More about priors

Bayesian Modeling for Socio-Environmental Data

N. Thompson Hobbs

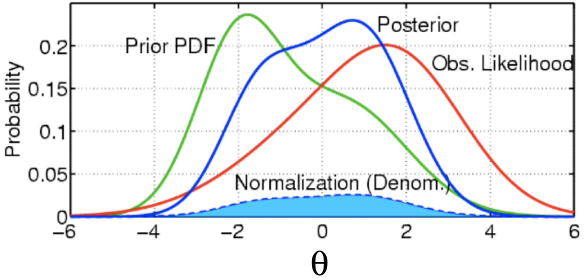
August 3, 2016

SESYNC

References for this lecture

- Hobbs and Hooten 2015, Section 5.4
- Seaman III, J. W. and Seaman Jr., J. W. and Stamey, J. D. 2012 Hidden dangers of specifying noninformative priors, The American Statistician 66, 77-84 (2012)
- McCarthy, M. A., R. Citroen, and S. C. McCall. 2008. Allometric scaling and Bayesian priors for annual survival of birds and mammals. American Naturalist 172:216-222.
- McCarthy, M. A., and P. Masters. 2005. Profiting from prior information in Bayesian analyses of ecological data. Journal of Applied Ecology 42:1012-1019.
- Seaman III, J. W. and Seaman Jr., J. W. and Stamey, J. D. 2012 Hidden dangers of specifying noninformative priors, The American Statistician 66, 77-84

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.

Influence of data and prior information

$$\text{beta}\left(\phi|y\right) = \frac{\text{binomial}\left(y|\phi,n\right)\text{beta}\left(\phi|\alpha_{prior},\beta_{prior}\right)}{[y]}$$

$$\alpha_{posterior} = \alpha_{prior} + y$$
$$\beta_{posterior} = \beta_{prior} + n - y$$

Influence of data and prior information

$$\text{gamma}\,(\lambda|\mathbf{y}) = \frac{\prod_{i=1}^{4}\text{Poisson}\,(y_i|\lambda)\,\text{gamma}\,(\lambda\,|\alpha_{prior},\beta_{prior})}{[\mathbf{y}]}$$

$$\alpha_{posterior} = \alpha_{prior} + \textstyle\sum_{i=1}^{4} y_i$$
$$\beta_{posterior} = \beta_{prior} + n$$

Roadmap

- Informative priors
- Vague priors
  - Scaling
  - Potential problems
  - Non-linear transformations of “vague” priors
  - Guidance

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?

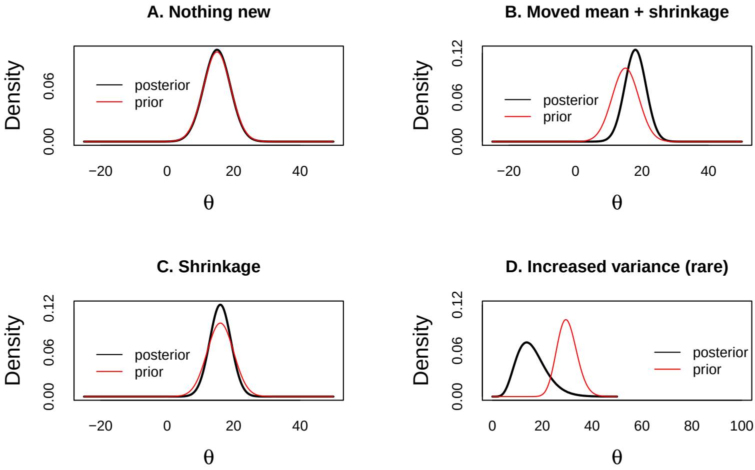
- Cultural– "All studies stand alone."
- Current texts mostly use vague priors (including ours!)
- Hard work!
- Worries about "excessive subjectivity"

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

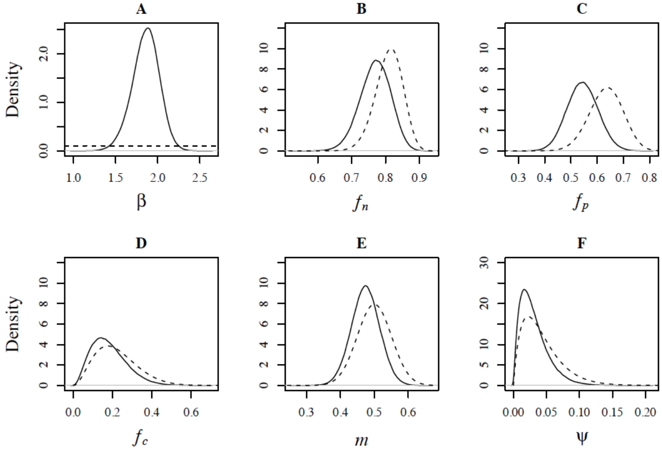


Presenting informative priors

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."

