Noninformative ("Default") Bayes

Rebecca C. Steorts
Bayesian Methods and Modern Statistics: STA 360/602

March 22, 2018

Agenda

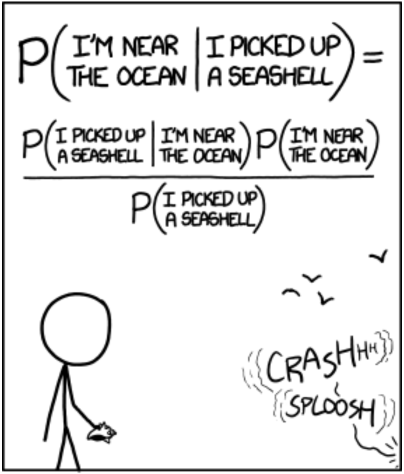
- Subjective prior
- Default prior
- Are they really noninformative?
- Invariance property
- Jeffreys' prior

- 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?

P R O B L E M.

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.

Propriety of the Posterior

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

- 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.

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

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

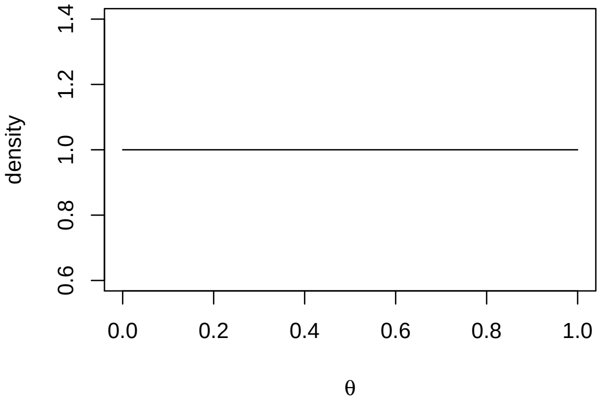


Figure 1: Unif(0,1) prior

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)?

Do this on your own using change of variables. To help answer the question, plot both the original and transformed priors.

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

What does invariance mean intuitively?

Let θ be our parameter of interest.

Transform to $g(\theta)$.

The transformation is said to be invariant (in distribution) if the distributions of θ and $g(\theta)$ have the same form (Normal, Beta, etc).

Furthermore, we could have invariance of parameters:

- location (mean),
- the scale (variance),
- or both (mean and variance).

Why is it nice to have such a property?

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).

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

Suppose we consider Jeffreys' prior, $p_J(\theta)$, where $X \sim \text{Bin}(n, \theta)$.

We calculate Jeffreys' prior by finding the Fisher information. The Fisher information tells us how much information the data gives us for certain parameter values.

- 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.

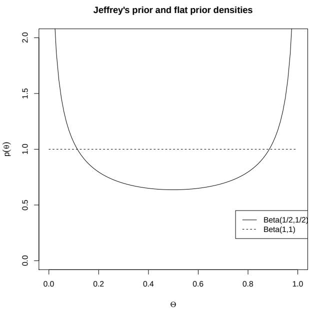


Figure 2: Beta(1/2,1/2) and Beta(1,1) priors.

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

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.

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).

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

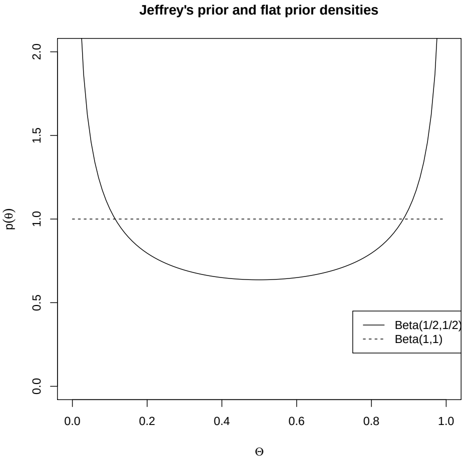


Figure 3: Jeffreys' prior and flat prior densities

Figure **??** 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 $\pi(\theta) \propto \dfrac{1}{\mathrm{Var}\theta}$ instead of $\pi(\theta) \propto \dfrac{1}{[\mathrm{Var}\theta]^{1/2}}$.
- 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, $\theta | x \sim \text{Beta}(x + 1/2, n - x + 1/2)$, which is a proper posterior since the prior is proper.

