# More about priors II

## Models for Socio-Environmental Data

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

June 11, 2019

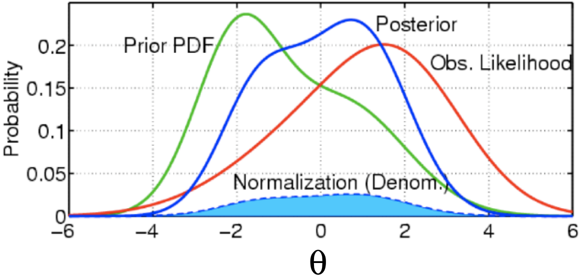
# References for this lecture


- 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)
- Northrup, J. M., and B. D. Gerber. 2018. A comment on priors for Bayesian occupancy models. PLoS ONE 13.
- Gelman, A. 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 1:1-19.
- Gelman, A., A. Jakulin, M. G. Pittau, and Y. S. Su. 2008. A weakly informative default prior distribution for logistic and other regression models. Annals of Applied Statistics 2:1360-1383.
- Gelman, A., and J. Hill. 2009. Data analysis using regression and multilevel / hierarchical models. Cambridge University Press, Cambridge, UK.

# Topics

- Priors for group-level variances in hierarchical models
- Priors for non-linear models illustrated with the inverse logit

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

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

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

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.

# 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 group-level variances in hierarchical models

The schools data

| school | estimate | sd |
|---|---|---|
| A | 28 | 15 |
| B | 8 | 10 |
| C | -3 | 16 |
| D | 7 | 11 |
| E | -1 | 9 |
| F | 1 | 11 |
| G | 18 | 10 |
| H | 12 | 18 |

# Hierarchical model

$$
\begin{aligned}
\theta_j &= \mu + \eta_j \\
y_j &\sim \text{normal}(\theta_j, \text{sd}_j) \\
\eta_j &\sim \text{normal}(0, \sigma_\theta^2) \\
\mu &\sim \text{normal}(0{,}100000) \\
\sigma_\theta^2 &\sim ?
\end{aligned}
$$

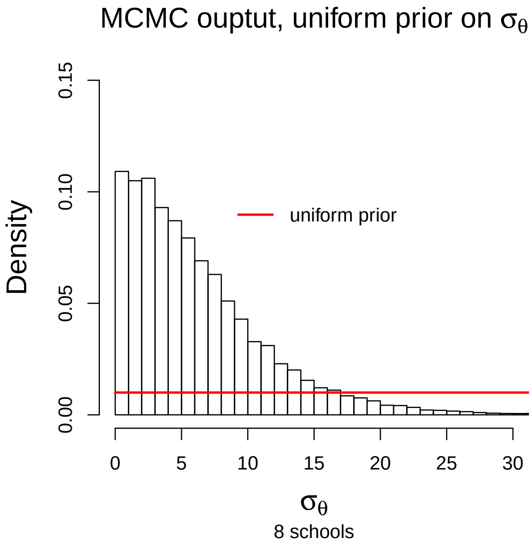
# Note that this is identical to:

$$
\begin{aligned}
y_j &\sim \text{normal}(\theta_j, \text{sd}_j) \\
\theta_j &\sim \text{normal}(\mu, \sigma_\theta^2) \\
\mu &\sim \text{normal}(0{,}100000) \\
\sigma_\theta^2 &\sim ?
\end{aligned}
$$

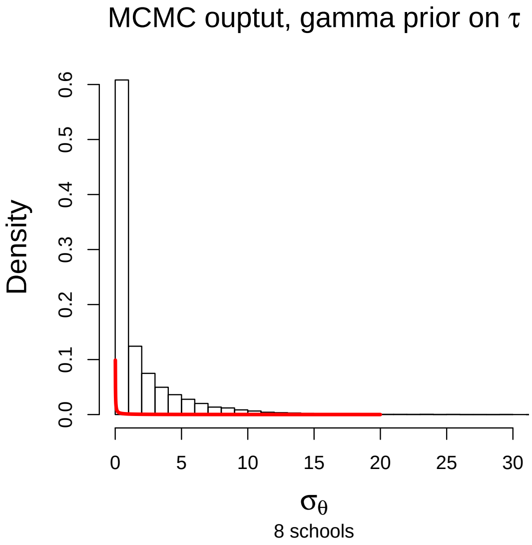
# If we had data on individual test scores...

