

Bayesian Multi-level Regression

Bayesian Models for Ecologists

Tom Hobbs

June 9, 2024



Middlebury



Funga



DEB 2042028

Lecture material

- ▶ Background
- ▶ Bayesian, multilevel models for grouped data
 - ▶ group level intercepts
 - ▶ group level intercepts with group level covariate
 - ▶ group level slopes and intercepts
 - ▶ an essential coding trick
- ▶ Priors on group level variances: Will be discussed in next priors lecture.

Recall that

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

MCMC allows us to discover the c .

The simple, Bayesian set-up

Deterministic model:

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

Stochastic model:

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

likelihood priors

The factored, joint distribution (aka joint conditional) provides a detailed blueprint for

1. Writing the full conditionals as the basis for Gibbs (for conjugate full conditionals) or Metropolis-Hastings sampling (for any full conditional).
2. Writing JAGS code.

Hierarchical models: “modeling parameters”

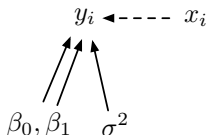
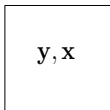
$$\begin{aligned}
 [\theta_1, \boldsymbol{\theta}_2, \alpha_1, \alpha_2, \sigma_1^2, \sigma_2^2 \mid y_{ij}] &\propto [\theta_1, \theta_2, \alpha_1, \alpha_2, \sigma_1^2, \sigma_2^2, y_{ij}] \\
 [\theta_1, \theta_2, \alpha_1, \alpha_2, \sigma_1^2, \sigma_2^2 \mid y_{ij}] &\propto [y_{ij} \mid g(\theta_1, \theta_{2,j}, x_{ij}), \sigma_1^2] \\
 &\times [\theta_{2,j} \mid h(\alpha_1, \alpha_2, u_j), \sigma_2^2] \\
 &\times [\theta_1][\alpha_1][\alpha_2][\sigma_1^2][\sigma_2^2]
 \end{aligned}$$

Draw the DAG.

x_{ij} and u_j are observed.

The problem we treated in simple Bayesian regression

A single “group”



$$g(\beta_0, \beta_1, x_i) = \beta_0 + \beta_1 x_i$$

$$[\beta_0, \beta_1, \sigma^2 \mid y_i] \propto [\beta_0, \beta_1, \sigma^2, y_i]$$

factoring rhs using DAG:

$$[\beta_0, \beta_1, \sigma^2 \mid y_i] \propto [y_i \mid g(\beta_0, \beta_1, x_i), \sigma^2][\beta_0][\beta_1][\sigma^2]$$

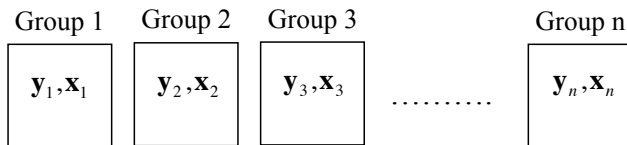
joint for all data :

$$[\beta_0, \beta_1, \sigma^2 \mid \mathbf{y}] \propto \prod_{i=1}^n [y_i \mid g(\beta_0, \beta_1, x_i), \sigma^2][\beta_0][\beta_1][\sigma^2]$$

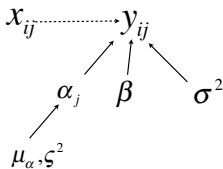
choose specific distributions:

$$\begin{aligned} [\beta_0, \beta_1, \sigma^2 \mid \mathbf{y}] &\propto \prod_{i=1}^n \text{normal}(y_i \mid g(\beta_0, \beta_1, x_i), \sigma^2) \\ &\times \text{normal}(\beta_0 \mid 0, 10000) \text{normal}(\beta_1 \mid 0, 10000) \\ &\times \text{uniform}(\sigma^2 \mid 0, 500) \end{aligned}$$

