STAT 525 Lecture 13

October 17, 2017

1 Hierarchical Models

1.1 Hierarchical / Multilevel methods applications

Example: Home Radon Measurements

EPA collected data on >80,000 homes in U.S. Goal: Estimate distribution of random levels at all counties $\approx 3,000$

Data: Houses within counties

1. Log radon for house i
2. Basement / first floor (house - level)
3. Soil uranium (Country - level)

Example: Election forecasting

Data: States within the US

1. D-party shares of votes for 11 past elections for each state.
2. Previous selection performance by state.
3. National economic trends.
4. National opinions polls.

Standard GLM predictions are reasonable.

- The corresponding prediction intervals were underestimated variability.
- State level errors are correlated.

1.2 Hierarchical / Multilevel methods uses

1. Learning about treatment effects that vary by group (example: patients within a hospital and students within a school.)
2. Using all the data as much as possible, if you perform inference on group with small samples size.
3. As far as prediction goes, you might have new units with existing group or new groups entirely.
4. You are looking at analysis of structured data. That is students within school or cluster sampling.
5. More efficient inference for regression parameter.
6. You can also include predictors at different levels.
7. You can accurately account for uncertainty (by modeling correlations.)

1.3 Additional layers of uncertainty

Priors on hyper-parameter

$$\begin{aligned} y &\sim p\left(y \mid \theta\right) \\ \theta &\sim p\left(\theta; \phi\right) \end{aligned}$$

or

$$\begin{aligned} y &\sim p\left(y \mid \theta\right) \\ \theta &\sim p\left(\theta \mid \phi\right) \\ \phi &\sim \pi\left(\phi\right) \end{aligned}$$

Example:

$$\begin{aligned} \theta \mid \tau_\mu &\sim N\left(0, \tau_\mu^{-1}\right) \\ \tau_\mu &\sim \text{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right) \end{aligned}$$

which is equivalent to the t distribution with ν degree of freedom when you try to marginalize τ_μ.

$$p\left(\theta\right) \sim t_\nu\left(0, 1\right)$$

Another way is that

$$\begin{aligned} y &\sim p\left(y \mid \theta\right) \\ \theta &\sim p\left(\theta \mid \phi\right) \\ \phi &\sim \pi\left(\phi\right) \end{aligned}$$

is equivalent to

$$\begin{aligned} y &\sim p\left(y \mid \theta\right) \\ \left(\theta, \phi\right) &\sim p\left(\theta, \phi\right) = p\left(\theta \mid \phi\right) \pi\left(\phi\right) \end{aligned}$$

Then the posterior distribution is

$$\begin{aligned} p\left(\theta, \phi \mid y\right) &\propto p\left(y \mid \theta, \phi\right) p\left(\theta, \phi\right) \\ &\propto p\left(y \mid \theta\right) p\left(\theta \mid \phi\right) \pi\left(\phi\right) \end{aligned}$$

Full condition:

$$\begin{aligned} p\left(\theta \mid y, \phi\right) &\propto p\left(y \mid \theta\right) p\left(\theta \mid \phi\right) \\ p\left(\phi \mid y, \theta\right) &\propto p\left(\theta \mid \phi\right) \pi\left(\phi\right) \end{aligned}$$

The above is a construction of a Gibbs sampler.

Alternatively, we can break this posterior distribution this way

$$p\left(\theta, \phi \mid y\right) = p\left(\theta \mid \phi, y\right) p\left(\phi \mid y\right)$$

1. Draw $\phi^* \sim p\left(\phi \mid y\right)$ (need to marginalize over θ)

2. Draw $\theta^* \sim p(\theta \mid y, \phi = \phi^*)$ (full conditional distribution)

Example: Gaussian $y \sim N\left(\theta, \phi^{-1}\right)$

How to compute $p(\phi \mid y)$?

$$
\begin{aligned}
p(\phi \mid y) &= \int p(\phi, \theta \mid y)\, d\theta \\
&\propto \int p(y \mid \theta)\, p(\theta \mid \phi)\, \pi(\phi)\, d\theta \\
&= \frac{p(\phi, \theta \mid y)}{p(\theta \mid \phi, y)}
\end{aligned}
$$

Exchangeability:

$$
\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_T)
$$

The joint density $p(\theta_1, \theta_2, \cdots, \theta_T)$ is invariant to permutations of $\{1, \cdots, T\}$

Marginalization exchangeable

$$
p(\boldsymbol{\theta} \mid \phi) = \prod_{j=1}^{T} p(\theta_j \mid \phi)
$$

we prove that θ is marginally exchangeable.

