The Normal Hierarchical Regression Model

m = 100 schools. Y_{ij} = Score of student i in school $j = 1, \dots, m$.

Chapter 8:

$\boxed{\mu_1}$ $\boxed{\mu_2}$ . . . . . . $\boxed{\mu_{100}}$

School 1 School 2 school 100

Is there a difference in average score? (between schools)

Chapter 11:

$\boxed{Y_{i1} = \beta_{01} + \beta_{11} X_{i1}}$ $\boxed{Y_{i2} = \beta_{02} + \beta_{12} X_{i2}}$ - - - - $\boxed{Y_{i100} = \beta_{0(100)} + \beta_{1(100)} X_{i100}}$

School 1 School 2. School 100

Are the regressions (relation SES and score) different between schools?

General model for any # of indep. variables. (1)

Y_j = vector of scores of n_j children in school $j = 1, \dots, m$. $m = 100$

X_j = Matrix ($n_j \times p$) with covariates school j

β_j = Regression coefficient vector for school j.

Model within group (within school)

$$Y_j \sim MVN(X_j \beta_j, \sigma^2 I). \quad j = 1, \dots, 100$$

σ^2 = error variance.

Notice: μ_j is now $X_j \beta_j$.

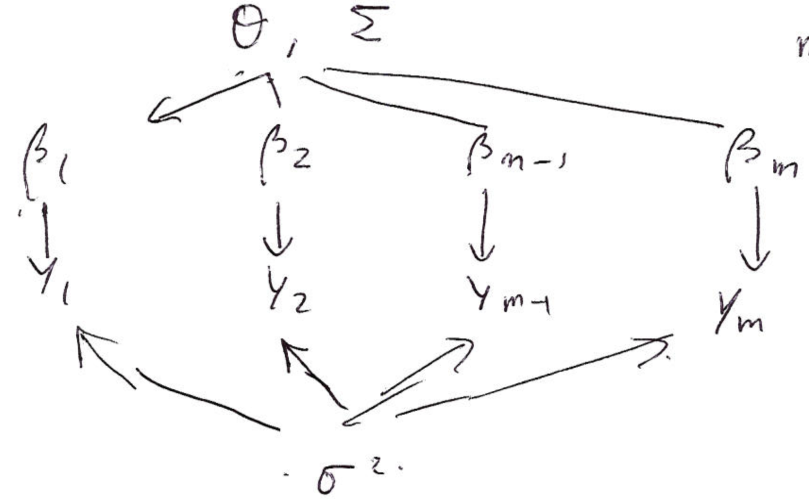
Model ~~within~~ between schools.

$$\beta_j \sim MVN(\theta, \Sigma).$$

θ = Population mean.

Σ = Population variance

Unknowns: $\theta, \Sigma, \beta_J, \sigma^2$



$m = 1, \dots, 100$

<u>Alternative parametrization</u>: Linear mixed <u>effect model</u>. (2)

$\beta_J = \theta + \gamma_J$. $\quad Y_{i,J} = \theta^T \vec{x}_{iJ} + \gamma^T \vec{x}_{iJ} + \varepsilon_{iJ}$.

$\gamma_1 \dots \gamma_m \sim$ iid $MVN(0, \Sigma)$.

θ = fixed effects $\quad \gamma_1 \dots \gamma_m$ = random effects.

Regressors could be different for fixed and random effects.

$$Y_{iJ} = \theta^T X_{iJ} + \gamma_J Z_{iJ} + \varepsilon_{iJ}.$$

<u>Regardless of parametrization</u>

Given Prior for θ, Σ, σ^2,

Having observed $\vec{Y}_1, \vec{Y}_2, \dots \vec{Y}_m$ $\quad m = 100$.

Posterior distribution

$$P(\beta_1, \dots, \beta_m, \theta, \Sigma, \sigma^2 \mid X_1, \dots X_m, Y_1, \dots Y_m)$$

Many parameters.

Full conditional Distributions

Information across groups is shared via
θ, Σ, and σ^2.

Conditional on θ, Σ, σ^2, the $\vec{\beta}_J$ elements
$\vec{\beta}_1, \vec{\beta}_2 \cdots \vec{\beta}_m$ are independent. $m = 100$

① $(\beta_J \mid y_J, X_J, \theta, \Sigma, \sigma^2) \sim MVN$ with

$$Var(\beta_J \mid y_J, X_J, \sigma^2, \theta, \Sigma) = \left(\Sigma^{-1} + X_J^T X_J / \sigma^2\right)^{-1}$$

$$E(\beta_J \mid y_J, X_J, \sigma^2, \theta, \Sigma) = \left(\Sigma^{-1} + X_J^T X_J / \sigma^2\right)^{-1} \left(\Sigma^{-1}\theta + X_J^T y_J / \sigma^2\right)$$

② $(\theta \mid \beta_1, \cdots, \beta_m, \Sigma) \sim MVN$ with (May want to look at chapter 7)

mean $\mu_m = \Lambda_m \left(\Lambda_0^{-1} \mu_0 + m \Sigma^{-1} \bar{\beta}\right)$

Variance $\Lambda_m = \left(\Lambda_0^{-1} + m \Sigma^{-1}\right)^{-1}$

Notice
$\bar{\beta}$ = average of β.
$= \frac{1}{m} \sum \vec{\beta}_J$.

