More on Hierarchical Models

PUBH 8442: Bayes Decision Theory and Data Analysis

Eric F. Lock
UMN Division of Biostatistics, SPH
elock@umn.edu

3/3/2021

Normal-normal-normal... model

- Assume

$$
\begin{aligned}
y_1, \ldots, y_n &\sim \text{Normal}(\theta_1, \sigma_1^2) \\
\theta_1 &\sim \text{Normal}(\theta_2, \sigma_2^2) \\
\theta_2 &\sim \text{Normal}(\theta_3, \sigma_3^2) \\
&\vdots \\
\theta_I &\sim \text{Normal}(\mu^*, \tau^2).
\end{aligned}
$$

- Then, $p(\theta_1) \sim \text{Normal}(\mu^*, \sigma_2^2 + \sigma_3^2 + \sigma_I^2 + \tau^2)$

$$\theta_l = \mu^* + Z_l, \quad Z_l \sim N(0, \tau^2)$$

$$\theta_{l-1} = \underbrace{\mu^* + Z_l}_{\theta_l} + Z_{l-1}, \quad Z_{l-1} \sim N(0, \sigma_l^2)$$

$$\vdots \qquad\qquad\qquad\qquad \vdots$$

$$\theta_1 = \underbrace{\mu^* + Z_l + Z_{l-1} + \dots + Z_2}_{\theta_2} + Z_1, \quad Z_1 \sim N(0, \sigma_2^2)$$

Hierarchical normal model (known variance)

- Assume $y_{ij} \sim \text{Normal}(\theta_i, \sigma^2)$ for
  - Groups $i = 1, \ldots, m$
  - Observations $j = 1, \ldots, n_i$ for group i
- Assume σ^2 is known, and $\theta_i' s$ are iid $\text{Normal}(\mu, \tau^2)$
- Assume τ^2 is known, and give prior $p(\mu)$ for μ.
- Note that for $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})$

$$p(\mathbf{y}_i \mid \theta_i, \sigma^2) \propto \text{Normal}(\bar{y}_i \mid \theta_i, \sigma_i^2)$$

where $\sigma_i^2 = \sigma^2 / n_i$. So it suffices to consider the group means $\bar{y}_i$.

$$\prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_{ij}-\theta_i)^2} \propto e^{-\frac{1}{2(\sigma^2/n)}(\bar{y}-\theta_i)^2}$$

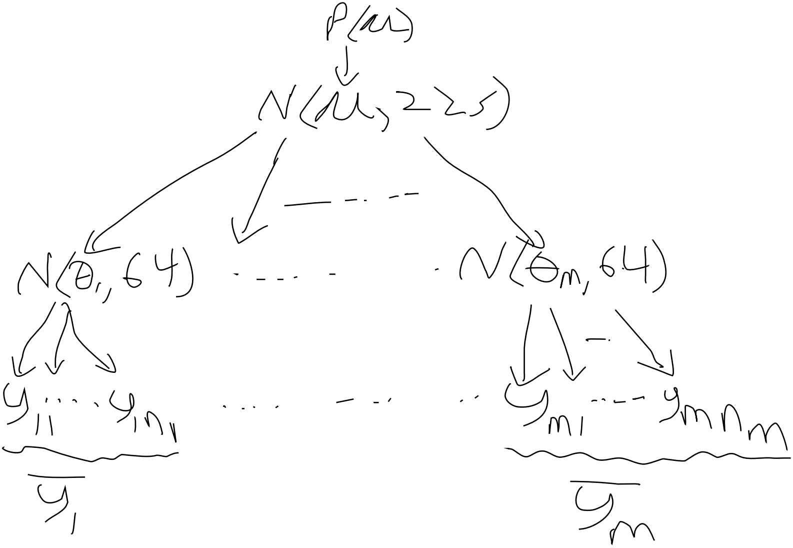
Hierarchical normal model (known variance)

- The joint posterior distribution is

$$
\begin{aligned}
p(\boldsymbol{\theta}, \mu \mid \mathbf{y}) &\propto p(\boldsymbol{\theta}, \mu, \mathbf{y}) \\
&= \prod_{i=1}^{m} \text{Normal}(\mathbf{y}_i \mid \theta_i, \sigma^2) \cdot \prod_{i=1}^{m} \text{Normal}(\theta_i \mid \mu, \tau^2) \cdot p(\mu) \\
&\propto \prod_{i=1}^{m} \text{Normal}(\bar{y}_i \mid \theta_i, \sigma_i^2) \cdot \prod_{i=1}^{m} \text{Normal}(\theta_i \mid \mu, \tau^2) \cdot p(\mu)
\end{aligned}
$$

Example: IQ scores

- Human IQs have variance 225 and are to be centered at 100
- Infer the "calibration" μ of a certain IQ test
- Sample of 20 individuals, each takes the test n_i times.
- Model:
  - $y_{ij} \sim \text{Normal}(\theta_i, 64)$ for $i = 1, \ldots, m$, $j = 1, \ldots, n_i$
  - $\theta_i \sim \text{Normal}(\mu, 225)$ for $i = 1, \ldots, m$
  - $p(\mu) = 1$ for all μ:
- Data: http://www.ericfrazerlock.com/IQData.csv


p(μ)
N(μ, 225)
N(θ1, 64)
N(θm, 64)
y11 ... y1n1
ym1 ... ymnm
ȳ1
ȳm

Example: IQ scores

- The posterior distribution for μ is

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

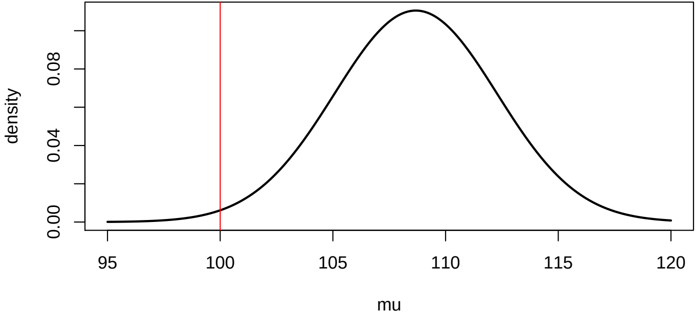
where

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

- Homework

Example: IQ scores

- Posterior for μ is Normal$(108.7, 13.0)$:



http://www.ericfrazerlock.com/More_on_Hierarchical_Models_Rcode2.r

- $P(\mu \geq 100 \mid \mathbf{y}) \approx 0.99$: suggests test is too generous