Vague priors are *provisional* in two ways:

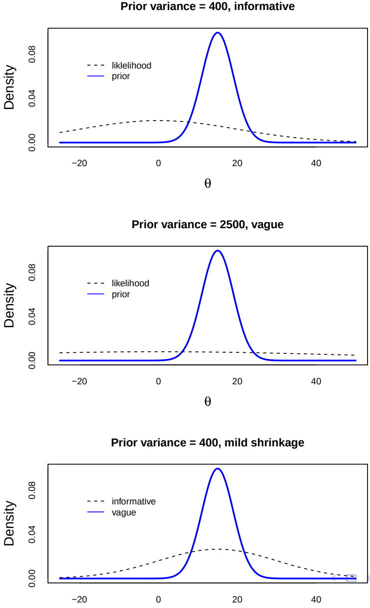
1. Operationally provisional: We try one. Does the output make sense? Are the posteriors sensitive to changes in parameters? Are there values in the posterior that are simply unreasonable? We may need to try another type of prior.
2. Strategically provisional: We use vague priors until we can get informative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

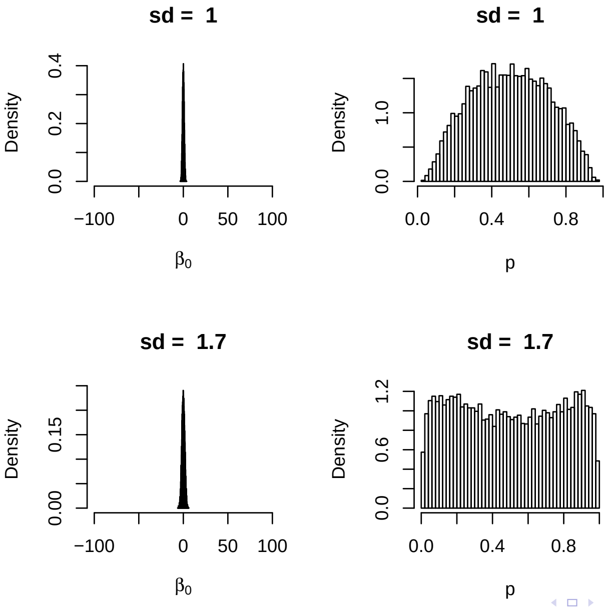


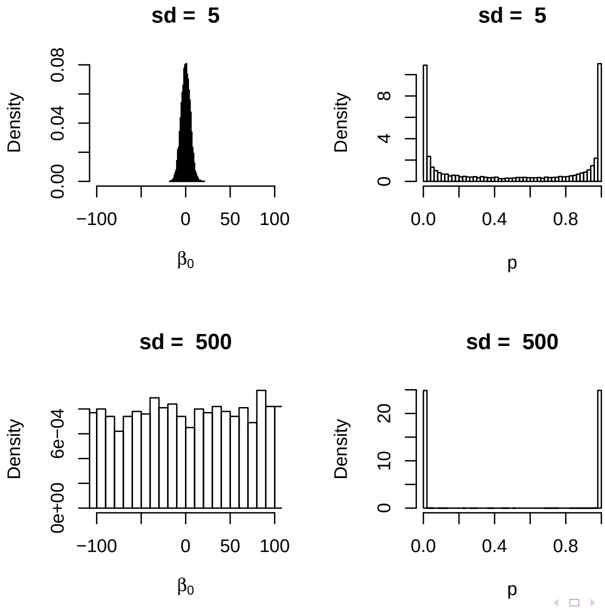
Problems with excessively vague priors

- 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: changing the parameters of “vague” priors meaningfully changes the posterior.
- Non-linear functions of parameters with vague priors have informative priors.

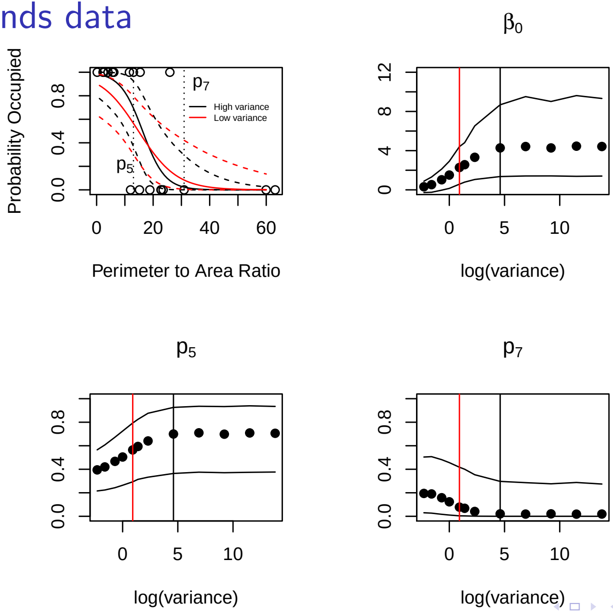
"Priors" on nonlinear functions of parameters

$$
\begin{aligned}
p_i = g(\boldsymbol{\beta}, x_i) &= \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \\
[\boldsymbol{\beta} | \mathbf{y}] &\propto \prod_{i=1}^{n} \text{Bernoulli}(y_i | g(\boldsymbol{\beta}, x_i)) \times \\
&\quad \text{normal}(\beta_0 | 0, 10000)\text{normal}(\beta_1 | 0, 10000)
\end{aligned}
$$

