The Jeffreys Prior

Paul Hewson

December 10, 2009

1 Choosing a prior

There are many things we might consider when choosing a prior. Some of our choices include:

- Conjugate prior: mathematical convenience (using a conjugate prior means we often have analytical expressions for the posterior)
- Informative priors: we have reason for strong prior belief as to the distribution of the parameter of interest and wish to capture this in the analysis.
- Non-informative priors: we have no prior belief and do not wish to influence the analysis with the prior

The trouble with non-informative priors, is that a statement that is uniformative on one scale (e.g. the uniform prior for a proportion) might take a very different shape when we transform. The logit function $\psi = \log\left(\frac{\pi}{1-\pi}\right)$, for example, is trying to transform a probability statement defined on $0 < \pi < 1$ to one defined on $-\infty < \psi < \infty$. The idea of the Jeffreys prior is to make sure when we do this, the statement of prior belief is the same on either scale.

2 Observed information and Fisher Information

We have met the observed information as:

$$I_O(\theta) = -\frac{\partial^2 \log L(\theta)}{\partial \theta^2}$$

The Fisher information can be defined as the Expectation of this, with respect to θ:

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

2.1 Example

If we consider the Binomial distribution, we have:

$$\log L(\pi) = \log \binom{n}{x} x \log(\pi) + (n-x)\log(1-\pi)$$

So the Score equation is given by:

$$\frac{\partial \log L(\pi)}{\partial \pi} = \frac{x}{\pi} - \frac{n-x}{1-\pi}$$

And so the observed information is given by

$$I(\theta) = -\frac{\partial^2 \log L(\theta)}{\partial \pi^2} = -\left(-\frac{x}{\pi^2} - \frac{n-x}{(1-\pi)^2}\right)$$

We want to find the expectation with respect to π. As we know that, for the Binomial distribution, $E[X] = n\pi$, we can substitute for x

$$\begin{aligned} I(\theta) = -E\left[\frac{\partial^2 \log L(\theta)}{\partial \pi^2}\right] &= -E\left(-\frac{x}{\pi^2} - \frac{n-x}{(1-\pi)^2}\right) \\ &= \frac{n\pi}{\pi^2} - \frac{n-n\pi}{(1-\pi)^2} \\ &= n\left(\frac{1}{\pi} - \frac{1}{1-\pi}\right) \\ &= \frac{n}{\pi(1-\pi)} \end{aligned}$$

We should note that defining the Fisher Information as the $(-1\times)$ expectation of the second differential only applies under the regularity condition $\int \frac{\partial^2 \log L(\theta)}{\partial \theta^2} dx = 0$. Otherwise, we use $E[(\frac{\partial \log L\theta}{\partial \theta})^2]$ to define the Fisher Information.

3 Jeffreys Prior

As stated above, the rationale for using the Jeffreys prior is that it is invariant to a reparameterisation.

Using the change of variables theorem, and knowledge of the Fisher Information we have (for a single parameter) that:

$$
\begin{aligned}
p(\psi) = p(\pi)|\frac{\partial \pi}{\partial \psi}| &\propto \sqrt{E\left[\left(\frac{\partial \log L}{\partial \theta}\right)^2\right]\left(\frac{\partial \theta}{\partial \psi}\right)^2} \\
&= \sqrt{E\left[\left(\frac{\partial \log L}{\partial \theta}\frac{\partial \theta}{\partial \psi}\right)^2\right]} \\
&= \sqrt{E\left[\left(\frac{\partial \log L}{\partial \psi}\right)^2\right]} \\
&= \sqrt{I(\psi)}
\end{aligned}
$$

If you look at $\pi^{a-1}(1-\pi)^{b-1}$ and convince yourself that when $a = \frac{1}{2}$ and $b = \frac{1}{2}$ we have $\sqrt{\frac{n}{\pi(1-\pi)}} \propto \pi^{a-1}(1-\pi)^{b-1}$ (I think we might have to regard n as another of these disposable constants to do this).