More about priors

ESS 575 Models for Ecological Data

N. Thompson Hobbs

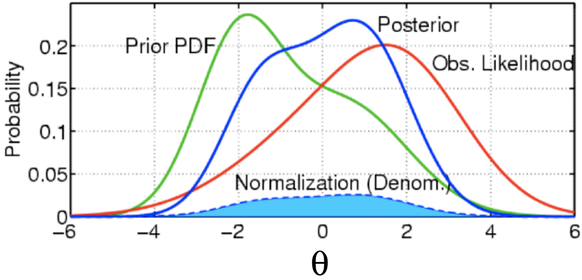
February 14, 2017

Graduate Degree Program in Ecology
Department of Ecosystem Science and Sustainability
Colorado State University

Learning outcomes

1. Explain basic principles of Bayesian inference.
2. Diagram and write out the posterior and joint distributions for Bayesian models.
3. Explain basics of Markov chain Monte Carlo (MCMC).
4. Use software for implementing MCMC methods .
5. Develop and implement hierarchical models.
6. Evaluate model fit.
7. Appreciate possibilities for model selection.
8. Understand papers and proposals using Bayesian methods.

Recall that the posterior distribution represents a balance between the information contained in the likelihood and the information contained in the prior distribution.



An informative prior influences the posterior distribution. A vague prior exerts minimal influence.

Roadmap

- Informative priors
- Vague priors
  - Scaling
  - Potential problems
  - Guidance
- Conjugate priors

Why use informative priors?

- A natural tool for synthesis and updating
- Speed convergence
- Reduce problems with identifyability
- Allows estimation of quantities that would otherwise be inestimable
- Reduces problems with sensitivity to transformation

They are a great tool! Why would you not use them?

Why are they not used more often?

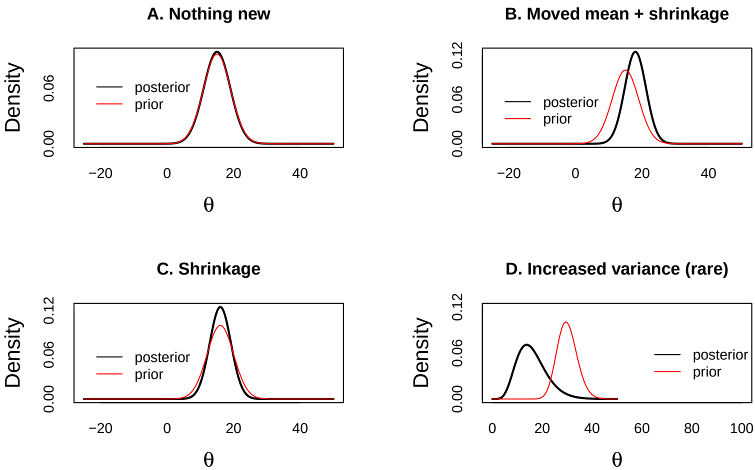
- Cultural– "All studies stand alone."
- Current texts mostly use vague priors (including ours!)
- Hard work!
- Worries about "excessive subjectivity"

How to develop?

- Strong scholarship is the basis of strong priors.
- Often need to use moment matching to convert means and standard deviations into parameters for priors.
- Pilot studies
- In biology, allometric relationships are a great source of informative priors on all sorts of parameters.[1]
- Build deterministic models with parameters with biological or socio-ecological definitions.
- Think about rescaling the data.

[1] See Peters. 1983. The Ecological Implications of Body Size. Cambridge University Press, Cambridge, U.K. and Pennycuick, C. J. 1992. Newton Rules Biology. Oxford University Press, Oxford U.K.

Interpreting posteriors relative to priors

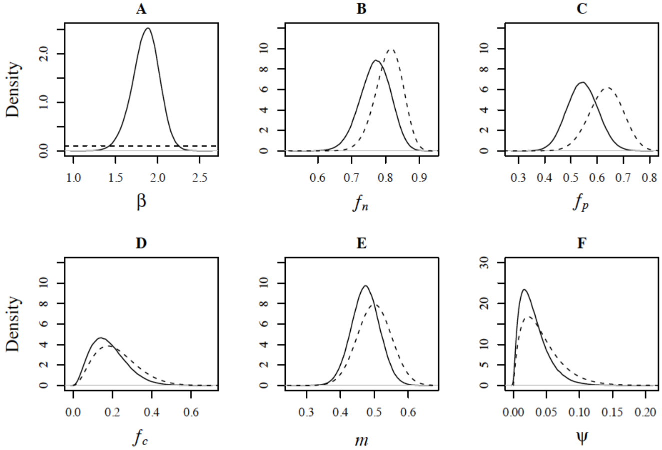


Presenting informative priors

Table 3: Prior distributions for parameters in model of brucellosis in the Yellowstone bison population. Sources are given for informative priors.

