STA 702: Random Effects

Merlise Clyde

Nov 17, 2022

Building Hierarchical Models

- Models for Gaussian Data with no Covariates

$$y_{ij} \sim ? \qquad i = 1, \ldots n; j = 1, \ldots, n_i$$

- i "block" - schools, counties, etc
- j observations within a block - students within schools, households within counties, etc
- potentially there may be within block dependence in the observations due to unmeasured covariates
- structure?

Models

- Naive model (baseline)

$$y_{ij} \stackrel{iid}{\sim} N(\mu, \sigma^2)$$

- issue: no systematic variation across blocks
- Fixed Effects model:

$$y_{ij} \stackrel{ind}{\sim} N(\mu_i, \sigma^2)$$

- Common reparameterization

$$y_{ij} \stackrel{ind}{\sim} N(\alpha + \beta_i, \sigma^2)$$

- μ intercept
- β_i shift for school
- Identifiability ?

Non-Identifiability

- Example: $y_i \sim N(\alpha + \beta_i, \sigma^2)$ overparameterized
- $\mu_i = \alpha + \beta_i$ and σ^2 are uniquely estimated, but not α or β_i
- $x_i \in \{1, \ldots, d\}$ factor levels

$$y_i \sim N(\mu + \sum_j \beta_j 1(x_i = j), \sigma^2)$$

  $\theta_j = \mu + \beta_j$ identifiable - d equations but $d+1$ unknowns

- Put constraints on parameters
  - $\alpha = 0$
  - $\beta_d = 0$
  - $\sum \beta_j = 0$

Bayesian Notion of Identifiability

- Bayesian Learning
- the posterior distribution differs from the prior
- **Note:** In general, it's good to avoid working with non-identifiable models;

Caveats:

- Forcing identifiability may involve (complex) constraints that bias parameter estimates and make MCMC less efficient
- sometimes purposely introduce non-identifiability to improve computation (parameter expansion PX)
- run non-identifiable model and focus on identifiable parameters or functions of them
- post-processing of output

Issue with Fixed Effect Approach

- What if n_i, number of observations per block, are small?
- Estimated uncertainty/variances are large based on MLE using group specific means
- What if blocks might be viewed as a sample from some larger population? Sample of schools?
- May want inference about the larger population and say things about future blocks!
- fixed effects do not allow us to say anything about blocks not in our sample!
- how to address this?

Random Effects

$$
\begin{aligned}
&y_{ij} = \alpha + \beta_i + \epsilon_{ij}, \qquad \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2) \\
&\beta_i \overset{iid}{\sim} N(0, \tau^2)
\end{aligned}
$$

- random effects β_j
- Random effect distributions should be viewed as part of the model specification
- We've specified the likelihood in a hierarchical manner to induce desirable structure
- unknown parameters are population parameters α, τ and σ^2
- Bayesians put prior distributions on α, τ and σ^2; frequentists don't!

Equivalent Model

$$y_i = (y_{i1}, y_{i2}, \dots, y_{in_i})$$

$$y_i \overset{ind}{\sim} N_{n_i}\left(\alpha 1_{n_i}, \begin{pmatrix} \sigma^2 + \tau & \tau & \dots & \tau \\ \tau & \ddots & & \tau \\ \vdots & & \ddots & \vdots \\ \tau & \dots & \tau & \sigma^2 + \tau \end{pmatrix} \right)$$

within-block correlation

- algorithmically we can use either the latent variable model or the collapsed (marginal) model for inferences;
- often latent variable is easier to work with!

