More About Priors 2

Alison Ketz (Notes adapted from Bailey Fosdick, Chris Che-Castaldo, Mary B. Collins, N. Thompson Hobbs)

June 10, 2024

DEB 2042028

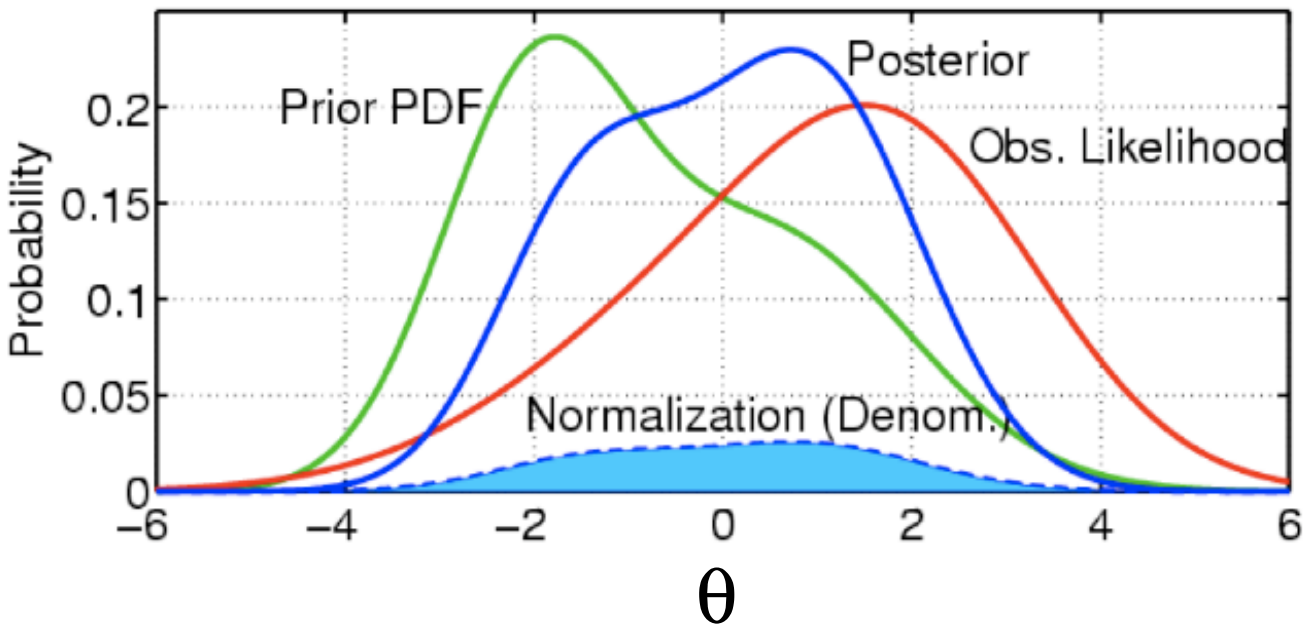
Outline

Examples when vague priors cause problems!

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

Review

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.

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, nonsubjective, locally uniform, objective, and, incorrectly, "non-informative"

Vague priors

Vague priors are *provisional* in two ways:

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

Problems with excessively vague priors

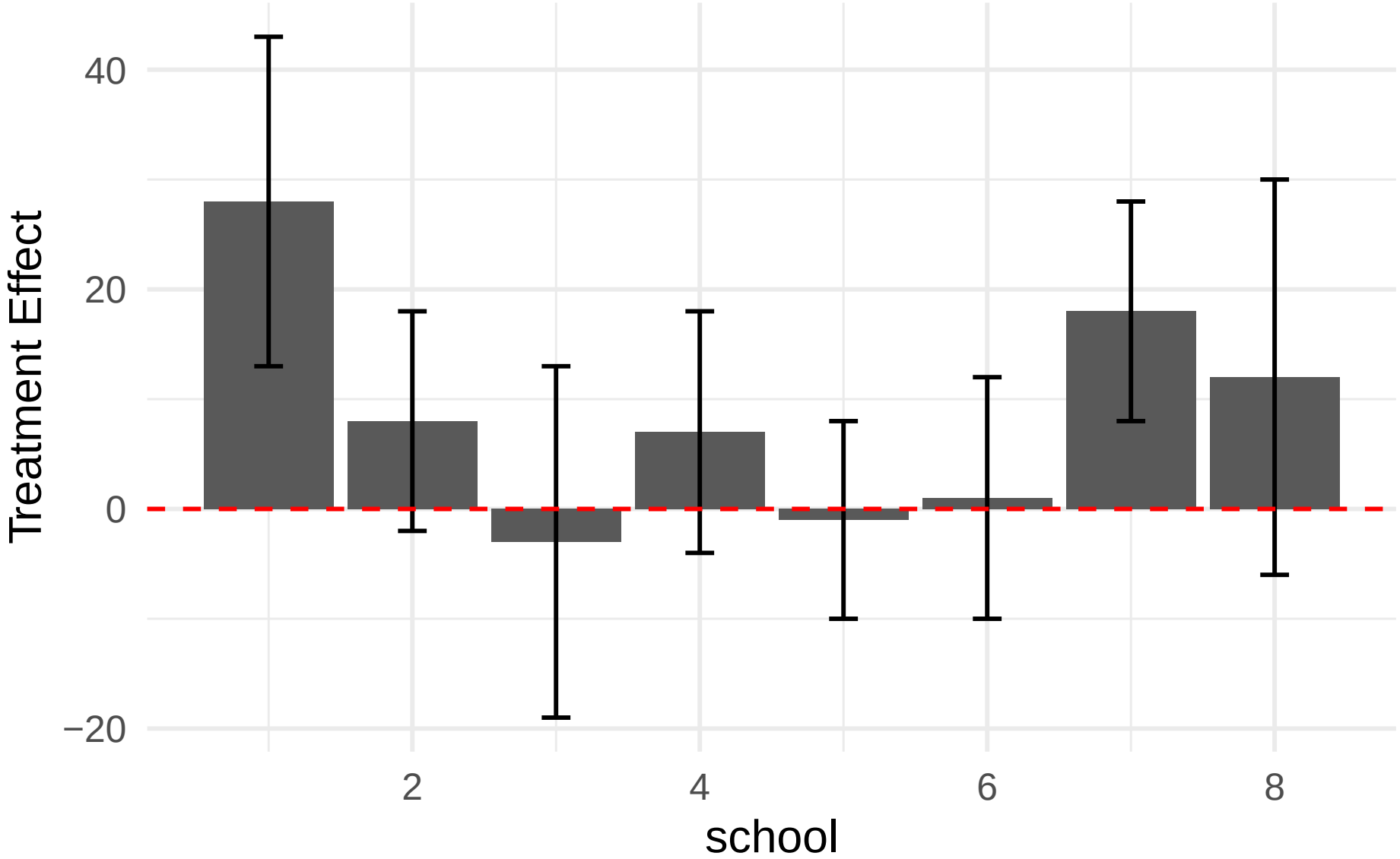
- **Computational:** failure to converge, slicer errors, failure to calculate log density, etc.
- **Cause pathological behavior in posterior distribution**, i.e, values are included that are unreasonable.
- **Sensitivity:** changing the parameters of "vague" priors meaningfully changes the posterior.
- **Non-linear functions of parameters with vague priors have informative priors.** AHH!

Ex: Educational testing experiments in schools (Gelman et al. 2013, sec 5.5)

A study was performed for the Educational Testing Service to analyze the effects of special coaching programs for SAT-V (Scholastic Aptitude Test-Verbal) in each of eight high schools. The outcome variable in each study was the score on a special administration of the SAT-V; the scores can vary between 200 and 800, with mean about 500 and standard deviation about 100.

