University of Nevada, Reno

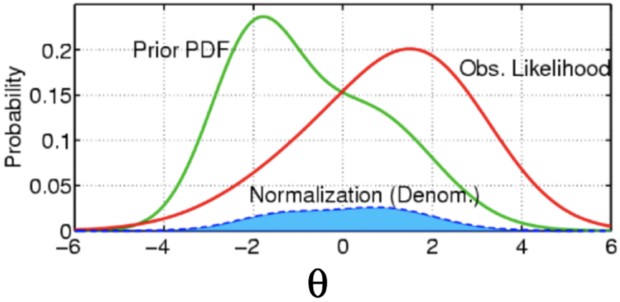
Lecture 11: Priors

Perry Williams, PhD

NRES 779

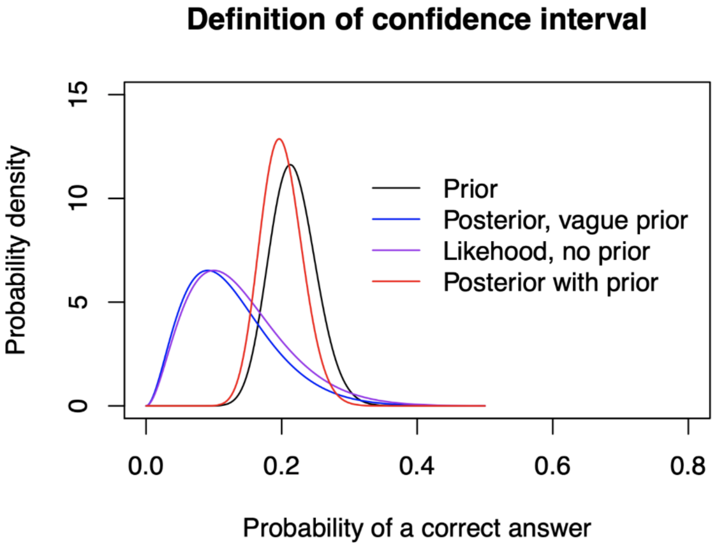
Bayesian Hierarchical Modeling in Natural Resources

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.

- Priors are optional in inference based on maximum likelihood (penalized likelihood, maximum *a posteriori* (MAP) estimation).
- Priors are mandatory in Bayesian inference. Why?


Definition of confidence interval
Probability density
15
10
5
0
Prior
Posterior, vague prior
Likehood, no prior
Posterior with prior
0.0
0.2
0.4
0.6
0.8
Probability of a correct answer

Road-map

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

Learning Objectives

- Understand the role of prior distributions (both vague and strong) in Bayesian inference
- Understand what a conjugate prior is and how to go about deriving them

Why Use Informative Priors?

- A natural tool for synthesis and updating
- Speed convergence of MCMC
- Reduce problems with identifiability
- Allows estimation of quantities that would otherwise be inestimable
- Much faster computations than possible with original data
- **Allows you to incorporate the history of science!!!**

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

Identifiability

- A model is identifiable if it is possible to learn the true values of this model's underlying parameters and latent quantities based on a "large number" of observations.
- Different values of the model's parameters must give rise to different distributions of the data.
- Non-identifiability occurs when parameters can "trade-off" and produce the same distribution of the data.

Priors

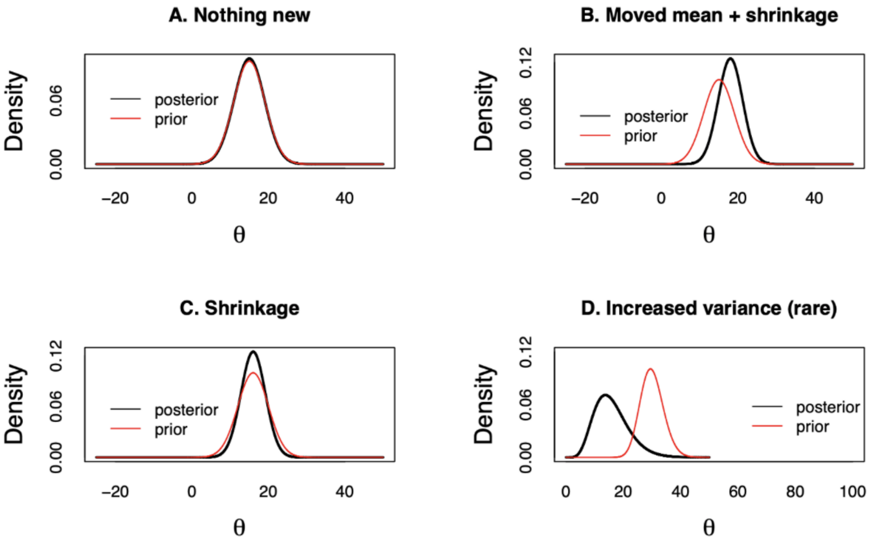
Why are they not used more often?