Simple Gibbs Sampler

$$\theta = (\alpha, \tau, \sigma^2, \beta_1, \ldots, \beta_n)$$

$$\begin{aligned} \alpha &\sim N(\alpha_0, V_0) \\ \tau^{-1} &\sim \textsf{Gamma}(a_\tau/2, b_\tau/2) \\ \sigma^{-2} &\sim \textsf{Gamma}(a_\sigma/2, b_\sigma/2) \end{aligned}$$

Full Conditionals:

$$\begin{aligned} \alpha \mid \tau, \sigma^2, \beta_1, \ldots \beta_n &\sim N(\hat{\alpha}, \hat{V}_n) \\ \hat{V}_n &= \left(\frac{1}{V_0} + \sum_i \frac{n_i}{\sigma^2} \right)^{-1} \\ \hat{\alpha} &= \frac{\frac{\alpha_0}{V_0} + \frac{\sum_i n_i \bar{y}_i^*}{\sigma^2}}{\hat{V}_n^{-1}} \end{aligned}$$

$$y_{ij}^* \equiv y_{ij} - \beta_i \qquad \bar{y}_i^* \equiv \frac{\sum_j (y_{ij} - \beta_i)}{n_i}$$

Full Conditional Continued

$$\sigma^{-2} \mid \alpha, \tau, \beta_1, \ldots, \beta_n \sim \textsf{Gamma}\left(\frac{a_\sigma + \sum_i n_i}{2}, \frac{b_\sigma + \sum_i \sum_j (y_{ij} - \alpha - \beta_i)^2}{2}\right)$$

$$\tau^{-1} \mid \alpha, \sigma^2, \beta_1, \ldots, \beta_n \sim \textsf{Gamma}\left(\frac{a_\tau + n}{2}, \frac{b_\tau + \sum_i \beta_i^2}{2}\right)$$

$$\begin{aligned}
\beta_j \mid \alpha, \tau, \sigma^2 &\overset{ind}{\sim} N(\hat{b}_i, \hat{V}_{\beta_i}) \\
\hat{V}_{\beta_i} &= \left(\frac{1}{\tau} + \frac{n_i}{\sigma^2}\right)^{-1} \\
\hat{b}_i &= \frac{\frac{0}{\tau} + \frac{n_i \bar{y}_i^*}{\sigma^2}}{\hat{V}_{\beta_i}^{-1}} \\
y_{ij}^{**} \equiv y_{ij} - \alpha \qquad \bar{y}_i^{**} &\equiv \frac{\sum_j (y_{ij} - \alpha)}{n_i}
\end{aligned}$$

Complications Relative to Usual Regression

1. Prior Choice
2. Mixing and its dependence on parameterization

- Early recommendation after Gibbs Sampler introduced non-informative priors

$$\begin{aligned}
\pi(\alpha) &\propto 1 \\
\pi(\sigma^{-2}) &\sim \textsf{Gamma}(\epsilon/2, \epsilon/2) \qquad \pi(\sigma^{-2}) \propto 1/\sigma^{-2} \text{ as } \epsilon \to 0 \\
\pi(\tau^{-1}) &\sim \textsf{Gamma}(\epsilon/2, \epsilon/2) \qquad \pi(\tau^{-1}) \propto 1/\tau^{-1} \text{ as } \epsilon \to 0
\end{aligned}$$

- Are full conditionals proper ?

- Is joint posterior proper ?

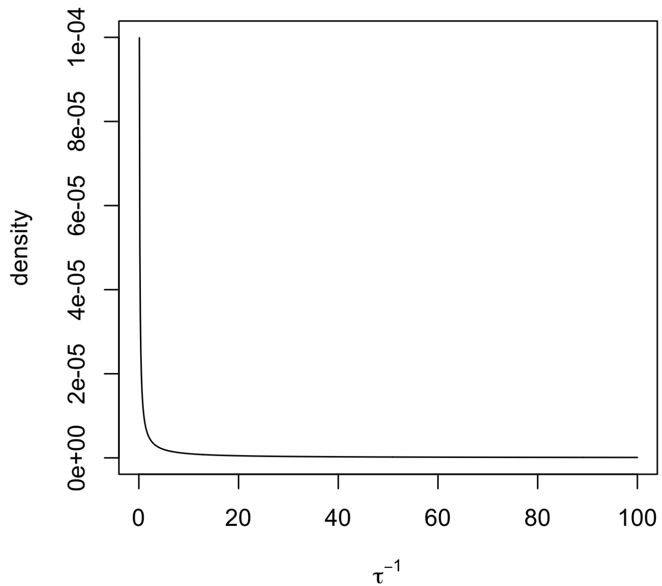
MCMC and Priors

- proper full conditionals
- joint is improper
- MCMC won't converge to the stationary distribution (doesn't exist)
- may not notice it!