The problem of multiple groups

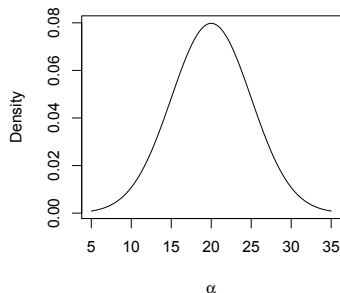
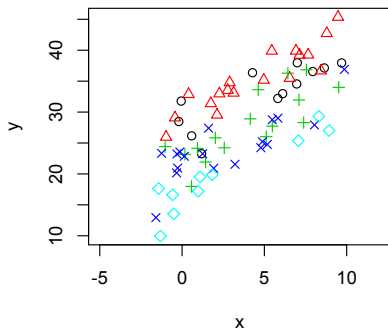


We can model the intercept (or slope):



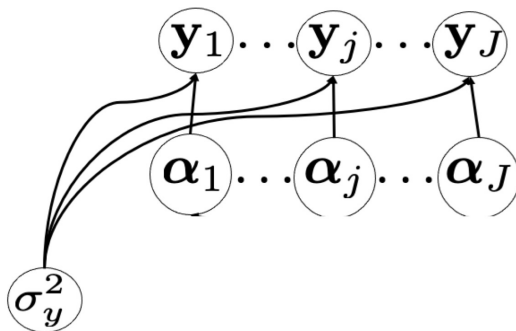
$$\begin{aligned}
 [\beta, \alpha, \sigma^2, \mu_\alpha, \zeta^2, \mathbf{y}] &\propto \prod_{i=1}^{n_j} \prod_{j=1}^J \text{normal}(y_{ij} | \alpha_j + \beta x_{ij}, \sigma^2) \\
 &\times \text{normal}(\alpha_j | \mu_\alpha, \zeta^2) \\
 &\times \text{normal}(\beta | 0, 10000) \text{normal}(\mu_\alpha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

We seek to understand the distribution of intercepts.



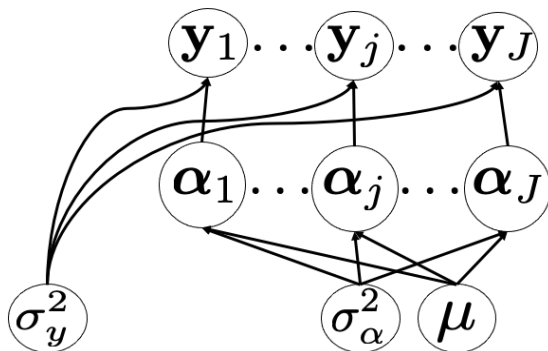
Borrowing strength: No pooling among groups

No pooled model, different mean for each group



Borrowing strength: "Partial Pooling"

Multi-level model

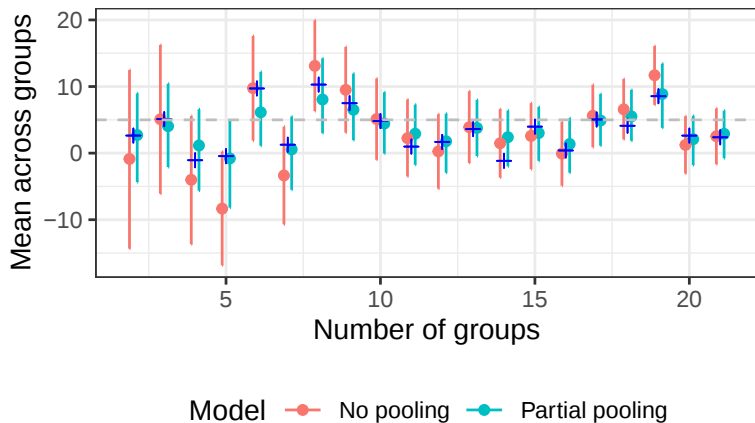


Borrowing strength: “Partial Pooling”

Multi-level model

$y_{ij} \sim \mathcal{N}(\alpha_j, \sigma_y^2), \text{ for } j = 1, \dots, J \text{ and } i = 1, \dots, n_j$	(Data)
$\alpha_j \sim \mathcal{N}(\mu, \sigma_\alpha^2), \text{ for } j = 1, \dots, J$	(Process)
$\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$	(Parameter)
$\sigma_y^2 \sim \mathcal{IG}(q_y, r_y)$	(Parameter)
$\sigma_\alpha^2 \sim \mathcal{IG}(q_\alpha, r_\alpha)$	(Parameter)