Parameter	Definition	Distribution	Mean	SD	Source
β	Rate of transmission (yr^{-1})	uniform(0,50)	25	14.3	vague
f_n	Number of offspring recruited per seronegative (susceptible) female	beta(77,18)	.81	.04	Fuller et al., 2007
f_p	Number of offspring recruited per seropositive (recovered) female	beta(37,20)	.64	.06	Fuller et al., 2007
f_c	Number of offspring recruited per seroconverting	beta(3.2,11)	.22	.10	Fuller et al., 2007

Presenting informative priors



A vague prior is a distribution with a range of uncertainty that is clearly wider than the range of reasonable values for the parameter (Gelman and Hill 2007:347).

- Also called: diffuse, flat, automatic, nonsubjective, locally uniform, objective, and, incorrectly, "non-informative."
- The best way to make a prior vague is to collect lots of good data!

Vague priors are *provisional* in two ways:

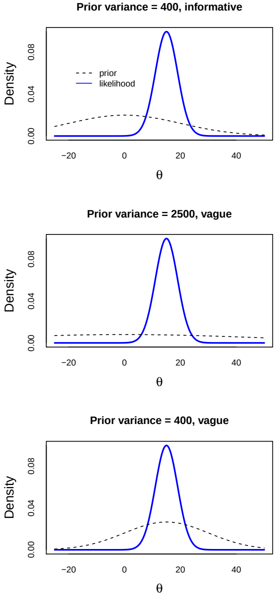
1. Operationally provisional: We try one. Does the output make sense? Are the posteriors sensitive to changes in parameters? Are there values in the posterior that are simply unreasonable? We may need to try another type of prior.
2. Strategically provisional: We use vague priors until we can get informative ones, which we prefer to use.

Scaling

Vague priors need to be scaled properly.

Suppose you specify a prior on a parameter, $\theta \sim \text{normal}(\mu = 0, \sigma^2 = 1000)$. Will this prior influence the posterior distribution?

Scaling vague priors



Problems with excessively vague priors

- Computational: failure to converge, slicer errors, failure to calculate log density, etc.
- Cause pathological behavior in posterior distribution, i.e, values are included that are unreasonable.
- Sensitivity: changes in parameters of “vague” priors meaningfully changes the posterior when data sets are small or with high variance.
- Non-linear functions of parameters with vague priors have informative priors.

Why conjugate priors?

- A easy way to find parameters of posterior distributions for simple problems.
- Critical to understanding Markov chain Monte Carlo (coming soon).

What are conjugate priors?

Assume we have a likelihood and a prior:

$$\overbrace{[\theta|y]}^{\text{posterior}} = \frac{\overbrace{[y|\theta]}^{\text{likelihood}}\ \overbrace{[\theta]}^{\text{prior}}}{[y]}.$$

If the form of the distribution of the posterior

$$[\theta|y]$$

is the same as the form of the distribution of the prior,

$$[\theta]$$

then the likelihood and the prior are said to be conjugates

$$\underbrace{[y|\theta][\theta]}_{\text{congugates}}$$

and the prior is called a conjugate prior for θ.

Derivation of beta-binomial conjugate relationship

We seek to the posterior distribution of the parameter ϕ, the probability of success on n trials with y successes:

$$[\phi|y] \propto \underbrace{\begin{pmatrix} y \\ n \end{pmatrix} \phi^{y}(1-\phi)^{n-y}}_{\text{binomial likelihood}} \underbrace{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \phi^{\alpha-1}(1-\phi)^{\beta-1}}_{\text{beta prior}}. \quad (1)$$

- Drop the normalizing constant:

$$[\phi|y] \propto \underbrace{\phi^{y}(1-\phi)^{n-y}}_{\text{binomial likelihood}} \underbrace{\phi^{\alpha-1}(1-\phi)^{\beta-1}}_{\text{beta prior}} \quad (2)$$

- Simplify and substitute:

$$[\phi|y] \propto \phi^{y+\alpha-1}(1-\phi)^{\beta+n-y-1} \quad (3)$$

Derivation of beta-binomial conjugate relationship

We seek to the posterior distribution of the parameter ϕ, the probability of success on n trials with y successes:

$$[\phi|y] \propto \underbrace{\left(\begin{array}{c} y \\ n \end{array}\right) \phi^{y}(1-\phi)^{n-y}}_{\text{binomial likelihood}} \underbrace{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \phi^{\alpha-1}(1-\phi)^{\beta-1}}_{\text{beta prior}}. \tag{1}$$

