

Milk production of cows - Mixed model example

- Data on milk production from n cows.
- n_j cows are daughters of bull j . There are J sires in dataset. Sires "group" the cows into J different "genetic" groups.
- Other covariates are herd and age of cow.
- Exchangeability: given sire, age and herd, cows are exchangeable.
- In classical statistics, herd and age are "fixed" effect, and sire is random effect. For us, all random, but we allocate flat priors to "fixed" parameters.
- Cows that are sired by same bull are more similar than those sired by different bulls: intraclass correlation induced by models with random effects.

Cows (cont'd)

- Mixed model:

$$y_{ij} = x_{ij}'\beta + s_j + \epsilon_{ij}$$
 with $s_j \sim N(0, \sigma_s^2)$, $\epsilon_{ij} \sim N(0, \sigma^2)$ and (s, ϵ) independent. $x_i = (\text{herd}, \text{age})$ are herd and age effects, and $(\beta, \sigma_s^2, \sigma^2)$ are unknown.
- Likelihood:

$$y_j \sim N(x_j'\beta, \sigma_s^2 + \sigma^2)$$
- In matrix form:

$$y \sim N(X\beta, \sigma^2 I + \sigma_s^2 ZZ')$$
 with $X: n \times p$, $Z: n \times q$.
- Intra-class correlation: correlation between milk production of cows sired by same bull:

$$\rho = \frac{\sigma_s^2}{\sigma_s^2 + \sigma^2}$$

Hierarchical Linear Models

- Much of this material already seen in Chapters 5 and 14
- Hierarchical linear models combine regression framework with hierarchical framework
- Unified approach to:
 - random effects models
 - mixed models
 - Set-up:
 - Sampling model:

$$y_{ij} | \beta, \Sigma_j \sim N(X_{ij}\beta, \Sigma_j)$$
 - Often $\Sigma_j = \sigma^2 I$
 - Prior dist. for J regression coefficients

$$\beta_j | \alpha, \Sigma_j \sim N(X_j \alpha, \Sigma_j)$$
 - Typically, $X_j = 1$, $\Sigma_j = \sigma_j^2 I$

Hierarchical Linear Models

- Hyperprior on K parameters α :

$$\alpha | \alpha_0, \Sigma_\alpha \sim N(\alpha_0, \Sigma_\alpha)$$
 with α_0, Σ_α known, often $p(\alpha) \propto 1$.
- Also need priors for Σ_j, Σ_j .
- Finally, might need to set priors for the hyperparameters in the priors for Σ_j, Σ_j . Typically assume known and fixed.

View as J regression experiments

- Model for j th experiment is

$$y_j | \beta_j, \sigma_j^2 \sim N(X_j \beta_j, \sigma_j^2)$$
 with $y_j = (y_{1j}, y_{2j}, \dots, y_{n_j j})'$.
- Putting all regression models together into a single model:

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_J \end{bmatrix} = X = \begin{bmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & X_J \end{bmatrix} \times \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_J \end{bmatrix}$$
- Priors and hyperpriors, for example:

$$\beta_j | \alpha, \Sigma_j \sim N(1\alpha, \Sigma_j)$$

$$p(\alpha, \Sigma_j) \propto 1$$

$$\sigma_j^2 | a, b \sim \text{Inv} - \chi^2(a, b)$$
- Implied model is

$$y_j | \alpha, \sigma_j^2, \Sigma_j \sim N(X_j \alpha, \sigma_j^2 I + X_j \Sigma_j X_j')$$
- Note: the hierarchy induces a correlation.