Borrowing strength



Some notation

$$\mu_{ij} = \beta_0 + \beta_1 x_{ij} + \varepsilon_j$$

$$\varepsilon_j \sim \text{normal}(0, \varsigma^2)$$

$$y_{ij} \sim \text{normal}(\mu_{ij}, \sigma^2)$$

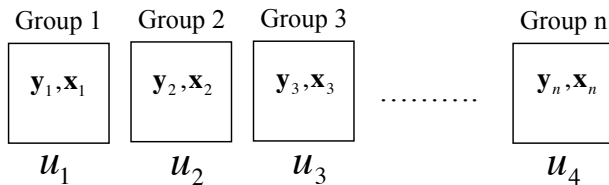
is algebraically equivalent to

$$\mu_{ij} = \beta_{0,j} + \beta_1 x_{ij}$$

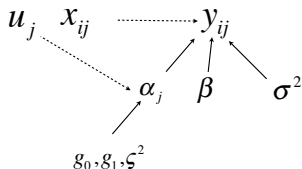
$$\beta_{0,j} \sim \text{normal}(\mu_{\beta_0}, \varsigma^2)$$

$$y_{ij} \sim \text{normal}(\mu_{ij}, \sigma^2)$$

Include data on groups.



We can model the intercept (or slope) as a function of group level data:



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

Modeling intercepts and slopes

A correlation matrix:

Correlations

	Weight in kg	Hours of Sleep	Exposure while Sleeping	Life Span
Weight in kg	1	-.307	.338	.302
Hours of Sleep	-.307	1	-.642	-.410
Exposure while Sleeping	.338	-.642	1	.360
Life Span	.302	-.410	.360	1

1

Let i index rows and j index columns. Recall that the correlation between two random variables is simply their covariance divided by the standard deviation of both variables, $\rho_{ij} = \frac{\text{cov}_{ij}}{\sigma_i \sigma_j}$. It is the *standardized* covariance. Standardization means it can take on values between -1 and 1.

¹<http://www.theanalysisfactor.com/covariance-matrices/>

Modeling intercepts and slopes

Let i index rows and j index columns. If we multiply this correlation matrix times $\sigma_i \sigma_j$ we obtain a *covariance* matrix:

Covariances

	body weight in kg	total sleep (hours/day)	sleep exposure index (1-5)	maximum life span (years)
body weight in kg	808485.128	-1313.960	488.116	5113.271
total sleep (hours/day)	-1313.960	21.222	-4.544	-36.297
sleep exposure index (1-5)	488.116	-4.544	2.575	10.661
maximum life span (years)	5113.271	-36.297	10.661	331.468

Covariance ranges from $-\infty$ to $+\infty$.

Modeling intercepts and slopes

Imagine a vector of 3 random variables, $(z_1, z_2, z_3)'$. The covariance between any two of these random variables is simply an unstandardized version of the correlation between them: it is correlation measured in the units of the random variables. The covariance matrix (aka variance-covariance matrix) of the random variable is:

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \text{Cov}_{1,2} & \text{Cov}_{1,3} \\ \text{Cov}_{2,1} & \sigma_2^2 & \text{Cov}_{2,3} \\ \text{Cov}_{3,1} & \text{Cov}_{3,2} & \sigma_3^2 \end{pmatrix} \quad (1)$$

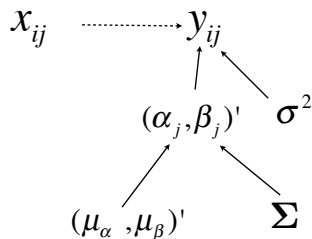
Generalizing, a $m \times m$ covariance matrix has the variances of the random variable on the diagonal and the covariance on the off diagonal. The covariance between random variable i and j is $\text{Cov}_{ij} = \rho \sigma_i \sigma_j$ where ρ is the correlation coefficient, which takes on values between -1 and 1 . Covariance can take on values between $-\infty$ and $+\infty$.