- Drop the normalizing constant:

$$[\phi|y] \propto \underbrace{\phi^{y}(1-\phi)^{n-y}}_{\text{binomial likelihood}} \underbrace{\phi^{\alpha-1}(1-\phi)^{\beta-1}}_{\text{beta prior}} \tag{2}$$

- Simplify and substitute:

$$[\phi|y] \propto \phi^{y+\alpha-1}(1-\phi)^{\beta+n-y-1} \tag{3}$$

Derivation of beta-binomial conjugate relationship

We seek to the posterior distribution of the parameter ϕ, the probability of success on n trials with y successes:

$$[\phi|y] \propto \underbrace{\begin{pmatrix} y \\ n \end{pmatrix} \phi^{y}(1-\phi)^{n-y}}_{\text{binomial likelihood}} \underbrace{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\phi^{\alpha-1}(1-\phi)^{\beta-1}}_{\text{beta prior}}. \quad (1)$$

- Drop the normalizing constant:

$$[\phi|y] \propto \underbrace{\phi^{y}(1-\phi)^{n-y}}_{\text{binomial likelihood}} \underbrace{\phi^{\alpha-1}(1-\phi)^{\beta-1}}_{\text{beta prior}} \quad (2)$$

- Simplify and substitute:

$$[\phi|y] \propto \phi^{y+\alpha-1}(1-\phi)^{\beta+n-y-1} \quad (3)$$

Derivation of beta-binomial conjugate relationship

- Let $\alpha_{new} = y + \alpha, \quad \beta_{new} = \beta + n - y$
- Multiply eq. 3 by the normalizing constant $\frac{\Gamma(\alpha_{new}+\beta_{new})}{\Gamma(\alpha_{new})\Gamma(\beta_{new})}$
- Voila, a new beta distribution informed by the prior and the data:

$$[\phi|y] = \frac{\Gamma(\alpha_{new}+\beta_{new})}{\Gamma(\alpha_{new})\Gamma(\beta_{new})}\phi^{\alpha_{new}-1}(1-\phi)^{\beta_{new}-1} \tag{4}$$

Derivation of beta-binomial conjugate relationship

- Let $\alpha_{new} = y + \alpha, \quad \beta_{new} = \beta + n - y$
- Multiply eq. 3 by the normalizing constant $\frac{\Gamma(\alpha_{new}+\beta_{new})}{\Gamma(\alpha_{new})\Gamma(\beta_{new})}$
- Voila, a new beta distribution informed by the prior and the data:

$$[\phi|y] = \frac{\Gamma(\alpha_{new}+\beta_{new})}{\Gamma(\alpha_{new})\Gamma(\beta_{new})}\phi^{\alpha_{new}-1}(1-\phi)^{\beta_{new}-1} \tag{4}$$

Derivation of beta-binomial conjugate relationship

- Let $\alpha_{new} = y + \alpha, \quad \beta_{new} = \beta + n - y$
- Multiply eq. 3 by the normalizing constant $\frac{\Gamma(\alpha_{new}+\beta_{new})}{\Gamma(\alpha_{new})\Gamma(\beta_{new})}$
- Voila, a new beta distribution informed by the prior and the data:

$$[\phi|y] = \frac{\Gamma(\alpha_{new} + \beta_{new})}{\Gamma(\alpha_{new})\Gamma(\beta_{new})}\phi^{\alpha_{new}-1}(1-\phi)^{\beta_{new}-1} \qquad (4)$$

Derivation of beta-binomial conjugate relationship

- Let $\alpha_{new} = y + \alpha, \quad \beta_{new} = \beta + n - y$
- Multiply eq. 3 by the normalizing constant $\frac{\Gamma(\alpha_{new}+\beta_{new})}{\Gamma(\alpha_{new})\Gamma(\beta_{new})}$
- Voila, a new beta distribution informed by the prior and the data:

$$[\phi|y] = \frac{\Gamma(\alpha_{new} + \beta_{new})}{\Gamma(\alpha_{new})\Gamma(\beta_{new})}\phi^{\alpha_{new}-1}(1-\phi)^{\beta_{new}-1} \tag{4}$$

Derivation of beta-binomial conjugate relationship

- Let $\alpha_{new} = y + \alpha, \quad \beta_{new} = \beta + n - y$
- Multiply eq. 3 by the normalizing constant $\frac{\Gamma(\alpha_{new}+\beta_{new})}{\Gamma(\alpha_{new})\Gamma(\beta_{new})}$
- Voila, a new beta distribution informed by the prior and the data:

$$[\phi|y] = \frac{\Gamma(\alpha_{new}+\beta_{new})}{\Gamma(\alpha_{new})\Gamma(\beta_{new})}\phi^{\alpha_{new}-1}(1-\phi)^{\beta_{new}-1} \qquad (4)$$

Some commonly used conjugates

Some Conjugate Priors

Likelihood	Prior and Posterior
Normal mean	Normal (assuming known variance)
Normal variance	Inverse gamma (assuming known mean)
Binomial	Beta
Poisson	Gamma
Multinomial	Dirichlet

Also see http://en.wikipedia.org/wiki/Conjugate_prior

http://www.johndcook.com/conjugate_prior_diagram.html#postpoisson

As we learned earlier, the gamma and the beta are continuous distributions. They are conjugate priors for discrete distributions. How can this be? Explain this seeming mismatch.

