# ST440/540 Applied Bayesian Analysis

## Lab activity for 2/3/2025

Announcements

- Quiz 5 due Friday, Feb 7
- Assignment 3 due Friday, Feb 7
- I will post Exam 1 on Friday and it is due on Feb 17

## A. HOMEWORK AND QUIZ SOLUTIONS

Q4: NC State classes have been cancelled due to snow in Y=5 of the past n=40 years (made up numbers). For planning purposes, we would like to estimate the true probability of a cancellation in a given year.

(a) Assuming university policy is constant over time and snowfall is independent and identically distributed across years, specify a likelihood for the data.

There are n iid binary trials so I choose likelihood Y|θ ~ Binomail(n,θ).

(b) Specify an uninformative conjugate prior.

The conjugate prior for the parameter θ in the likelihood in (a) is a Beta(a,b). To make it uninformative (uniform), I select a=b=1.

(c) Give (do not derive) the posterior distribution.

As shown in the video, the posterior is θ|Y ~ Beta(Y+a,n-Y+b).

(d) What are the posterior mean and standard deviation? How do they compare with the sample proportion, Y/n = $\hat{\theta}$= 0.125 and its approximate standard error, $\sqrt{\hat{\theta}(1-\hat{\theta})/n}$ = 0.052 (i.e., are they pretty similar, yes or no)?

The posterior mean and SD are 0.143 and 0.053, which are fairly similar to the sample proportion 0.125 and its standard error 0.052.

```
> Y <- 5
> n <- 40
> a <- b <- 1
> A <- Y+a
> B <- n-Y+b
>
> A/(A+B)
[1] 0.1428571
> sqrt(A*B/((A+B)*(A+B)*(A+B+1))) # From https://en.wikipedia.org/wiki/Beta_distribution
[1] 0.05336338
```

## B. STUDENT QUESTIONS

(1) What is the kernel of a distribution?

It is the portion that involves the parameter of interest (θ) and excludes other constants. That is, if the distribution for θ can be written is

p(θ) = [stuff that doesn't depend on θ]*[stuff that does depend on θ],

then [stuff that does depend on θ] is the kernel. For example, if p(θ) = $4\lambda\theta^4$ the kernel is $\theta^4$.

(2) Is the kernel in this context the same as what is used in kernel density estimation?

There may be some way to connect the two, but I think of them as separate.

(3) Will we get into how we deal with likelihoods that aren't reliant on θ? Like, if Y in the question below was the amount of snowfall, and θ was the probability that Y was greater than a certain number, how do we deal with that (since I would think that Y would thus follow a Gamma distribution)?

If the likelihood doesn't involve a parameter at all this means the data are not informative about the parameter and there is no value in analyzing the data if the goal is to learn about the parameter. But I think your question is what to do if the quantity if interest isn't exactly a parameter in the model. In your example, I might say log(snowfall) ~ Normal(μ,1). Then Prob(snowfall>threshold) = Prob[log(snowfall)>log(threshold)] = 1-F[log(threshold) - μ] = θ, where F is the standard normal CDF. We could than analyze the data using the model log(snowfall) ~ Normal(μ,1) to get the posterior for μ, and then convert this to a posterior for θ. This is easy if we have Monte Carlo samples for μ because we can just convert each sample to θ = 1-F[log(threshold) - μ] and we now have Monte Carlo samples for the parameter of interest, θ.

(4) Please explain the code for the Smoking Example code on Slide 21 of Lecture Notes for Chapter 2.1 Conjugate Priors. I think the Beta Distribution represents the Prior, and the Negative Binomial represents the Likelihood, but then where is the Posterior? Also, what is the reason we use theta1 and theta2 for Beta but then 1-theta1 and 1-theta2 for NB?

The posterior is also a beta distribution as given in the third bullet of Slide 19. The 1- θ is there because the R function is written in terms of the probability of failure rather than the probability of success.

(5) I did not follow the lecture on Slide 8 of Conjugate Priors concerning Improper Priors, can you please explain Improper Priors at the next lab?

An improper prior is one that doesn't integrate to one. For example, if θ has support θ>0 and we pick prior π(θ) = θ then $\int_0^\infty \pi(\theta)d\theta = \infty$ so the prior is improper. As we will discuss later in the course, it is possible that the posterior is a proper distribution so improper priors are sometimes used. In this week's video on objective Bayesian analysis, there will be several examples of improper priors.

(6) Do the concepts of sampling and parametric error change between a Bayesian and frequentist perspective? If so, how would I translate between the two?

(7) Does parametric uncertainty exist in frequentist approaches?

Yes, this is what the frequentist standard error is measuring as in the quiz solution above.

(8) In future do you prefer work done in this panel or is it okay to attach a google document with images to my work

I prefer you type the result into the text box.

