem 1

1.1 Part 1: Hierarchical parametric model

Suppose y_{ijt} denotes the measurement of blood pressure of the j-th patient in the i-th hospital at time-point t where, $t = 1, \dots, T_{ij}$, $j = 1, \dots, n_i$ and $i = 1, \dots, 15$. The set of covariates include age, smoking status and others which to my belief are not varying over time, If X_{ij} encodes an intercept and the independent variables corresponding to the j-th patient in the i-th hospital and time, then we can consider a parametric model as follows assuming that the observations can be modelled using a mixed effects model. For $t = 1, \dots, T_{ij}$, $j = 1, \dots, n_i$ and $i = 1, \dots, 15$,

$$
\begin{aligned}
y_{ijt} \,|\, \boldsymbol{\beta}, \mathbf{b}_{ij}, \alpha_i, \sigma_\epsilon^2 &\overset{\text{iid}}{\sim} \mathcal{N}\left(X_{ij}^\top \boldsymbol{\beta} + \alpha_i + b_{ij}^{(0)} + b_{ij}^{(1)} t, \sigma_\epsilon^2\right) \\
\mathbf{b}_{ij} \,|\, \mathbf{D} &\sim \mathcal{N}(\mathbf{0}, \mathbf{D}), j = 1, \dots, n_i, i = 1, \dots, 15 \\
\alpha_i \,|\, \sigma_\alpha^2 &\sim \mathcal{N}(0, \sigma_\alpha^2), i = 1, \dots, 15 \\
\boldsymbol{\beta} \,|\, \mu_\beta, \mathbf{V} &\sim \mathcal{N}(\mu_\beta, \mathbf{V});\ \sigma_\epsilon^2 \sim \mathcal{IG}(a_1, b_1)
\end{aligned}
$$

1.2 Part 2: Heterogeneity

To make the model allow the distribution of the varying parameters to be unknown to allow uncertainty in variability (heterogeneity) among patients as well as patients within study centers, in addition to the model above we can treat the prior parameters like $\{\sigma_\alpha^2, \mathbf{V}, \mathbf{D}\}$ as unknown and consider a hyperprior on these parameters such as follows.

$$
\begin{aligned}
\sigma_\alpha^2 &\sim \mathcal{IG}(a_2, b_2) \\
\mathbf{V} &\sim \mathcal{IW}(q_1, k_1 \mathbf{I}) \\
\mathbf{D} &\sim \mathcal{IW}(q_2, k_2 \mathbf{I})
\end{aligned}
$$

We can choose the hyperparameters q_1, q_2 and k_1, k_2 in alignment with the prior belief of the amount of correlation that can be present in the corresponding covariance matrices of $\boldsymbol{\beta}$ and $\mathbf{b}_{ij}$.

1.3 Part 3: Gibbs sampler

The model above can be written in the following notations for the j-th patient of the i-th hospital in the following way.

$$\mathbf{y}_{ij} \sim \mathcal{N}(\mathbf{X}_{ij}^\top \boldsymbol{\beta} + \alpha_i \mathbf{1} + \mathbf{Z}_{ij}^\top \mathbf{b}_{ij}, \sigma_\epsilon^2 \mathbf{I})$$

where, $\mathbf{X}_{ij} = \mathbf{1} \otimes X_{ij}$ and $\mathbf{Z}_{ij}$ is appropriately considered with an intercept and the time points. If aggregated over i and j, we can write the model as

$$\mathbf{y} \,|\, \boldsymbol{\alpha}, \mathbf{b}, \sigma_\epsilon^2 \sim \mathcal{N}(\mathbf{X}^\top \boldsymbol{\beta} + \boldsymbol{\alpha} + \mathbf{Z}^\top \mathbf{b}, \sigma_\epsilon^2 \mathbf{I})$$

where the vectors $\boldsymbol{\alpha}$ and $\mathbf{b}$ are appropriately stacked. An appropriate semiparametric model can be given as follows.

