

22S:138
Bayesian Statistics

Intro to Hierarchical Normal Linear Models

Lecture 16
 Nov. 1, 2004

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

Hierarchical normal linear models

- combine
 - hierarchical models
 - linear regression

Review of assumptions of linear regression

- Homoscedasticity
- Linearity
- Independence
- Normality
- Existence

Example: AIDS study ACTG116B/117

- randomized, controlled, double-blind clinical trial
- patients with at least 16 weeks of prior treatment with zidovudine (ZDV)
- each patient was randomized to one of 3 treatments
 - continued ZDV
 - 2 different dose levels of dd* (another antiretroviral drug)
- primary endpoint: progression to a new AIDS-defining event or death
- primary results published in Kahn et al. (1992)
- CD4 counts measured on all patients
 - at study entry (week 0) and at weeks 2, 8, 12, 16, 24, 32, 40, 48, 56, and 64

– your homework dataset consists of CD4 counts taken up to week 24 from a subset of the patients in 1 treatment group

Research question and statistical models

- two parameters of interest are average change in CD4 count per week in patients on each of the two treatments

– $\beta_{1,A}$, $\beta_{1,B}$, and $\beta_{1,C}$

- one approach: simple linear regression applied separately to patients in each treatment group

$$y_{ij} | \beta_{0,g}, \beta_{1,g}, \sigma^2 \sim N(\beta_{0,g} + \beta_{1,g} t_{ij}, \sigma^2)$$

where $g = A, B, \text{ or } C$

- which assumption of linear regression is violated?
- what are the likely consequences of the violation of this assumption?

Another possibility: separate linear regressions for each patient

- would result in poor estimation of individual slopes and intercepts since there are few data values for each patient
- then question arises of how to compute overall slope for treatment group
 - average?
 - weighted average?

Hierarchical normal linear model

- a compromise between
 - pooling all the data into one simple linear regression model
 - * would violate independence assumption
 - separate linear regressions for each patient
 - * would result in poor estimation of individual slopes and intercepts since there are few data values for each patient

- notation:
 - y_{ij} –(transformed) CD4 count measured on patient i at week t_{ij}
- stage 1: likelihood
 - for each patient i , $i = 1, \dots, N$

$$y_{ij} | \alpha_{0i}, \alpha_{1i}, \tau_y^2 \sim N(\alpha_{0i} + \alpha_{1i}t_{ij}, \tau_y^2)$$
 - where τ_y^2 is the precision of the points around the patient-specific regression line
- stage 2, formulation 1
 - $\alpha_{0i} | \beta_0, \tau_{\alpha_0}^2 \sim N(\beta_0, \tau_{\alpha_0}^2)$
 - $\alpha_{1i} | \beta_1, \tau_{\alpha_1}^2 \sim N(\beta_1, \tau_{\alpha_1}^2)$
- stage 2, multivariate formulation

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

11

- * Wishart is multivariate generalization of gamma
- * ρ is the “degrees of freedom”
 - determines the degree of certainty you have about the mean
 - for a Wishart distribution to be prior, ρ must be \geq dimension of the matrix
 - ρ is equivalent to prior sample size
- * Wishart distribution is parameterized several different ways
- * WinBUGS does not use the same parameterization as GCSR table
- * in WinBUGS parameterization

$$\Sigma_\alpha^{-1} \sim \text{Wishart}(R[2, 2], \rho)$$

implies that $E(\Sigma_\alpha)$ is ρR^{-1}

- third stage, first formulation
 - $\beta_0 \sim N(\mu_0, \tau_0^2)$
 - $\beta_1 \sim N(\mu_1, \tau_1^2)$
 - $\tau_y^2 \sim G(a_y, b_y)$
 - $\tau_{\alpha_0}^2 \sim G(a_{\alpha_0}, b_{\alpha_0})$
 - $\tau_{\alpha_1}^2 \sim G(a_{\alpha_1}, b_{\alpha_1})$
- third stage, multivariate formulation
 - $\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} | \begin{bmatrix} \mu_0 \\ \mu_1 \end{bmatrix} \Sigma_0 \sim N_2 \left(\begin{bmatrix} \mu_0 \\ \mu_1 \end{bmatrix}, \Sigma_0^{-1} \right)$
 - $\tau_y^2 \sim G(a_y, b_y)$
 - $\Sigma_\alpha^{-1} \sim \text{Wishart}(R[2, 2], \rho)$
 - where ρ is the degrees of freedom (scalar)
 - * equivalent prior sample size
 - * must be greater than or equal to dimension of matrix in order for Wishart to be proper
 - R is prior guess at order of *covariance* matrix Σ_α

