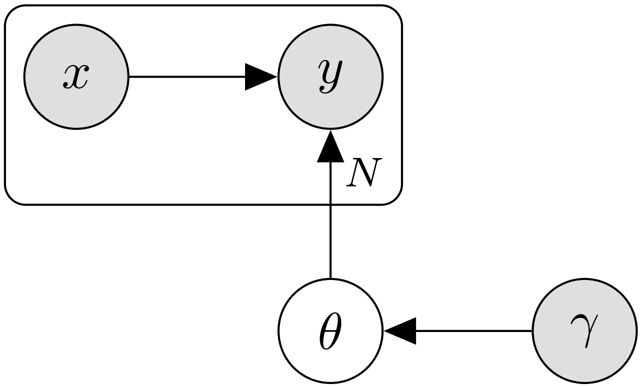
Theoretical notes on setting priors for Bayesian regression models

Setting priors for Bayesian regression models

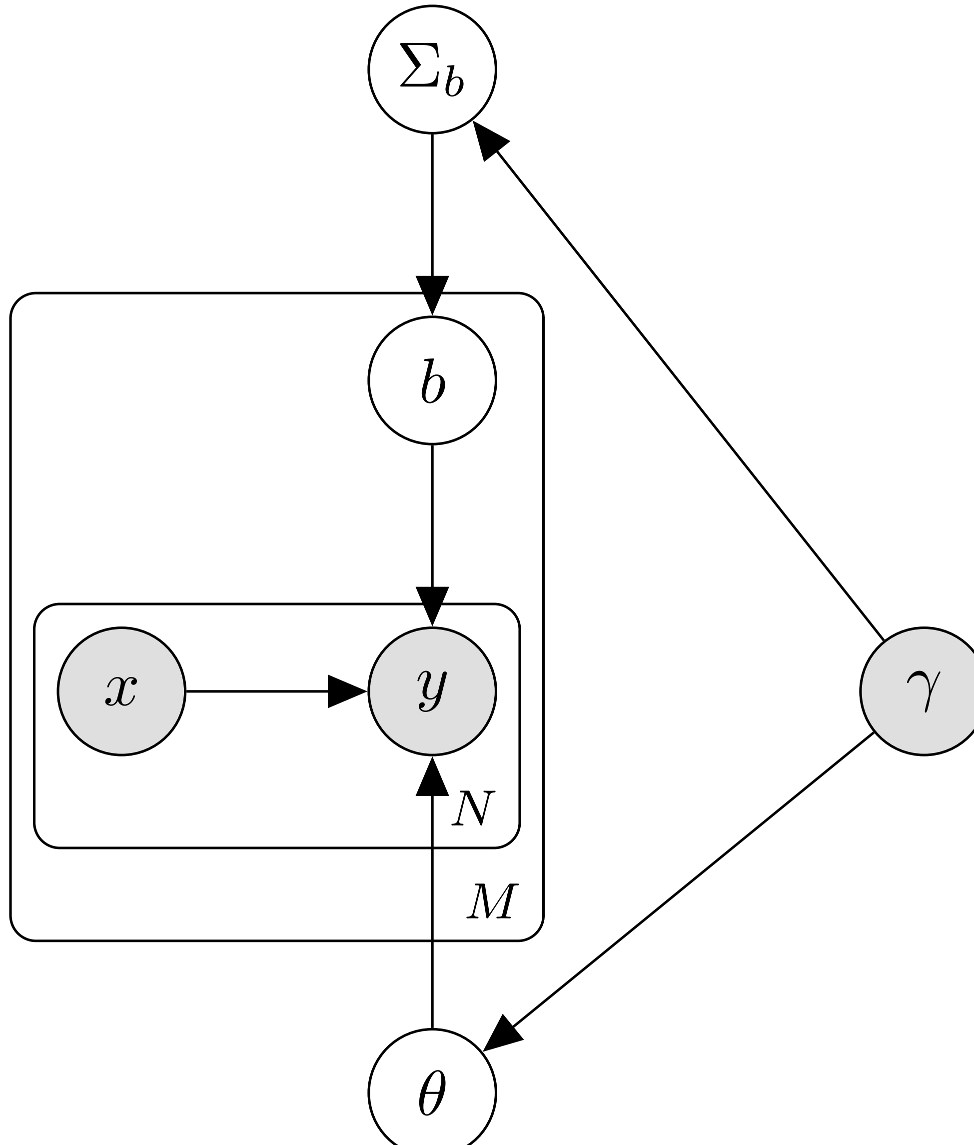
Single-level model



General principle: focus on using priors that do not unduly bias the posterior with respect to the scientific question

(in some sense, use "just enough of a prior to get the Bayesian crank turning")

Multi-level model



"Flat" priors versus "uninformative" priors

- Simple case: Bernoulli coin flip with parameter $p \in [0,1]$
- Innocent-seeming choice: use a flat, or uniform prior

Probability ***density*** *function*

$$P(p) = \begin{cases} 1 & p \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

- But...suppose I prefer to think about my Bernoulli coin flip in terms of **odds ratios** $r = \dfrac{p}{1-p}$
- What is the equivalent probability density on r?
- To figure this out, you need to know the **change of variables formula for probability density functions**