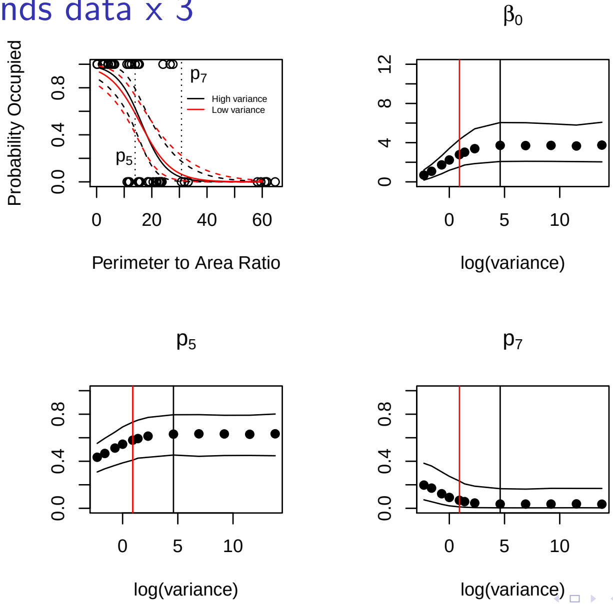

sd = 1
Density
0.0 0.2 0.4
−100 0 50 100
β0
sd = 1
Density
0.0 1.0
0.0 0.4 0.8
p
sd = 1.7
Density
0.00 0.15
−100 0 50 100
β0
sd = 1.7
Density
0.0 0.6 1.2
0.0 0.4 0.8
p


sd = 5
Density
0.00
0.04
0.08
−100
0
50
100
β0
sd = 5
Density
0
4
8
0.0
0.4
0.8
p
sd = 500
Density
0e+00
6e−04
−100
0
50
100
β0
sd = 500
Density
0
10
20
0.0
0.4
0.8
p

Islands data



Islands data x 3

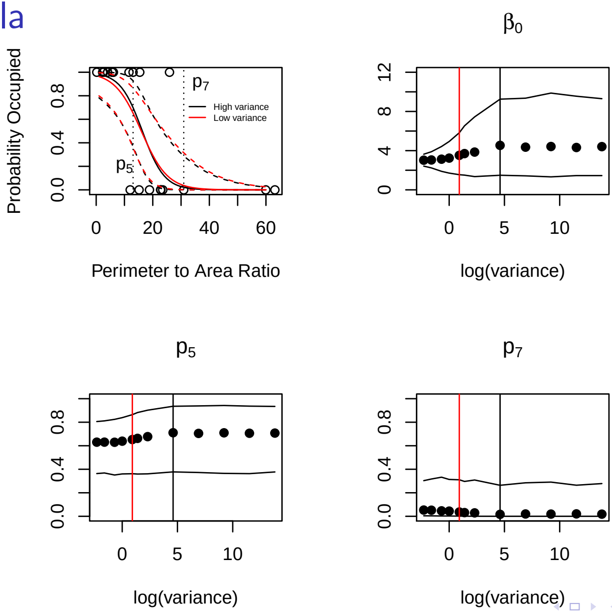


Slightly more informed priors with original data

$$
\begin{aligned}
\beta_0 &\sim \text{normal}(3, \sigma^2) \\
\beta_1 &\sim \text{normal}(-1, \sigma^2)
\end{aligned}
$$

We center β_0 on 3 using the reasoning that large islands are almost certainly (p=.95 at PA = 0) occupied. Choosing a negative value for the slope make sense because we *know* the probability of occupancy goes down as islands get smaller.

Voila


Probability Occupied
0.0
0.4
0.8
p7
p5
High variance
Low variance
0
20
40
60
Perimeter to Area Ratio
β0
12
8
4
0
5
10
log(variance)
p5
0.0
0.4
0.8
0
5
10
log(variance)
p7
0.0
0.4
0.8
0
5
10
log(variance)

More guidance

- Choose vague priors thoughtfully, particularly when data are limited.
- Use uniform priors on σ for group level models when there are 4 or more groups. Consider half-Cauchy priors with scale parameter set at reasonable estimate of σ with fewer groups.
- Know that priors that are vague for parameters can influence non-linear functions of parameters. This influence can be minimized if vague priors are centered in the vicinity of the central tendency of the posteriors of parameters.
- Always use informative priors when you can. You know more than you think you do.