Module 5: Objective Bayes

Rebecca C. Steorts

Agenda

- Subjective versus objective priors
- The uniform prior
- Transforming the uniform prior
- Invariance
- Jeffreys' prior
- Binomial example for Jeffreys' prior

Subjective versus Objective priors

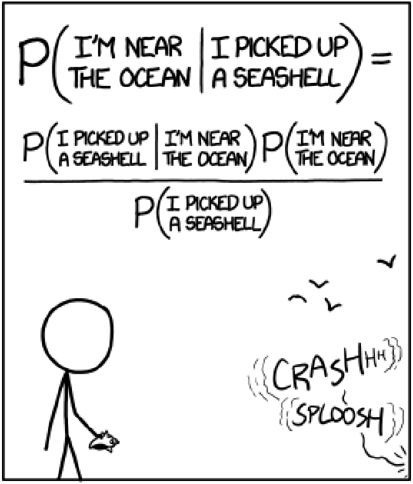
- Ideally, we would like a *subjective prior*: a prior reflecting our beliefs about the unknown parameter of interest.
- What are some examples in practice when we have subjective information?
- When may we not have subjective information?

Subjective versus Objective priors

In dealing with real-life problems you may run into problems such as

- not having past historical data,
- not having an expert opinion to base your prior knowledge on (perhaps your research is cutting-edge and new), or
- as your model becomes more complicated, it becomes hard to know what priors to put on each unknown parameter.
- What do we do in such situations?

That rule Bayes



STATISTICALLY SPEAKING, IF YOU PICK UP A SEASHELL AND *DON'T* HOLD IT TO YOUR EAR, YOU CAN PROBABLY HEAR THE OCEAN.

What did Bayes say exactly?

PROBLEM.

Given the number of times in which an unknown event has happened and failed: *Required* the chance that the probability of its happening in a ſingle trial lies ſomewhere between any two degrees of probability that can be named.

Translation

- Billiard ball W rolled on a line of length one, with a uniform probability of stopping anywhere.
- W stops at p
- Second ball O then rolled n times under the same assumptions.
- X denotes the number of times the ball O stopped on the left of W

Derive the posterior distribution of p given X, when $p \sim U[0, 1]$ and $X \mid p \sim \text{Binomial}(n, p)$

Such priors on p are said to be uniform or flat.

A “flat" prior

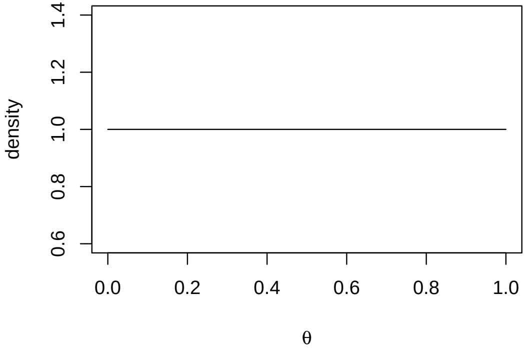
Let’s talk about what people really mean when they use the term “flat," since it can have different meanings.

Often statisticians will refer to a prior as being flat, when a plot of its density actually looks flat, i.e., uniform.

$$\theta \sim \text{Unif}(0, 1).$$

Why do we call it flat? It’s assigning equal weight to each parameter value. Does it always do this?

Uniform(0,1) prior



Uniform(0,1) prior (continued)

What happens if we consider though the transformation to $\phi = 1/\theta$.

Is our prior still flat (does it place equal weight at every parameter value)?

Uniform prior versus the Transformed Prior

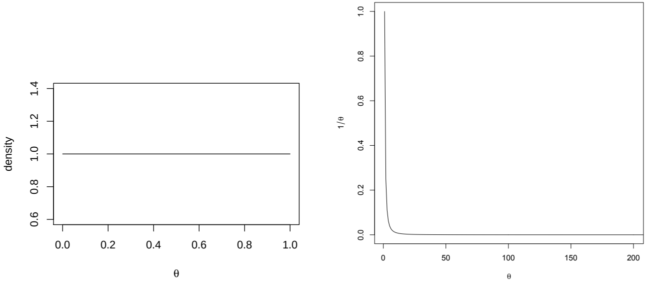


Figure 1: Comparison of the Uniform prior and the transformed prior on θ.

Invariance

You will find the the uniform prior is not invariant to transformations.

What does invariance mean intuitively?

Invariance

Let θ be our parameter of interest.

Transform to $g(\theta)$.

The transformation is said to be invariant
θ and $g(\theta)$ have the same distributions form (Normal, Beta, etc) up to a normalizing constant k.