12

Priors on precision matrices

- WinBUGS requires parameterizing models including an unknown variance/covariance matrix of a multivariate normal distribution in terms of the *precision matrix* (inverse of the variance/covariance matrix).
- The Wishart distribution is the conjugate prior for the precision matrix of a multivariate normal distribution with known mean.
- It is the standard choice of prior for precision matrices in realistic multivariate-normal-based models with means (and possibly many other parameters) unknown because it leads to a Wishart full conditional distribution for the precision matrix that simplifies MCMC-based model fitting.
- The two parameters of the Wishart distribution are a mean matrix and a scalar parameter called the degrees of freedom.

multiple parameterizations

Confusingly, several different parameterizations of the Wishart density appear in the literature. If X denotes a $p \times p$ symmetric, positive definite random matrix, R is a fixed $p \times p$ symmetric, positive definite matrix, ν is a strictly positive scalar, and the p.d.f. of X is

$$p(X|R, \nu) \propto |R|^{\frac{\nu}{2}} |X|^{\frac{\nu-p-1}{2}} \exp\left[-\frac{1}{2}tr(RX)\right] \quad (1)$$

then the references below define the two parameters as follows:

Reference	Parameterization
[?]	$X \sim dwish(R, \nu)$
[?]	$X \sim dwish(R^{-1}, \nu)$
[?]	$X \sim dwish(R^{-1}, \nu - p + 1)$

In what follows, we use the WinBUGS parameterization. The Wishart distribution is proper if $\nu \geq p$. If $X \sim dwish(R, \nu)$, then the moments are as follows:

$$\begin{aligned} E(X_{ij}) &= \nu(R^{-1})_{ij} \\ Var(X_{ij}) &= \nu[(R^{-1})_{ij}^2 + (R^{-1})_{ii}(R^{-1})_{jj}] \\ Cov(X_{ij}, X_{kl}) &= \nu[(R^{-1})_{ik}(R^{-1})_{jl} + (R^{-1})_{il}(R^{-1})_{jk}] \end{aligned}$$

Note that the gamma distribution is a special (one-dimensional) case of the Wishart. If X and R are scalars, and the p.d.f of X is proportional to $x^{\frac{\nu}{2}-1} \exp(-\frac{Rx}{2})$ then

$$W(R, \nu) = G\left(\frac{\nu}{2}, \frac{R}{2}\right)$$

WinBUGS does not allow the use of its Wishart distribution with one-dimensional matrices, however.

If $X \sim dwish(R, \nu)$, then X^{-1} has an *inverse Wishart* distribution: $X^{-1} \sim IW(R, \nu)$, where

$$E((X_{ij}^{-1})) = \frac{R_{ij}}{\nu - p - 1}$$

The inverse Wishart distribution is always proper; however, it has a degenerate form if $\nu < p$, and obviously the first moment is negative or infinite unless $\nu > p + 1$.

Since statisticians and subject-matter experts tend to be better able to think in terms of variances and correlations rather than of elements of precision matrices, the following way of specifying a prior on a covariance matrix, say Σ , in WinBUGS is attractive:

1. Let R equal the prior guess for the mean of the $p \times p$ *variance/covariance* matrix Σ .
2. Choose a degrees-of-freedom parameter ν ($> p + 1$) that roughly represents an “equivalent prior sample size” – your belief in R as the value of Σ is as strong as if you had seen ν previous vectors with sample covariance matrix R .
3. Define a matrix $S = (\nu - p - 1)R$.
4. In WinBUGS, put the following Wishart prior on the corresponding precision matrix Σ^{-1} :

```
Sigmainv[,] ~ dwish( S[,], nu )
```

5. then

- $E(\Sigma_{i,j}) = R_{i,j}$
- the variance of the prior will be decreasing in ν
- $E((\Sigma)_{i,j}^{-1}) = \frac{\nu}{\nu-p-1}(R^{-1})_{i,j}$

- HNLMs “borrow strength” from other subjects’ data to help in estimating subject-specific intercepts and slopes
 - useful when there are few data points for some or all individual subjects

Advantages of hierarchical normal linear models

- compared to
 - pooling the data (one single linear regression)
 - separate linear regressions for each patient
- HNLMs accommodate correlations within subject
- HNLMs accommodate differences between subjects with respect to
 - intercept
 - slope
- HNLMs provide appropriate posterior credible sets for the population intercept and slope

Requirements for posterior to be proper

- must have proper priors on
 - $\tau_{\alpha_0}^2$ and $\tau_{\alpha_1}^2$ or
 - Σ_{α}^{-1}
- at least 1 subject must have > 2 measurements

Centering covariates in hierarchical normal linear models

- 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
 - $\tau_{\alpha_0}^2$ or Σ_α