- For each of the 8 schools ($J = 8$), we have an estimated treatment effect y_j and a standard error of the effect estimate (sd_j). The treatment effects in the study were obtained by a linear regression on the treatment group using PSAT-M and PSAT-V scores as control variables.
- "As there was no prior belief that any of the schools were more or less similar or that any of the coaching programs would be more effective, we can consider the treatment effects as exchangeable."
(https://www.tensorflow.org/probability/examples/Eight_Schools)

Data



Hierarchical model

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

What is the interpretation of σ_θ^2?

What prior distributions might we consider for σ_θ^2?

Hierarchical model

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

is equivalent to

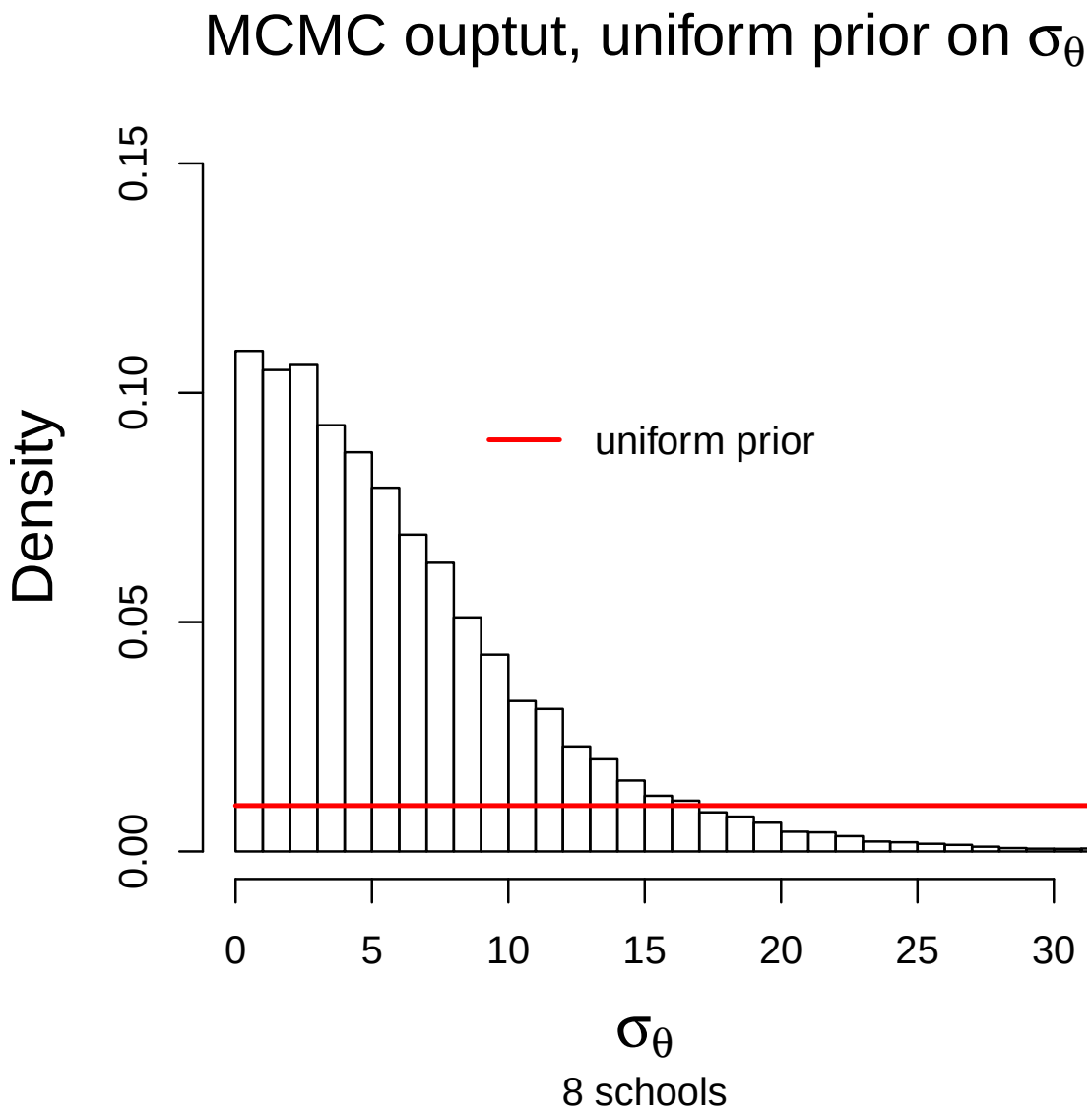
$$
\begin{aligned}
y_j &\sim \text{normal}(\theta_j, \varsigma_j^2) \\
\theta_j &\sim \text{normal}(\mu, \sigma_\theta^2) \\
\mu &\sim \text{normal}(0,100000) \\
\sigma_\theta^2 &\sim ?
\end{aligned}
$$

What if we had individual test scores? What else do we need? (Draw DAG/Write Model)

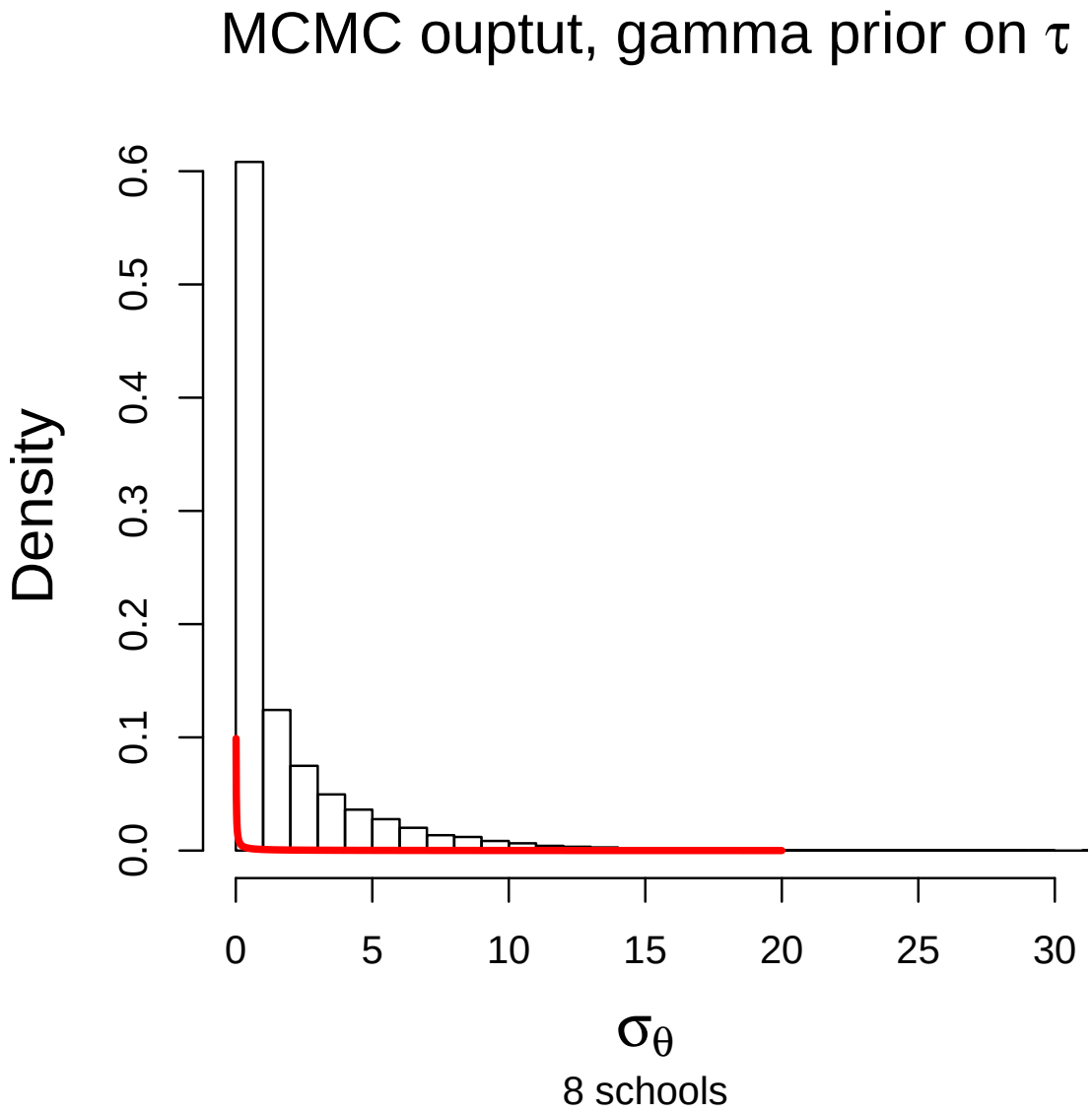
If we had individual test scores. . .

