Bayesian Tools for Synthesis of Ecological Data
Priors from Multiple Studies

Tom Hobbs

December 6, 2022

Department of Ecosystem
Science and Sustainability

Objective of videos

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

Sequence

Slides for this video can be found at

https://github.com/nthobbs50/Bayesian_synthesis

The 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?

Method 1: Weighted means

We have reviewed J studies and we seek a prior on the parameter $\overset{\text{all}}{\theta}$ reported in all studies. We can use

$$\begin{bmatrix} \overset{\text{all}}{\theta} \end{bmatrix} \sim \sum_{1}^{J} w_j [\theta_j]$$

$$\sum_{i=1}^{J} w_j = 1$$

where the w_m are weights and $w_m \geq 0$. If we think all studies should have equal weight then $w_j = \frac{1}{J}$. It is also common to weight by sample size such that $w_j = \frac{n_j}{\sum_{j=1}^{J} n_j}$.

This method treats the different studies as fixed effects. Our inference is limited to the five specific studies we analyzed.

Weighted means method example

Suppose we had five studies reporting results of regression and we want a prior on the intercepts reflecting all studies $\overset{\text{all}}{\beta_0}$. We assume the studies were conducted equally well. We would use

$$\overset{\text{all}}{\beta_0} = \sum_{j=1}^{5} \frac{1}{5} \text{normal}(\beta_{0,j}, \sigma_j^2)$$

Hierarchical method for prior across studies, $\overset{\text{all}}{\theta}$

The weighted means approach treats each study as fixed. We might want to assume that each study is drawn from a distribution of studies with mean $\overset{\text{all}}{\theta}$ and variance $\overset{\text{all}}{\sigma^2}$

$$y_j \sim [y_j \mid \theta_j, \sigma_j^2]$$
$$\theta_j \sim [\theta_j \mid \overset{\text{all}}{\theta}, \overset{\text{all}}{\sigma^2}]$$

where y_j is the observed, published value of θ from study j and θ_j is the true, unobserved mean from study j. The quantity σ_j^2 is the variance of the distribution of y_j, which will most often be the published standard error squared.

This method treats the different studies as random effects. We can make inference to all possible realizations of studies from which the published studies were drawn.

Hierarchical method

Suppose we had five studies reporting results of regression and we want a prior on the intercepts reflecting all studies using the hierarchical approach. We would use

$$
\left[\overset{\text{all}}{\beta_0}, \overset{\text{all}}{\sigma^2} \,\middle|\, \mathbf{y}, \sigma^2 \right] \propto \prod_{j=1}^{5} \text{normal}(y_j \mid \beta_{0,j}, \sigma_j^2)[\beta_{0,j}] \overbrace{\left[\overset{\text{all}}{\beta_0}\right] \left[\overset{\text{all}}{\sigma^2}\right]}^{\text{vague}}
$$

where y_j published intercept in study j and σ_j^2 is the square of the standard deviation of the intercept in study j. We would use the mean and standard deviation of the MCMC output as the parameters for a normal prior on $\overset{\text{all}}{\beta_0}$ reflecting the information in all five studies

Example: A prior on food intake η rate by large herbivores

Data from the literature:

Species	Mean Intake (g/kg BW/day)	SE	n	Source
Mule deer				
(Odocoileus hemionus)	20.6	2.4	87	Alldredge eta al. 1974
Elk (Cevus elaphus)	31.1	1.79	2	Jiang and Hudson 1992
Moose (Alces alces)	23.8	11.4	2	Renecker and Hudson 1985
Domestic sheep (Ovis aires)	24.3	1.03	4	Donnelly et al. 1974
Black-tailed deer				
(Odocoileus hemionus)	15	1.73	6	Ulapa et al 2020

Weighted method for prior on η

$$20.6 \sim \text{gamma}\left(\frac{\eta_1^2}{2.4^2}, \frac{\eta_1}{2.4^2}\right)$$

$$31.1 \sim \text{gamma}\left(\frac{\eta_2^2}{1.79}, \frac{\eta_2}{1.79^2}\right)$$

$$23.8 \sim \text{gamma}\left(\frac{\eta_3^2}{11.4^2}, \frac{\eta_3}{11.4^2}\right)$$

$$24.3 \sim \text{gamma}\left(\frac{\eta_4^2}{1.03^2}, \frac{\eta_4}{1.03^2}\right)$$

$$15 \sim \text{gamma}\left(\frac{\eta_5^2}{1.73^2}, \frac{\eta_5}{1.73^2}\right)$$

$$[\overset{\text{all}}{\eta}] = \sum_{i=j}^{5} w_j \eta_j$$

η_j is the true mean intake from study j. The numbers on the left hand side of the $\sim$ are the observed mean intake rates. The numbers in the denominators are the squared standard errors from each study. $\overset{\text{all}}{\eta}$ is the mean intake across the five intake studies. The $w_i = \frac{1}{5}$ for equal weighting or $w_j = \frac{n_j}{\sum_{j=1}^{J} n_j}$ to weight by sample size.

