Module 2: Bayesian Hierarchical Models

Francesca Dominici

Michael Griswold

The Johns Hopkins University

Bloomberg School of Public Health

Key Points from yesterday

- "Multi-level" Models:
  - Have covariates from many levels and their interactions
  - Acknowledge correlation among observations from within a level (cluster)
- Random effect MLMs condition on unobserved "latent variables" to describe correlations
- Random Effects models fit naturally into a Bayesian paradigm
- Bayesian methods combine prior beliefs with the likelihood of the observed data to obtain posterior inferences

Bayesian Hierarchical Models

- Module 2:
  - Example 1: School Test Scores
    - The simplest two-stage model
    - WinBUGS
  - Example 2: Aww Rats
    - A normal hierarchical model for repeated measures
    - WinBUGS

Example 1: School Test Scores

Testing in Schools

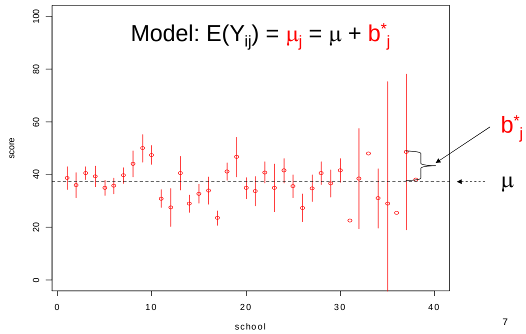
- Goldstein et al. (1993)
- Goal: differentiate between `good' and `bad' schools
- Outcome: Standardized Test Scores
- Sample: 1978 students from 38 schools
  - MLM: students (obs) within schools (cluster)
- Possible Analyses:
  1. Calculate each school's observed average score
  2. Calculate an overall average for all schools
  3. Borrow strength across schools to improve individual school estimates

Testing in Schools

- Why borrow information across schools?
  - Median # of students per school: 48, Range: 1-198
  - Suppose small school (N=3) has: 90, 90,10 (avg=63)
  - Suppose large school (N=100) has avg=65
  - Suppose school with N=1 has: 69 (avg=69)
  - Which school is 'better'?
  - Difficult to say, small N $\Rightarrow$ highly variable estimates
  - For larger schools we have good estimates, for smaller schools we may be able to borrow information from other schools to obtain more accurate estimates
  - How? Bayes

Testing in Schools: "Direct Estimates"

Mean Scores & C.I.s for Individual Schools



Fixed and Random Effects

- Standard Normal regression models: $\varepsilon_{ij} \sim N(0,\sigma^2)$

  1. $Y_{ij} = \mu + \varepsilon_{ij} \longrightarrow \hat{\mu}_j = \bar{X}$ (overall avg)
  2. $Y_{ij} = \mu_j + \varepsilon_{ij} \longrightarrow \hat{\mu}_j = \bar{X}_j$ (shool avg)

     $= \mu + b^*_j + \varepsilon_{ij}$ $\qquad = \bar{X} + \hat{b}^*_j = \bar{X} + (\bar{X}_j - \bar{X})$ Fixed Effects

Fixed and Random Effects

- Standard Normal regression models: $\varepsilon_{ij} \sim N(0,\sigma^2)$

  1. $Y_{ij} = \mu + \varepsilon_{ij} \longrightarrow \hat{\mu}_j = \bar{X}$ (overall avg)
  2. $Y_{ij} = \mu_j + \varepsilon_{ij} \longrightarrow \hat{\mu}_j = \bar{X}_j$ (shool avg)

     $= \mu + b^*_j + \varepsilon_{ij}$ $\qquad = \bar{X} + \hat{b}^*_j = \bar{X} + (\bar{X}_j - \bar{X})$ Fixed Effects

- A random effects model:

  3. $Y_{ij} \mid b_j = \mu + b_j + \varepsilon_{ij}$, with: $b_j \sim N(0,\tau^2)$ Random Effects

Represents **Prior** beliefs about similarities between schools!