$$
\begin{aligned}
\theta_j &= \mu + \eta_j \\
y_{ij} &\sim \text{normal}(\theta_j, \sigma_j^2) \\
\eta_j &\sim \text{normal}(0, \sigma_\theta^2) \\
\mu &\sim \text{normal(0,100000)} \\
\sigma_\theta^2 &\sim ?
\end{aligned}
$$

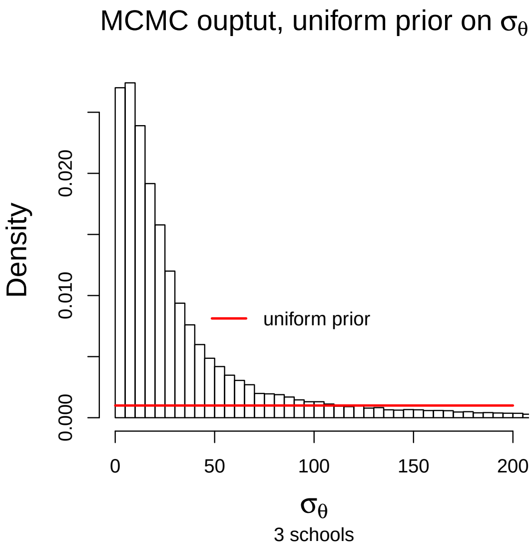
# $\sigma_\theta \sim \text{uniform}(0, 100)$, $\tau = \frac{1}{\sigma^2}$, 8 schools



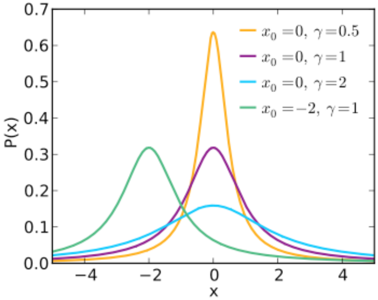
# $\tau \sim \text{gamma}(.001, .001)$, 8 schools



# $\sigma_\theta \sim \text{uniform}(0, 100)$, $\tau = \frac{1}{\sigma^2}$, 3 schools



# The Cauchy distribution




$$[z|\gamma, z_0] = \frac{1}{\pi\gamma\left[1+\left(\frac{z-z_0}{\gamma}\right)^2\right]}$$

$z_0$ = location

$\gamma$ = scale

Represents ratio of two normally distributed random variables

# A weakly informative prior on $\sigma_\theta$

half-Cauchy prior:

$$\sigma_\theta \sim \text{Cauchy}(0,\gamma)\text{T}(0,)$$

The scale parameter $\gamma$ is chosen based on experience to be a bit higher than we would expect for the standard deviation of the underlying $\theta_j$'s. This puts a weak constraint on $\sigma_\theta$. An equivalent formulation is the half t distribution,

$$\sigma_\theta \sim t(0,\gamma^2,1)\text{T}(0,) \tag{1}$$

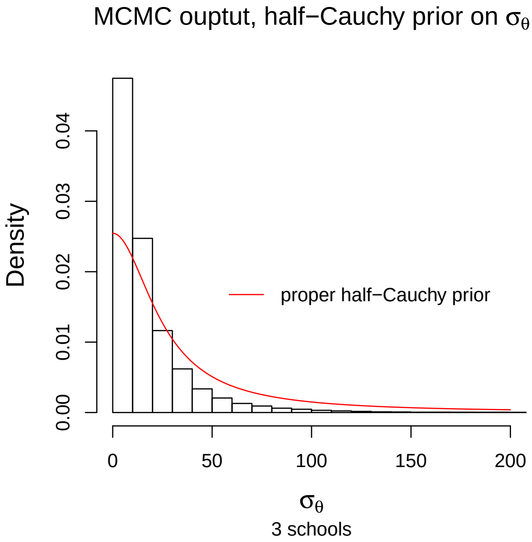
which can be coded in JAGS using

```
simga_theta ~ dt(0,1/gamma^2,1)(T,)
```

or, alternatively,

```
tau_theta ~ dscaled.gamma(gamma,1)
sigma_theta = 1/sqrt(tau_theta)
```

