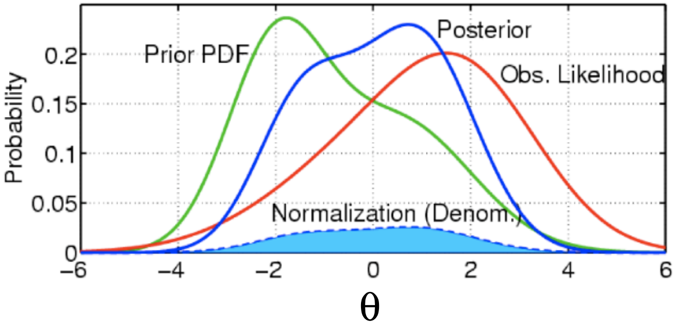
# More About Priors

## Bayesian Modeling for Socio-Environmental Data

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

August 2017

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

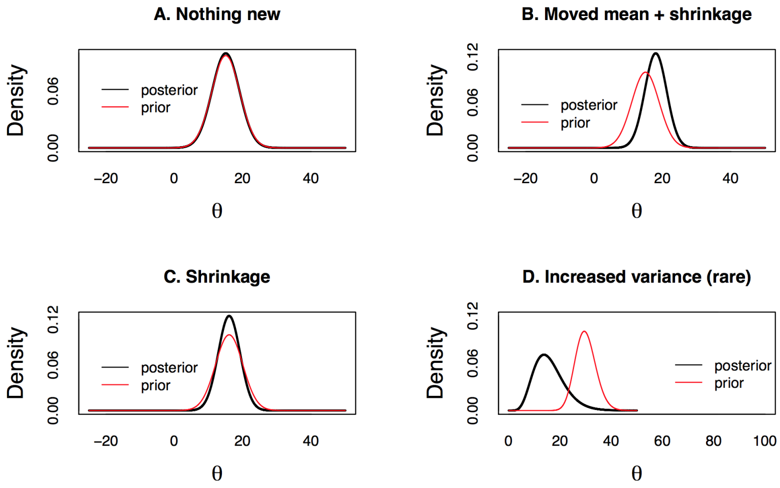
# Why don't we find people using informative priors more often?

- *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?

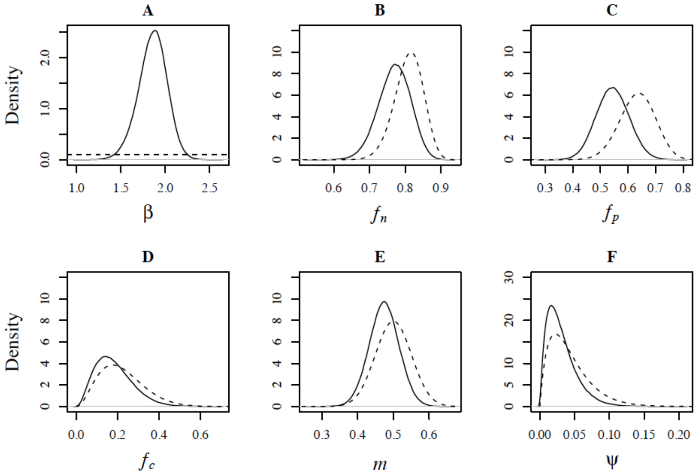


# Communicating your use of informative priors

*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 (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 |

# 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” rather, we favor difuse, 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!

# 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. Gamma$(.001, 001)$ can really be problematic)

# Conjugacy

- In special cases the posterior, $[\theta|y]$, has the same form as the prior, $[\theta]$.
- 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:

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

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

## Deriving the Beta-Binomial Conjugacy Relationship

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

(2)

(3)

## Deriving the Beta-Binomial Conjugacy Relationship

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

$$[\phi|y,n] \propto \phi^{y}(1-\phi)^{n-y}\phi^{\alpha-1}(1-\phi)^{\beta-1} \tag{2}$$

$$\tag{3}$$

## Deriving the Beta-Binomial Conjugacy Relationship

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

$$[\phi|y,n] \propto \phi^{y}(1-\phi)^{n-y}\phi^{\alpha-1}(1-\phi)^{\beta-1} \tag{2}$$

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

## Deriving the Beta-Binomial Conjugacy Relationship

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

Let, $\alpha_{new} = y + \alpha$ and $\beta_{new} = \beta + n - y$.

