Homework 4

PUBH 8442: Bayes Decision Theory and Data Analysis

Include any code used to generate answers at the end of your assignment.

1. Under the normal-normal hierarchical model described in the Hierarchical Models slide set, show that the posterior distribution for μ is

$$p(\mu \mid \mathbf{y}) = \text{Normal}(\hat{\mu}, V_\mu)$$

where

$$\hat{\mu} = \frac{\sum_{i=1}^m (\sigma_i^2 + \tau^2)^{-1}\bar{y}_i}{\sum_{i=1}^m (\sigma_i^2 + \tau^2)^{-1})} \text{ and } V_\mu = \left[\sum_{i=1}^m (\sigma_i^2 + \tau^2)^{-1}\right]^{-1}.$$

Under the normal-normal hierarchical model,

$$\bar{y}_i|\theta_i \sim N(\theta_i, \sigma_i^2)$$
$$\theta_i|\mu \sim N(\mu, \tau^2)$$

where σ^2, τ^2 are known and $p(\mu) = 1$.

Note that

$$\bar{y}_i = \theta_i + \epsilon_1, \quad \epsilon_1 \sim N(0, \sigma_i^2)$$
$$\theta_i = \mu + \epsilon_2, \quad \epsilon_2 \sim N(0, \tau^2)$$
$$\bar{y}_i = \mu + \epsilon_2 + \epsilon_1 = \mu + \epsilon^*, \quad \epsilon^* \sim N(0, \sigma_i^2 + \tau^2)$$

where $\epsilon_1 \perp \epsilon_2$. Thus, $\bar{y}_i \sim N(\mu, \sigma_i^2 + \tau^2)$. So, by

$$\begin{aligned} p(\mu|\mathbf{y}) &\propto p(\mu) \cdot p(\mathbf{y}|\mu) \\ &\propto 1 \cdot \prod_{i=1}^m N(\bar{y}_i|\mu, \sigma_i^2 + \tau^2) \\ &\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^m \frac{1}{\sigma_i^2 + \tau^2}(\mu - \bar{y}_i)^2\right\} \\ &\propto N(\hat{\mu}, V_\mu) \end{aligned}$$

where

$$\hat{\mu} = \frac{\sum_{i=1}^m (\sigma_i^2 + \tau^2)^{-1}\bar{y}_i}{\sum_{i=1}^m (\sigma_i^2 + \tau^2)^{-1}}$$
$$V_\mu = \left\{\sum_{i=1}^m (\sigma_i^2 + \tau^2)^{-1}\right\}^{-1}$$

(Alternatively, we could show the result by repeated applications of the normal-normal model: first condition on $\mathbf{y}_1$, then update the normal posterior for μ by conditioning on $\mathbf{y}_2$, etc.)

2. (a) conditional posterior distribution for each $p(\beta_k|\mathbf{y}, \beta_0)$:
Note that

$$\mathbf{y}_k = X_k\beta_k + \epsilon_1, \qquad \epsilon_1 \sim N(0, \sigma^2\mathbf{I})$$
$$\beta_k = \beta_0 + \epsilon_2 \qquad \epsilon_2 \sim N(0, \sigma^2\tau_k^2\mathbf{I})$$

where $\epsilon_1 \perp \epsilon_2$.
If $\beta_0 = 0$, this is a direct application of the result in Bayesian Linear Models slide 17. Note that

$$\mathbf{y}_k - X_k\beta_0 = X_k\epsilon_2 + \epsilon_1$$

so we can apply the result on slide 17 to derive the posterior of ϵ_2 for $\mathbf{y}_k^* = \mathbf{y}_k - X_k\beta_0$:

$$p(\epsilon_2|\sigma^2, \tau_k^2, \beta_0, \mathbf{y}) = \text{Normal}\left(\tilde{\beta}, \sigma^2 V_\beta\right)$$

where $\tilde{\beta} = (X_k^T X_k + (\tau_k^2\mathbf{I})^{-1})^{-1}(X^T\mathbf{y}_k^*)$ and $V_\beta = (X_k^T X_k + (\tau_k^2\mathbf{I})^{-1})^{-1}$. It follows that

$$\beta_k \sim \text{Normal}\left(\beta_0 + \tilde{\beta}, \sigma^2 V_\beta\right).$$

(b) conditional posterior distribution for $p(\beta_0|\mathbf{y}, \{\beta_k\}_{k=1}^K)$.

