ST559 Homework 3

Nick Sun

May 4, 2020

Abstract

Answer questions 5, 7, 17, 22 from *Bayesian Data Analysis* and also find Jeffrey's Prior for $X|\theta \sim$ Negative Binomial random variable where θ is the probability of success.

1 Q5

Posterior distribution is a compromise between prior information and data. Let y be the number of heads in n spins of the coin whose probability is θ.

1.1 part a.

If your prior distribution for θ is uniform on the range $[0, 1]$ derive your prior predictive distribution for y for each $k = 0, 1, \dots, n$.

$$\begin{aligned}
Pr(y = k) &= \int_0^1 Pr(y = k|\theta)d\theta \\
&= \int_0^1 \binom{n}{k} \theta^k (1-\theta)^{n-k} d\theta \\
&= \binom{n}{k} \int_0^1 \theta^k (1-\theta)^{n-k} d\theta \\
&= \frac{n!}{(n-k)!k!} \frac{(n-k)!k!}{(n+1)!} \\
&= \frac{1}{n+1}
\end{aligned}$$

We used the fact that $\int_0^1 \theta^k (1-\theta)^{n-k} d\theta$ is the kernel of a Beta distribution with $\alpha = k + 1$ and $\beta = n - k + 1$.

1.2 part b.

Suppose you assign a Beta(α, β) prior distribution for θ and then you observe y heads out of n spins. Show algebraically that your posterior mean of θ always lies between the prior mean $\frac{\alpha}{\alpha+\beta}$ and the observed relative frequency of heads $\frac{y}{n}$.

The posterior can be calculated as

$$\binom{n}{y}\theta^y(1-\theta)^{n-y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \propto \theta^{y+\alpha}(1-\theta)^{n-y+\beta-1}$$

This posterior is equal to Beta$(y+\alpha, n-y+\beta)$. The posterior mean for this distribution is $\frac{y+\alpha}{\alpha+n+\beta}$.

In order to show that $\frac{y+\alpha}{\alpha+n+\beta}$ is in between $\frac{y}{n}$ and $\frac{\alpha}{\alpha+\beta}$, we can show that the posterior mean is a convex combination between the prior and sample means. If we consider the weight between the means to be γ, we can calculate:

$$\begin{aligned}
\frac{\alpha+y}{\alpha+\beta+n} &= \gamma\left(\frac{\alpha}{\alpha+\beta}\right) + (1-\gamma)\left(\frac{y}{n}\right) \\
&= \frac{y}{n} + \gamma\left(\frac{\alpha n - y(\alpha+\beta)}{(\alpha+\beta)n}\right) \\
\gamma &= \left(\frac{(\alpha+y)n - (\alpha+\beta+n)y}{(\alpha+\beta+n)n}\right)\left(\frac{(\alpha+\beta)n}{\alpha n - y(\alpha+\beta)}\right) \\
\gamma &= \frac{\alpha+\beta}{\alpha+\beta+n}
\end{aligned}$$

The last equality shows that γ is between $[0, 1]$, therefore the posterior mean is in between the sample mean and the prior mean.

1.3 part c.

Show that if the prior distribution on θ is uniform, the posterior variance of θ is always less than the prior variance.

Let's take the standard uniform distribution to start $U(0, 1)$. It is known the the variance of the standard uniform distribution is $\frac{1}{12}(b-a)^2$, so the variance of the standard uniform is $\frac{1}{12}$.

We showed in part (a) that the posterior distribution with a uniform prior will be $Beta(k+1, n-k+1)$. This gives us a posterior variance of

$$\begin{aligned}
\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} &= \frac{(k+1)(n-k+1)}{(n+2)^2(n+3)} \\
&= \frac{k+1}{n+2}\frac{n-k+1}{n+2}\frac{1}{n+3}
\end{aligned}$$

The smallest possible n we can have is $n = 1$. We see that the first two terms in this quantity sum to 1, so at most, both of these values can be $\frac{1}{2}$ and their product would be $\frac{1}{4}$. The last term in this quantity cannot be bigger than $\frac{1}{4}$ since $n \geq 0$. Therefore, the posterior variance will always be less than or equal to $\frac{1}{12}$.

1.4 part d.