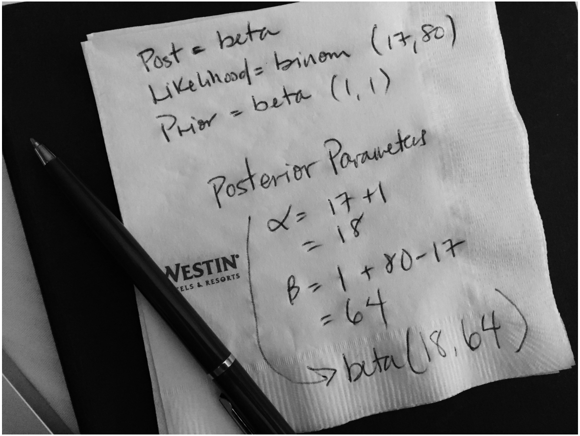
Multiply by normalizing constant $\frac{\Gamma(\alpha_{new}+\beta_{new})}{\Gamma(\alpha_{new})\Gamma(\beta_{new})}$ we find the posterior of $\phi$ to be a Beta, with parameters $\alpha_{new}$ and $\beta_{new}$.

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

# Conjugate priors

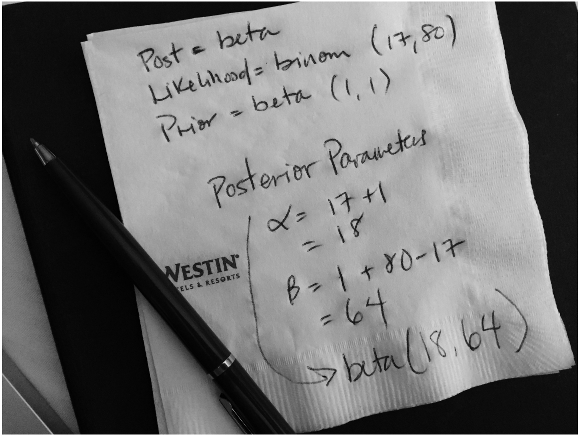
*Table A.3:* Table of conjugate distributions

| Likelihood | Prior distribution | Posterior distribution |
|---|---|---|
| $y_i \sim \text{binomial}\,(n, \phi)$ | $\phi \sim \text{beta}\,(\alpha, \beta)$ | $\phi \sim \text{beta}\,(\sum y_i + \alpha, n - \sum y_i + \beta)$ |
| $y_i \sim \text{Bernoulli}(\phi)$ | $\phi \sim \text{beta}\,(\alpha, \beta)$ | $\phi \sim \text{beta}\,(\sum_{i=1}^{n} y_i + \alpha, \sum_{i=1}^{n}(1 - y_i) + \beta)$ |
| $y_i \sim \text{Poisson}\,(\lambda)$ | $\lambda \sim \text{gamma}\,(\alpha, \beta)$ | $\lambda \sim \text{gamma}\,(\alpha + \sum_{i=1}^{n} y_i,\ \beta + n)$ |
| $y_i \sim \text{normal}(\mu, \sigma^2)$ <br> $\sigma^2$ is known. | $\mu \sim \text{normal}\,(\mu_0, \sigma_0^2)$ | $\mu \sim \text{normal}\left(\frac{\left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^{n} y_i}{\sigma^2}\right)}{\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)}, \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}\right)$ |
| $y_i \sim \text{normal}(\mu, \sigma^2)$ <br> $\mu$ is known. | $\sigma^2 \sim$ <br> inverse gamma$(\alpha, \beta)$ | $\sigma^2 \sim$ <br> inverse gamma $\left(\alpha + \frac{n}{2}, \beta + \frac{\sum_{i=1}^{n}(y_i - \mu)^2}{2}\right)$ |
| $y_i \sim \text{lognormal}(\mu, \sigma^2)$ , <br> $\mu$ is known | $\sigma^2 \sim$ <br> inverse gamma$(\alpha, \beta)$ , | $\sigma^2 \sim$ <br> inverse gamma $\left(n/2 + \alpha, \frac{(\log(y_i) - \mu)^2}{2} + \beta\right)$ |
| $y_i \sim \text{lognormal}\,(\mu, \sigma^2)$ <br> $\sigma^2$ 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
α = 17 + 1
= 18
β = 1 + 80 - 17
= 64
> beta (18, 64)
WESTIN

# 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, as illustrated next.


Post = beta
Likelihood = binom (17, 80)
Prior = beta (1, 1)
Posterior Parameters
α = 17 + 1
= 18
β = 1 + 80 - 17
= 64
=> beta(18, 64)
WESTIN

# 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.