Noninformative (“Default”) Bayes

Rebecca C. Steorts

Bayesian Methods and Modern Statistics: STA 360/601

Lecture 6

Exam I

- Exam Thursday, Feb 11th in class. Be early to class so that you can start you exam on time.
- You will need pencil and paper. No calculators, no computers, no cell phones, etc permitted. No notes permitted.
- The exam will cover material through Module 4. This includes all readings.
- Assignment 2 solutions will be posted shortly.
- Assignment 3 has been posted with 2 suggested problems to work on (and you will get credit for them).
- There was an optional homework problem with Module 3, Part I. The solutions have been posted.
- Lab next week: Review sessions to prepare for the exam.

Exam I

- Intro to Bayes. What is it and why do we use it?
- Decision theory - loss, risk (all three of them).
- Hierarchical modeling - conjugacy, priors, posteriors, likelihood.
- Consistency, posterior predictive, credible intervals.
- Objective Bayes

Exam I: Expect 4 – 6 problems. You will need to really know the material to get through this exam.

Today's menu

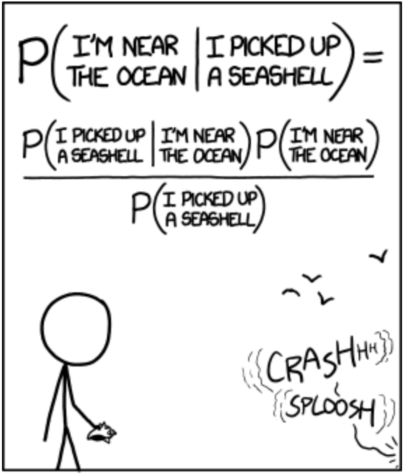
- Subjective prior
- Default prior
- Are they really noninformative?
- Invariance property
- Jeffreys' prior

- 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?

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 (courtesy of Christian Robert)!

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.

Comment: Since many of the objective priors are improper, so we must check that the posterior is proper.

Propriety of the Posterior

- If the prior is proper, then the posterior will *always* be proper.
- If the prior is improper, you must check that the posterior is proper.

A flat prior (my longer translation....)

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.

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

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

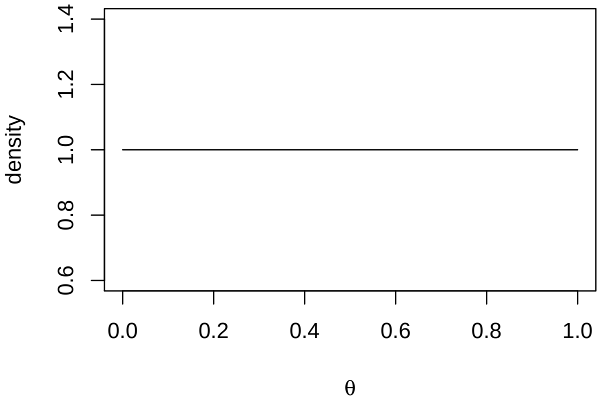


Figure 1: Unif(0,1) prior

What happens if we consider though the transformation to $1/\theta$. Is our prior still flat (does it place equal weight at every parameter value)?

Hint: Use change of variables from calculus.

Criticism of the Uniform Prior

- The Uniform prior of Bayes (and Laplace) and has been criticized for many different reasons.
- We will discuss one important reason for criticism and not go into the other reasons since they go beyond the scope of this course.
- In statistics, it is often a good property when a rule for choosing a prior is *invariant* under what are called one-to-one transformations.
- Invariant basically means unchanging in some sense.
- The invariance principle means that a rule for choosing a prior should provide equivalent beliefs even if we consider a transformed version of our parameter, like p^2 or $\log p$ instead of p.

Jeffreys' Prior

One prior that is invariant under one-to-one transformations is Jeffreys' prior.

What does the invariance principle mean?

Suppose our prior parameter is θ, however we would like to transform to ϕ.

Define $\phi = f(\theta)$, where f is a one-to-one function, meaning that f is injective.

