22S:138
Specifying priors on variance components
Intro to Generalized Linear Models

Lecture 17
Oct. 26-31, 2005

Kate Cowles
374 SH, 335-0727
kcowles@stat.uiowa.edu

Priors on variance components

- stage 2, formulation 1

$$\alpha_{0i}|\beta_0, \tau^2_{\alpha_0} \sim N(\beta_0, \tau^2_{\alpha_0})$$
$$\alpha_{1i}|\beta_1, \tau^2_{\alpha_1} \sim N(\beta_1, \tau^2_{\alpha_1})$$

- stage 2, multivariate formulation

$$\begin{bmatrix} \alpha_{0i} \\ \alpha_{1i} \end{bmatrix} | \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \Sigma_\alpha \sim N_2\left(\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \Sigma_\alpha^{-1} \right)$$

- $\tau^2_{\alpha_0}$ is precision of subject-specific intercepts
- $\tau^2_{\alpha_1}$ is precision of subject-specific slopes

Possible approaches to specifying priors on variance components

- fit a frequentist mixed model to a different but related dataset and get estimated variances of slopes and intercepts
- use results of someone else's study
- elicit expert opinion
- use common sense

Common-sense approach

- What do intercepts mean?
  - birthweight of baby rat in grams
  - $\sqrt[4]{CD4count}$ at study entry
- What would be a plausible but very small value for this quantity?
- Plausible but unusually large value?
- (large - small) $\simeq$ 4 standard deviations
- baby rats example
  - guess $\sigma_{\alpha_0} = 10$
  - guess $\sigma_{\alpha_1} = \sqrt{0.1}$
  - corresponding single number guess for $\tau^2_{\alpha_0}$
  - corresponding single number guess for $\tau^2_{\alpha_1}$

- How would you use this to construct Gamma priors?

- How did Gelfand *al.* use this to construct Wishart prior for second formulation?

$$\begin{aligned} R &= \Sigma_\alpha^{-1} \\ R &\sim dwish(Omega[], 2) \\ Omega &= \begin{bmatrix} 200 & 0 \\ 0 & 0.2 \end{bmatrix} \\ Omega^{-1} &= \begin{bmatrix} 1/200 & 0 \\ 0 & 1/0.2 \end{bmatrix} \\ E(R^{-1}) &= 2\ Omega^{-1} \\ &= \begin{bmatrix} 1/100 & 0 \\ 0 & 1/0.1 \end{bmatrix} \end{aligned}$$

- Better to have larger degrees of freedom and use procedure we outlined in previous lecture

Centering covariates and priors on variance components

- correlations between α_{i0}'s and α_{i1}s
- even if covariate is centered, random effects do not become independent in HNLM
- centering is straightforward if all subjects have the same set of values of predictor variable
- question: if subjects have different values of the predictor, should we center around
  - subject-specific average $(x_{ij} - x.bar_i)$
  - overall average $(x_{ij} - x.bar)$
- centering affects the priors on
  - β_0
  - $\tau^2_{\alpha_0}$ or Σ_α

Generalized linear models (GLMs)

- goal is to determine relationship between predictor variables and a response variable that is not normally distributed
  - may not even be quantitative
- example: vote on library proposal in Iowa City 2 years ago
  - subjects are individual voters
  - response variables Y_i are whether person voted for or against (coded 1 and 0)
  - predictors might be gender, income level, whether person has children, etc.

Why normal linear regression would not work for example

- $$E(Y_i) = \beta_0 + \beta_1 income_i$$
- there would always be values of income for which results would be negative or > 1
- residuals could not be normally distributed since real Y_i's are all 0's and 1's

Solution: logistic regression

- (special case of GLM)
- Y_i's actually have what distribution

- $E(Y_i) \;=$
- Choose an appropriate transformation of $E(Y_i)$ that can take on values anywhere on real line
- one good choice is logit

$$logit(p_i) \;=\; log(\left(= \frac{p_i}{1-p_i}\right)$$

- model *transformed* expectations as linear function of predictors

$$logit(p_i) \;=\; \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots$$

- there is no separate variance parameter at the level of the likelihood (why)

- so likelihood is completely specified in WinBUGS as

```
Y[i]  ~ dbern(p[i])
logit(p[i]) <- beta[0] + beta[1] * x[i,1]
```

and priors are needed only on βs

more general case of GLM

- conditional on model parameters and values of predictor variables, response variables Y_i are independent draws from distributions in a natural exponential family
- we are most interested in 2 natural exponential families
  - bernoulli (or binomial)
  - Poisson
- both are 1-parameter families
- $E(Y_i) \;=\; g^{-1}(\beta_0 \;+\; \beta_1 x_{1i} \;+\; \beta_2 x_{2i}$ where g is a link function
  - monotonic and differentiable
  - log is most commonly used link function for Poisson regression