Bayesian Multi-level Regression

Bayesian Modeling for Socio-Environmental Data

N. Thompson Hobbs

August 6, 2016

SESYNC

Lecture material

- Bayesian, multilevel models for grouped data
  - group level intercepts
  - group level intercepts with group level covariate
  - group level slopes and intercepts
- Some useful tricks
- Priors on group level variances (later, in lab)

The simple, Bayesian set-up

Deterministic model:

$$g(\boldsymbol{\theta}, x_i)$$

Stochastic model:

$$\underbrace{[\boldsymbol{\theta}, \boldsymbol{\sigma}^2 | y_i]}_{\text{posterior}} \propto \overbrace{\underbrace{[y_i | g(\boldsymbol{\theta}, x_i), \boldsymbol{\sigma}^2]}_{\text{likelihood}} \underbrace{[\boldsymbol{\theta}]}_{\text{priors}}}^{\text{joint}}$$

Draw the DAG.

Recall that

$$\underbrace{[\boldsymbol{\theta}, \sigma^2 | y_i]}_{\text{posterior}} \propto \overbrace{[y_i, \boldsymbol{\theta}, \sigma^2]}^{\text{joint}}$$

Hierarchical models: "modeling parameters"

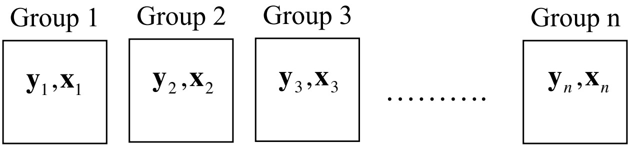
$$
\begin{aligned}
[\theta_1, \boldsymbol{\theta}_2, \boldsymbol{\alpha}, y_i, \boldsymbol{\sigma}^2] &\propto [y_{ij} | g(\theta_1, \theta_{2,j}, x_{ij}), \boldsymbol{\sigma}_1^2] \\
&\times [\theta_{2,j} | h(\alpha_1, \alpha_2, u_j), \boldsymbol{\sigma}_2^2] \\
&\times [\theta_1, \boldsymbol{\theta}_2, \boldsymbol{\alpha}, \boldsymbol{\sigma}^2]
\end{aligned}
$$

Draw the DAG.

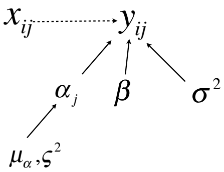
Steps in Bayesian analysis

1. Compose joint distribution of observed and unobserved quantities.
2. Factor joint distribution into sensible parts.
3. Use factored joint distribution to write:
    - 3.1 JAGS code *or*
    - 3.2 Own MCMC sampler
        - 3.2.1 Write full-conditional distributions
        - 3.2.2 Choose sampling method for each full-conditional
4. Check model
5. Make inference

The problem

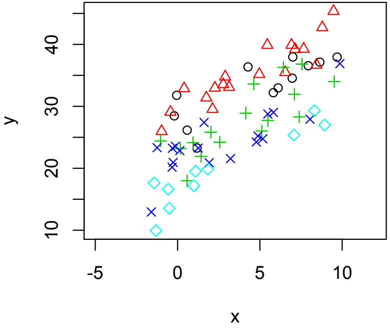

Group 1
Group 2
Group 3
Group n
$\mathbf{y}_1, \mathbf{x}_1$
$\mathbf{y}_2, \mathbf{x}_2$
$\mathbf{y}_3, \mathbf{x}_3$
....
$\mathbf{y}_n, \mathbf{x}_n$

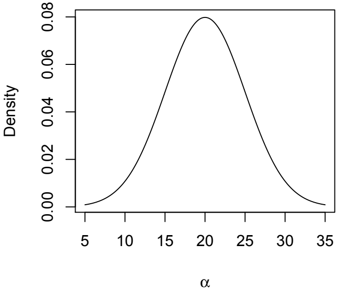
We can model the intercept:



$$
\begin{aligned}
[\beta, \boldsymbol{\alpha}, \sigma^2, \mu_\alpha, \varsigma^2, |\mathbf{y}] &\propto \prod_{i=1}^{n_j} \prod_{j=1}^{J} \text{normal}\left(y_{ij} | \alpha_j + \beta x_{ij}, \sigma^2\right) \\
&\times \text{normal}\left(\alpha_j | \mu_\alpha, \varsigma^2\right) \\
&\times \text{normal}\left(\beta | 0, 10000\right) \text{normal}\left(\mu_\alpha | 0, 1000\right) \\
&\times \text{inverse gamma}\left(\sigma^2 | .001, .001\right) \text{inverse gamma}\left(\varsigma^2 | .001, .001\right)
\end{aligned}
$$

We seek to understand the distribution of intercepts.





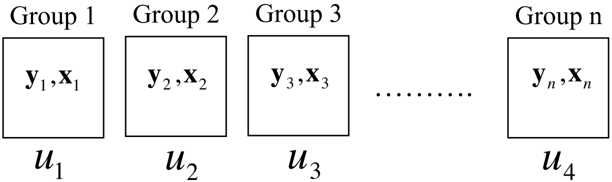
Some notation

$$\mu_{ij} = \beta_0 + \beta_1 x_{ij} + \alpha_j$$
$$y_{ij} \sim \text{normal}(\mu_{ij}, \sigma^2)$$
$$\alpha_j \sim \text{normal}(0, \varsigma^2)$$

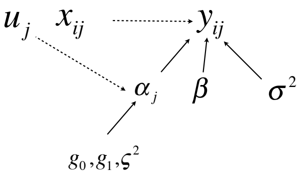
is identical to:

$$\mu_{ij} = \alpha_j + \beta_1 x_{ij}$$
$$y_{ij} \sim \text{normal}(\mu_{ij}, \sigma^2)$$
$$\alpha_j \sim (\mu_\alpha, \varsigma^2)$$

Include data on groups.