Fixed and Random Effects

- Standard Normal regression models: $\varepsilon_{ij} \sim N(0,\sigma^2)$

  1. $Y_{ij} = \mu + \varepsilon_{ij} \longrightarrow \hat{\mu}_j = \bar{X}$ (overall avg)
  2. $Y_{ij} = \mu_j + \varepsilon_{ij} \longrightarrow \hat{\mu}_j = \bar{X}_j$ (shool avg)

     $= \mu + b^*_j + \varepsilon_{ij}$ $\qquad = \bar{X} + \hat{b}^*_j = \bar{X} + (\bar{X}_j - \bar{X})$ Fixed Effects

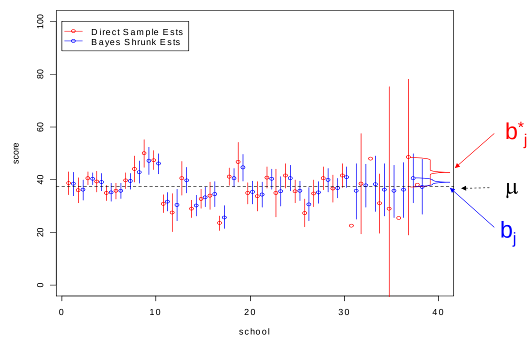
- A random effects model:

  3. $Y_{ij} \mid b_j = \mu + b_j + \varepsilon_{ij}$, with: $b_j \sim N(0,\tau^2)$ Random Effects

     $$\hat{\mu}_j = \bar{X} + \hat{b}_j^{blup} = \bar{X} + \left[\frac{\tau^2}{\frac{\sigma^2}{n}+\tau^2}\right]\hat{b}^*_j = \bar{X} + \left[\frac{\tau^2}{\frac{\sigma^2}{n}+\tau^2}\right](\bar{X}_j - \bar{X})$$

  - Estimate is part-way between the model and the data
  - Amount depends on variability (σ) and underlying truth (τ)

Testing in Schools: Shrinkage Plot



Testing in Schools: Winbugs

- *Data: i=1..1978 (students), s=1…38 (schools)*
- *Model*:
  - $Y_{is} \sim$ Normal(α_s , σ^2_y)
  - $\alpha_s \sim$ Normal(α , σ^2_α) (priors on school avgs)

Note: WinBUGS uses *precision* instead of *variance* to specify a normal distribution!

- *WinBUGS*:
  - $Y_{is} \sim$ Normal(α_s , τ_y) with: $\sigma^2_y = 1 / \tau_y$
  - $\alpha_s \sim$ Normal(α , τ_α) with: $\sigma^2_\alpha = 1 / \tau_\alpha$

Testing in Schools: Winbugs

- *WinBUGS Model*:
  - $Y_{is} \sim$ Normal(α_s , τ_y) with: $\sigma^2_y = 1 / \tau_y$
  - $\alpha_s \sim$ Normal(α , τ_α) with: $\sigma^2_\alpha = 1 / \tau_\alpha$
  - $\tau_y \sim \Gamma(0.001,0.001)$ (prior on precision)
- Hyperpriors
  - Prior on mean of school means
    - $\alpha \sim$ Normal(0 , 1/1000000)
  - Prior on precision (inv. variance) of school means
    - $\tau_\alpha \sim \Gamma(0.001,0.001)$
- Using "Vague" / "Noninformative" Priors

Testing in Schools: Winbugs

- *Full WinBUGS Model*:
  - $Y_{is} \sim$ Normal(α_s , τ_y) with: $\sigma^2_y = 1 / \tau_y$
  - $\alpha_s \sim$ Normal(α , τ_α) with: $\sigma^2_\alpha = 1 / \tau_\alpha$
  - $\tau_y \sim \Gamma(0.001,0.001)$
  - $\alpha \sim$ Normal(0 , 1/1000000)
  - $\tau_\alpha \sim \Gamma(0.001,0.001)$

Testing in Schools: Winbugs

- *WinBUGS Code*:

```
model
  {
       for( i in 1 : N ) {
                Y[i] ~ dnorm(mu[i],y.tau)
                mu[i] <- alpha[school[i]]
                }
       for( s in 1 : M ) {
                alpha[s] ~ dnorm(alpha.c, alpha.tau)
                }
       y.tau ~ dgamma(0.001,0.001)
       sigma <- 1 / sqrt(y.tau)
       alpha.c ~ dnorm(0.0,1.0E-6)
       alpha.tau ~ dgamma(0.001,0.001)
  }
```