Give an example of a Beta(α, β) prior distribution and data y, n in which the posterior variance of θ is higher than the prior variance.

After playing around with the formula for the variance of a beta distribution in R, the following combination of $n = 2$, $y = 2$, prior $\alpha = 1$ and prior $\beta = 10$, we get the prior variance to be $\approx$ .00688 (variance of a Beta(1,10) distribution) and the posterior variance to be .0126 (variance of a Beta(3,10) distribution).

2 Q7

Let's investigate noninformative prior densities.

2.1 part a.

For the binomial likelihood $y \sim Bin(n, \theta)$ show that $p(\theta) \propto \theta^{-1}(1-\theta)^{-1}$ is the uniform prior distribution for the natural parameter of the exponential family.

We note here that the binomial distribution can be reparameterized in terms of the natural parameter η:

$$\begin{aligned}\binom{n}{k}\theta^k(1-\theta)^{n-k} &= \binom{n}{k}\left(\frac{\theta}{1-\theta}\right)^k(1-\theta)^n \\ &= \binom{n}{k} exp\left(klog\frac{\theta}{1-\theta}\right)(1-\theta)^n\end{aligned}$$

So our natural parameter is $log\frac{\theta}{1-\theta}$. If we use a univariate transformation, we get the following pdf for θ in terms of the uniform distribution of η:

$$\begin{aligned}&p(\theta) = p\left(\frac{e^\eta}{1+e^\eta}\right)\|\frac{d}{d\theta}log\left(\frac{\theta}{1-\theta}\right)\| \\ &p(\theta) = p\left(\frac{e^\eta}{1+e^\eta}\right)\|\frac{d}{d\theta}log\theta - log(1-\theta)\| \\ &p(\theta) \propto \theta^{-1}(1-\theta)^{-1}\end{aligned}$$

2.2 part b.

Show that if $y = 0$ or n, the resulting posterior distribution is improper.

If $y = 0$, then the posterior distribution that we obtained above will be $p(\theta|y) \propto \theta^{-1}(1-\theta)^{n-1}$. Notice that when θ is small, this integral will not converge since there is an infinite integral near $\theta = 0$.

If $y = n$, then the posterior distribution will be $p(\theta|y) \propto \theta^{n-1}(1-\theta)^{-1}$ and we run into a similar infinite integral when θ is large, around $\theta = 1$.

3 Q17

Unlike the central posterior interval, the highest posterior interval is not invariant to transformation. For example, suppose that given σ^2 the quantity nv/σ^2 is distributed as χ^2_n and that σ has the improper noninformative prior density $p(\sigma) \propto \sigma^{-1}, \sigma > 0$

3.1 part a.

Prove that the corresponding prior density for σ^2 is $p(\sigma^2) \propto \sigma^{-2}$

Let's define $v = \sigma^2$ and note that $\sigma = \sqrt{v}$. Now we can use a univariate transformation:

$$\begin{aligned}
p(v) &= p(\sqrt(v))\|\frac{d}{dv}\sqrt{v}\| \\
&\propto \frac{1}{\sqrt{v}}\frac{1}{2\sqrt{v}} \\
&\propto \frac{1}{v} = \frac{1}{\sigma^2}
\end{aligned}$$

3.2 part b.

Show that the 95% highest posterior density region for σ^2 is not the same as the region obtained by squaring the endpoints of a posterior interval for σ

First we need to get the posterior for $p(\sigma|data)$ and $p(\sigma^2|data)$. Recognizing that $\frac{nv}{\sigma^2}\|\sigma^2 \sim \chi^2_n$, we can calculate the posterior probaiblities as:

$$\begin{aligned}
p(\sigma|data) &\propto (\sigma^2)^{1/2-n/2} exp\left(\frac{-nv}{2\sigma^2}\right) \\
\text{and } p(\sigma^2|data) &\propto (\sigma^2)^{-1-n/2} exp\left(\frac{-nv}{2\sigma^2}\right)
\end{aligned}$$

Now let's suppose that we have (a, b) is the 95% highest density interval for $p(\sigma^2|data)$. If the proposition is true, then we have $(\sqrt{a}, \sqrt{b})$ as the 95% highest density interval for $p(\sigma|data)$ as well.