$$\begin{aligned}
y_{ijt} \mid \boldsymbol{\beta}, \mathbf{b}_{ij}, \alpha_i, \sigma_\epsilon^2 &\overset{\text{iid}}{\sim} \mathcal{N}\left(X_{ij}^\top \boldsymbol{\beta} + \alpha_i + b_{ij}^{(0)} + b_{ij}^{(1)} t, \sigma_\epsilon^2\right) \\
\mathbf{b}_{ij} &\sim \mathrm{DP}(\phi_b, G_b), j = 1, \ldots, n_i, i = 1, \ldots, 15 \\
\alpha_i &\sim \mathrm{DP}(\phi_\alpha, G_\alpha), i = 1, \ldots, 15 \\
\boldsymbol{\beta} &\sim \mathrm{DP}(\phi_\beta, G_\beta);\ \sigma_\epsilon^2 \sim \mathcal{IG}(a_1, b_1) \\
\text{where, } \phi_{(\cdot)} &\sim \mathrm{Gamma}(c_., \lambda_.) \\
G_b &= \mathcal{N}_2(\mathbf{0}, \mathbf{D});\ G_\alpha = \mathcal{N}(0, \sigma_\alpha^2);\ G_\beta = \mathcal{N}_p(\mathbf{0}, \mathbf{V}) \\
\sigma_\alpha^2 &\sim \mathcal{IG}(a_2, b_2) \\
\mathbf{V} &\sim \mathcal{IW}(q_1, k_1 \mathbf{I}) \\
\mathbf{D} &\sim \mathcal{IW}(q_2, k_2 \mathbf{I})
\end{aligned}$$

In the Gibbs sampler for the parametric model, the full conditionals of the parameters $\boldsymbol{\beta}, \{\alpha_i\}_{i=1}^{15}, \{\mathbf{b}_{ij}\}$ will be either univariate or multivariate normal whereas, for the semiparametric model, the atoms for the corresponding parameters will be sampled from a similar normal distribution and the weights from a beta distribution.

2 Problem 2

2.1 Part 1: Simulation of data

The function `normmixture` generates a finite sample from a mixture of k normal distribution given vectors of mixture probability $\mathbf{p} = \{p_i\}_{i=1}^k$, component means $\boldsymbol{\mu} = \{\mu_i\}_{i=1}^k$ and variances $\boldsymbol{\tau}^2 = \{\tau_i^2\}_{i=1}^k$. Here, $\mathbf{p} = (0.1, 0.5, 0.4)$, $\boldsymbol{\mu} = (-1, 0, 1)$ and $\boldsymbol{\tau}^2 = (0.2^2, 1, 0.4^2)$.

2.2 Part 2: Frequentist density estimation

The following density estimation is carried out by the `density()` routine in `R`. The function `truemixnorm` calculates the actual density of a given mixture of normal distributions.

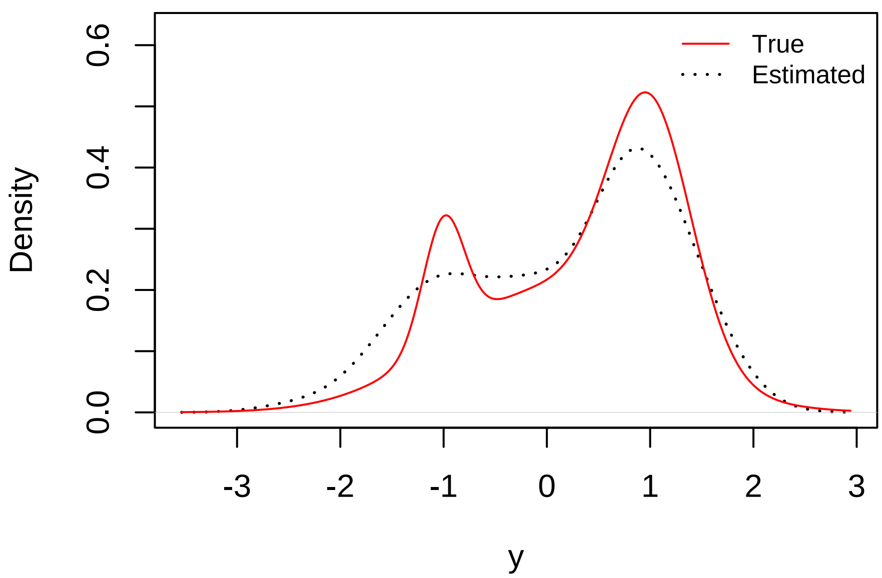

Figure 1: Density estimation based on 100 samples from a mixture-normal using the density() routine in R