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

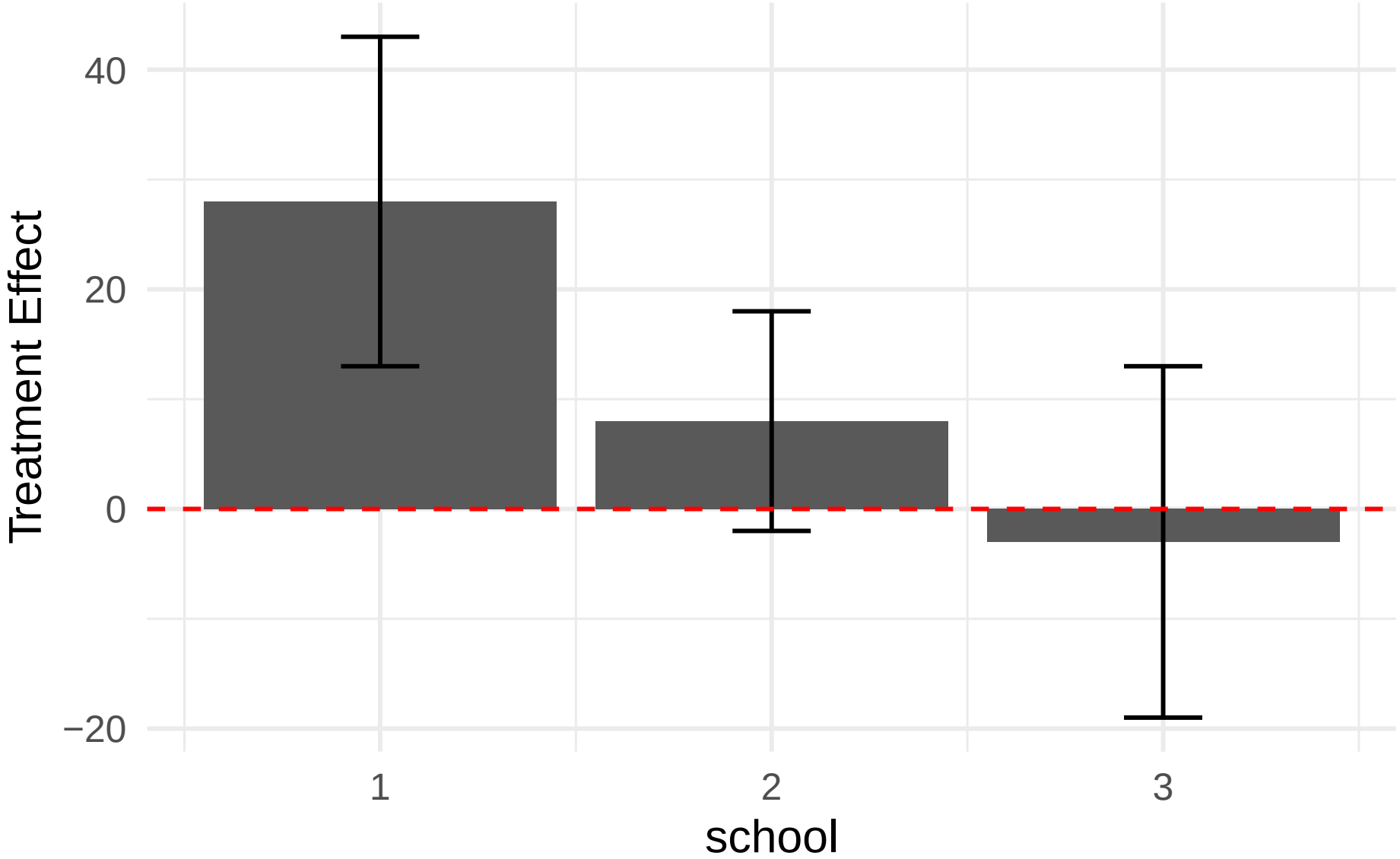
$\sigma_\theta \sim \text{uniform}(0, 1000)$, $\tau = \frac{1}{\sigma^2}$, 8 schools



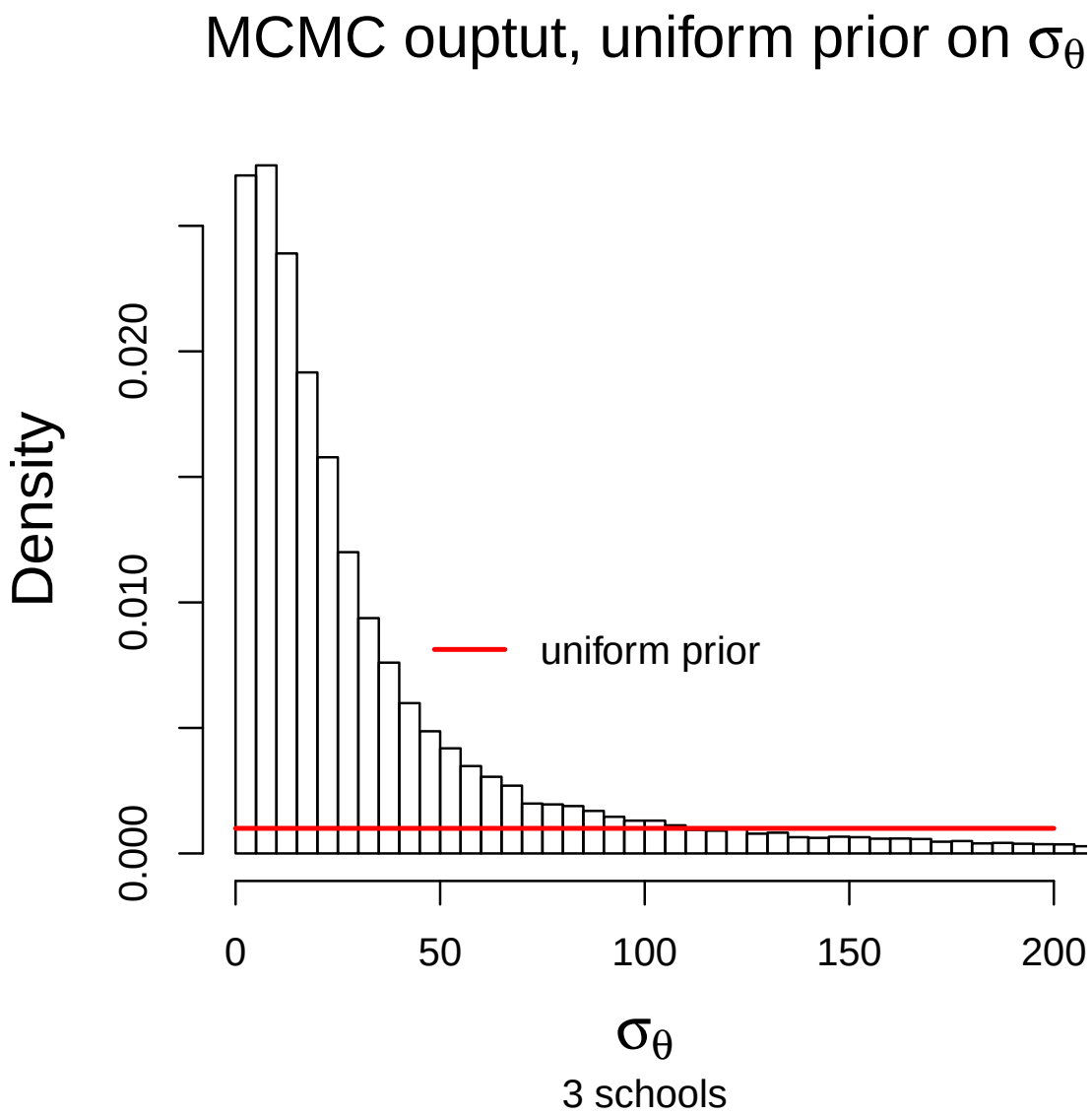
$\tau \sim$ gamma(.001, .001), 8 schools