# A more reasonable posterior

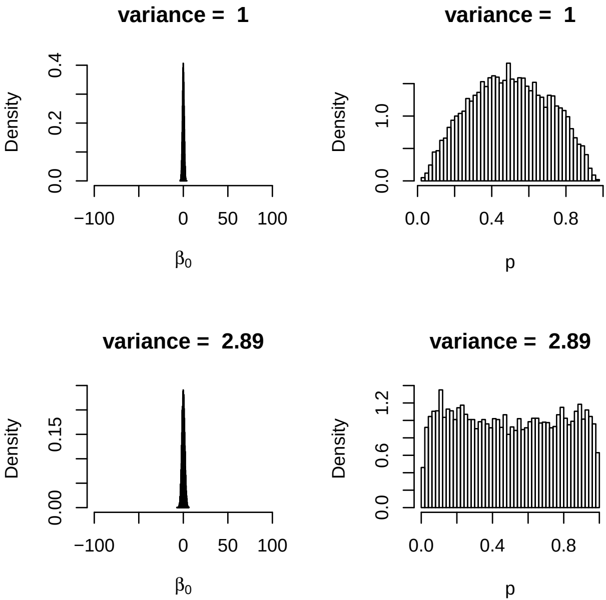


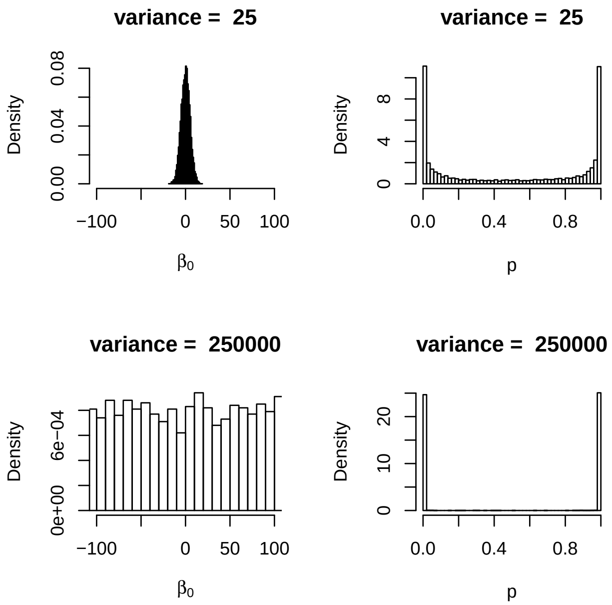
# Guidance

- Uniform priors on $\sigma$ are recommended over gamma priors on group level variances in hierarchical models with at least 4-5 groups.
- When groups are $\leq 4$, a half-Cauchy prior can usefully constrain the posterior of group level $\sigma$'s.
- This illustrates that it can be useful to use weakly informative priors when vague priors produce posteriors with unreasonable values.

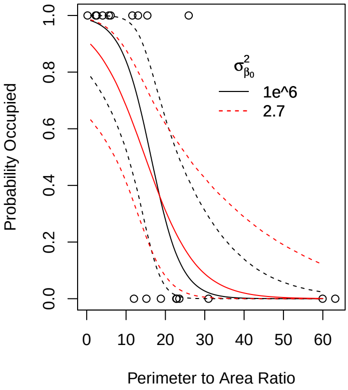
# "Priors" on nonlinear functions of parameters

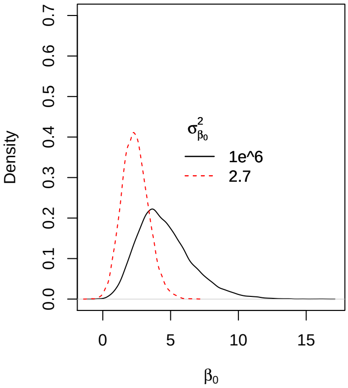
$$
\begin{aligned}
p_i = g(\boldsymbol{\beta}, x_i) &= \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}} \\
[\boldsymbol{\beta}|\mathbf{y}] &\propto \prod_{i=1}^{n} \text{Bernoulli}(y_i|g(\boldsymbol{\beta}, x_i)) \times \\
& \quad \text{normal}(\beta_0|0, 10000)\text{normal}(\beta_1|0, 10000)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


variance = 1
Density
0.0
0.2
0.4
−100
0
50
100
$\beta_0$
variance = 1
Density
0.0
1.0
0.0
0.4
0.8
p
variance = 2.89
Density
0.00
0.15
−100
0
50
100
$\beta_0$
variance = 2.89
Density
0.0
0.6
1.2
0.0
0.4
0.8
p