Example 2: Aww, Rats…
A normal hierarchical model for repeated measures

Improving individual-level estimates

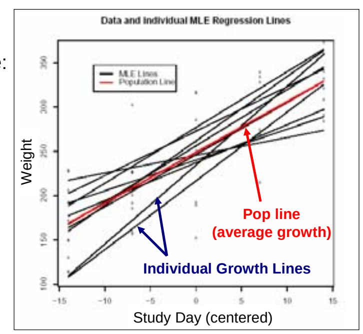
- Gelfand *et al* (1990)
- 30 young rats, weights measured weekly for five weeks
- Dependent variable (Y_{ij}) is weight for rat "i" at week "j"
- Data:

	Weights Y_{ij} of rat i on day x_j				
	$x_j = 8$	15	22	29	36
Rat 1	151	199	246	283	320
Rat 2	145	199	249	293	354
…….					
Rat 30	153	200	244	286	324

- Multilevel: weights (observations) within rats (clusters)

Individual & population growth

- Rat "i" has its own expected growth line:
  $E(Y_{ij}) = b_{0i} + b_{1i}X_j$
- There is also an overall, average population growth line:
  $E(Y_{ij}) = \beta_0 + \beta_1 X_j$



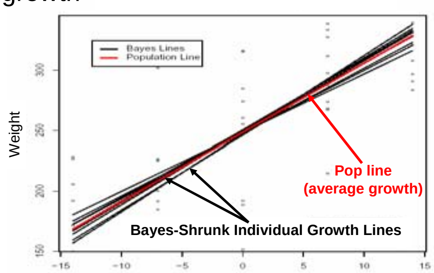
Improving individual-level estimates

- Possible Analyses
  1. Each rat (cluster) has its own line:
     intercept= b_{i0}, slope= b_{i1}
  2. All rats follow the same line:
     $b_{i0} = \beta_0$, $b_{i1} = \beta_1$
  3. A compromise between these two:
     Each rat has its own line, BUT…
     the lines come from an assumed distribution

$$E(Y_{ij} \mid b_{i0}, b_{i1}) = b_{i0} + b_{i1}X_j$$

"Random Effects" $\begin{cases} b_{i0} \sim N(\beta_0, \tau_0^2) \\ b_{i1} \sim N(\beta_1, \tau_1^2) \end{cases}$

A compromise:
Each rat has its own line, but information is borrowed across rats to tell us about individual rat growth



Rats: Winbugs (see help: Examples Vol I)

- *WinBUGS Model*:

$$Y_{ij} \sim \text{Normal}(\alpha_i + \beta_i(x_j - x_{bar}), \tau_c)$$

$$\alpha_i \sim \text{Normal}(\alpha_c, \tau_\alpha)$$

$$\beta_i \sim \text{Normal}(\beta_c, \tau_\beta)$$

$\alpha_c, \tau_\alpha, \beta_c, \tau_\beta, \tau_c$ are given independent ``noninformative" priors.

Rats: Winbugs (see help: Examples Vol I)

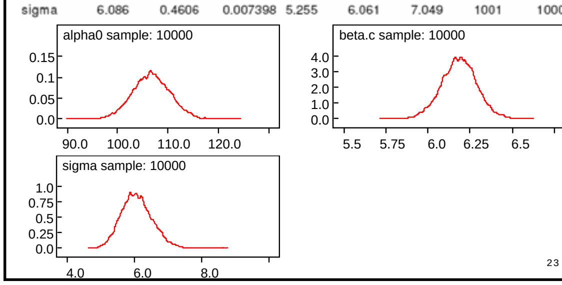
- *WinBUGS Code*:

```
model
{
   for( i in 1 : N ) {
      for( j in 1 : T ) {
         Y[i , j] ~ dnorm(mu[i , j],tau.c)
         mu[i , j] <- alpha[i] + beta[i] * (x[j] - xbar)
      }
      alpha[i] ~ dnorm(alpha.c,alpha.tau)
      beta[i] ~ dnorm(beta.c,beta.tau)
   }
   tau.c ~ dgamma(0.001,0.001)
   sigma <- 1 / sqrt(tau.c)
   alpha.c ~ dnorm(0.0,1.0E-6)
   alpha.tau ~ dgamma(0.001,0.001)
   beta.c ~ dnorm(0.0,1.0E-6)
   beta.tau ~ dgamma(0.001,0.001)
   alpha0 <- alpha.c - xbar * beta.c
}
```


Rats: Winbugs (see help: Examples Vol I)

- *WinBUGS Results:* 10000 updates

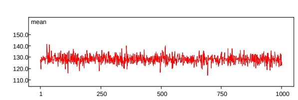
	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
alpha0	106.6	3.655	0.04079	99.44	106.5	113.8	1001	10000
beta.c	6.185	0.1061	0.00132	5.975	6.185	6.394	1001	10000
sigma	6.086	0.4606	0.007398	5.255	6.061	7.049	1001	10000