(9) Can you do another step-by-step example on how to pick a likelihood, prior, and get your posterior? I still struggle to pick/spend a lot of time on these.

Sure, see the first problem below.

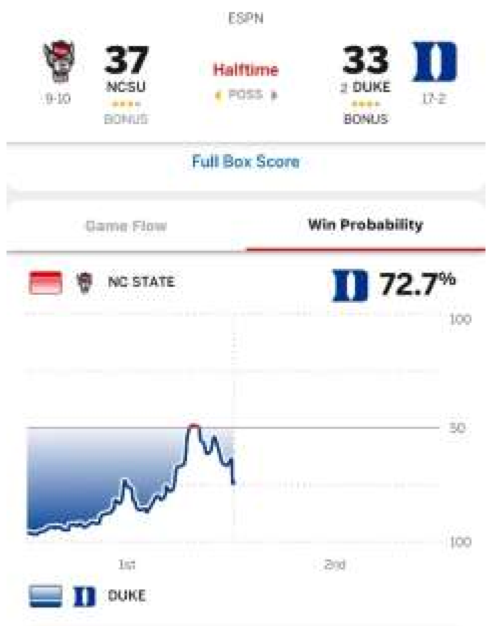
(10) There's no "correct" way to select a prior. How about some examples of wrong ways to select priors?

This is a good question. It is always fun to break things! See the first problem below.

## C. DISCUSSION QUESTIONS

(0) Explain the plot below using Bayesian jargon.

Based on the teams' records before the game (NCSU was 9-10, Duke was 17-2), the prior probability of Duke winning was almost 100%. The data collected during the first half of the game (Pack up 37-33) favors NCSU, but when combining the prior and first-half data the posterior probability of Duke winning was still 73% (they went on to win the game by 10 points).



(1) You are managing a factory and a collectively-bargained agreement says the injury rate should not exceed 4 injuries per month. In this past 6 months there have been 25 injuries, 4, 5, 2, 7, 1 and 6 respectively. Outline an analysis to determine whether the injury rate exceeds the agreed upon level.

(a) Specify a likelihood for the data

Assuming the months are iid, we can combine the data as Y=25 events in N=6 months and assume the likelihood is $Y|\theta \sim \text{Poisson}(N\theta)$.

(b) Identify a conjugate prior and derive the posterior

The likelihood is $f(Y|\theta) = (N\theta)^Y \exp(-N\theta) / Y!$ which is the kernel of a gamma for θ. Therefore, the conjugate prior is $\theta \sim \text{Gamma}(a,b)$ which gives posterior $\theta|Y \sim \text{Gamma}(Y+a,N+b)$ since

$$p(\theta|Y) \propto f(Y|\theta)\pi(\theta) \quad \propto \quad [\theta^Y \exp(-N\theta)] * [\theta^{a-1} \exp(-b\theta)] \quad \propto \quad \theta^{(Y+a)-1} \exp[-(N+b)\theta].$$

(c) Select an uninformative prior for this analysis

If we select a=0.5 and b=0.05 the prior median and 95% interval below spans the plausible values for this problem.

```
> round(qgamma(c(0.025,0.50,0.975),0.5,0.05),3)
[1]  0.010  4.549 50.239
```

(d) Select a “wrong” prior for this analysis

The prior $\theta \sim \text{Uniform}(0,4)$ has support [0,4], so the posterior support is also [0,4], and this would rule out the hypothesis to be tested. Other wrong priors include $\theta \sim \text{Normal}(0,10)$ and $\theta \sim \text{Poisson}(4)$.

(e) Describe how you would summarize the posterior.

The posterior is $\theta|Y \sim \text{Gamma}(Y+a,N+b)$. Plugging in values for the problem gives $\theta|Y \sim \text{Gamma}(25.5,6.05)$. The posterior is plotted below, and the posterior probability that $P(\theta>4|Y) = 0.58$ so there is not a lot of evidence that the injury rate exceeds the agreed upon level

```
> theta <- seq(0,10,.10)
> plot(theta,dgamma(theta,25.5,6.05))
> 1-pgamma(4,25.5,6.05)
[1] 0.5775358
```

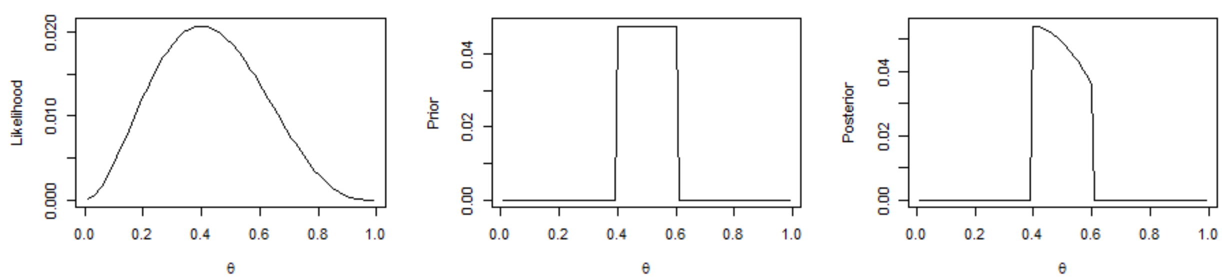
(2) Say $Y|\theta \sim$ Binomial$(n, \theta)$ and the prior is $\theta \sim$ Uniform$(0.4,0.6)$. Is this prior conjugate? How would you compute the posterior?

