ST 740: Noninformative Prior Distributions

Alyson Wilson

Department of Statistics
North Carolina State University

September 11, 2013

Other Choices of Prior for the Binomial Sampling Distribution

Haldane proposed using a $\text{Beta}(0, 0)$ prior distribution.

- This is an improper prior.
- It is equivalent to choosing a uniform distribution on $\log(\frac{p}{1-p})$.
- The posterior is $\text{Beta}(y, n - y)$, which has a posterior mean of $\frac{y}{n}$.
- When is the posterior improper?
- For a $\text{Beta}(\alpha, \beta)$ prior distribution, α is often interpreted as the "prior number of successes," β as "the prior number of failures," and $\alpha + \beta$ as the "prior sample size." (Why?) The Haldane prior obviously has the smallest possible number of "prior successes and failures."

Jeffreys' Prior

The prior distribution for a single parameter θ 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.

Invariance under Reparameterization

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

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. Reparameterize first, and then use the same rule directly to determine the prior distribution for η.

We'd like both of these approaches to lead to the same distribution for η.

Jeffreys' Priors

Example

Start with a binomial sampling distribution and let $\theta = \log(\frac{p}{1-p})$.

$$f(y \mid \theta) = \binom{n}{y} \left(\frac{\exp(\theta)}{1+\exp(\theta)} \right)^{y} \left(\frac{1}{1+\exp(\theta)} \right)^{n-y}$$

We want to use a Jeffreys' (noninformative) prior for θ.

$$\begin{aligned} \log(f(y \mid \theta)) &= \log(\binom{n}{y}) + y \log(\frac{\exp(\theta)}{1+\exp(\theta)}) + \\ &\quad (n-y) \log(\frac{1}{1+\exp(\theta)}) \\ &= \log(\binom{n}{y}) + y\theta - n \log(1+\exp(\theta)) \end{aligned}$$

Jeffreys' Priors

Example

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

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

so we take

$$\begin{aligned} \pi^*(\theta) &\propto \sqrt{\frac{\exp(\theta)}{(1+\exp(\theta))^2}} \\ &\propto \frac{\exp(0.5\theta)}{1+\exp(\theta)} \end{aligned}$$

Jeffreys' Prior

Invariance

Now suppose that we change variables in the *Jeffreys' prior* that we derived for p.

$p = \frac{\exp(\theta)}{1+\exp(\theta)}$, so $\left|\frac{dp}{d\theta}\right| = \frac{\exp(\theta)}{(1+\exp(\theta))^2}$.

Changing variables in the Jeffrey's prior for p gives

$$\begin{aligned}
\pi(\theta) &= \pi_p\left(\frac{\exp(\theta)}{1+\exp(\theta)}\right)\left|\frac{dp}{d\theta}\right| \\
&\propto \left(\frac{\exp(\theta)}{1+\exp(\theta)}\right)^{-0.5}\left(\frac{1}{1+\exp(\theta)}\right)^{-0.5}\frac{\exp(\theta)}{(1+\exp(\theta))^2} \\
&\propto \frac{\exp(0.5\theta)}{1+\exp(\theta)}
\end{aligned}$$

More on Fisher's Information

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

Since the log-likelihood $\log(L(\theta \mid y))$ differs from $\log(f(y \mid \theta))$ only by a constant, all of their derivatives are equal. Thus information can be equivalently defined as

$$
I(\theta) = -E\left[\frac{\partial^2 \log(L(\theta \mid y))}{\partial\theta^2}\right]
$$

Also note that if we have n observations,

$$
I(\theta \mid y_1, \ldots, y_n) = nI(\theta \mid y)
$$

More on Fisher's Information

What happens if we start with $\log(L(\theta \mid y))$ and reparameterize with $\eta = g(\theta)$? By the chain rule,

$$\frac{\partial \log(L(\eta \mid y))}{\partial \eta} = \frac{\partial \log(L(\theta \mid y))}{\partial \theta} \frac{\partial \theta}{\partial \eta}$$

If we square and take expectations over y (noticing that $\frac{\partial \theta}{\partial \eta}$ does not depend on y), we have

$$I(\eta \mid y) = I(\theta \mid y) \left(\frac{\partial \theta}{\partial \eta} \right)^2$$

Review of Change of Variables

$$\begin{aligned} y &= g(x), \text{ where } g(x) \text{ is a one-to-one transformation} \\ x &= g^{-1}(y) \end{aligned}$$

If $p_X(x)$ is the probability density function for X, we have

$$p_Y(y) = p_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|$$

Jeffreys' Prior

If we take square roots of

$$I(\eta \,|\, y) = I(\theta \,|\, y) \left(\frac{\partial \theta}{\partial \eta} \right)^2$$

we have

$$\sqrt{I(\eta \,|\, y)} = \sqrt{I(\theta \,|\, y)} \left| \frac{\partial \theta}{\partial \eta} \right|$$

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. Reparameterize first, and then use the same rule directly to determine the prior distribution for η.