③ $(\Sigma \mid \theta, \beta_1, \cdots \beta_m) \sim \text{Inverse Wishart}\left(n_0 + m, (S_0 + S_\theta)^{-1}\right)$

$$S_\theta = \sum_{J=1}^{m} (\beta_J - \theta)(\beta_J - \theta)^T$$

④ $\sigma^2 \sim IG\left(\left[\nu_0 + \sum n_J\right]/2, \left[\nu_0 \sigma_0^2 + SSR\right]/2\right)$

$$SSR = \sum_{J=1}^{m} \sum_{i=1}^{n_J} \left(y_{i,J} - \beta_J^T x_{i,J}\right)^2$$

YOU INPUT

Note: all items $\Lambda_0, \mu_0, \eta_0, S_0, \nu_0, \bar{\sigma}_0, m$ must be provided (not computed by Gibbs or Metropolis sampling).

→ μ_0 = prior expectation of θ.

(i) = Average of ordinary least squared regression estimates

or

(ii) If you have prior info from literature, put your own numbers

In (i),

$$\mu_0 = \begin{pmatrix} \bar{\hat{\beta}}_0 \\ \vdots \\ \bar{\hat{\beta}}_p \end{pmatrix} = \begin{pmatrix} \sum_{J=1}^{m} \frac{\hat{\beta}_{0(J)}}{m} \\ \vdots \\ \sum_{J=1}^{m} \frac{\hat{\beta}_{p(J)}}{m} \end{pmatrix}$$

Say, for school 1 you get

$\hat{\beta}_0 = 0.2 \quad \hat{\beta}_1 = 3 \quad \hat{\beta}_2 = 2$

for school 2 you get

$\hat{\beta}_0 = 0.1 \quad \hat{\beta}_1 = 1.5 \quad \hat{\beta}_3 = 1$.

Then $\bar{\hat{\beta}}_0 = \frac{0.2 + 0.1}{2} \quad \bar{\hat{\beta}}_1 = \frac{3 + 1.5}{2} \quad \bar{\hat{\beta}}_2 = \frac{2+1}{2}$.

$\rightarrow \Lambda_0$ = prior variance of θ

$\Rightarrow$ Think that we have $m = 100$ sets of $\hat{\beta}$, i.e. 100 $\hat{\beta}_1$'s for each school.
100 $\hat{\beta}_2$'s, etc.

$$Var(\hat{\beta}_1) = \frac{\sum_{j=1}^{m} \left(\hat{\beta}_{1(j)} - \bar{\hat{\beta}}_1\right)^2}{m-1}$$

$$Var(\hat{\beta}_2) = \frac{\sum_{j=1}^{m} \left(\hat{\beta}_{1(j)} - \bar{\hat{\beta}}_2\right)^2}{m-1}$$

$$Cov(\hat{\beta}_1, \hat{\beta}_2) = \sum \left(\hat{\beta}_{1(j)} - \bar{\hat{\beta}}_1\right)\left(\hat{\beta}_{2(j)} - \bar{\hat{\beta}}_2\right) \cdot \left(\frac{1}{m-1}\right)$$

So you calculate all this,

$$\begin{pmatrix} Var(\hat{\beta}_1) & Cov(\hat{\beta}_1, \hat{\beta}_2) & \cdots & Cov(\hat{\beta}_1, \hat{\beta}_p) \\ Cov(\hat{\beta}_2, \hat{\beta}_1) & Var(\hat{\beta}_2) & \cdots & \vdots \\ Cov(\hat{\beta}_3, \hat{\beta}_1) & Cov(\hat{\beta}_3, \hat{\beta}_2) & Var(\hat{\beta}_3) & \\ \vdots & & & \ddots \\ Cov(\hat{\beta}_p, \hat{\beta}_1) & Cov(\hat{\beta}_p, \hat{\beta}_2) & \cdots & Var(\hat{\beta}_p) \end{pmatrix}$$

This and the μ_0 given represent information of a person with unbiased but weak priors

→ η_0 = # of parameters in the regression + 2
$= p+2$.

This makes the $E(\Sigma) = S_0$ but the prior distribution is diffuse.

→ S_0 = also equal to Λ_0.

→ σ_0^2 = average of within group sample variance

i.e. Compute $s_1^2, s_2^2, \ldots, s_m^2$.

$\sum_{j=1}^{m} \frac{s_j^2}{m}$ is your σ_0^2.

recall $s_j^2 = \frac{\sum (y_{ij} - \bar{y}_j)^2}{n_j - 1}$.

→ $\nu_0 = 1$

Please, read section 11.1 – 11.3.

What if the Normal Model For Y is not good?

Can we do regression still?

Yes, Generalized linear model. Assume $Y \sim$ some distribution (Poisson, Binomial). and use a link to regression

e.g. $Y_{iJ} \sim \text{Poisson}(\lambda_{iJ})$.

$$\log(\lambda_J) = \beta_0 + \beta_1 X_1 + \cdots \beta_p X_p + \varepsilon_p.$$

Some θ, Σ have similar full conditional distributions but not the β's.

Need Metropolis Hastings.

What is Metropolis-Hasting?

Chapter 10.

and 11.4 – end of chapter 10.