Diffuse But Proper

$$\alpha \sim N(0, 10^{-6})$$
$$\pi(\sigma^{-2}) \sim \mathsf{Gamma}(10^{-6}, 10^{-6})$$
$$\pi(\tau^{-1}) \sim \mathsf{Gamma}(10^{-6}, 10^{-6})$$

- Nearly improper priors lead to terrible performance! highly sensitive to just how vague the prior is!



Alternative Priors

- Choose a flat or heavy tailed prior for random effect standard deviation $\tau^{1/2}$

$$
\begin{aligned}
y_{ij} &= \alpha + \beta_i + \epsilon_{ij} & & & y_{ij} &= \alpha + \lambda\eta_i + \epsilon_{ij} \\
& & \Leftrightarrow & & & \\
\beta_i &\overset{iid}{\sim} N(0, \tau) & & & \eta_i &\overset{iid}{\sim} N(0, 1)
\end{aligned}
$$

- Reparameterization

$$\eta_i = \frac{\beta_i}{\tau^{1/2}} \Rightarrow \frac{\beta_i}{\lambda} \sim N(0,1)$$

- $\pi(\lambda) \propto 1(\lambda > 0)$ (improper prior)
- $\pi(\lambda) \propto 1(\lambda > 0) N(0,1)$ folded standard normal (half-normal)
- $\pi(\lambda) \propto 1(\lambda > 0) N(0, 1/\psi) \qquad \psi \sim \text{Gamma}(\nu/2, \nu/2)$ folded t or half t

Proper Posterior

Work with

$$\pi(\mu, \tau, \sigma^2 \mid y) \propto \pi(\mu, \tau, \sigma^2) \prod_{i=1}^{n} N\left(y_i; \alpha 1_{n_i}, \begin{pmatrix} \sigma^2 + \tau & \tau & \dots & \tau \\ \tau & \ddots & & \tau \\ \vdots & & \ddots & \vdots \\ \tau & \dots & \tau & \sigma^2 + \tau \end{pmatrix}\right)$$

- take $\pi(\mu, \tau^{1/2}, \sigma^2) \propto \sigma^{-2}\, \mathsf{t}_1^+(\tau^{1/2}; 0, 1)$
- take $\pi(\mu, \tau^{1/2}, \sigma^2) \propto \sigma^{-2}$
- Show joint posterior is proper !
- See Gelman 2005 discussion of Draper paper in Bayesian Analysis

Propriety

- need expression for likelihood; requires determinant and inverse of intra-class correlation matrix! Write covariance as $\sigma^2 I_{n_i} + \tau n_1 P_1$ and find spectral decomposition to provide determinant and inverse!
- integrate out α (messy)
- determine if integrals are finite (what happens at 0 and infinity ?)
- look at special case when n_i are all equal.

Bayes Estimates of Variances

- Avoids issues when estimate of variance is on the boundary of the parmaeter space
- Do not have to use asymptotics to construct CI!

Linear Mixed Effects

$$y_{ij} = X_{ij}^T B + z_{ij}^T \beta_i + \epsilon_{ij}$$

- Fixed effects $X_{ij}^T B$
- Random effects $z_{ij}^T \beta_i$ with $\beta_i \overset{iid}{\sim} N(0, \Psi)$
- Designed to accomodate correlated data due to nested/hierarchical structure/repeated measurements
- students w/in schools; patients w/in hospitals
- As before not inherently Bayesian! It's just a model/likelihood specification!
- If θ is population parameters, $\theta = (B, \Psi, \sigma^2)$, find the marginal distribution for y_i given θ!