STAT 525 Lecture 14

October 17, 2017

1 Hierarchical Models

Topics

1. Hierarchical Poisson
2. Hierarchical Gaussian
3. Poisson for variance components
4. Hierarchical EF

1.1 The Hospital Example in lecture 13:

$$
\begin{aligned}
[y_h \mid \lambda_h] &\overset{\text{indep}}{\sim} \text{Poisson}\left(x_h \lambda_h\right) \\
[\lambda_h \mid \alpha, \mu] &\overset{\text{indep}}{\sim} \text{Gamma}\left(\alpha, \alpha/\mu\right) \\
[\lambda_h \mid y, \alpha, \mu] &\sim \text{Gamma}\left(\alpha + y_h, \frac{\alpha}{\mu} + x_h\right)
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbb{E}\left[\lambda_h \mid y_h, \alpha, \mu\right] &= \left(1 - \kappa_h\right) \frac{y_h}{x_h} + \kappa_h \mu \\
\kappa_h &= \frac{\alpha}{\alpha + x_h \mu}
\end{aligned}
$$

Two version of Gibbs sampling

$$
\left[\lambda_h \mid y, \mu\right], \left[\mu \mid \left\{\lambda_h\right\}, \boldsymbol{y}\right]
$$

$$
\left[\mu, \left\{\lambda\right\}_h \mid y\right] = \left[\left\{\lambda_h\right\} \mid \boldsymbol{y}, \mu\right] \left[\mu \mid \boldsymbol{y}\right]
$$

1.2 Hierarchical Gaussian Models

$$
\begin{aligned}
y_{ij} &= \theta_j + \epsilon_{ij} \\
\theta_j &= \mu + \upsilon_j \\
\epsilon_{ij} &\sim N\left(0, \sigma_y^2\right) \\
\upsilon_j &\sim N\left(0, \sigma_\theta^2\right)
\end{aligned}
$$

$$
\begin{aligned}
J &= \#\text{ of experiments} \\
n_j &= \#\text{ replicates with in experiment } j
\end{aligned}
$$

$$
[y_{ij} \mid \theta_j, \sigma_y] \overset{\text{indep}}{\sim} N\left(\theta_j, \sigma_y^2\right)
$$

where $i = 1, \cdots, n_j$ and $j = 1, \cdots, J$.

$$
[\bar{y}_{\cdot j} \mid \theta_j, \sigma_y] \sim N\left(\theta_j, \sigma_y^2/n_j\right)
$$

where

$$
\bar{y}_{\cdot j} = \sum_{i=1}^{n_j} y_{ij}
$$

Prior for θ_j

$$
[\theta_j \mid \mu, \sigma_\theta] \overset{\text{indep}}{\sim} N\left(\mu, \sigma_\theta^2\right)
$$

Prior for $(\sigma_y, \mu, \sigma_\theta)$

1. σ_y is standard (Jeffrey; I-G)
2. μ is standard (flat-prior, diffuse conjugate)
3. σ_θ is challenging.

Full conditionals: let $\sigma_j = \sigma_y/\sqrt{n_j}$ (se for group j)

$$
[\bar{y}_{\cdot j} \mid \theta_j, \sigma_y] \sim N\left(\theta_j, \sigma_j^2\right)
$$

Therefore,

$$
[\theta_j \mid y, \mu, \sigma_\theta, \sigma_j] \sim N\left(Q_{\theta_j}^{-1}\ell_{\theta_j}, Q_{\theta_j}^{-1}\right)
$$

where

$$
\begin{aligned}
Q_{\theta_j} &= n_j/\sigma_y^2 + \sigma_\theta^{-2} = \sigma_j^{-2} + \sigma_\theta^{-2} \\
\ell_{\theta_j} &= \sigma_y^{-2}\sum_{i=1}^{n_j} y_{ij} + \sigma_\theta^{-2}\mu \\
&= \sigma_y^{-2}\bar{y}_{\cdot j} + \sigma_\theta^{-2}\mu
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left[\theta_j \mid y, \mu, \sigma_\theta, \sigma_j\right] &= \frac{\sigma_y^2 \sigma_\theta^2}{n_j \sigma_\theta^2 + \sigma_y^2} \left(\sigma_y^{-2} \sum_i y_{ij} + \sigma_\theta^{-2} \mu\right) \\
&= \frac{\left(\sigma_y^2 / n_j\right) \sigma_\theta^2}{\sigma_\theta^2 + \left(\sigma_y^2 / n_j\right)} \left(\sigma_y^{-2} n_j \bar{y}_{\cdot j} + \sigma_\theta^{-2} \mu\right) \\
&= \frac{\sigma_j^2 \sigma_\theta^2}{\sigma_\theta^2 + \sigma_j^2} \left(\sigma_j^{-2} \bar{y}_{\cdot j} + \sigma_\theta^{-2} \mu\right) \\
&= \left(1 - \kappa_{\theta_j}\right) \bar{y}_{\cdot j} + \kappa_{\theta_j} \mu
\end{aligned}
$$