$$
\begin{aligned}
p(\boldsymbol{\theta}) &= \int p(\boldsymbol{\theta} \mid \phi)\, p(\phi)\, d\phi \\
&= \int \left\{ \prod_{j=1}^{T} p(\theta_j \mid \phi) \right\} p(\phi)\, d\phi
\end{aligned}
$$

1.4 Heart transplant data

Goal: Model success rate of heart transplant surgeries in U.S. for many hospitals jointly

Data: for hospitals $h = 1, \cdots, H = 94$

$$
\begin{aligned}
y_h &= \#\text{ deaths within 30 days of surgery (for hospital } h) \\
X_h &= \text{ units of exposure (\# of patients)} \\
\lambda_h &= \text{ mortality rate per unit of exposure}
\end{aligned}
$$

Model

$$
\begin{aligned}
y_h \mid X_h &\overset{\text{iid}}{\sim} \text{Poisson}(X_h \lambda_h) \\
E[y_h \mid X_h] &= \lambda_h
\end{aligned}
$$

No Pooling:

Model λ_h independently e.g.

$$
\lambda_h \overset{\text{iid}}{\sim} \text{Gamma}(\alpha_h, \beta_h)
$$

Individual MLE

$$
\hat{\lambda}_h = \frac{y_h}{X_h}
$$

1. can be poor estimates when X_h small or $y_h = 0$
2. can be highly variable
3. ignore likely dependence among $\{\lambda_h\}$

<u>Complete Pooling:</u> All hospital shares the same $\lambda = \lambda_1 =, \cdots, = \lambda_H$

$$\begin{aligned} y_h \mid \lambda &\overset{\text{iid}}{\sim} \text{Poisson}(\lambda) \\ E[y_h \mid X_h] &= \lambda \end{aligned}$$

pooled MLE

$$\hat{\lambda} = \frac{\sum_{h=1}^{H} y_h}{\sum_{h=1}^{H} X_h}$$

1. Ignore hospital to hospital variation
2. Strong assumption

<u>Hierarchical Model:</u> Partially Pooling

Consider the exchangeable priors

$$\begin{aligned} [\lambda_h \mid \alpha, \mu] &\overset{\text{indep.}}{\sim} \text{Gamma}(\alpha, \alpha/\mu) \\ E[\lambda_h \mid \alpha, \mu] &= \mu \\ \text{Var}[\lambda_h \mid \alpha, \mu] &= \mu^2/\alpha \end{aligned}$$

<u>The posterior for λ_h</u>

$$[\lambda_h \mid y_h, \alpha, \mu] \sim \text{Gamma}\left(\alpha + y_h, \frac{\alpha}{\mu} + X_h\right)$$

The posterior mean is

$$\begin{aligned} E[\lambda_h \mid y_h, \alpha, \mu] &= \frac{\alpha + y_h}{a/\mu + X_h} \\ &= (1 - \kappa_h) \frac{y_h}{X_h} + \kappa_h \mu \end{aligned}$$

where

$$\kappa_h = \frac{\alpha}{\alpha + X_h \mu} \in [0, 1]$$

<u>The posterior for μ</u>

$$\left[\mu^{-1}\right] \sim \text{Gamma}(\alpha_\mu, \beta_\mu)$$

$$\begin{aligned} \left[\mu^{-1} \mid \{\lambda_h\}, \alpha\right] &\propto \left[\mu^{-1}\right]^{\alpha_\mu - 1} \exp\left(-\beta_\mu \left[\mu^{-1}\right]\right) \prod_{h=1}^{H} \frac{(\alpha/\mu)^\alpha}{\Gamma(\alpha)} \lambda_h^{\alpha-1} \exp\left(-[\alpha/\mu]\lambda_h\right) \\ &\propto \left[\mu^{-1}\right]^{\alpha_\mu + \alpha H - 1} \exp\left\{-\left(\beta_\mu + \alpha \sum_{h=1}^{H} \lambda_h\right)\left[\mu^{-1}\right]\right\} \\ &\sim \text{Gamma}\left(\alpha_\mu + \alpha H, \beta_\mu + \alpha \sum_{h=1}^{H} \lambda_h\right) \end{aligned}$$

Then the posterior mean is

$$E\left[\mu \mid \{\lambda_h\}, \alpha\right] = \frac{\beta_\mu + \alpha \sum_{h=1}^{H} \lambda_h}{\alpha_\mu + \alpha H - 1}$$

So if $\beta_\mu \to 0$, $\alpha_\mu = 1$, then

$$E\left[\mu \mid \cdots\right] \approx \frac{\sum_{h=1}^{H} \lambda_h}{H}$$