2.3 Part 3: Blocked Gibbs Sampler

A Bayesian nonparametric approach to model the data ($\mathbf{y} = (y_1, \ldots, y_n)$) is a Dirichlet process mixture which is given below.

$$
\begin{aligned}
y_i \,|\, P &\overset{\text{ind}}{\sim} \int \phi(\cdot \,|\, \mu, \tau^2) dP(\mu, \tau^2) \\
P &\sim \text{DP}(\alpha, G_0)
\end{aligned}
$$

where, ϕ denotes the Gaussian kernel. Assuming a finite mixture of a DPM_H with $H = 20$ mixture components, we can write a hierarchical model as follows.

$$
\begin{aligned}
y_i \,|\, r_i = h &\sim \mathcal{N}(\mu_h, \tau_h^2) \text{ with } \Pr(r_i = h) = w_h, h = 1, \ldots, H \\
\text{where, } w_h &= v_h \prod_{\ell < h} (1 - v_\ell), h < H \text{ and } v_H = 1 \\
v_h &\sim \text{Beta}(1, \alpha), h = 1, \ldots, H \\
\mu_h \,|\, \tau_h^2 &\sim \mathcal{N}(\mu_0, \kappa), \;\; \tau_h^2 \sim \mathcal{IG}(a_\tau, b_\tau)
\end{aligned}
$$

where, $\mathbf{r} = (r_1, \ldots, r_n) \in \{1, \ldots, H\}^n$ denotes the cluster membership indicator vector. Suppose $\mathbf{v} = (v_1, \ldots, v_{H-1})$, $\mathbf{m} = (m_1, \ldots, m_H)$ and $\boldsymbol{\tau}^2 = (\tau_1^2, \ldots, \tau_n^2)$ and $\boldsymbol{\theta} = \{\boldsymbol{m}, \tau^2\}$. Given the above approximate model, the blocked Gibbs sampler algorithm to sample from $\pi(\mathbf{r}, \mathbf{w}, \boldsymbol{\theta} \,|\, \mathbf{y})$ proceeds as follows.

Algorithm 1: Blocked Gibbs sampler for posterior $\pi(\mathbf{r}, \mathbf{w}, \boldsymbol{\theta} \,|\, \mathbf{y})$

Result: posterior samples $(\mathbf{r}^{(t)}, \mathbf{w}^{(t)}, \mathbf{m}^{(t)}, \boldsymbol{\tau}^{2(t)})_{t=0}^T$
Initialize: Start with $\mathbf{w}^{(0)} = (1/H, \ldots, 1/H)$ and suitable random $\mathbf{r}^{(0)}, \mathbf{m}^{(0)}, \boldsymbol{\tau}^{2(0)}$.

while $t = 0, 1, \ldots, T$ **do**

1. Sample $\mathbf{r}^{(t+1)}$ with
$$\Pr(r_i^{(t+1)} = h) \propto w_h^{(t)} \phi(y_i \,|\, m_h^{(t)}, \tau_h^{2(t)}), h = 1, \ldots, H.$$
2. Sample $\mathbf{w}^{(t+1)}$ by the stick-breaking construction $w_h^{(t+1)} = v_h^{(t+1)} \prod_{\ell=1}^{h-1} (1 - v_\ell^{(t+1)})$ for $h = 1, \ldots, H$ after sampling $\mathbf{v}^{(t+1)}, h = 1, \ldots, H-1$ as
$$v_h^{(t+1)} \sim \text{Beta}(A_h^{(t)} + 1, B_h^{(t)} + \alpha)$$
where, $A_h^{(t)} = \sum_{i=1}^n \mathbf{1}(r_i^{(t)} = h)$ and $B_h^{(t)} = \sum_{i=1}^n \mathbf{1}(r_i^{(t)} > h)$.
3. Let $S_h^{(t)} = \{i : r_i^{(t)} = h\}$. Then sample $\mathbf{m}^{(t+1)}$ as follows. For $h = 1, \ldots, H$,
$$m_h^{(t+1)} \sim \mathcal{N}\left(\frac{\frac{|S_h^{(t)}|}{\tau_h^{2(t)}} \bar{y}_{S_h^{(t)}} + \frac{\mu_0}{\kappa}}{\frac{|S_h^{(t)}|}{\tau_h^{2(t)}} + \frac{1}{\kappa}}, \frac{1}{\frac{|S_h^{(t)}|}{\tau_h^{2(t)}} + \frac{1}{\kappa}} \right).$$
4. Sample $\boldsymbol{\tau}^{2(t+1)}$ as following. For $h = 1, \ldots, H$,
$$\tau_h^{2(t+1)} \sim \mathcal{IG}\left(a_\tau + \frac{|S_h^{(t)}|}{2}, b_\tau + \frac{1}{2} \sum_{i \in S_h^{(t)}} \left(y_i - m_h^{(t)} \right)^2 \right).$$