Intra-class correlation

- Random effects introduce correlations
- Suppose that observations come from J groups or clusters so that $y = (y_1, y_2, \dots, y_J)$, and $y_j = (y_{1j}, y_{2j}, \dots, y_{n_j j})'$ as above.
- Model: $y \sim N(\alpha, \Sigma_y)$
- Let $\text{var}(y_{it}) = \eta^2$, and let

$$\text{cov}(y_{it}, y_{it'}) = \rho \eta^2$$
, for same group

$$\text{cov}(y_{it}, y_{it'}) = 0$$
, for different group
- For $\rho \geq 0$, now consider model $y \sim N(X\beta, \sigma^2 I)$, with X an $n \times J$ matrix of group indicators.
- If $\beta \sim N(\alpha, \sigma^2 I)$ and if we let $\eta^2 = \sigma^2 + \sigma_s^2$, then $\rho = \sigma_s^2 / (\sigma^2 + \sigma_s^2)$ and the two model formulations are equivalent.
- To see that models are equivalent, do $p(y) = \int p(y, \beta) d\beta$.

Example: growth curves in rats

- From Ceiland et al., 1990, JASA.
- CIBA-GEIGY measured the growth of 30 rats weekly, for five weeks. Interest is in growth curve.
- Assume linear growth (rats are young) and let:
 - y_j : weight of i th rat in j th week, $i = 1, \dots, 30$, $j = 1, \dots, 5$
 - $x_i = (8, 15, 22, 29, 36)$ days
- Each rat gets her "own" curve if α_i and β_i are random effects parameters.

$$y_{ij} = \alpha_i + \beta_i(x_{ij} - \bar{x}_i) + \epsilon_{ij}$$

Rats (cont'd)

- Likelihood:

$$y_{ij} \sim N(\mu_i, \sigma^2)$$
 with $\mu_i = \alpha_i + \beta_i(x_{ij} - \bar{x}_i)$.
- Population distributions:

$$\alpha_i \sim N(\alpha_0, \sigma_\alpha^2)$$

$$\beta_i \sim N(\beta_0, \sigma_\beta^2)$$
- Priors:

$$\sigma^2 \sim \text{Inv} - \chi^2(\nu, \sigma_0^2)$$

$$\alpha_0, \beta_0 \sim N(0, 0.1, 10000)$$

$$\sigma_\alpha^2, \sigma_\beta^2 \sim \text{Inv} - \chi^2$$
- Priors for $\sigma_\alpha^2, \sigma_\beta^2$ can be as non-informative as possible by having very small degrees of freedom parameter. Same for prior for σ^2 if desired.
- A more reasonable formulation is to model α_i, β_i as dependent in the population distribution.

Intra-class correlation

- Positive intra-class correlations can be accommodated with a random effects model where class membership is reflected by indicators whose regression coefficients have the population distribution

$$\beta \sim N(\mathbf{1}\alpha, \sigma_\beta^2 I)$$

- This is general formulation for several more general models

- Mixed effects models:

$$p(\beta_1, \dots, \beta_k) \propto 1 \rightarrow \text{"fixed"} \text{ effects}$$

$$p(\beta_{j_1}, \dots, \beta_j) \propto N(\mathbf{1}\alpha, \sigma_\beta^2 I) \rightarrow \text{random effects}$$

Mixed effects models

- A more general version of the mixed model has different random effects that generate different sets of intra-class correlations:

$$p(\beta_i) \propto 1, \quad i = 1, \dots, I$$

$$b_{j_1} | \alpha_1, \sigma_1^2 \sim N(\mathbf{1}\alpha_1, \sigma_1^2 I), \quad j_1 = 1, \dots, J_1$$

⋮

$$\beta_{j_k} | \alpha_k, \sigma_k^2 \sim N(\mathbf{1}\alpha_k, \sigma_k^2 I), \quad j_k = 1, \dots, J_k$$

- The J components of β are divided into K clusters
- Exchangeability at the level of the observations is achieved by conditioning on the indicators that define the clusters or groups

Computation

- Hierarchical linear models have nice structure for computation.
- With conjugate prior, recall that:
 - Observations are N
 - Regression parameters are N
 - Variance components (or variance matrices) are $Inu - \chi^2$ (or Wishart).
- All conditional distributions are of standard form:
 - For location parameters (regression coefficients, means of priors and hyperpriors), conditionals are normal
 - For scale parameters, conditionals are also $Inu - \chi^2$, even if prior is improper.
- For one example, go back to earlier lecture on Gibbs sampling and example therein.