Why use conjugate priors when specifying models?

- It is not necessary to use conjugate priors when simulating the posterior distribution, which we will learn how to do soon. For example, you can use uniform distributions whenever you need an uninformative prior.
- However, conjugate priors will accelerate MCMC (more about that soon).
- For simple models, you can use conjugate priors to obtain the posterior distribution in closed form, without any simulation, as illustrated next.

Return to the beta-binomial

The conjugate prior for a binomial likelihood binomial$(y|\phi)$ is beta$(\phi|\alpha,\beta)$. It specifies prior information about the probability of a success $[\phi]$.

$$[\phi|y] = \text{beta}\left(\phi|\underbrace{\overbrace{\alpha}^{\text{The prior }\alpha} + y}_{\text{The new }\alpha}, \underbrace{\overbrace{\beta}^{\text{Ther prior }\beta} + n - y}_{\text{The new }\beta}\right) \tag{5}$$

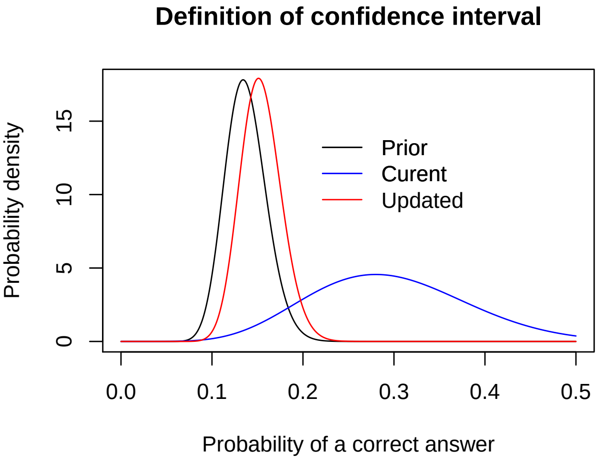
Use CSP data to update posterior distribution of correct answers

- Prior distribution based on data from Data from previous ESS 575 classes, Swedish Agricultural University, Woods Hole Research Center, Conservation Science Partners:
  - $\alpha_{\text{prior}} = 32$ successes $+ 1$
  - $\beta_{\{\text{prior}\}} = 184$ failures $+ 1$
- Data from ESS 575, Spring 2017
  - 8 successes
  - 19 failures
- Updated posterior on ϕ the probability of a correct answer:

$$\phi \sim \text{beta}(\alpha_{\text{prior}} + 8, \beta_{\text{prior}} + 19) \tag{6}$$

$$\phi \sim \text{beta}(41, 204) \tag{7}$$

Updating


Definition of confidence interval
Prior
Curent
Updated
Probability density
0
5
10
15
0.0
0.1
0.2
0.3
0.4
0.5
Probability of a correct answer

Gamma-Poisson conjugate relationship

The conjugate prior distribution for a Poisson likelihood is gamma$(\lambda|a,b)$. Given n observations y_i of new data, the posterior distribution of λ is

$$[\lambda|\mathbf{y}] = \text{gamma}\left(\lambda|\underbrace{\overbrace{a}^{\text{The prior } a} + \sum_{i=1}^{n} y_i}_{\text{The new } a}, \underbrace{\overbrace{b}^{\text{The prior } b} + n}_{\text{The new } b}\right). \qquad (8)$$

Normal mean as random variable, normal variance known

If the likelihood for the data is normal$(y_i|\mu,\sigma^2)$ with σ^2 known and the prior on μ is normal$(\mu|\mu_0,\sigma_0^2)$ then the posterior distribution of μ is:

$$\mu \sim \text{normal}\left(\frac{\left(\frac{\mu_0}{\sigma_0^2}+\frac{\sum_{i=1}^{n} y_i}{\sigma^2}\right)}{\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)}, \left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)^{-1}\right) \tag{9}$$

If the likelihood for the n observations (y_i) is normal$(y_i|\mu,\sigma^2)$ with μ known and the prior on σ^2 is inverse gamma$(\sigma^2|\alpha,\beta)$ then the posterior distribution of σ^2 is:

$$\sigma^2 \sim \text{inverse gamma}\left(\alpha+\frac{n}{2},\beta+\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}\right) \tag{10}$$