

# More About Priors I

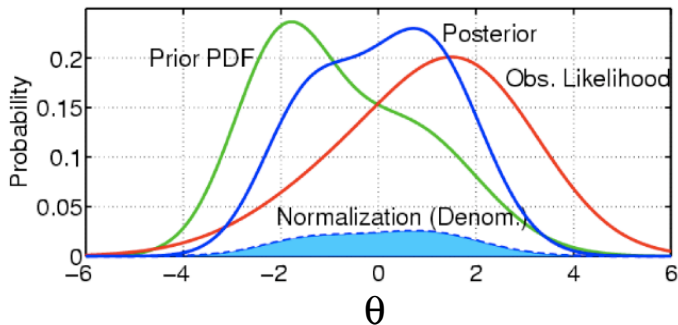
## Models for Socio-Environmental Data

Chris Che-Castaldo, Mary B. Collins, N. Thompson Hobbs

May 23, 2019



The posterior distribution uses information from both the likelihood and prior distributions. *An informative prior is more influential than an vague prior.*



# Outline

- Informative priors
- Vague priors
- Conjugate priors

# Why use informative priors?

- They speed up convergence
- They reduce problems with identifiability
- They can allow you to estimate difficult/impossible quantities

# Why don't we find people using informative priors more often?

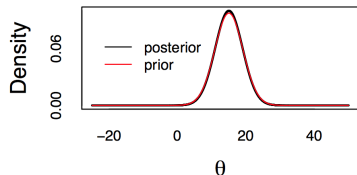
- *Cultural*: “All studies stand alone” argument
- Texts often use vague priors (including H&H)
- Hard work!
- Concerns about “excessive subjectivity”

# If you wanted to use an informative prior, how would you do it?

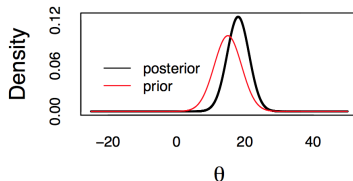
- Strong scholarship is the basis for strong priors
- Moment match, converting means and standard deviations to usable parameters
- Pilot studies
- Allometric relationships
- Deterministic models with parameters that have specific meaning

# How much does an informative prior influence the posterior?

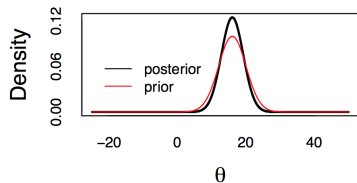
**A. Nothing new**



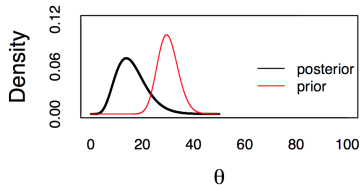
**B. Moved mean + shrinkage**



**C. Shrinkage**



**D. Increased variance (rare)**



# Communicating your use of informative priors

*Ecological Monographs*, 85(4), 2015, pp. 525–556  
© 2015 by the Ecological Society of America

## State-space modeling to support management of brucellosis in the Yellowstone bison population

N. THOMPSON HOBBS,<sup>1,5</sup> CHRIS GEREMIA,<sup>2</sup> JOHN TREANOR,<sup>2</sup> RICK WALLEN,<sup>2</sup> P. J. WHITE,<sup>2</sup> MEVIN B. HOOTEN,<sup>3</sup> AND  
JACK C. RHYAN<sup>4</sup>