where

$$
\kappa_{\theta_j} = \frac{\left(\sigma_y^2 / n_j\right)}{\sigma_\theta^2 + \left(\sigma_y^2 / n_j\right)} = \frac{\sigma_j^2}{\sigma_\theta^2 + \sigma_j^2}
$$

We explain the above as follows:

$$
\begin{aligned}
\sigma_\theta &\to \infty \text{ no pooling} \\
\sigma_\theta &\to 0 \text{ complete pooling}
\end{aligned}
$$

Assume

$$
p(\mu) \propto 1
$$

Full condition

$$
\left[\mu \mid y, \{\theta_j\}, \sigma_\theta, \sigma_y\right] \sim N\left(Q_\mu^{-1} \ell_\mu, Q_\mu^{-1}\right)
$$

where

$$
\begin{aligned}
Q_\mu &= J \sigma_\theta^{-2} \\
\ell_\mu &= \sigma_\theta^{-2} \sum_{j=1}^{J} \theta_j = \sigma_\theta^{-2} J \bar{\theta}
\end{aligned}
$$

Note that

$$
\left[\bar{y}_{\cdot j} \mid \mu, \sigma_\theta\right] \overset{\text{indep}}{\sim} N\left(\mu, \sigma_j^2 + \sigma_\theta^2\right)
$$

Now

$$
\left[\mu \mid y, \sigma_\theta, \sigma_j\right] \sim N\left(\tilde{Q}_\mu^{-1} \tilde{\ell}_\mu, \tilde{Q}_\mu^{-1}\right)
$$

where

$$
\begin{aligned}
\tilde{Q}_\mu &= \sum_{j=1}^{J} \frac{1}{\sigma_j^2 + \sigma_\theta^2} \\
\tilde{\ell}_\mu &= \sum_{j=1}^{J} \frac{\bar{y}_{\cdot j}}{\sigma_j^2 + \sigma_\theta^2}
\end{aligned}
$$

Two Algorithms

1. Consider varying σ_θ
2. Gibbs sampler
   (a) $\left[\theta_j \mid \mu, y, \sigma_\theta^2, \sigma_y^2\right]$
   (b) $\left[\mu \mid y, \{\theta_j\}, \sigma_\theta^2, \sigma_y^2\right]$
3. $\left[\{\theta_j\}, \mu \mid y, \sigma_\theta^2, \sigma_y^2\right] = \left[\{\theta_j\} \mid \mu, y, \sigma_\theta^2, \sigma_y^2\right] \times \left[\mu \mid y, \sigma_\theta^2, \sigma_y^2\right]$

1.3 Prior for σ_θ (See 5.7)

Choice matters when J small or σ_θ is small (group level variation small)

Setting improper prior as limit of proper priors.

Similar to conjugate $\approx$ Jeffreys.

Uniform $\sigma_\theta \sim \text{Uniform}\,(0, A)$ as $A \to \infty$ (proper post whenever $J \geq 3$)

Inversed-Gamma $\sigma_\theta^{-2} \sim \text{Gamma}\,(\epsilon, \epsilon)$ as $\epsilon \to 0$ (improper post)

If $\epsilon \to 0$ the IG is non-informative but improper. If σ_θ is small, choice of ϵ can be highly informative.

Uniform Priors

$$\log \sigma_\theta \sim \text{Uniform}\,(-A, A)$$

1. Improper as $A \to \infty$
2. for fixed A, prior depends heavily on A

Marginal Likelihood

$$p\,(y \mid \sigma_\theta) \to K$$

as $\sigma_\theta \to 0$, then $0 < K < \infty$.

- The data can't role out group variance of zeros.
- prior can't put infinite mass near zero.

Uniform on σ_θ

1. finite integral near zero
2. proper for $J \geq 3$
3. miss-calibration toward larger values

Uniform on σ_θ^2

1. Proper for $J \geq 4$
2. more extreme miss-calibration

Half t, Half normal, half Cauchy, conditional conjugate

$$\sigma_\theta \sim C^+\,(0, A)$$

as $A \to \infty$, $\sigma_\theta \sim \text{Uniform}\,(0, \infty)$