Jeffreys' and Conjugacy

- In general, they are not conjugate priors.
- For example, with a Gaussian model $X \sim N(\mu, \sigma^2)$, it can be shown that $\pi_J(\mu) = 1$ and $\pi_J(\sigma) = \frac{1}{\sigma}$, which do not look anything like a Gaussian or an inverse gamma, respectively.
- However, it can be shown that Jeffreys priors are limits of conjugate prior densities.
- For example, a Gaussian density $N(\mu_o, \sigma_o^2)$ approaches a flat prior as $\sigma_o^2 \to \infty$, while the inverse gamma $\sigma^{-(a+1)} e^{-b/\sigma} \to \sigma^{-1}$ as $a, b \to 0$.

Limitations of Jeffreys'

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)]

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

Think we're moving from the variance to the precision!

$$\phi = \frac{1}{\theta} \implies \theta = \frac{1}{\phi}. \tag{1}$$

Then using a change of variables transformation,

$$|\frac{\partial \theta}{\partial \phi}| = \frac{1}{\phi^2}. \tag{2}$$

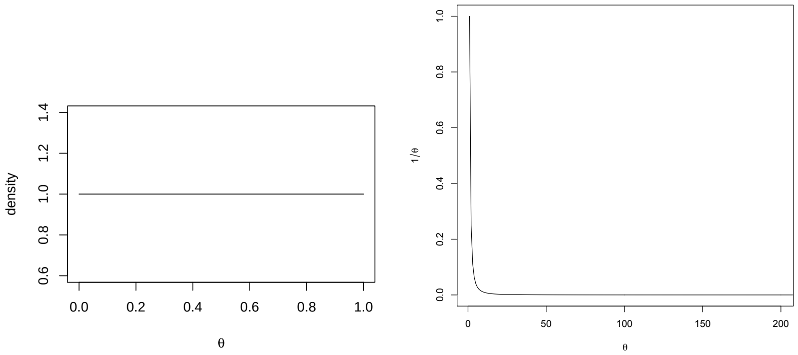


Figure 4: Comparison of the Uniform prior and the transformed prior on θ.

Recall Jeffrey's prior

Let's go back over Jeffreys' when $X \sim \text{Bin}(n, \theta)$.

How do we calculate Jeffreys' prior?

Specifically, we showed that

$$p_J(\theta) \propto \text{Beta}(1/2, 1/2).$$

Recall Jeffrey's prior

Recall

$$p_J(\theta) = \sqrt{\frac{n}{\theta(1-\theta)}}.$$

What happens if we consider though the transformation to $\phi = \dfrac{1}{\theta}$?

$$\begin{aligned} p_J(\phi) &= p_J(1/\phi) \times |\frac{\partial \theta}{\partial \phi}| && (3)\\ &= \sqrt{\frac{n}{\frac{1}{\phi}(1-\frac{1}{\phi})}} \times \frac{1}{\phi^2} && (4)\\ &\propto \frac{\phi}{\sqrt{\phi - 1}} \times \frac{1}{\phi^2} && (5)\\ &\propto \frac{1}{\phi\sqrt{\phi-1}}, \;\; \text{where } \; 1 \leq \phi < \infty. && (6) \end{aligned}$$

Is the resulting transformation invariant? Yes!

Can write the posterior as a Beta distribution. What is the update?

It's invariant because $\phi = \dfrac{1}{\theta}$ is a one-to-one function.

Now consider the transformation $\theta = \frac{1}{\sqrt{\phi}}$.

$$|\frac{\partial\theta}{\partial\phi}| = \frac{1}{2\phi^{3/2}}. \tag{7}$$

$$\begin{aligned}
p_J(1/\sqrt{\phi}) &= p_J(1/\sqrt{\phi}) \times |\frac{\partial\theta}{\partial\phi}| && (8)\\
&= (\phi)^{-1/4}\left(\frac{\phi^{1/2}-1}{\phi^{1/2}}\right)^{-1/2} \times \frac{1}{2\phi^{3/2}} && (9)\\
&\propto \phi^{3/2}(\phi^{1/2}-1)^{-1/2} \text{ where } 1 \leq \phi < \infty. && (10)
\end{aligned}$$

Is the resulting transformation invariant? Why or why not? Can write the posterior as a Beta distribution. What is the update?

Take home message with Jeffreys'

Consider parameter θ and transformation $g(\theta)$.

Jeffreys' prior is invariant under parameterization means that Jeffreys' prior corresponds to $g(\theta)$ is the *same* as applying a change of "measure" or distribution to the Jeffreys' prior for θ.

Said differently, if $\theta = g(\theta)$, then 1 and 2 are the same below. Let J represent a Jeffreys' prior.

1.
$$J_\phi = J_\theta \times |\frac{\partial g^{-1}(\theta)}{\partial \theta}|$$

2.
$$J_\phi = \sqrt{I(\phi)},$$

where $I(\phi)$ is the Fisher information of ϕ.