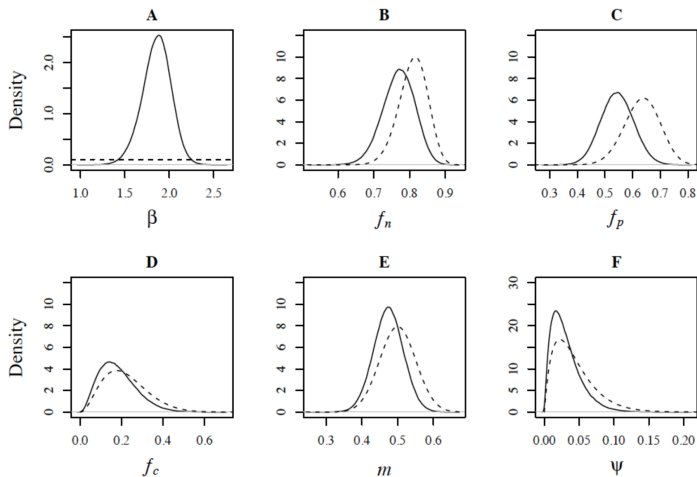


# Communicating your use of informative priors

*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
$\beta$	Rate of transmission ( $\text{yr}^{-1}$ )	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 female	beta(3.2,11)	.22	.10	Fuller et al., 2007

# Communicating your use of 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).*

# Vague Priors

- Avoid calling a prior “uninformative” or “non-informative” rather:
  - ▶ diffuse
  - ▶ flat
  - ▶ automatic
  - ▶ nonsubjective
  - ▶ locally uniform
  - ▶ objective

# Issues With Vague Priors

- Computational: failure to converge, slicer errors, failure to calculate log density, etc.
- Sensitivity: changes in parameters of “vague” priors meaningfully changes the posterior when data sets are small or when they have high variance (e.g.  $\tau \sim \text{gamma}(.001, .001)$  can really be problematic, this will come up in the multilevel modeling lab)

# Conjugacy

- In special cases the posterior,  $[\theta|y]$ , has the same distributional form as the prior,  $[\theta]$ .
  - ▶ For example, if you had a prior,  $\text{gamma}(\alpha, \beta)$ , your posterior would be  $\text{gamma}(\alpha_{\text{new}}, \beta_{\text{new}})$
- In these cases, the prior and the posterior are said to be *conjugate*.

## Conjugacy is important for two reasons:

- 1 Conjugacy minimizes computational work and, in more complicated cases, allows us to break down calculations into manageable chunks.
- 2 Conjugacy plays an important role in Markov chain Monte Carlo (more on this later).

# Deriving the Beta-Binomial Conjugacy Relationship

We know that the beta distribution is a conjugate prior for the binomial likelihood.

- Consider calculating the posterior distribution for the parameter  $\theta$ .
- $\theta$  is the probability of a success conditional on  $n$  trials and  $y$  observed successes.



# Deriving the Beta-Binomial Conjugacy Relationship

Using Bayes theorem:

$$[\phi \mid y, n] \propto \underbrace{\binom{n}{y} \phi^y (1 - \phi)^{n-y}}_{\text{binomial likelihood}} \underbrace{\frac{\gamma(\alpha + \beta)}{\gamma(\alpha)\gamma(\beta)} \phi^{\alpha-1} (1 - \phi)^{\beta-1}}_{\text{beta prior}}$$

where  $\alpha$  and  $\beta$  are beta prior parameters.

## Deriving the Beta-Binomial Conjugacy Relationship

$$[\phi \mid y, n] \propto \binom{n}{y} \phi^y (1 - \phi)^{n-y} \frac{\gamma(\alpha + \beta)}{\gamma(\alpha)\gamma(\beta)} \phi^{\alpha-1} (1 - \phi)^{\beta-1}$$

## Deriving the Beta-Binomial Conjugacy Relationship

$$[\phi \mid y, n] \propto \binom{n}{y} \phi^y (1 - \phi)^{n-y} \frac{\gamma(\alpha + \beta)}{\gamma(\alpha)\gamma(\beta)} \phi^{\alpha-1} (1 - \phi)^{\beta-1}$$

$$[\phi \mid y, n] \propto \phi^y (1 - \phi)^{n-y} \phi^{\alpha-1} (1 - \phi)^{\beta-1}$$