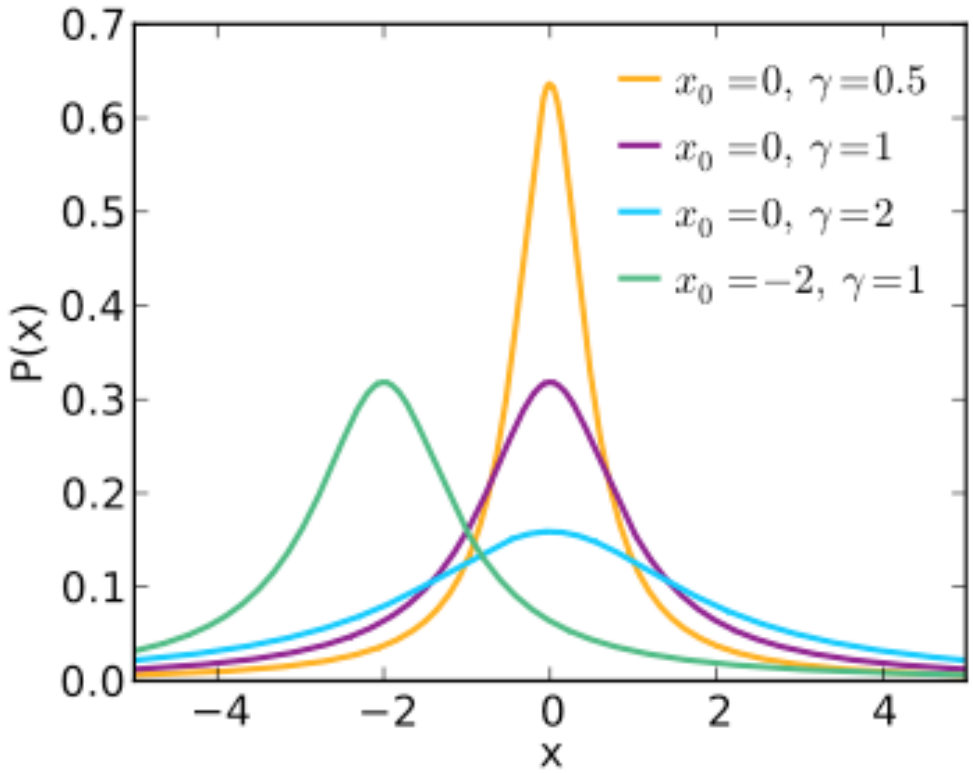
Data - 3 schools



$\sigma_\theta \sim \text{uniform}(0, 1000)$, $\tau = \frac{1}{\sigma^2}$, 3 schools – YIKES!



The Cauchy distribution



$$[x|\gamma, x_0] = \frac{1}{\pi\gamma\left[1+\left(\frac{x-x_0}{\gamma}\right)^2\right]}$$

x_0 = location, γ = scale

A weakly informative prior on σ_θ

half-Cauchy prior:

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

The scale parameter γ is chosen based on experience to be a bit higher than we would expect for the standard deviation of the underlying θ_j's. This puts a weak constraint on σ_θ. An equivalent formulation is the half t-distribution,

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

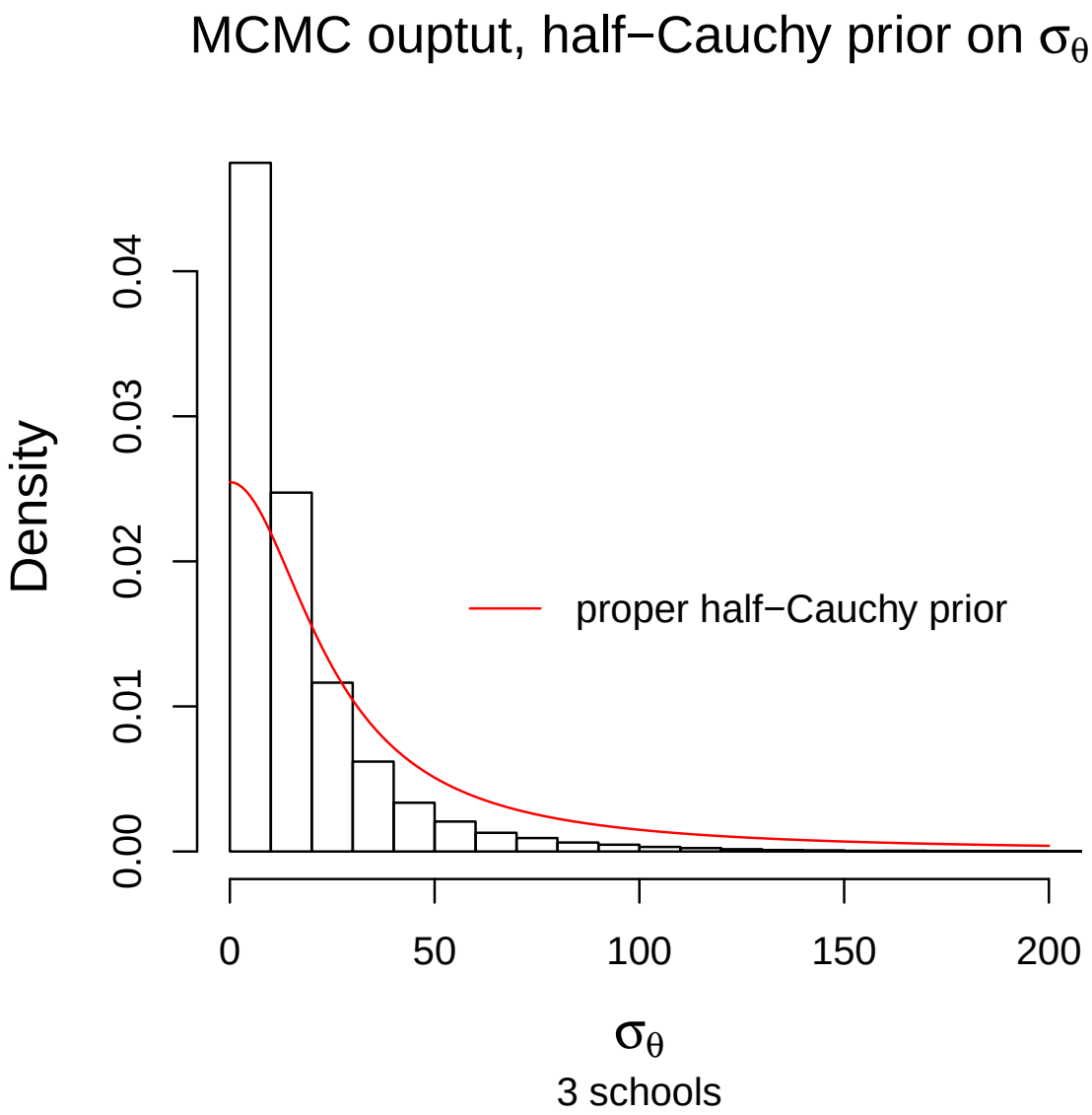
which can be coded in JAGS using

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

or, alternatively,

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

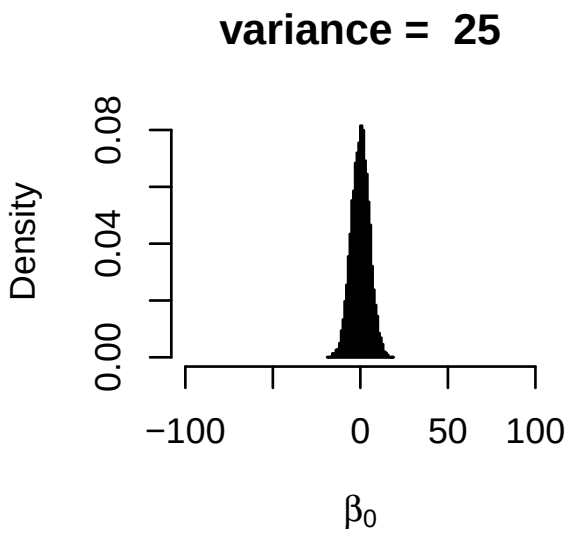
A more reasonable posterior ($\gamma = 25$)