Why is invariance important?

Suppose θ is the true population height in inches!

However, we receive some data the data is now in cm.

Instead of reformatting the data, we could just transform the parameter.

Also, we would hope that our prior is not sensitive to a slight change in our parameter (inches, cm).

Invariance

One prior that is nice to work with was discovered by Jeffreys' and is invariant under one-to-one transformations of the parameter.[1]

[1] This is not true for the Uniform(0,1) prior.

Jeffreys' prior

Define

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

where $I(\theta)$ is called the Fisher information. Then *Jeffreys' prior* is defined to be

$$p_J(\theta) = \sqrt{I(\theta)}.$$

Jeffreys' and the Binomial Likelihood

Consider

$$Y \sim \text{binomial}(n, \theta).$$

Derive Jeffrey's prior.

Log-likelihood

The likelihood is

$$p(y|\theta) = \binom{n}{y}\theta^{y}(1-\theta)^{n-y}$$

, which implies that the log-likelihood is

$$\log p(y|\theta) = \log\binom{n}{y} + y\log\theta + (n-y)\log(1-\theta) \qquad (1)$$

Partial derivatives

Recall that $\log p(y|\theta) = \log \binom{n}{y} + y \log \theta + (n-y) \log(1-\theta)$. It follows that the first and second partial derivatives with respect to θ.

$$\frac{\partial \log p(y|\theta)}{\partial \theta} = \frac{y}{\theta} - \frac{n-y}{1-\theta} \tag{2}$$

and

$$\frac{\partial^2 \log p(y|\theta)}{\partial \theta^2} = \frac{y}{-\theta^2} - \frac{n-y}{(1-\theta)^2}. \tag{3}$$

Fisher information

Given $E(y|\theta) = n\theta$ and $\frac{\partial^2 \log p(y|\theta)}{\partial\theta^2} = \frac{y}{-\theta^2} - \frac{n-y}{(1-\theta)^2}$.

$$E[\frac{\partial^2 \log p(y|\theta)}{\partial\theta^2}|\theta] = \frac{E(y|\theta)}{-\theta^2} - \frac{n - E(y|\theta)}{(1-\theta)^2} \tag{4}$$

$$= \frac{n\theta}{-\theta^2} - \frac{n - n\theta}{(1-\theta)^2} \tag{5}$$

$$= -\frac{n}{\theta} - \frac{n}{(1-\theta)} \tag{6}$$

$$I(\theta) = -E[\frac{\partial^2 \log p(y|\theta)}{\partial\theta^2}|\theta] = \frac{n}{\theta} + \frac{n}{(1-\theta)} = \frac{n}{\theta(1-\theta)}. \tag{7}$$

Jeffreys' and the Binomial Likelihood

Thus, the Jeffreys' prior for the binomial model is

$$p_J(\theta) = \sqrt{I(\theta)} = \sqrt{n}\theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}} \propto \text{Beta}(\frac{1}{2}, \frac{1}{2})$$

Comparison with Jeffreys' prior and the Uniform(0,1) prior

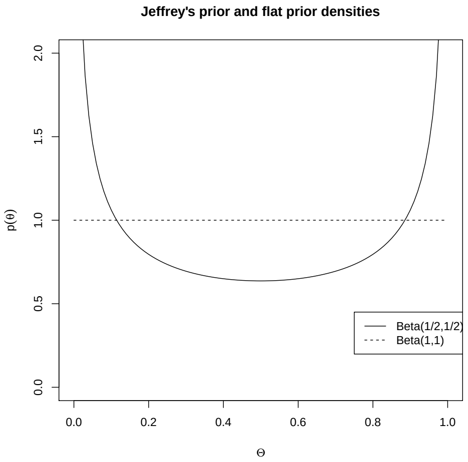


Figure 2: Comparison of the prior density $\pi_J(\theta)$ with that for a flat prior, which is equivalent to a Beta(1,1) distribution.

Limitations of Jeffreys' prior

Jeffreys' priors work well for single-parameter models, but not for models with multidimensional parameters. By analogy with the one-dimensional case, one might construct a naive Jeffreys prior as the joint density:

$$\pi_J(\theta) = |I(\theta)|^{1/2},$$

where $|\cdot|$ denotes the determinant and the (i, j)th element of the Fisher information matrix is given by

$$I(\theta)_{ij} = -E\left[\frac{\partial^2 \log p(X|\theta)}{\partial\theta_i\partial\theta_j}\right].$$

[For more reading: See PhD notes: Objective Bayes Chapter on reference priors, Gelman, et al. (2013)]