## Deriving the Beta- Binomial Conjugacy Relationship

$$[\phi \mid y, n] \propto \binom{n}{y} \phi^y (1 - \phi)^{n-y} \frac{\gamma(\alpha + \beta)}{\gamma(\alpha)\gamma(\beta)} \phi^{\alpha-1} (1 - \phi)^{\beta-1}$$

$$[\phi \mid y, n] \propto \phi^y (1 - \phi)^{n-y} \phi^{\alpha-1} (1 - \phi)^{\beta-1}$$

$$[\phi \mid y, n] \propto \phi^{y+\alpha-1} (1 - \phi)^{\beta+n-y-1}$$

## Deriving the Beta-Binomial Conjugacy Relationship

$$[\phi \mid y, n] \propto \phi^{y+\alpha-1} (1-\phi)^{\beta+n-y-1}$$

Let  $\alpha_{new} = y + \alpha$  and  $\beta_{new} = \beta + n - y$ . Multiply by normalizing constant:

$$\frac{\gamma(\alpha_{new} + \beta_{new})}{\gamma(\alpha_{new})\gamma(\beta_{new})},$$

and the posterior of  $\phi$  is a beta distribution:

$$[\phi \mid y, n] = \frac{\gamma(\alpha_{new} + \beta_{new})}{\gamma(\alpha_{new})\gamma(\beta_{new})} \phi^{\alpha_{new}-1} (1-\phi)^{\beta_{new}-1},$$

with parameters  $\alpha_{new}$  and  $\beta_{new}$ .

# Conjugate priors

Table A.3: Table of conjugate distributions

Likelihood	Prior distribution	Posterior distribution
$y_i \sim \text{binomial}(n, \phi)$	$\phi \sim \text{beta}(\alpha, \beta)$	$\phi \sim \text{beta}(\sum y_i + \alpha, n - \sum y_i + \beta)$
$y_i \sim \text{Bernoulli}(\phi)$	$\phi \sim \text{beta}(\alpha, \beta)$	$\phi \sim \text{beta}(\sum_{i=1}^n y_i + \alpha, \sum_{i=1}^n (1 - y_i) + \beta)$
$y_i \sim \text{Poisson}(\lambda)$	$\lambda \sim \text{gamma}(\alpha, \beta)$	$\lambda \sim \text{gamma}(\alpha + \sum_{i=1}^n y_i, \beta + n)$
$y_i \sim \text{normal}(\mu, \sigma^2)$ $\sigma^2$ is known.	$\mu \sim \text{normal}(\mu_0, \sigma_0^2)$	$\mu \sim \text{normal}\left(\frac{\left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^n y_i}{\sigma^2}\right)}{\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)}, \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}\right)$
$y_i \sim \text{normal}(\mu, \sigma^2)$ $\mu$ is known.	$\sigma^2 \sim$ inverse gamma( $\alpha, \beta$ )	$\sigma^2 \sim$ inverse gamma( $\alpha + \frac{n}{2}, \beta + \frac{\sum_{i=1}^n (y_i - \mu)^2}{2}$ )
$y_i \sim \text{lognormal}(\mu, \sigma^2)$ , $\mu$ is known	$\sigma^2 \sim$ inverse gamma( $\alpha, \beta$ )	$\sigma^2 \sim$ inverse gamma( $n/2 + \alpha, \frac{(\log(y_i) - \mu)^2}{2} + \beta$ )
$y_i \sim \text{lognormal}(\mu, \sigma^2)$ $\sigma^2$ is known	$\mu \sim \text{normal}(\mu_0, \sigma_0^2)$	$\mu \sim \text{normal}\left(\frac{\left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^n \log y_i}{\sigma^2}\right)}{\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)}, \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}\right)$

Post = beta  
Likelihood = binom (17, 80)  
Prior = beta (1, 1)

Posterior Parameters

$$\alpha = 17 + 1 \\ = 18$$

$$\beta = 1 + 80 - 17 \\ = 64$$

$$\rightarrow \text{beta}(18, 64)$$

# Why Use Conjugacy

- It is not necessary, conjugate priors will accelerate MCMC.
- For simple models, you can use conjugate priors to obtain the posterior distribution in closed form, without any simulation.



# Things to remember

- There is no such thing as a uninformative prior, but certain priors influence the posterior distribution more than others.
- Informative priors, when properly justified, can be useful.
- Strong data overwhelms a prior.
- Priors represent current knowledge (or lack of), which is updated with data.
- We encourage you to think of vague priors as a provisional starting point.

Lab exercises.