# ST 740: Noninformative Prior Distributions

Alyson Wilson

Department of Statistics
North Carolina State University

September 4, 2013

# Noninformative Prior Distributions

A non-informative prior distribution

- represents the idea that "nothing (or more realistically, very little) is known a priori,"
- has little impact on the posterior distribution, and
- lets the data "speak for themselves."

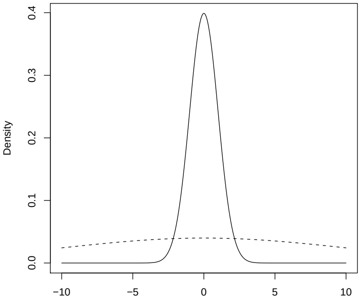
In certain situations, this may describe what you want to do.

# Important Concepts

- Locally uniform
  Prior distributions that satisfy the idea that the likelihood is "peaked" as compared to the prior distribution are called *vague*, *locally uniform*, *diffuse*, or *flat*.
- Noninformative
  *Noninformative* is often used to denote a prior derived using some mathematical criteria (to be discussed shortly). But the term is also commonly used to describe the class of "know nothing priors," as are *uninformative* or *objective*.
- Reference
  A *reference* prior has many definitions: sometimes *locally uniform* (Box and Tiao 1973), sometimes *noninformative*, sometimes "default" (Kass and Wasserman 1996).

# Locally Uniform Priors

From Box and Tiao (1973), "a prior which is dominated by the likelihood is one which does not change *very much* over the region in which the likelihood is appreciable and does not assume large values outside that range." They refer to a prior distribution which has these properties as a *locally uniform* prior. "For such a prior distribution we can approximate the result of Bayes' formula (theorem) by substituting a constant for the prior distribution."



# Flat Priors

What happens when the prior distribution is flat?

The posterior distribution looks like the normalized likelihood. Why?

$$\begin{aligned}
\text{posterior} &\propto \text{likelihood} \times \text{prior} \\
\text{posterior} &\approx \text{likelihood} \times \text{constant} \\
\text{posterior} &\approx \text{likelihood}
\end{aligned}$$

# Flat Priors

**Examples**

- Normal distribution with a large variance.
- Uniform distribution with a large range.

# Diffuse Prior for Binomial Sampling Distribution

$$\begin{aligned} Y \mid p &\sim \text{Binomial}(n, p) \\ p &\sim \text{Beta}(\alpha, \beta) \end{aligned}$$

Suppose that we choose a Beta$(1, 1)$ (equivalently, a Uniform$(0, 1)$) prior for $p$.

Choosing this prior is known as *Bayes' postulate*. It seems intuitively reasonable that we treat all values the same.

*The difficulty of applying Bayes' theorem is that the probabilities of the different causes are seldom known, in which case it may be postulated that they are all equal (sometimes known as postulating the equidistribution of ignorance).*

# Diffuse Prior for Binomial Sampling Distribution

This prior is conjugate, and our posterior is $\text{Beta}(1+y, 1+n-y)$. The posterior mode is

$$\begin{aligned}
\text{Mode}[p \mid y] &= \frac{\alpha - 1}{\alpha + \beta - 2} \\
&= \frac{1 + y - 1}{1 + y + 1 + n - y - 2} \\
&= \frac{y}{n}
\end{aligned}$$

# Why not use the uniform distribution as a reference prior?

- A uniform prior distribution may not be all that uninformative.
- A uniform prior is not invariant under transformation.

# Why not use the uniform distribution as a reference prior?

Consider:

- $\eta = \exp\{\theta\}$, so that $\theta = \log\{\eta\}$ is the inverse transformation.
- $\frac{d\theta}{d\eta} = \frac{1}{\eta}$ is the Jacobian
- Then, if $\pi(\theta)$ is prior for $\theta$, $\pi^*(\eta) = \eta^{-1}\pi(\log\eta)$ is corresponding prior for transformation.
- For $\pi(\theta) \propto c$, $\pi^*(\eta) \propto \eta^{-1}$.

It seems inconsistent that simply transforming the parameter somehow changes our "lack of knowledge."

# Jeffreys' Prior

The prior distribution for a single parameter $\theta$ is approximately noninformative if it is taken proportional to the square root of Fisher's information measure.

$$\begin{aligned} \pi(\theta) &\propto I(\theta)^{1/2} \\ I(\theta) &= -E\left[\frac{\partial^2 \log(f(y \mid \theta))}{\partial\theta^2}\right] \end{aligned}$$

This choice of prior is invariant under one-to-one parameter transformations.