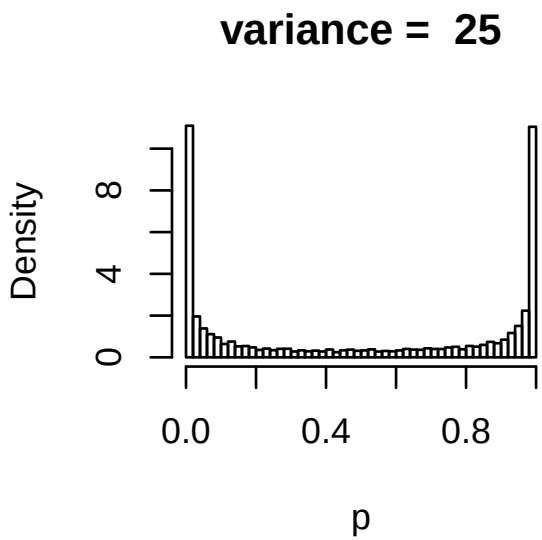


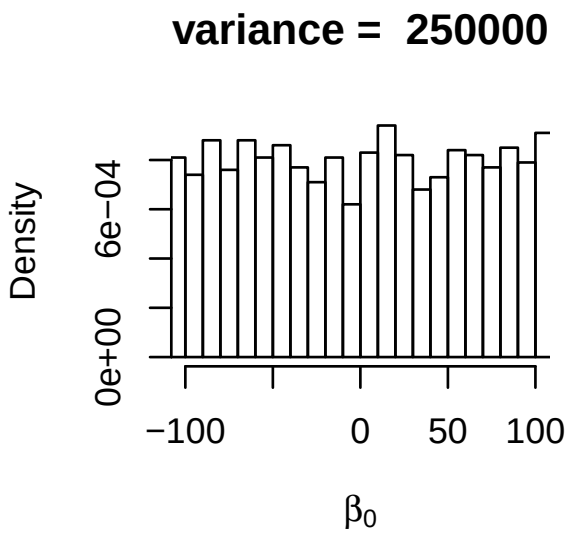
Priors on nonlinear functions of parameters

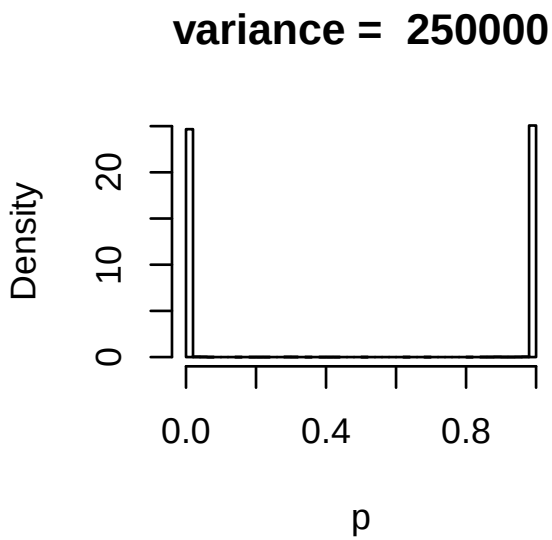
$$
\begin{aligned}
p_i = g(\beta, x_i) &= \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}} \\
[\beta|\mathbf{y}] &\propto \prod_{i=1}^{n} \text{Bernoulli}(y_i|g(\beta, x_i)) \times \\
&\quad \text{normal}(\beta_0|0, 250000)\text{normal}(\beta_1|0, 250000)
\end{aligned}
$$

Priors on nonlinear functions of parameters

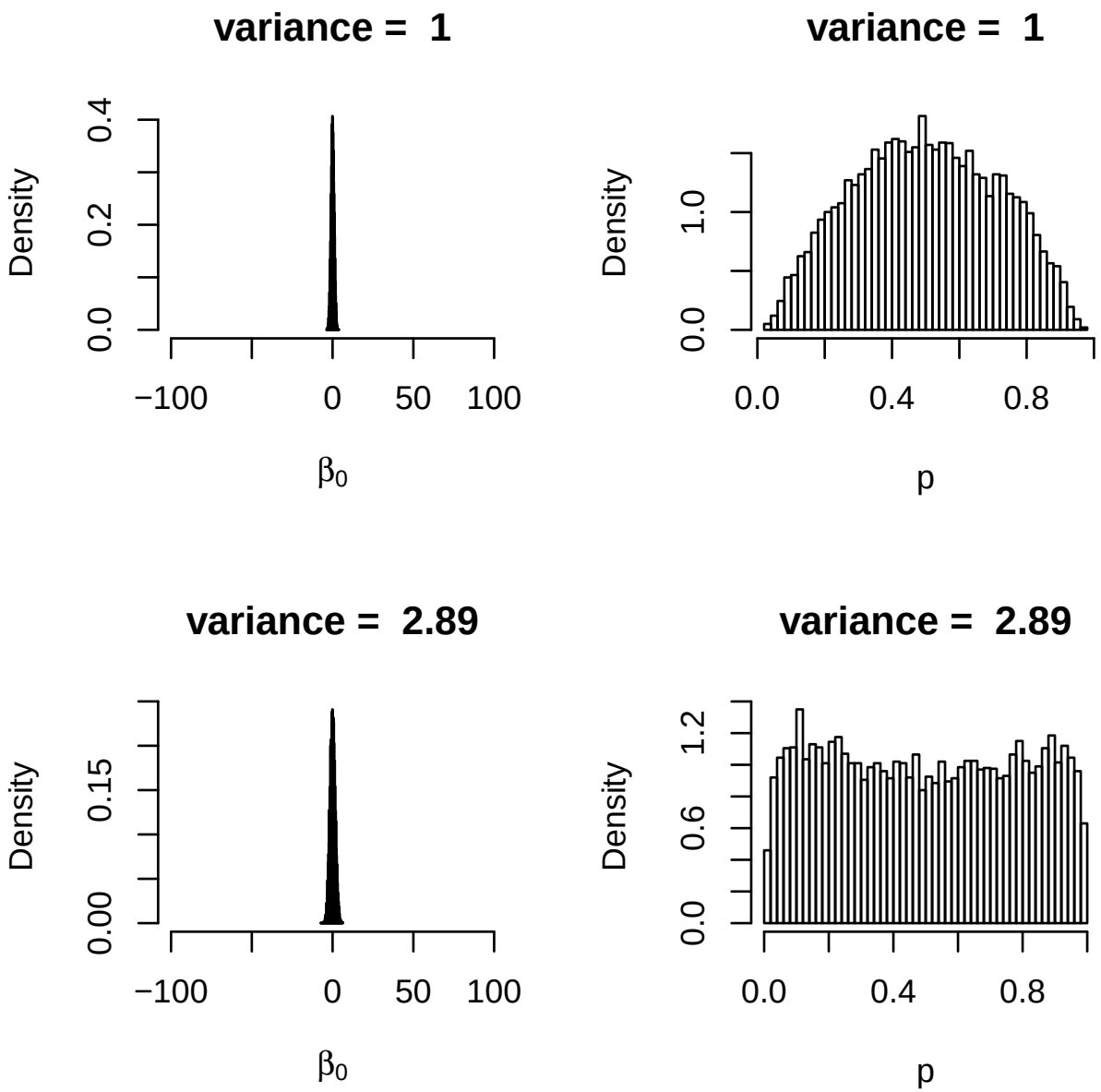








Priors on nonlinear functions of parameters



When vague priors cause problems. . . Ex: Lizards on Islands

The probability of occupancy of islands p by lizards as a function of the ratio of the islands' perimeter to area ratios (Polis et al., 1998). Recall the data set from the JAGS lab:

$$g(\beta_0, \beta_1, x_i) = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}$$