end

The function `bnp_blockgibbs` runs a blocked Gibbs sampler and outputs a chain containing relevant parameters.

```
Running Blocked Gibbs sampler with options:
```

```
n.iter =  1111 , Burn-in =  111.1
Prior: DPM(α, G_0)
Hyperpriors: G_0 = N(μ_0 ,κ)
        μ_0  =  0 , κ =  1 ;
         α =  1 ; a_τ =  1 , b_τ =  1 .
==========================================
Elapsed 1.37 secs
```

2.4 Part 4: Slice Sampler

I spent a lot of time and tried to figure out the algorithm as given in Walker (2007) and Kalli, Griffin, and Walker (2011). I could not understand the algorithms for updating the collection of random variables $\{v_h\}_{h\geq 1}$ at which the algorithms given in these two papers differ. While Walker (2007) uses the truncated beta construction, Kalli, Griffin, and Walker (2011) uses the same update rule as we have seen in the block sampler. I understand that we do not need to generate only finitely many (v_h, m_h, τ_h^2) at each iteration and continue for $h = 1, \ldots, N^*$ where $N^* = \max_h N_i$ and $N_i = \min\{k : \sum_{h=1}^k w_h > 1 - u_i\}$. Hence, if we sample N^* many weights, then the sampling of $\mathbf{r}$ is guaranteed. Walker (2007) mentions a k^* twice in his algorithm - once, when sampling $p(v_j \mid v_{-j}, \cdots)$ and next, when sampling the cluster indicator variables for calculating the set $A_w(u_i)$; are they the same k^*?

2.5 Part 5: Bayesian nonparametric density estimates

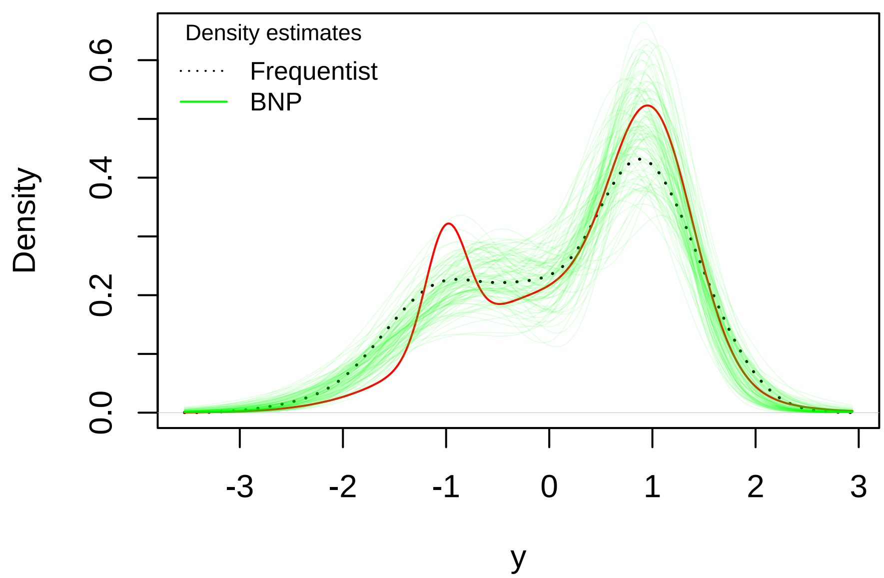

Figure 2: Blocked Gibbs sampler: Posterior samples of the predictive density given the generated 100 samples

References

Kalli, Maria, Jim E. Griffin, and Stephen G. Walker. 2011. "Slice Sampling Mixture Models." *Statistics and Computing* 21 (1): 93–105. https://doi.org/10.1007/s11222-009-9150-y.