# Jeffreys Prior for Binomial

$$Y \sim \text{Binomial}(n, p)$$

$$f(y \mid p) = \binom{n}{y} p^y (1-p)^{n-y}$$

$$\log(f(y \mid p)) = y \log(p) + (n-y)\log(1-p) + \text{constant}$$

$$\frac{\partial^2 \log(f(y \mid p))}{\partial p^2} = -\frac{y}{p^2} - \frac{n-y}{(1-p)^2}$$

# Jeffreys' Prior for Binomial

Taking expectations,

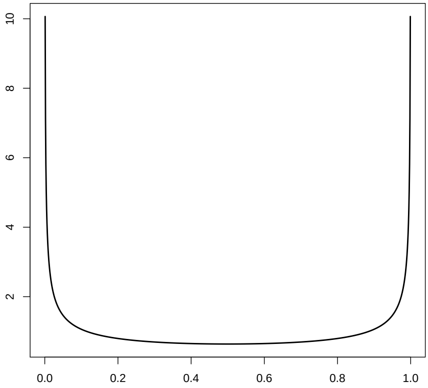
$$
\begin{aligned}
I(p) &= \frac{E[y]}{p^2} + \frac{n - E[y]}{(1-p)^2} \\
&= \frac{np}{p^2} + \frac{n - np}{(1-p)^2} \\
&= \frac{n}{p} + \frac{n}{(1-p)^2} \\
&= \frac{n}{p(1-p)}
\end{aligned}
$$

This implies that our prior should be

$$\pi(p) \propto p^{-\frac{1}{2}}(1-p)^{-\frac{1}{2}}$$

which is a Beta($\frac{1}{2}, \frac{1}{2}$).

# Jeffreys' Prior for Binomial




$$\text{Beta}\left(\frac{1}{2}, \frac{1}{2}\right)$$

# Invariance under Reparameterization

Intuition: Suppose that we have a parameter $\theta$ and a one-to-one transformation of the parameter $\eta = h(\theta)$. There are two ways we can think about determining a prior distribution for $\eta$.

1. Use a rule to determine a prior distribution $\pi(\theta)$ then use the *change of variables* technique to determine the distribution $\pi^*(\eta)$.
2. Use the same rule directly to determine the prior distribution for $\eta$.

We'd like both of these approaches to lead to the same distribution for $\eta$.

# Prior for Normal Mean

If $y \sim N(\theta, \sigma^2)$, the conjugate prior distribution for $\theta$ is $N(\mu_1, v_1^2)$, and posterior is $N(\mu_2, v_2^2)$, where

$$\begin{aligned}
\mu_2 &= \frac{\mu_1\sigma^2 + n\bar{y}v_1^2}{\sigma^2 + nv_1^2} \\
&= \left(\frac{nv_1^2}{\sigma^2 + nv_1^2}\right)(\bar{y}) + \left(\frac{\sigma^2}{\sigma^2 + nv_1^2}\right)(\mu_1) \\
v_2^2 &= \frac{\sigma^2 v_1^2}{\sigma^2 + nv_1^2} \\
&= \frac{1}{\frac{1}{v_1^2} + \frac{n}{\sigma^2}}
\end{aligned}$$

As $v_1^2 \to \infty$, $\mu_2 \to \bar{y}$ and $v_2^2 \to \frac{\sigma^2}{n}$.

We get this same limiting result using $\pi(\theta) \propto 1$.

# Jeffreys Prior for Normal Mean

Assume $\sigma^2$ is known.

$$Y \sim \text{Normal}(\mu, \sigma^2)$$

$$f(y \mid \mu) = \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp(-\frac{n}{2\sigma^2}(\bar{y} - \mu)^2)$$

$$\log(f(y \mid \mu)) = -\frac{n}{2\sigma^2}(\bar{y} - \mu)^2 + \text{constant}$$

$$\frac{\partial^2 \log(f(y \mid \mu))}{\partial \mu^2} = -\frac{n}{\sigma^2}$$

# Jeffreys Prior for Normal Mean

This expression does not depend on $y$, so $I(\mu) = \frac{n}{\sigma^2} = \text{constant}$.

This implies that our prior should be $\pi(\mu) \propto 1$ or uniform on the real line.

# Improper Prior

Notice that the uniform prior on the real line is *improper*:

$$\pi(\mu)d\mu = \int d\mu = \infty$$

yet in the normal mean case leads to proper posterior for $\mu$. This is not always true. If you use an improper prior, it is important to check to be sure your posterior distribution is *proper*, which means that the posterior integrates to 1.