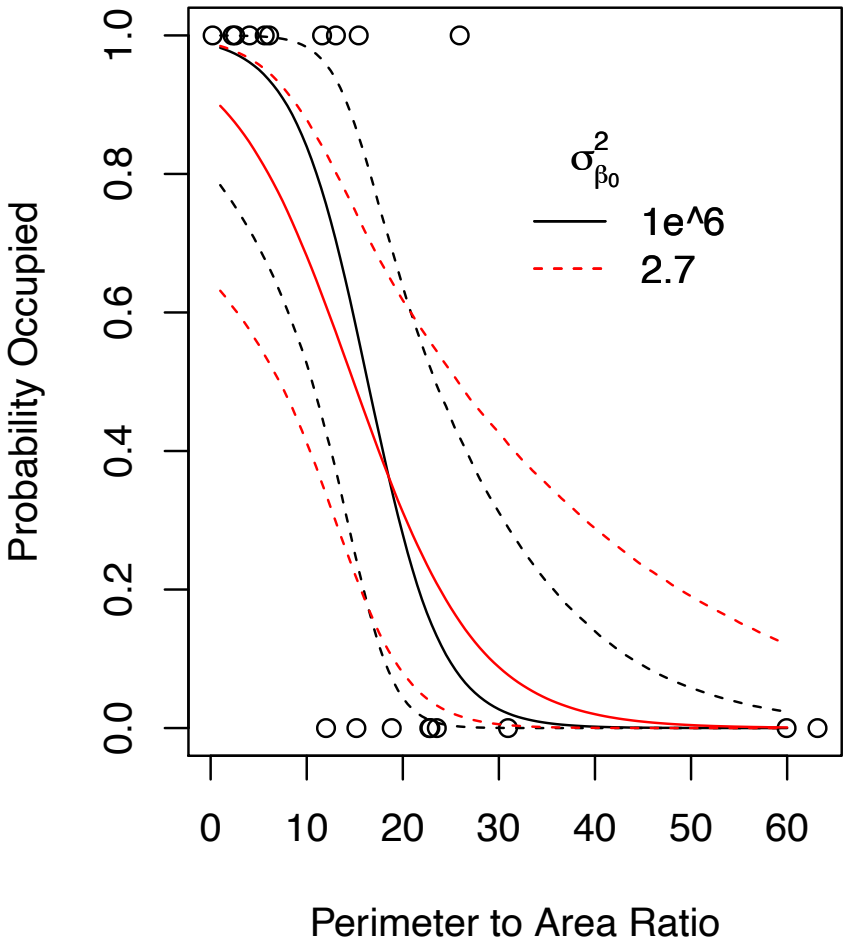
$$y_i \sim \text{Bernoulli}(p_i = g(\beta_0, \beta_1, x_i))$$

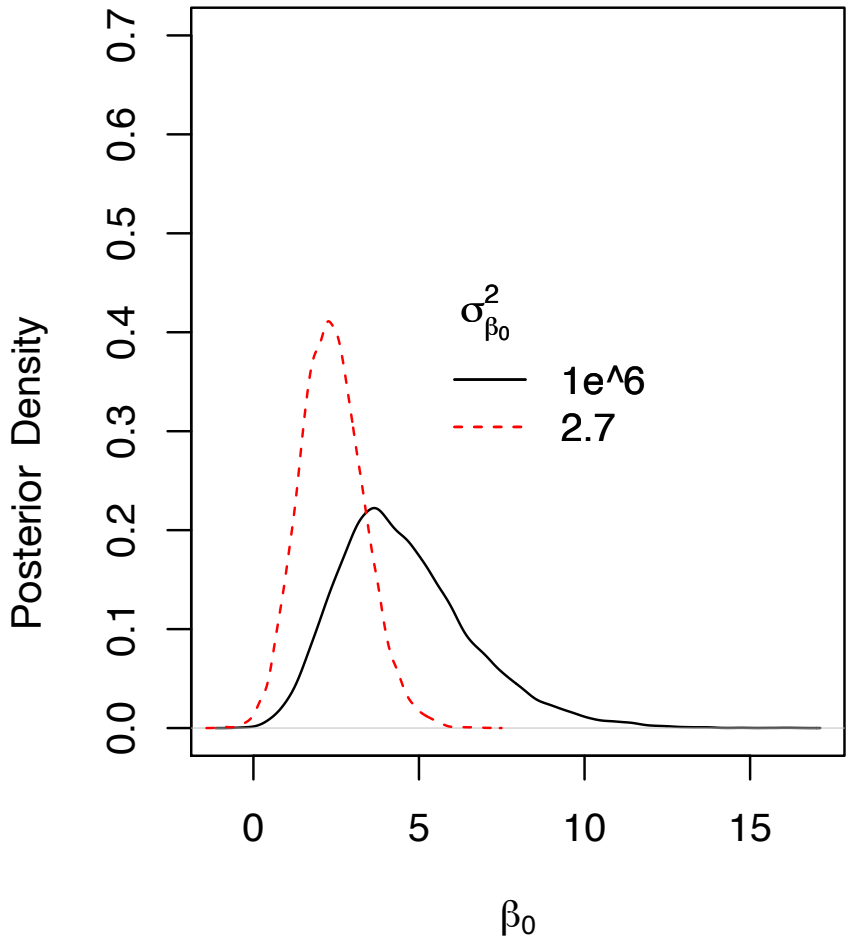
$$\beta_0 \sim \text{normal}(0, \sigma^2_{\beta_0})$$

$$\beta_1 \sim \text{normal}(0, \sigma^2_{\beta_1})$$

How should we specify $\sigma^2_{\beta_0}$?