Jeffreys' prior says that if θ has the distribution specified by Jeffreys' prior for θ, then $f(\theta)$ will have the distribution specified by Jeffreys' prior for ϕ. We will clarify by going over two examples to illustrate this idea.

Example: Uniform

Note, for example, that if θ has a Uniform prior, Then one can show $\phi = f(\theta)$ will not have a Uniform prior (unless f is the identity function).

Show this at home. Hint: use change of variables.

Example: Jeffreys'

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

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

For homework you will prove that the uniform prior in not invariant to transformation but that Jeffrey's is.

Example: Jeffreys'

Suppose

$$X|\theta \sim \text{Binomial}(n, \theta).$$

Let's calculate the posterior using Jeffreys' prior. To do so we need to calculate $I(\theta)$. Ignoring terms that don't depend on θ, we find

$$\begin{aligned}
\log p(x|\theta) &= x \log(\theta) + (n-x)\log(1-\theta) \implies \\
\frac{\partial \log p(x|\theta)}{\partial \theta} &= \frac{x}{\theta} - \frac{n-x}{1-\theta} \\
\frac{\partial^2 \log p(x|\theta)}{\partial \theta^2} &= -\frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2}
\end{aligned}$$

Example: Jeffreys'

Since, $E(X) = n\theta$, then

$$I(\theta) = -E\left[-\frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2}\right] = \frac{n\theta}{\theta^2} + \frac{n-n\theta}{(1-\theta)^2} = \frac{n}{\theta}\frac{n}{(1-\theta)} = \frac{n}{\theta(1-\theta)}.$$

This implies that

$$\begin{aligned} p_J(\theta) &= \sqrt{\frac{n}{\theta(1-\theta)}} \\ &\propto \text{Beta}(1/2, 1/2). \end{aligned}$$

- Here, $p_J(\theta) \propto \text{Beta}(1/2, 1/2)$.
- Let's consider the plot of this prior. Flat here is a purely abstract idea.
- In order to achieve objective inference, we need to compensate more for values on the boundary than values in the middle.

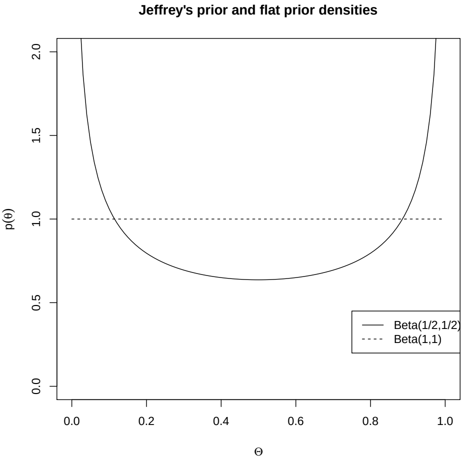


Figure 2: Jeffreys' prior and the Uniform (0,1) prior

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

- We see that the data has the least effect on the posterior when the true $\theta = 0.5$, and has the greatest effect near the extremes, $\theta = 0$ or 1.
- Jeffreys' prior compensates for this by placing more mass near the extremes of the range, where the data has the strongest effect.
- We could get the same effect by (for example) letting the prior be $\pi(\theta) \propto \frac{1}{\text{Var}\theta}$ instead of $\pi(\theta) \propto \frac{1}{[\text{Var}\theta]^{1/2}}$.
- However, the former prior is not invariant under reparameterization, as we would prefer.

We then find that

$$\begin{aligned}
p(\theta \mid x) &\propto \theta^x (1-\theta)^{n-x} \theta^{1/2-1} (1-\theta)^{1/2-1} \\
&= \theta^{x-1/2} (1-\theta)^{n-x-1/2} \\
&= \theta^{x-1/2+1-1} (1-\theta)^{n-x-1/2+1-1}.
\end{aligned}$$

Thus, $\theta|x \sim$ Beta$(x+1/2, n-x+1/2)$, which is a proper posterior since the prior is proper.