WinBUGS Diagnostics:

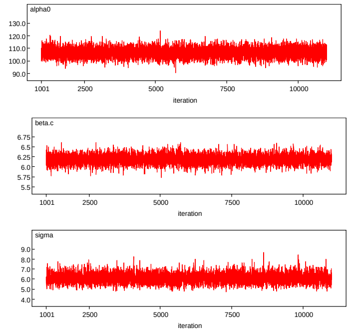
- MC error tells you to what extent simulation error contributes to the uncertainty in the estimation of the mean.
- This can be reduced by generating additional samples.

- Always examine the trace of the samples.
- To do this select the history button on the Sample Monitor Tool.
- Look for:
  - Trends
  - Correlations



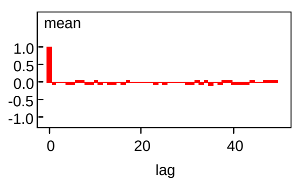
Rats: Winbugs (see help: Examples Vol I)

- *WinBUGS Diagnostics: history*



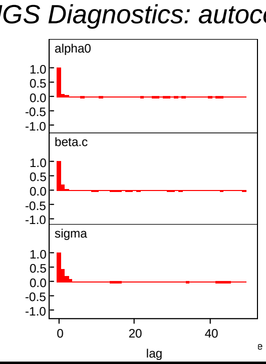
WinBUGS Diagnostics:

- Examine sample autocorrelation directly by selecting the ‘auto cor’ button.
- If autocorrelation exists, generate additional samples and thin more.

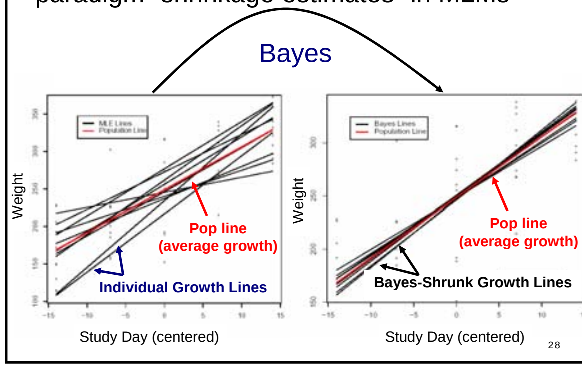


Rats: Winbugs (see help: Examples Vol I)

- *WinBUGS Diagnostics: autocorrelation*



WinBUGS provides machinery for Bayesian paradigm “shrinkage estimates” in MLMs



School Test Scores Revisited

Testing in Schools revisited

- Suppose we wanted to include covariate information in the school test scores example
- Student-level covariates
  - Gender
  - London Reading Test (LRT) score
  - Verbal reasoning (VR) test category (1, 2 or 3)
- School -level covariates
  - Gender intake (all girls, all boys or mixed)
  - Religious denomination (Church of England, Roman Catholic, State school or other)

Testing in Schools revisited

- Model

$$Y_{ij} \sim \text{Normal}(\mu_{ij}, \tau_{ij})$$
$$\mu_{ij} = \alpha_{1j} + \alpha_{2j}\,\text{LRT}_{ij} + \alpha_{3j}\,\text{VR1}_{ij} + \beta_1\,\text{LRT}_{ij}^2 + \beta_2\,\text{VR2}_{ij} + \beta_3\,\text{Girl}_{ij} + \beta_4\,\text{Girls' school}_j + \beta_5\,\text{Boys' school}_j + \beta_6\,\text{CE school}_j + \beta_7\,\text{RC school}_j + \beta_8\,\text{other school}_j$$
$$\log \tau_{ij} = \theta + \phi\,\text{LRT}_{ij}$$

- Wow! Can YOU fit this model?
- Yes you can!
- See WinBUGS>help>Examples Vol II for data, code, results, etc.
- More Importantly: Do you understand this model?

Bayesian Concepts

- Frequentist: Parameters are "the truth"
- Bayesian: Parameters have a distribution
- "Borrow Strength" from other observations
- "Shrink Estimates" towards overall averages
- Compromise between model & data
- Incorporate prior/other information in estimates
- Account for other sources of uncertainty
- Posterior $\propto$ Likelihood * Prior