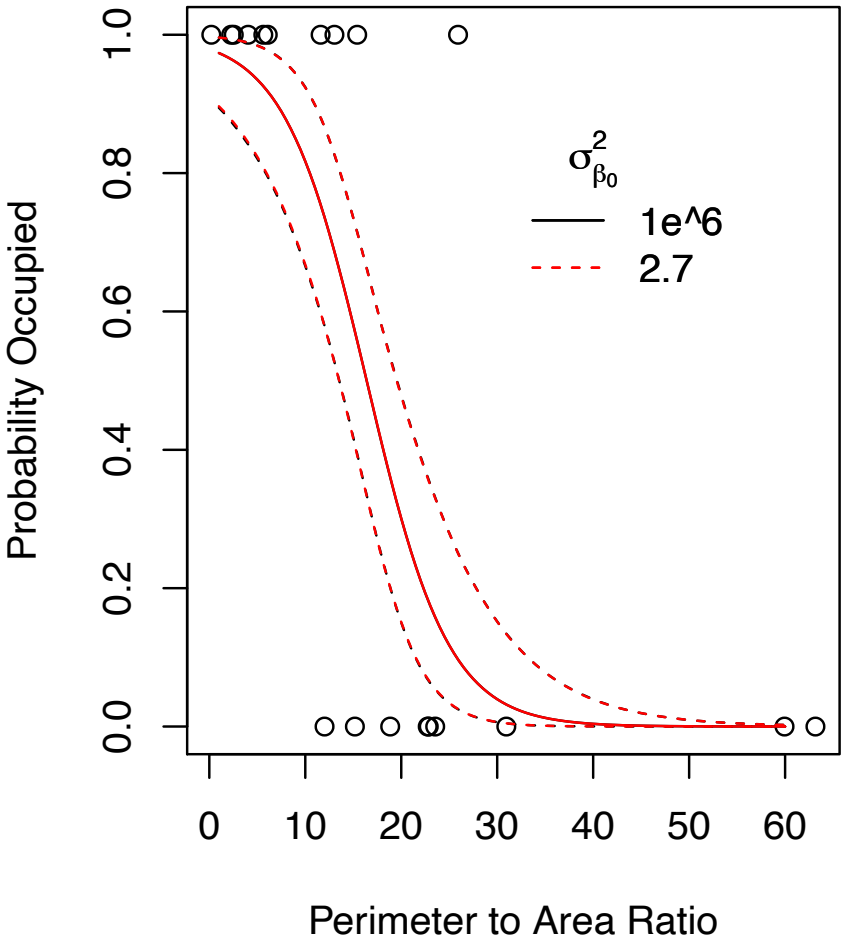
Default variance on β_0 prior

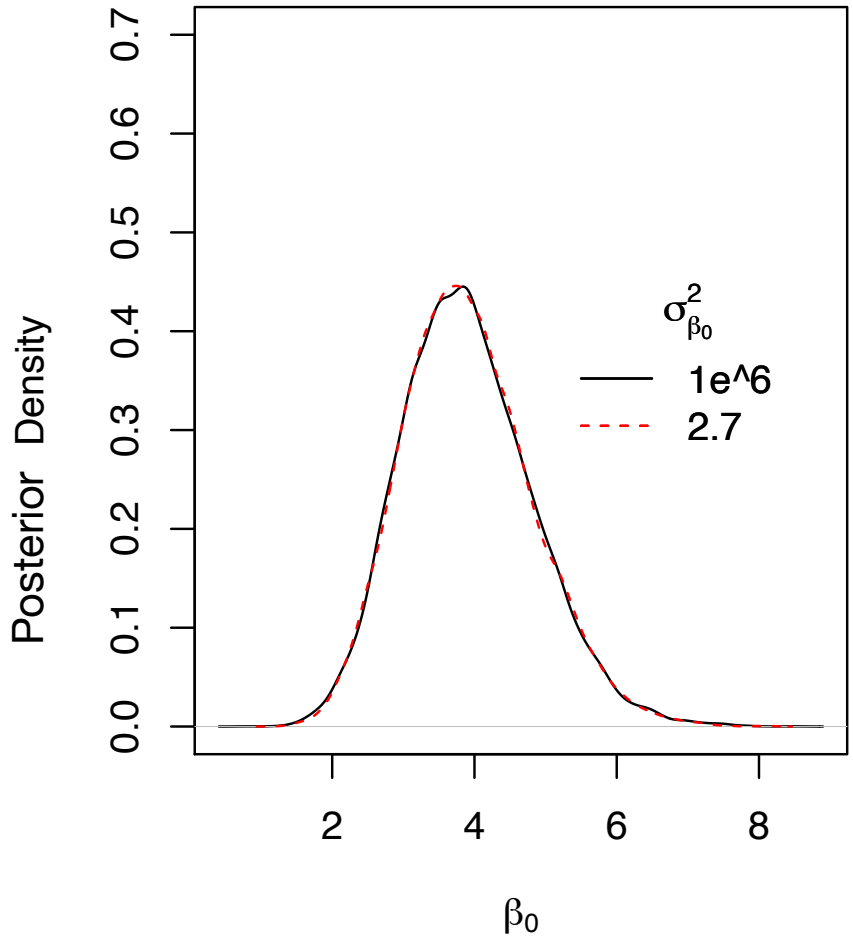




Posterior inference on β_0 impacted by prior!

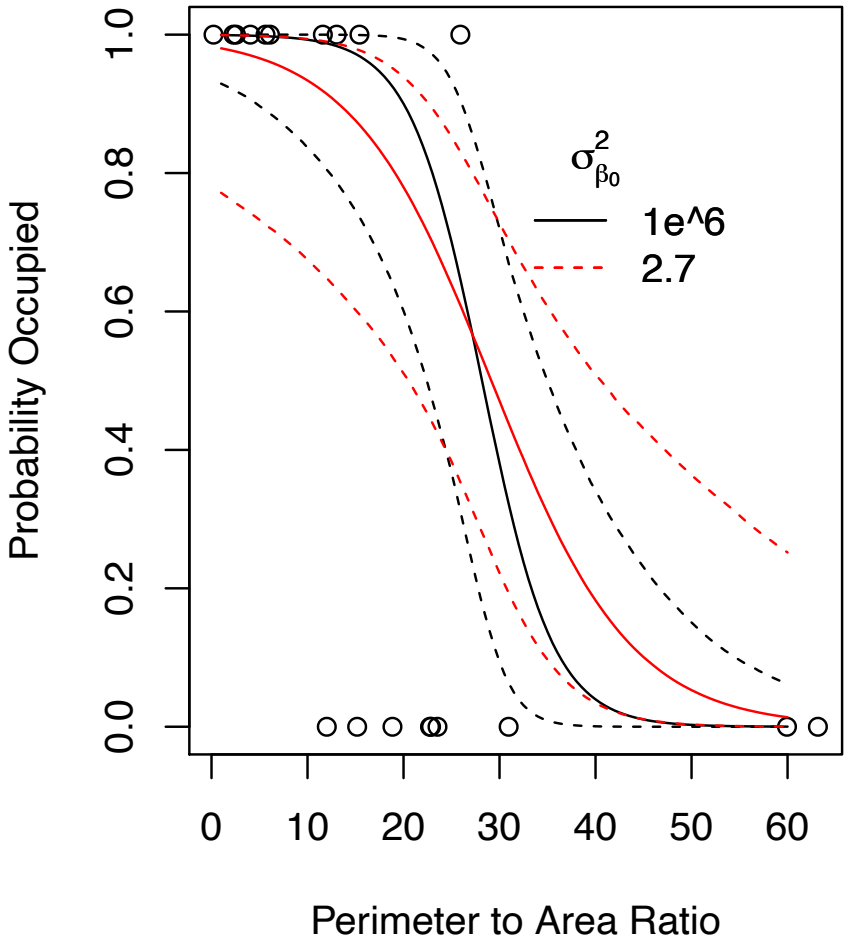
What if we had more data?

