Bayesian models 1 - Prior distribution

By Matthew Stephens (edited by Heejung Shim)

Priors

Formally there are two different ways to think about how to specify the prior distribution, which we might call “subjective prior specification”, and “objective prior specification”.

Subjective specification: the prior is chosen by making a serious attempt to quantify a state of knowledge prior to performing the experiment. This approach is most often used when the prior actually contains non-trivial information that helps inform the analysis.

Priors

Objective specification: the prior is chosen to satisfy some objective criterion. Examples of criterion that are sometimes used are:

1. the conclusions of the analysis (i.e. posterior distributions of quantities of interest) should be invariant to scale of measurement (e.g. multiplying all data by a constant as would happen if we switched from inches to centimeters).
2. more generally, the conclusions of the analysis should be invariant to choice of parameterization. (Jeffrey's priors are the most common example.)

Priors

In general we should think of the choice of prior as part of the model building process, just as the choice of likelihood for the observed data requires a choice.

Influential priors are not necesarily bad, but we need to pay attention to the magnitude of their influence.

In practice, the main challenge in prior specification is avoiding specifying a prior that is both i) unrealistic and ii) has a severe effect on the inference.

Priors

Note also there are many other terms people use for priors, some of which may be useful, others may not.

non-informative priors: this isn't a useful term. No prior is "non-informative". It isn't clear what it means.

flat priors: this usually means $p(\theta)$ is a constant, but you might as well say that explicitly.

diffuse priors: this is a qualitative description, usually used to indicate that no serious attempt was made to quantify subjective information, and usually implying that the author believes the results will be relatively insensitive to this choice.

Priors

improper prior: this is used to refer to a function $p(\theta)$ that does not integrate to a finite quantity, but which is used in place of the prior distribution in Bayes Theorem, to compute the posterior by $p(\theta|D) \propto p(D|\theta)p(\theta)$.

The idea is that even if $p(\theta)$ does not integrate to a finite quantity, if $\int p(D|\theta)p(\theta)d\theta$ is finite then the above produces a "proper" posterior.

1. For example, suppose $x|\theta \sim N(\theta, 1)$. Assume an "improper prior" on θ with $p(\theta) \propto 1$. This is improper because $\int p(\theta) d\theta$ is infinite, and so $p(\theta)$ is not proportional to a density. However, if we just "plug in" this prior to Bayes Theorem we get $p(\theta|x) \propto \exp(-(\theta - x)^2/2)$, which implies that $\theta|x \sim N(x, 1)$. That is we get a "proper" posterior.
2. Improper priors can never be "subjective" priors. This is because they are not a distribution and so cannot be interpreted as using probability to represent uncertainty.
3. Improper priors often result from attempts to define Objective priors.

Jeffreys priors

Jeffreys wanted to come up with a "rule" for obtaining an objective prior that had the following property: if you applied the rule to θ you would get the same prior for any (monotone differentiable) 1-1 transformation of θ, $h(\theta)$ say, as if you applied the rule to $\phi = h(\theta)$. In this sense the prior (or the rule) is "transformation invariant".

Jeffreys priors

Definition: Recall that for a model with parameter space $\Theta \subseteq \mathbf{R}$, the Fisher information is

$$I(\theta) = E\left(\frac{d\log(f(x \mid \theta))}{d\theta}\right)^2$$

where $f(x \mid \theta)$ is the sampling distribution and the expectation is taken over $f(x \mid \theta)$. Under regularity conditions,

$$I(\theta) = -E\left(\frac{d^2\log(f(x \mid \theta))}{d\theta^2}\right).$$

In such a setting, the *Jeffreys Prior* for θ is defined by $\pi_\theta(\theta) \propto I(\theta)^{1/2}$, to be proportional to the square root of the Fisher Information at θ.

Jeffreys priors

Note that in general the Jeffreys prior may be improper.

Note that by the chain rule,

$$I(\theta) = I(h(\theta))\Big(\frac{dh}{d\theta}\Big)^2.$$

If θ has the Jeffreys prior and h is a monotone differentiable function of θ, the prior induced on $\phi = h(\theta)$ by the Jeffreys prior on θ is

$$\pi_\phi(\phi) = \pi_\theta(\theta)\Big|\frac{d\theta}{d\phi}\Big| \propto I(\theta)^{1/2}\Big|\frac{dh}{d\theta}\Big|^{-1} = I(h(\theta))^{1/2}.$$

Thus the Jeffreys priors are invariant under reparameterization.

Jeffreys priors

Example:
Suppose $x \sim \text{Binomial}(n, p)$. Then:

$$f(x \mid \theta) = \binom{n}{x} p^x (1-p)^{n-x}.$$

What is the Jeffreys prior for p?

Jeffreys priors

Solution:

$$\frac{d^2 \log(f(x \mid p))}{dp^2} = -\frac{x}{p^2} - \frac{(n-x)}{(1-p)^2}.$$

Thus

$$\begin{aligned} I(p) &= E\left(\frac{x}{p^2} + \frac{n-x}{(1-p)^2}\right) \\ &= \left(\frac{n}{p} + \frac{n}{1-p}\right) = \frac{n}{p(1-p)}. \end{aligned}$$

Thus the Jeffreys prior for p is

$$\pi(p) \propto [p(1-p)]^{-1/2},$$

which is the Beta$(\frac{1}{2}, \frac{1}{2})$ density (and hence proper).