Covariance matrix for two parameter model

Imagine that we have $j = 1, \dots, J$ groups with multiple observations within groups and we fit a two parameter linear model to each group, finding J intercepts and slopes. We denote the vector of intercepts as $\boldsymbol{\alpha}$ and the vector of slopes as $\boldsymbol{\beta}$. We can calculate the variance for each vector $(\sigma_{\alpha}^2, \sigma_{\beta}^2)$ as well as the correlation between the vectors ρ . The variance covariance matrix is thus:

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

where $\text{Cov}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \text{Cov}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \rho \sigma_{\alpha} \sigma_{\beta}$

Modeling intercepts *and* slopes

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

$$\Sigma = \begin{pmatrix} \sigma_\alpha^2 & \rho \sigma_\alpha \sigma_\beta \\ \rho \sigma_\alpha \sigma_\beta & \sigma_\beta^2 \end{pmatrix}$$

Modeling intercepts *and* slopes

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

Modeling intercepts *and* slopes for more than one slope

$$\begin{aligned}
[\boldsymbol{\beta}, \boldsymbol{\mu}_{\beta}, \sigma_{\text{reg}}^2, \mathbf{y}] &\propto \prod_{j=1}^J \prod_{i=1}^{n_j} \text{normal}(y_{ij} | \mathbf{x}'_{ij} \boldsymbol{\beta}_j, \sigma_{\text{reg}}^2) \\
&\times \text{MVN} \left(\begin{pmatrix} \beta_{0j} \\ \beta_{1j} \\ \beta_{2j} \\ \vdots \\ \beta_{mj} \end{pmatrix} \middle| \begin{pmatrix} \mu_{\beta_0} \\ \mu_{\beta_1} \\ \mu_{\beta_2} \\ \vdots \\ \mu_{\beta_m} \end{pmatrix}, \boldsymbol{\Sigma} \right) \\
&\times \text{priors on } \boldsymbol{\mu}_{\beta}, \sigma_{\text{reg}}^2, \boldsymbol{\Sigma}
\end{aligned}$$

Modeling intercepts and slopes for > 1 slope

The Wishart distribution:

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{|\mathbf{x}|^{(n-p-1)/2} e^{-\text{tr}(\mathbf{V}^{-1}\mathbf{x})/2}}{2^{\frac{np}{2}} |\mathbf{V}|^{n/2} \Gamma_p(\frac{n}{2})}$$

A vague prior on Σ :

$$\Sigma \sim \text{Wishart}(\mathbf{S}, m + 1) \quad (3)$$

where m is the number of coefficients including the intercept and $\mathbf{S}$ is an $m \times m$ matrix with ones on the diagonal and zeros on the off diagonals.

Example code: `Sigma ~ dwish(S,y.Nvar + 1)`

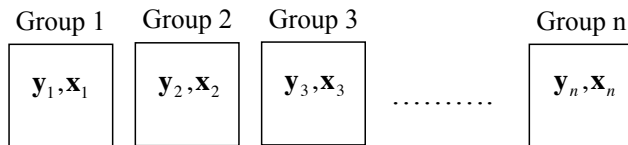
Compute σ 's and ρ as derived quantities of the elements of Σ .

Remember, the Sigma in JAGS uses precisions not variances. For informed priors, see the `eivtools` package in R.

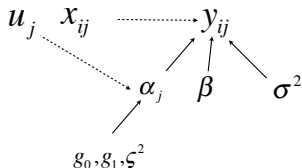
Guidance

- ▶ The Wishart distribution is an easy, useful way to impose reasonably vague priors on covariance matrices. See Gelman and Hill 2009, pages 376-380.
 - ▶ My experience is that these priors are vague for the means but somewhat informative for the variances and for the correlation.
 - ▶ rjags glm module has a scaled Wishart that seems to work better. See http://mmeredith.net/blog/2020/Correlated_priors.htm and Chung et al. 2015 J of Ed. and Behav. Statistics 2015, Vol. 40 136-157
- ▶ It is also entirely feasible, if tedious, to expand the two parameter case (done in lab) to include more than one slope.

Multiple groups



An essential coding trick: Indexing groups



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

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)
  }
}
```

dplyr code for making indexes

```
##Make indices for year, site, and treatment
y = your_data %>%
  site.index =
  as.integer(as.factor(as.character(site_id))),
  treat.index =
  as.integer(as.factor(as.character(treat)))) %>%
  #make indices for plants within plots
  group_by(site_id, treat) %>% mutate(plant.index =
  as.integer(as.factor(as.character(willid)))) %>%
  ungroup()
#check
unique(y$site_id)
unique(y$site.index)
```