Corresponding JAGS code

```
model{
for(j in 1:length(y)){
 #weighted equally
 w[j] = 1/length(y)
 #weighted by sample size were n is vector of data from table
 #w[j]=n[j]/sum(n[])
 eta[j] ~ dunif(0,500)
 #moment match gamma distribution
 y[j] ~ dgamma(eta[j]^2/sigma[j]^2, eta[j]/sigma[j]^2)
}
eta.all = sum(eta * w)
}
```

Include `eta.all` in you MCMC output. The mean and standard deviation of `eta.all` can be moment matched to form a gamma prior on η using data from all five studies.

Hierarchical method for prior on η

$$20.6 \sim \text{gamma}\left(\frac{\eta_1^2}{2.4^2}, \frac{\eta_1}{2.4^2}\right)$$

$$31.1 \sim \text{gamma}\left(\frac{\eta_2^2}{1.79^2}, \frac{\eta_2}{1.79^2}\right)$$

$$23.8 \sim \text{gamma}\left(\frac{\eta_3^2}{11.4^2}, \frac{\eta_3}{11.4^2}\right)$$

$$24.3 \sim \text{gamma}\left(\frac{\eta_4^2}{1.03^2}, \frac{\eta_4}{1.03^2}\right)$$

$$15 \sim \text{gamma}\left(\frac{\eta_5^2}{1.73^2}, \frac{\eta_5}{1.73^2}\right)$$

$$\eta_j \sim \text{gamma}\left(\frac{\mu_\eta^2}{\sigma_\eta^2}, \frac{\mu_\eta}{\sigma_\eta^2}\right)$$

η_j is the true mean intake from study j. The numbers on the left hand side of the $\sim$ are the observed mean intakes. The numbers in the denominators are the squared standard errors from each study. μ_η is the mean intake across all possible realizations of intake studies.

Corresponding JAGS code

```
model{
mu_eta ~ dunif(0,500)
sigma.eta ~ dunif(0,50)
for(j in 1:length(y)){
 #moment match gamma distribution
 eta[j] ~ dgamma(mu_eta^2/sigma.eta^2, mu_eta/sigma.eta^2)
 y[j] ~ dgamma(eta[j]^2/sigma[j]^2, eta[j]/sigma[j]^2)
 }
}
```

Include `mu_eta` in your MCMC output. Use the mean of `mu_eta` and the standard deviation of `mu_eta` from the MCMC output to moment match input for a gamma prior on η reflecting information from all studies.

Comparison of methods
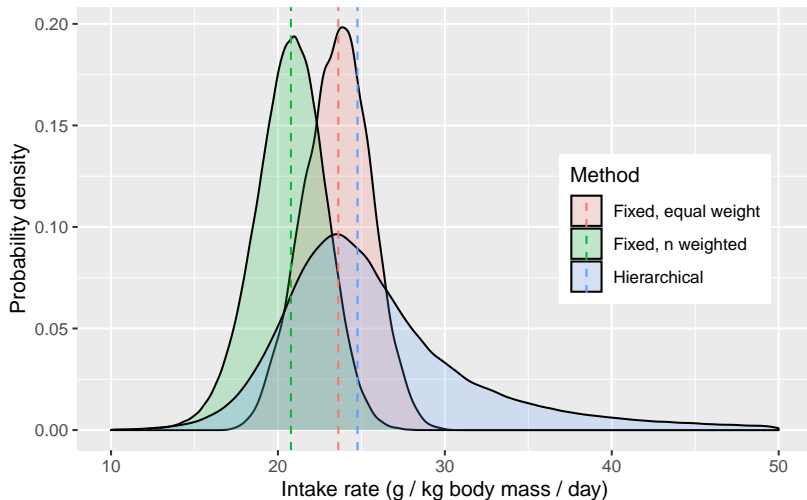
Guidance on which method to use

▶ Fixed: When you seek a highly informative prior that will exert a strong influence on the posterior.
  ▶ Equal weight: If you believe that all studies were done equally well and should play equal roles in inference.
  ▶ Weighted by sample size: When some studies have very small samples and you seek to give greater influence to studies that included large samples.
▶ Hierarchical: When you seek a prior that is more diffuse than would be obtained with a the fixed method, thereby allowing the current data to have greater influence on the posterior.

Take home

Prior distributions are powerful tools for synthesizing results from ecological studies. Three different methods can be used to find parameters of priors using results from multiple studies.