- Cultural reasons. "All studies stand alone"
- Current texts mostly use vague priors
- Hard work!
- Worries about "excessive subjectivity"

How to Develop Good Priors

- Strong scholarship is the basis of strong priors.
- Pilot studies
- In biology, allometric relationships are a great source of informative priors on all sorts of parameters.
- Build deterministic models with parameters defined by processes.
- Often need to use moment matching to convert means and standard deviations into parameters for priors.

Interpreting Posteriors Relative to Priors

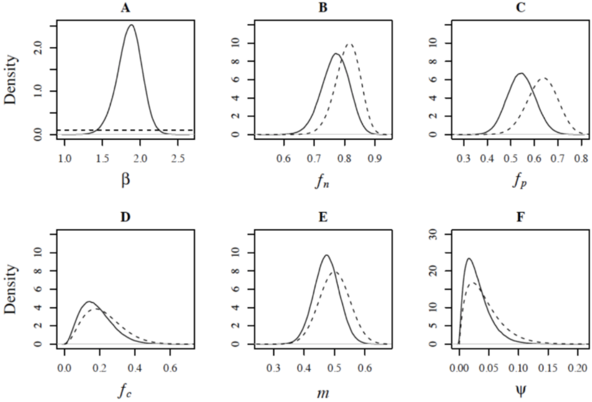


Presenting Informative Priors

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



Vague 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, non-subjective, locally uniform, objective, and, incorrectly, "non-informative."
- The best way to make a prior vague is to collect lots of good data!

Vague Priors

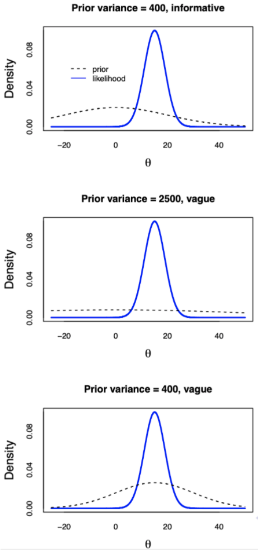
1. Operationally provisional: We try one. Does the output make sense? Are the posteriors sensitive to ranges in parameters? Are there values in the posterior that are simply unreasonable? We may need to try another type of prior.
2. Strategically provisional: We use vague priors until we can get informative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

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:

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

- 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:

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

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:

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

- 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:

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

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:

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

- Drop the normalizing constant:

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

- Simplify:

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

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:

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

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:

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

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

Some commonly used conjugates

> 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.

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

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

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

$$\boldsymbol{\mu} \sim \text{normal}\left(\frac{\left(\frac{\boldsymbol{\mu}_0}{\boldsymbol{\sigma}_0^2} + \frac{\sum_{i=1}^{n} y_i}{\boldsymbol{\sigma}^2}\right)}{\left(\frac{1}{\boldsymbol{\sigma}_0^2} + \frac{n}{\boldsymbol{\sigma}^2}\right)}, \left(\frac{1}{\boldsymbol{\sigma}_0^2} + \frac{n}{\boldsymbol{\sigma}^2}\right)^{-1}\right)$$

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$(\sigma|\alpha, \beta)$ the the posterior distribution of σ^2 is:

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

Take Home

- Prior distributions are a required, exceedingly useful component of Bayesian models.
- Good scholars inform their priors.
- Conjugate priors are an essential part of some MCMC algorithms and can be analytically useful for simple problems.