Walker, Stephen G. 2007. "Sampling the Dirichlet Mixture Model with Slices." *Communications in Statistics - Simulation and Computation* 36 (1): 45–54. https://doi.org/10.1080/03610910601096262.

Algorithm 2: Slice sampler for posterior $\pi(\mathbf{u}, \mathbf{r}, \mathbf{w}, \boldsymbol{\theta} \,|\, \mathbf{y})$

Result: Posterior samples $(\mathbf{u}^{(t)}, \mathbf{r}^{(t)}, \mathbf{w}^{(t)}, \mathbf{m}^{(t)}, \boldsymbol{\tau}^{2(t)})_{t=0}^{T}$
Initialize: Start with suitable random $\mathbf{v}^{(0)}, \mathbf{r}^{(0)}, \mathbf{m}^{(0)}, \boldsymbol{\tau}^{2(0)}$.

while $t = 0, 1, \ldots, T$ **do**

1. For $i = 1, \ldots, n$, sample u_i as below.
$$u_i^{(t)} \sim \mathrm{U}(0, w_{r_i^{(t)}})$$
2. Let $S_h^{(t)} = \{i : r_i^{(t)} = h\}$. Then sample $\mathbf{m}^{(t+1)}$ as follows. For $h = 1, \ldots, H$,
$$m_h^{(t+1)} \sim \mathcal{N}\left(\frac{\frac{|S_h^{(t)}|}{\tau_h^{2(t)}} \bar{y}_{S_h^{(t)}} + \frac{\mu_0}{\kappa}}{\frac{|S_h^{(t)}|}{\tau_h^{2(t)}} + \frac{1}{\kappa}}, \frac{1}{\frac{|S_h^{(t)}|}{\tau_h^{2(t)}} + \frac{1}{\kappa}} \right).$$
3. Sample $\boldsymbol{\tau}^{2(t+1)}$ as following. For $h = 1, \ldots, H$,
$$\tau_h^{2(t+1)} \sim \mathcal{IG}\left(a_\tau + \frac{|S_h^{(t)}|}{2}, b_\tau + \frac{1}{2} \sum_{i \in S_h^{(t)}} \left(y_i - m_h^{(t)} \right)^2 \right).$$
4. Find $r^* = \max \mathbf{r}^{(t)}$. For $h = 1, 2, \ldots, r^*$, first calculate c_h and d_h where,
$$c_h = \max_{r_i^{(t)} = h} \left\{ \frac{u_i^{(t)}}{\prod_{\ell < h}(1 - v_\ell)} \right\} \text{ and, } d_h = \max_{r_i^{(t)} > h} \left\{ \frac{u_i^{(t)}}{v_{r_i^{(t)}} \prod_{l < r_i^{(t)}, l \neq h}(1 - v_l)} \right\}.$$
Then, draw a uniform random variable U on $(0, 1]$, and then for $h = 1, \ldots, r^*$, set $v_h = F^{-1}(U)$, where
$$F^{-1}(u) = 1 - [(1 - y)(1 - c_h)^\alpha + y(1 - d_h)^\alpha]^{1/\alpha}$$
is the inverse of the cdf on $c_h < v_h < d_h$, given by
$$F(v_h) = \frac{(1 - c_h)^\alpha - (1 - v_h)^\alpha}{(1 - c_h)^\alpha - (1 - d_h)^\alpha}.$$
5. From $\{v_h, h = 1, \ldots, r^*\}$, calculate $\{w_h^{(t+1)}, h = 1, \ldots, r^*\}$.
6. For $i = 1, \ldots, n$, find the set $A_w(u_i^{(t)}) = \{h : w_h > u_i^{(t)}\}$. For each i, draw $r_i^{(t+1)}$ as
$$\Pr(r_i^{(t+1)} = h \,|\, \cdots) \propto \mathbf{1}\left(h \in A_w(u_i^{(t)})\right) \phi(y_i \,|\, m_h^{(t+1)}, \tau_h^{2(t+1)})$$

end