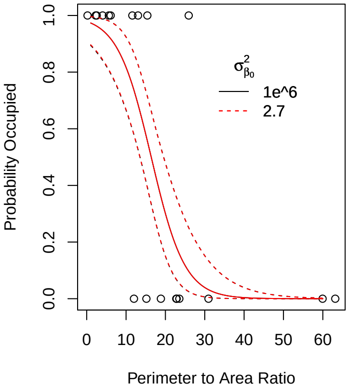

variance = 25
Density
0.00 0.04 0.08
−100 0 50 100
β0
variance = 25
Density
0 4 8
0.0 0.4 0.8
p
variance = 250000
Density
0e+00 6e−04
−100 0 50 100
β0
variance = 250000
Density
0 10 20
0.0 0.4 0.8
p

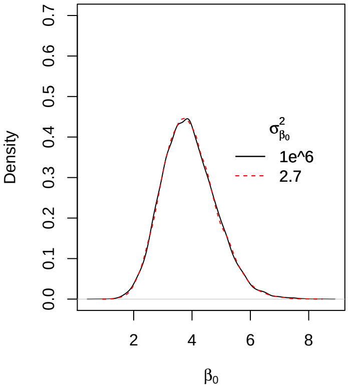
# Islands data



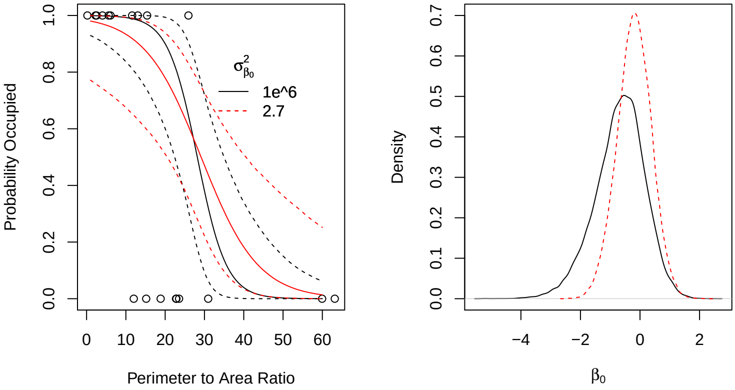



# Islands data x 4



# Standardize the original data

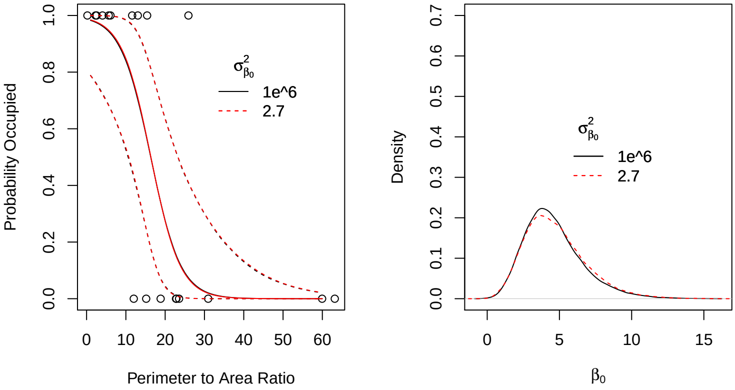


# *Slightly* more informed priors with original data

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

We center $\beta_0$ 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.

# Weakly informative priors on parameters



# Guidance

- Know that priors that are vague for parameters can influence non-linear functions of parameters.
- Explore sensitivity of all non-linear models to priors.
- Always use informative priors when you can.
- Always standardize data for non-linear models.
- Set variance ≊2.7 for normal priors on parameters in inverse logit models (precision $\approxeq .37$). Set means at "reasonable" values if possible.
- Use Cauchy prior on the coefficients, i.e.,$\beta_i \sim \text{Cauchy}(0, 2.5)$ on standardized data. Implemented in JAGS using `beta[i] ~ dt(0, 1/2.5^2, 1)`.See Gelman et al. 2008 for details.[3]

[3] Gelman, A., A. Jakulin, M. G. Pittau, and Y. S. Su. 2008. A weakly informative default prior distribution for logistic and other regression models. Annals of Applied Statistics 2:1360-1383.

# Guidance

- Always use informed priors when you can. All of the problems we discussed go away if priors are informed.
- Group level variances for fewer than four or five groups will often need sensibly informed half-Cauchy priors.
- Vague priors for non-linear models should be centered on reasonable values. Always examine sensitivity of marginal posteriors to variation in priors for non-linear models.