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:

    $$y|\beta, \Sigma_y \sim N(X\beta, \Sigma_y)$$

    Often $\Sigma_y = \sigma^2 I$
  - Prior dist. for J regression coefficients

    $$\beta|\alpha, \Sigma_\beta \sim N(X_\beta \alpha, \Sigma_\beta)$$

    Typically, $X_\beta = 1, \;\; \Sigma_\beta = \sigma_\beta^2 I$

Hierarchical Linear Models

- Hyperprior on K parameters α:

  $$\alpha|\alpha_0, \Sigma_\alpha \sim N(\alpha_0, \Sigma_\alpha)$$

  with α_0, Σ_α knwown, often $p(\alpha) \propto 1$.
- Also need priors for Σ_y, Σ_β.
- Finally, might need to set priors for the hyperparameters in the priors for Σ_y, Σ_β. Typically assume known and fixed.

Example: growth curves in rats

- From Gelfand 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_{ij}: weight of ith rat in jth week, $ii = 1, ..., 30$, $j = 1, ..., 5$

  $x_i = (8, 15, 22, 29, 36)$ days

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

- Each rat gets her "own" curve if α_i and β_i are random effects parameters.

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:

$$\alpha_i \sim N(\alpha_0, \sigma^2_\alpha)$$
$$\beta_i \sim N(\beta_0, \sigma^2_\beta)$$

- Priors:

$$\sigma^2 \sim Inv - \chi^2(\nu, \sigma^2_0)$$
$$\alpha_0, \beta_0 \sim N(0.01, 10000)$$
$$\sigma^2_\alpha, \sigma^2_\beta sim Inv - \chi^2$$

- Priors for $\sigma^2_\alpha, \sigma^2_\beta$ 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.

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:

  $$y_{ij} = x_i'\beta + s_j + e_{ij}$$

  with $s_j \sim N(0, \sigma_s^2)$, $e_{ij} \sim N(0, \sigma^2)$ and (s, e) independent. $x_i = (herd, age)$ are herd and age effects, and $(\beta, \sigma_s^2, \sigma^2)$ are unknown.
- Likelihood:

  $$y_{ij} \sim N(x_i'\beta, \sigma_s^2 + \sigma^2)$$

- In matrix form:

  $$y\ N(X\beta, \sigma^2 I + \sigma_s^2 ZZ')$$

  with $X : n \times p$, $Z : n \times q$.
- Intra-class correlation: correlation between milk production of cows sired by same bull:

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

View as J regression experiments

- Model for jth experiment is

$$y_j|\beta_j, \sigma_j^2 \sim N(X_j\beta_j, \sigma_j^2)$$

  with $y_j = (y_{1j}, y_{2j}, ..., y_{n_j j})$.

- Putting all regression models together into a single model:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_J \end{bmatrix} = X = \begin{bmatrix} X_1 & 0 \ldots 0 \\ 0 & X_2 \ldots 0 \\ \vdots & \vdots \ldots \vdots \\ 0 \ldots & \ldots X_J \end{bmatrix} \times \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_J \end{bmatrix}$$

- Priors and hyperpriors, for example:

$$\begin{aligned} \beta_j|\alpha, \Sigma_\beta &\sim N(1\alpha, \Sigma_\beta) \\ p(\alpha, \Sigma_\beta) &\propto 1 \\ \sigma_j^2|a, b &\sim Inv-\chi^2(a, b) \end{aligned}$$

- Implied model is

$$y_j|\alpha, \sigma_j^2, \Sigma_\beta \sim N(X_j\alpha, \sigma_j^2 I + X_j\Sigma_\beta X_j')$$

- Note: the hierarchy induces a correlation.

Intra-class correlation

- Random effects introduce correlations

- Suppose that observations come from J groups or clusters so that $y = (y_1, y_2, ..., y_J)$, and $y_j = (y_{1j}, y_{2j}, ..., y_{n_j j}$ as above.

- Model: $y \sim N(\alpha, \Sigma_y)$

- Let $var(y_{ij}) = \eta^2$, and let

$$\begin{aligned} cov(y_{ij}, y_{kj}) &= \rho\eta^2, \text{ for same group} \\ cov(y_{ij}, y_{kl}) &= 0, \text{ for different group} \end{aligned}$$

- 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 $\beta \sim N(\alpha, \sigma_\beta^2 I)$ and if we let $\eta^2 = \sigma^2 + \sigma_\beta^2$, then $\rho = \sigma_\beta^2/(\sigma^2 + \sigma_\beta^2)$ and the two model formulations are equivalent.

- To see that models are equivalent, do $p(y) = \int p(y, \beta) d\beta$.

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

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

- This is general formulation for several more general models
- Mixed effects models:

$$\begin{aligned} p(\beta_1, ..., \beta_{J_1}) &\propto 1 \rightarrow \text{ ``fixed'' effects} \\ p(\beta_{J_{1+1}}, ..., \beta_J) &\propto N(1\alpha, \sigma_\beta^2 I) \rightarrow \text{random effects} \end{aligned}$$

Mixed effects models

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

$$\begin{aligned} p(\beta_i) &\propto 1, \;\; i = 1, ..., I \\ b_{j_1} | \alpha_1, \sigma_1^2 &\sim N(1\alpha_1, \sigma_1^2 I), \;\; j_1 = 1, ..., J_1 \\ &\vdots \\ \beta_{j_k} | \alpha_k, \sigma_k^2 &\sim N(1\alpha_k, \sigma_k^2 I), ;; j_k = 1, ..., J_k \end{aligned}$$

- 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 $Inv-\chi^2$ (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 $Inv-\chi^2$, even if prior is improper.
- For one example, go back to earlier lecture on Gibbs sampling and example therein.