STAT 431 — Applied Bayesian Analysis — Course Notes

More About Priors

Spring 2019

Sensitivity Analysis

How do the results (estimates, posterior probabilities, intervals) vary depending on what prior is used?

Practical Approach:

Choose several reasonable priors (including noninformative and subjective), compute results for each, and assess them for agreement.

R Example 5.1:

Population Proportion — Sensitivity Analysis

Improper Priors

An **improper density** $p(x)$ satisfies

- $p(x) \geq 0$ for all x
- $\sum_x p(x) \; = \; \infty$ (discrete) or
  $\int p(x)\, dx \; = \; \infty$ (continuous)
  (i.e. it cannot be *normalized* to a **proper** density)

When an improper density is used as a ("noninformative") prior density, it is called an **improper prior**.

Eg: Population proportion π (binomial model)

$$p(\pi) \;=\; \frac{1}{\pi(1-\pi)} \qquad 0 < \pi < 1$$

(like a beta density with "$\alpha = 0$" and "$\beta = 0$")

Improper because

$$\int_0^1 p(\pi)\, d\pi \;=\; \infty$$

[Draw density ...]

Warning: Improper priors may lead to improper posteriors!

Eg: (continued)

If

$$p(\pi) \;=\; \frac{1}{\pi(1-\pi)} \qquad 0 < \pi < 1$$

and either $y = 0$ or $y = n$, then can show that $p(\pi \mid y)$ is improper!

Alternative: Use “vague” priors — proper, but close to improper.

Eg: Use $\mathrm{beta}(\alpha, \beta)$ with α and β “near” zero.

Jeffreys Priors

Consider data $\boldsymbol{y}$, parameter θ, and model defined by density

$$p(\boldsymbol{y} \mid \theta)$$

assumed to be continuously differentiable in θ for all $\boldsymbol{y}$.

The **Fisher information** is

$$I(\theta) \;\; = \;\; -\,\mathrm{E}\left(\frac{\partial^2}{\partial\theta^2} \ln p(\boldsymbol{Y} \mid \theta) \;\middle|\; \theta \right)$$

(Note: $p(\boldsymbol{Y} \mid \theta)$ is the *random* likelihood $L(\theta; \boldsymbol{Y})$)

Eg: Binomial model ($\theta = \pi \in (0, 1)$)

$$p(y \mid \pi) \;\; = \;\; \binom{n}{y} \pi^y (1-\pi)^{n-y} \qquad\qquad y = 0, 1, \ldots n$$

$$\ln p(y \mid \pi) \;\; = \;\; \ln \binom{n}{y} \; + \; y \ln \pi \; + \; (n-y) \ln(1-\pi)$$

$$\frac{\partial^2}{\partial \pi^2} \ln p(y \mid \pi) \;\; = \;\; -\frac{y}{\pi^2} \; - \; \frac{n-y}{(1-\pi)^2}$$

$$\mathrm{E}\left(\frac{\partial^2}{\partial \pi^2} \ln p(Y \mid \pi) \;\middle|\; \pi \right) \;=\; \mathrm{E}\left(-\frac{Y}{\pi^2} \;-\; \frac{n-Y}{(1-\pi)^2} \;\middle|\; \pi \right)$$

$$
\begin{aligned}
\mathrm{E}\left(\frac{\partial^2}{\partial \pi^2} \ln p(Y \mid \pi) \;\middle|\; \pi \right) &= \mathrm{E}\left(-\frac{Y}{\pi^2} - \frac{n-Y}{(1-\pi)^2} \;\middle|\; \pi \right) \\
&= -\frac{\mathrm{E}(Y \mid \pi)}{\pi^2} - \frac{n - \mathrm{E}(Y \mid \pi)}{(1-\pi)^2}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}\left(\frac{\partial^2}{\partial\pi^2}\ln p(Y \mid \pi) \;\Bigm|\; \pi\right) &= \mathrm{E}\left(-\frac{Y}{\pi^2} - \frac{n-Y}{(1-\pi)^2} \;\Bigm|\; \pi\right) \\
&= -\frac{\mathrm{E}(Y \mid \pi)}{\pi^2} - \frac{n - \mathrm{E}(Y \mid \pi)}{(1-\pi)^2} \\
&= -\frac{n\pi}{\pi^2} - \frac{n - n\pi}{(1-\pi)^2}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}\left(\frac{\partial^2}{\partial \pi^2} \ln p(Y \mid \pi) \;\middle|\; \pi\right) &= \mathrm{E}\left(-\frac{Y}{\pi^2} - \frac{n-Y}{(1-\pi)^2} \;\middle|\; \pi\right) \\
&= -\frac{\mathrm{E}(Y \mid \pi)}{\pi^2} - \frac{n - \mathrm{E}(Y \mid \pi)}{(1-\pi)^2} \\
&= -\frac{n\pi}{\pi^2} - \frac{n - n\pi}{(1-\pi)^2} \\
&= -\frac{n}{\pi} - \frac{n}{1-\pi} = -\frac{n}{\pi(1-\pi)}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}\left(\frac{\partial^2}{\partial \pi^2} \ln p(Y \mid \pi) \;\middle|\; \pi \right) &= \mathrm{E}\left(-\frac{Y}{\pi^2} - \frac{n-Y}{(1-\pi)^2} \;\middle|\; \pi \right) \\
&= -\frac{\mathrm{E}(Y \mid \pi)}{\pi^2} - \frac{n - \mathrm{E}(Y \mid \pi)}{(1-\pi)^2} \\
&= -\frac{n\pi}{\pi^2} - \frac{n - n\pi}{(1-\pi)^2} \\
&= -\frac{n}{\pi} - \frac{n}{1-\pi} \;=\; -\frac{n}{\pi(1-\pi)}
\end{aligned}
$$

So

$$
I(\pi) = \frac{n}{\pi(1-\pi)}
$$

The **Jeffreys prior** density for a model with parameter θ is

$$p(\theta) \propto \sqrt{I(\theta)}$$

Note:

- Requires Fisher information to exist for all θ.
- May be improper.

Often regarded as “noninformative.”

Eg: (binomial continued)

$$p(\pi) \quad \propto \quad \sqrt{\frac{n}{\pi(1-\pi)}} \quad \propto \quad \pi^{-1/2}(1-\pi)^{-1/2}$$

[Draw ...]

Recognize as kernel of a $\mathrm{beta}(1/2, 1/2)$.
So the Jeffreys prior for the binomial model is $\mathrm{beta}(1/2, 1/2)$.

Important Property:

A Jeffreys prior is **invariant to reparameterization**:

If $\phi = g(\theta)$ is a (smooth) **reparameterization**, then the Jeffreys priors for ϕ and θ give equivalent posterior distributions (under transformation of variables).

(See Cowles, Sec. 5.3.3 for proof.)

Eg: A common reparameterization of the binomial model is

$$\phi \;=\; \mathrm{logit}(\pi) \;=\; \ln\left(\frac{\pi}{1-\pi}\right) \qquad (0 < \pi < 1)$$

For multiple parameters, there is a generalization of Jeffreys priors.

However, they are not always recommended in the multi-parameter case.