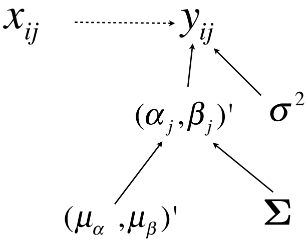


We can model the intercept as a function of group level data:



$$
\begin{aligned}
\left[\boldsymbol{\alpha}, \beta, \boldsymbol{\sigma}^2, \mathbf{g}, \varsigma^2, |\mathbf{y}\right] \propto & \prod_{i=1}^{n_j} \prod_{j=1}^{J} \text{normal}\left(y_{ij} | \boldsymbol{\alpha}_j + \boldsymbol{\beta} x_{ij}, \boldsymbol{\sigma}^2\right) \\
& \times \text{normal}\left(\boldsymbol{\alpha}_j | g_0 + g_1 u_j, \varsigma^2\right) \\
& \times \text{normal}\left(\boldsymbol{\beta} | 0, .001\right) \text{normal}\left(g_0 | 0, 1000\right) \text{normal}\left(g_1 | 0, 1000\right) \\
& \times \text{inverse gamma}\left(\boldsymbol{\sigma}^2 | .001, .001\right) \text{inverse gamma}\left(\varsigma^2 | .001, .001\right)
\end{aligned}
$$

Modeling intercepts *and* slopes



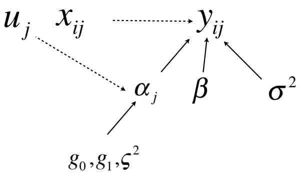
$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim \text{multivariate normal}\left(\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \boldsymbol{\Sigma} \right)$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \varsigma_\alpha^2 & \text{Cov}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \\ \text{Cov}(\boldsymbol{\alpha}, \boldsymbol{\beta}) & \varsigma_\beta^2 \end{pmatrix}, \text{Cov}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \rho \varsigma_\alpha \varsigma_\beta$$

Modeling intercepts *and* slopes

$$
\begin{aligned}
[\boldsymbol{\alpha},\boldsymbol{\beta},\mu_\alpha,\mu_\beta,\sigma^2,\varsigma_\alpha,\varsigma_\beta,\rho|\mathbf{y}] &\propto \prod_{j=1}^{J}\prod_{i=1}^{n_j}\text{normal}(y_{ij}|\alpha_j+\beta_j x_{ij},\sigma^2)\\
&\times \text{MVN}\left(\left(\begin{array}{c}\alpha_j\\ \beta_j\end{array}\right)\middle|\left(\begin{array}{c}\mu_\alpha\\ \mu_\beta\end{array}\right),\boldsymbol{\Sigma}\right)\\
&\times \text{priors on } \mu_\alpha,\mu_\beta,\sigma^2,\varsigma_\alpha,\varsigma_\beta,\rho
\end{aligned}
$$

Indexing groups



$$
\begin{aligned}
&\left[\boldsymbol{\alpha}, \beta, , \boldsymbol{\sigma}^2, \mathbf{g}, \varsigma^2, |\mathbf{y}\right] \propto \prod_{i=1}^{n_j} \prod_{j=1}^{J} \text{normal}\left(y_{ij} | \alpha_j + \beta x_{ij}, \sigma^2\right) \\
&\times \text{normal}\left(\alpha_j | g_0 + g_1 u_j, \varsigma^2\right) \\
&\times \text{normal}\left(\beta | 0, .001\right) \text{normal}\left(g_0 | 0, 1000\right) \times \text{normal}\left(g_1 | 0, 1000\right) \\
&\times \text{inverse gamma}\left(\sigma^2 | .001, .001\right) \text{inverse gamma}\left(\varsigma^2 | .001, .001\right)
\end{aligned}
$$

Indexing groups

```
> u
[1]  6.215579  8.716296 10.064460 11.292387 14.504154 14.734861
[7] 18.356877 18.910133
```

```
> head(y[,1:4])
     group i        x[i]     y[i]
[1,]     1 1 -0.00266051 13.48934
[2,]     1 2  4.54802848 22.29538
[3,]     1 3  9.86832462 29.03655
[4,]     1 4  0.99869789 18.61136
[5,]     1 5  1.27733200 20.59178
[6,]     1 6  4.32915675 25.37082
> tail(y[,1:4])
       group   i     x[i]     y[i]
[108,]     8 108 4.543959 38.93163
[109,]     8 109 1.287844 34.65796
[110,]     8 110 6.642313 40.62259
[111,]     8 111 7.404183 40.46518
[112,]     8 112 8.252571 41.47995
[113,]     8 113 9.558780 46.14771
```

Indexing groups

```
model{
beta ~ dnorm(0,.0001)
sigma ~ dunif(0,50)
tau.p <- 1/sigma^2
g0 ~ dnorm(0,.0001)
g1 ~ dnorm(0,.0001)
varsigma ~ dunif(0,50)
tau.g <- 1/varsigma^2
 for (i in 1:length(y)){
  mu[i] <- alpha[group[i]]+ beta*x[i]
  y[i] ~ dnorm(mu[i],tau.p)
  }
  for(j in 1:n.group){
  mu.g[j] <- g0 + g1*u[j]
  alpha[j]~dnorm(mu.g[j],tau.g)
    }
```