Change of variables for probability density

- Consider a continuous random variables X, and another continuous random variable $Y = g(X)$
- How do we write the probability density function $p_Y(Y)$ as a function of $p_X(X)$? It turns out that:

$$p_Y(Y) = p_X(g(X)) \left| \frac{dX}{dY} \right|$$

- The rightmost term is required for properness of $p_Y(Y)$

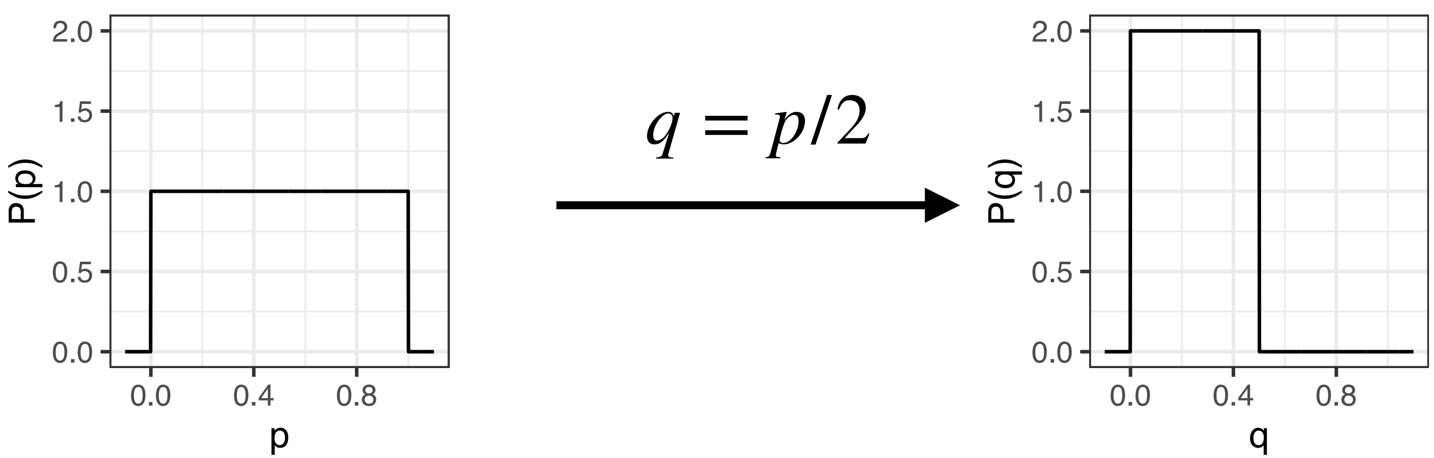
Transforming Bernoulli success parameter

- Simple example: let p be uniform on $[0,1]$ and $q = p/2$

$$P_q(q) = P_p(p)\left|\frac{dp}{dq}\right|$$

$$\left|\frac{dp}{dq}\right| = 2$$

- so $P_q(q) = 2P_p(p)$



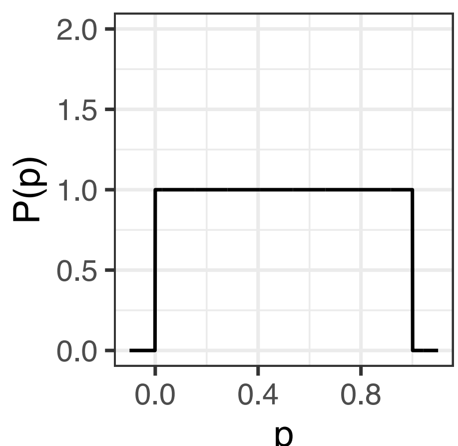
Transforming Bernoulli success parameter

- Now consider the odds transform $r = \dfrac{p}{1-p}$

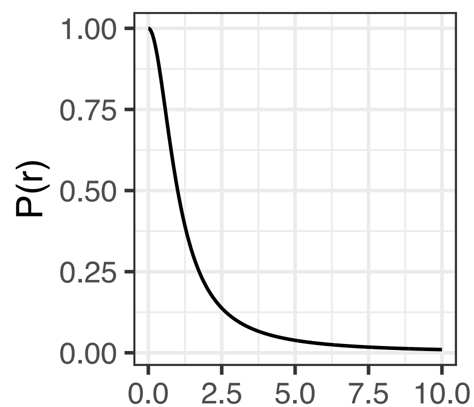
$$P_r(r) = P_p(p)\left|\frac{dp}{dr}\right|$$

$$\left|\frac{dp}{dr}\right| = \frac{1}{1+r^2}$$

- so $P_q(q) = P_p(p)/(1+r^2)$



$$r = p/(1-p) \longrightarrow$$

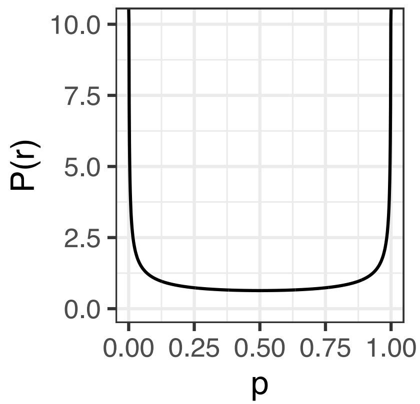


The Jeffreys prior as "uninformative"

- The **Jeffreys** prior is a principled choice of "uninformative" prior, in a technical sense: the mass assigned to any volume is parameterization-invariant
- We set $\pi_J(\theta) \propto \sqrt{I(\theta)}$ – $I(\theta)$ is the **Fisher information**:

$$I(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log p(x \mid \theta)\right)^2\right]$$

- Note that this **depends on the likelihood function!**
- It turns out that for Bernoulli likelihood this leads to a Beta(0.5,0.5) prior



Practical "uninformative" priors

- Unfortunately, the Jeffreys prior is not always **proper**
- For example, consider Gaussian data. The Jeffreys prior turns out to be:

$$P(\mu) \propto 1$$

$$P(\sigma) \propto \frac{1}{\sigma}$$

- These results are still very useful!
- But in practice, we need to **bound** the priors to practical ranges
- Additionally, the Jeffreys prior is not even always **analytically tractable**
- Therefore, we often must be more heuristic in prior choice