We start with the joint posterior distribution:

$$p(\mathbf{y}, \sigma^2, \beta_0, \beta_k) = N(\beta_0|\mathbf{0}, \sigma^2\tau_0^2\mathbf{I}) \times \prod_{k=1}^{K} N(\beta_k|\beta_0, \sigma^2\tau_k^2\mathbf{I}) \times \prod_{k=1}^{K} N(\mathbf{y}_k|X_k\beta_k, \sigma^2\mathbf{I})$$

We only need to consider terms that contain β_0. All terms not including β_0 can be considered constants. This gives us:

$$\begin{aligned} p(\beta_0|\mathbf{y}, \{\beta_k\}_{k=1}^K) &\propto N(\beta_0|\mathbf{0}, \sigma^2\tau_0^2\mathbf{I}) \times \prod_{k=1}^{K} N(\beta_k|\beta_0, \sigma^2\tau_k^2\mathbf{I}) \\ &= \prod_{i=1}^{p} N(\beta_{0i}|\mathbf{0}, \sigma^2\tau_0^2) \times \prod_{k=1}^{K} N(\beta_{ki}|\beta_{0i}, \sigma^2\tau_k^2). \\ &= p(\beta_{0i}|\{\beta_k\}_{k=1}^K) \end{aligned}$$

So the conditional posterior does not depend on $\mathbf{y}$ and the coefficients β_{0i} are independent. Applying the univariate normal-normal result gives $p(\beta_{0i}|\{\beta_k\}_{k=1}^K) = N(C_i c_i, C_i)$ where

$$C_i = \left((\sigma^2\tau_0^2)^{-1} + \sum_{k=1}^{K}(\sigma^2\tau_k^2)^{-1}\right)^{-1}$$
$$c_i = \left(\sum_{k=1}^{K}\beta_{ki}^T(\sigma^2\tau_k^2)^{-1}\right)^T$$

In matrix form, $p(\beta_0|\mathbf{y}, \{\beta_k\}_{k=1}^K) \sim N(Cc, C)$ where

$$C = \left((\sigma^2\tau_0^2\mathbf{I})^{-1} + \sum_{k=1}^{K}(\sigma^2\tau_k^2\mathbf{I})^{-1}\right)^{-1}$$
$$c = \left(\sum_{k=1}^{K}\beta_k^T(\sigma^2\tau_k^2\mathbf{I})^{-1}\right)^T$$

(c) conditional posterior distribution for a new observation from a new group $p(y_{k+1}|\mathbf{y}, \beta_0, \{\beta_k\}_{k=1}^K)$. For this problem we consider β_0 and X_k to be constants and known. There is only 1 observation in this group. The easiest method to solve this problem relies on the use of writing out the equations.

$$\begin{aligned} y_{k+1} &= X_{k+1}\beta_{k+1} + \epsilon_1 & \epsilon_1 &\sim N(0, \sigma^2) \\ \beta_{k+1} &= \beta_0 + \epsilon_2 & \epsilon_2 &\sim N(0, \sigma^2\tau_k^2\mathbf{I}) \end{aligned}$$

again where $\epsilon_1 \perp \epsilon_2$. This fact is important because it tells us that we do not need to consider covariance terms when we plug β_{k+1} into the equation for y_{k+1}. Then, we arrive at

$$y_{k+1} = X_{k+1}(\beta_0 + \epsilon_2) + \epsilon_1$$

Now, we simply take the expected value and variance with respect to y_{k+1} for the quantity of the right hand side. Specifically, we find $p(y_{k+1}|\mathbf{y}, \beta_0, \{\beta_k\}_{k=1}^K) \sim N(\mu_y, \Sigma_y)$ where

$$\mu_y = X_{k+1}\beta_0$$

$$\Sigma_y = X_{k+1}(\sigma^2\tau_{k+1}^2\mathbf{I})X_{k+1}^T + \sigma^2$$