For $p(\sigma|data)$, we get the following relation for the densities as a and b.

$$\begin{aligned}
a^{-1/2-n/2} exp\left(\frac{-nv}{2a}\right) &= b^{-1/2-n/2} exp\left(\frac{-nv}{2b}\right) \\
\text{or equivalently } \left(-\frac{1}{2} - \frac{n}{2}\right) log(a) - \frac{nv}{2a} &= \left(-\frac{1}{2} - \frac{n}{2}\right) log(b) - \frac{nv}{2b}
\end{aligned}$$

For $p(\sigma^2|data)$, we get the following relation for the densities a and b.

$$\begin{aligned}
a^{-1/2-n/2} exp\left(\frac{-nv}{2a}\right) &= b^{-1/2-n/2} exp\left(\frac{-nv}{2b}\right) \\
\left(-1 - \frac{n}{2}\right) log(a) - \frac{nv}{2a} &= \left(-1 - \frac{n}{2}\right) log(b) - \frac{nv}{2b}
\end{aligned}$$

If we solve for a in terms of b and substitute it back in, we get $a = b$. This is not possible since this interval is supposed to be the 95% highest posterior interval. Therefore, the highest posterior density of σ^2 cannot be just the square of the interval of σ.

4 Q22

A study is performed to estimate the effect of a simple training program on free throws. A random sample of 100 college students is recruited into the study. Each student first shootes 100 free-throws to establish a baseline success probability. Each student then takes 50 practice shots each day for a month. At the end of that time, he or she take 100 shots for a final measurement. Let θ be the avearge improvement in success probability.

Give three prior distributions for θ, explaining each in a sentence.

A noninformative prior It is known for the Binomial distribution parameter p that a noninformative prior would be Jeffrey's Prior $Beta(1/2, 1/2)$. This distribution is essentially a bell curve limited to and centered on the unit interval. We can apply similar logic to θ where we want a bell curve limited between the beforehand success probability π with $1 - \pi$ as an upper bound since we cannot have an overall success probability that is greater than 1. Our noninformative prior would be a Beta distribution centered at the midpoint of π and $1 - \pi$.

A subjective prior based on your best knowledge A subjective prior might be calculated from past knowledge about this training program, for example, that most students on average improved 10% with a standard deviation of 2%.

A weakly informative prior A weakly informative prior could be based on our subjective prior. For example, if we think that the average improvement rate that we have seen in the past is 10% with a standard deviation of 2%, a weakly informative prior might increase the standard deviation so that there is more uncertainty in the prior.

5 Additional Question about Negative Binomial Distribution

If $X|\theta \sim$ Negative Binomial with parameters n, θ where θ is the probability of success, find Jeffrey's Prior for θ.

The first thing we need to find is the Fisher information about θ given in the data. Parameterizing the negative binomial as follows:

$$p(X|n, \theta) = \binom{x-1}{n-1} \theta^n (1-\theta)^{x-n} \text{ for } x = n, n+1, \ldots$$

We get the following for $I(\theta)$:

$$\begin{aligned}
\log P(X|n, \theta) &= log\left(\binom{x-1}{n-1}\right) nlog(\theta) + (x)log(1-\theta) \\
\frac{\partial}{\partial\theta}\log P(x|n, \theta) &= \frac{n}{\theta} - \frac{x-n}{1-\theta} \\
\frac{\partial^2}{\partial^2\theta}\log P(x|n, \theta) &= -\frac{n}{\theta^2} - \frac{x-n}{(1-\theta)^2} \\
I(\theta) &= \frac{n}{\theta^2(1-\theta)}
\end{aligned}$$

Note that we made use of the fact that in this parameterization, the negative binomial distribution has $E_\theta[X] = \frac{n(1-\theta)}{\theta}$.

Jeffrey's prior will be proportional to $I(\theta)^{1/2}$. In that case, we have that Jeffrey's prior of the Negative Binomial distribution will be $\propto \theta^{-1}\theta^{-1/2}$. This is similar to a beta distribution, but since the beta requires that $\alpha, \beta > 0$, it is not exactly a beta distribution.

We can contrast this with Jeffrey's prior for the binomial distribution which was $\propto \theta^{-1/2}(1-\theta)^{-1/2}$. This is a proper $\text{Beta}(1/2, 1/2)$ distribution.