Bayesian Tools for Synthesis of Ecological Data

Analysis of Data from Multiple Studies

Tom Hobbs

December 2, 2022

Graduate Degree Program in ECOLOGY

Department of Ecosystem Science and Sustainability

Colorado State University®

Objective of videos

Learn Bayesian methods for synthesizing existing data and published findings to gain new insight in ecology

Objective: Learn Bayesian methods for synthesizing existing data and published findings to gain new insight in ecology

Previous problem: Combining results from multiple studies

Synthesis of findings from many published studies is our goal. These results might include sample or experimental means and variances, regression coefficients, and other quantities of interest. How do we use Bayesian methods to form prior distributions from multiple studies?

The problem here: Combining *data* from multiple studies

Synthesis of findings from many published studies is our goal. These results might include sample or experimental means and variances, regression coefficients, and other quantities of interest. How do we use Bayesian methods to compose a comprehensive analysis using multiple data sets?

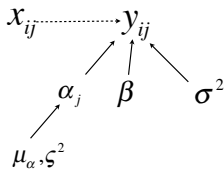# Regression models reflecting data from multiple studies

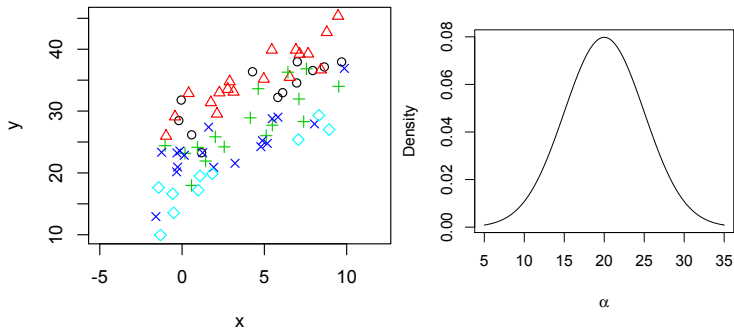Study 1   Study 3   Study J
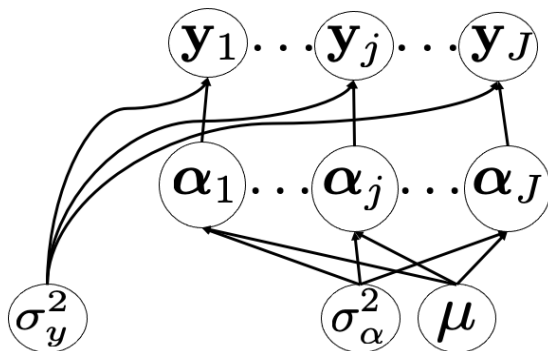
We can model the intercept (or slope):

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

We seek to understand the distribution of intercepts.

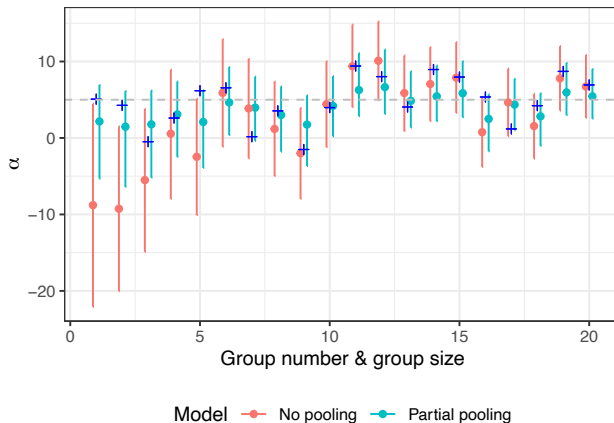## Borrowing strength

Multi-level model

Borrowing strength

Multi-level model

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

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

$$\mu_{ij} = \beta_{0,j} + \beta_1 x_{ij}$$
$$\beta_{0j} \sim \text{normal}(\mu_{\beta_0}, \varsigma^2)$$
$$y_{ij} \sim \text{normal}(\mu_{ij}, \sigma^2)$$
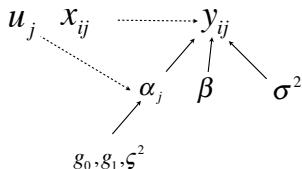
Include data at whole study scale

Study 1   Study 3   Study J

u_1 u_2 u_J

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

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

Modeling intercepts and slopes requires modeling covariance

Imagine a vector of 3 random variables, $(z_i, z_2, z_3)'$ 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:

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

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 $j = 1,...,J$ 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:

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

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

Modeling intercepts *and* slopes
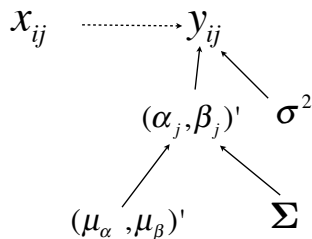
$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim \text{MVN} \left(\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \boldsymbol{\Sigma} \right) \quad \text{MVN} = \text{multivariate normal}$$

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

Modeling intercepts *and* slopes

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

Modeling intercepts *and* slopes for more than one slope

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

Modeling intercepts and slopes for > 1 slope

The Wishart distribution:

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

A vague prior on $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma} \sim \operatorname{Wishart}(\mathbf{S}, m+1) \tag{3}$$

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 $\sigma's$ and ρ as derived quantities of the elements of $\boldsymbol{\Sigma}$. 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 example problem) to include more than one slope.
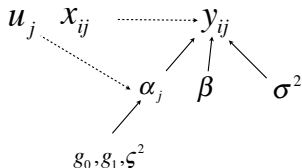
Multiple groups

Study 1   Study 3   Study J

$$u_1 \qquad u_2 \qquad u_J$$

An essential coding trick: Indexing groups

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

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)
```

Take home

Bayesian hierarchical models are powerful tools for synthesizing results from from multiple sets of data on the same ecological process. Borrowing strength provides a natural weighting of influence on estimates of parameters. The studies with the lowest variance have the greatest effect on those parameters.

Objective of videos

Learn Bayesian methods for synthesizing existing data and published findings to gain new insight in ecology