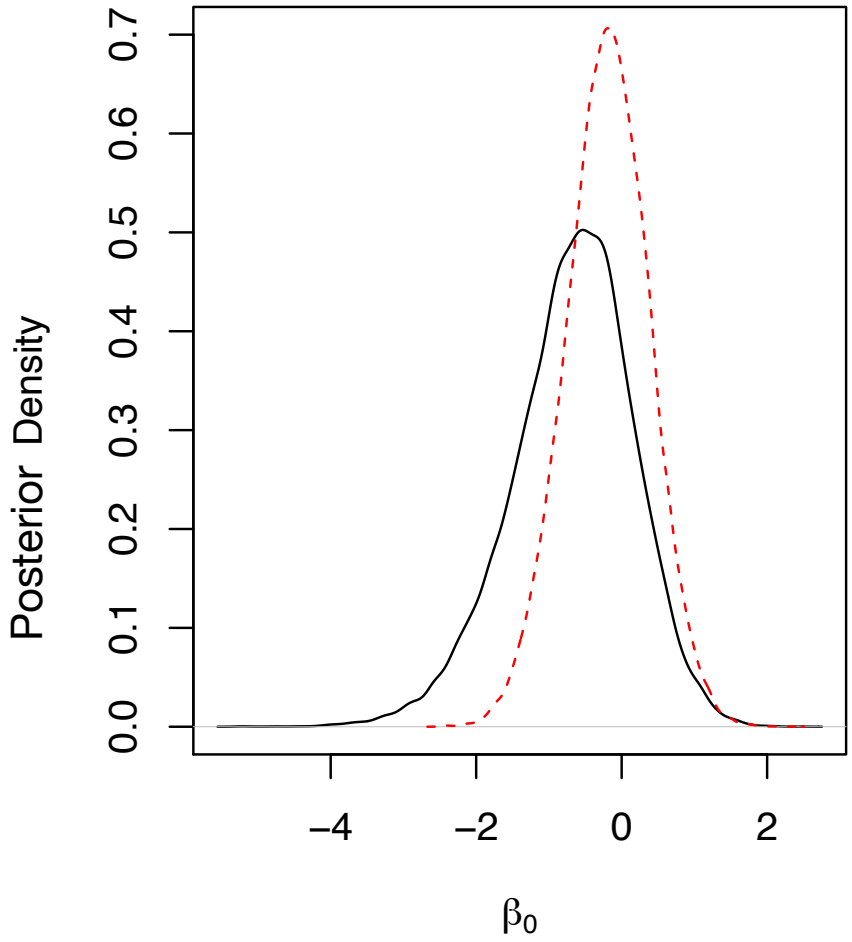




Now the data overwhelms the prior :) But we don't have this much data...

What if we standardize the data?





Posterior distributions look more similar but still notably different.

Slightly more informed priors with original data

Recall our original priors on the regression coefficients:

$$\beta_0 \sim \text{normal}(0, \sigma^2_{\beta_0})$$

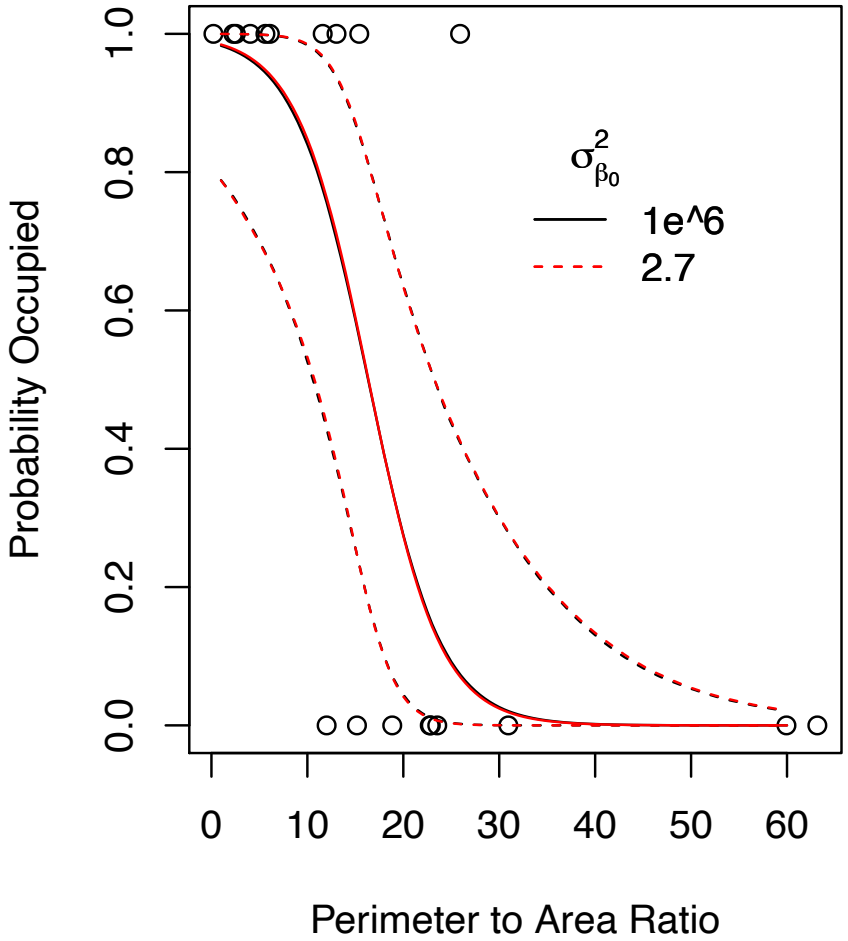
$$\beta_1 \sim \text{normal}(0, \sigma^2_{\beta_1})$$

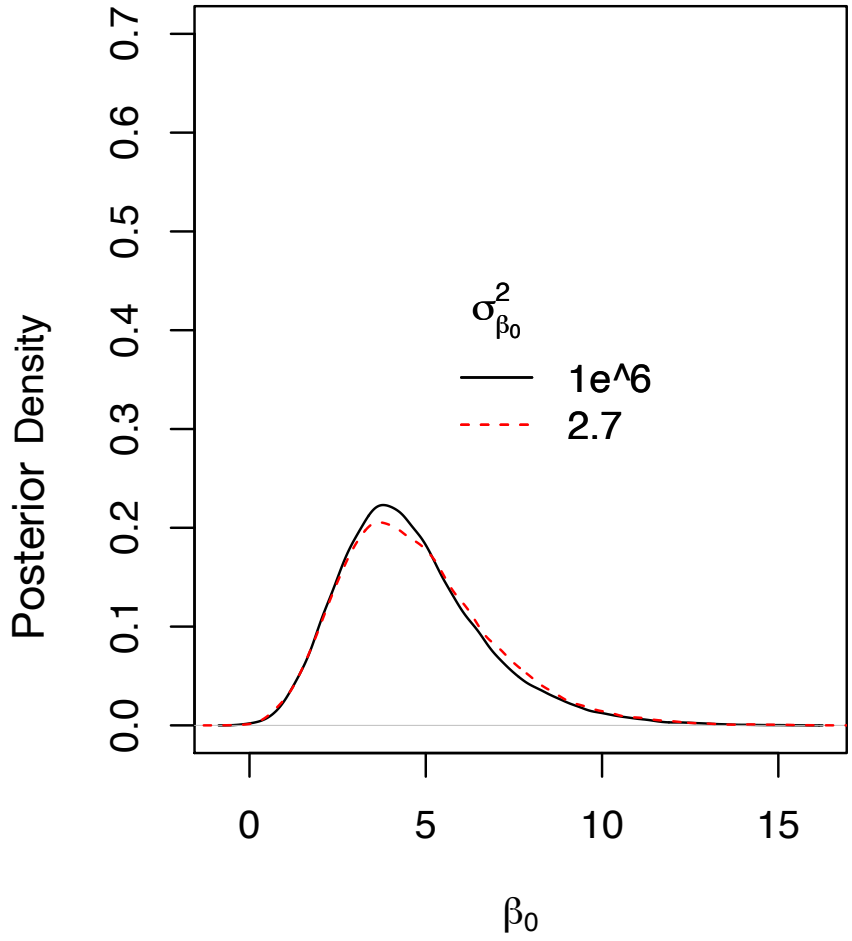
Now consider

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

- We center 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.

Weakly informative priors on parameters





Posterior distributions on β_0 align nicely!

Guidance

- Always use informed priors when you can.
- Always examine sensitivity of marginal posteriors to variation in priors for non-linear models.
- Vague priors for non-linear models should be centered on reasonable values.
- Consider standardizing data for non-linear models.
- Use Cauchy prior on group-level variances when only a few groups. See Gelman et al. 2008 for details.[1]
- Group level variances for fewer than four or five groups will often need sensibly informed half-Cauchy priors.

[1] 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.

References for this Lab

- 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.
- Banner, KM, Irvine, KM, Rodhouse, TJ. The use of Bayesian priors in Ecology: The good, the bad and the not great. Methods Ecol Evol. 2020; 11: 882– 889.