Write the prior as $\pi(\theta) = 5*I(0.4< \theta < 0.6)$ where $I(x)$ equals 1 is x is true and zero otherwise. The posterior is then

$$p(\theta|Y) \ \alpha \ f(Y|\theta)\pi(\theta) \quad \alpha \quad [\theta^{Y} (1-\theta)^{N-Y}] * I(0.4< \theta < 0.6).$$

The prior is conjugate if and only if the posterior is also uniform, but it's not uniform because it is not flat in $\theta$. Even though the posterior is not conjugate, we can still approximate it graphically like we did before we knew about conjugacy.

```
> theta <- seq(0.01,0.99,0.01)
> Y     <- 2
> n     <- 5
>
> like  <- dbinom(Y,n,theta)
> prior <- dunif(theta,0.4,0.6)
> post  <- like*prior
>
> par(mfrow=c(1,3))
> plot(theta,like/sum(like))
> plot(theta,prior/sum(prior))
> plot(theta,post/sum(post))
```



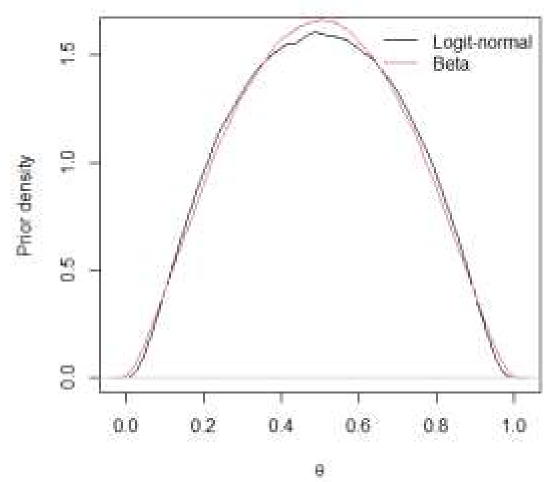
(3) Say Y|θ ~ Binomial(n, θ) and previous literature suggests that the log odds, $\gamma = \log(\theta/(1-\theta))$ should have standard normal prior (the inverse relationship is $\theta = \exp(\gamma)/[1+\exp(\gamma)] = \text{expit}(\gamma)$ may be helpful).

(a) How could you (approximately) convert his information to be a conjugate prior for θ?

You could find the Beta(a,b) that matches the mean and SD. This seems pretty good.

```
gamma  <- rnorm(1000000)
theta1 <- exp(gamma)/(1+exp(gamma))
theta2 <- rbeta(1000000,2.4,2.4)
mean(theta1);sd(theta1)
[1] 0.5000447
[1] 0.2081487
mean(theta2);sd(theta2)
[1] 0.4996944
[1] 0.2075064
plot(density(theta1),xlab=expression(theta),ylab="Prior density")
lines(density(theta2),col=2)

legend("topright",c("Logit-normal","Beta"),col=1:2,lty=1,bty="n")
```

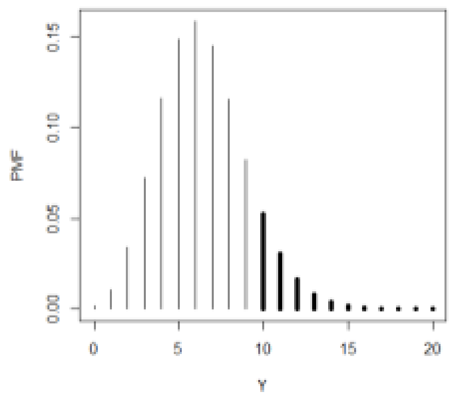



(b) How would you compute the posterior distribution θ using the normal prior for γ?

Option 1: Write the model as Y| γ ~ binomial(n, expit(γ)) with prior γ ~ Normal(0,1). From this analysis, obtain posterior samples of γ and convert each to θ = expit(γ) go obtain sample for θ.

Option 2: It turns out θ = expit(γ) follows a known distribution, the logit-normal distribution, so you could use this as the prior PDF in the analysis Y| θ ~ binomial(n, θ) with prior θ ~ LogitNormal(0,1).

(4) A survey of n=10 people found that participants had a sample mean of $\bar{Y}$ = 6.4 online subscriptions. The analysis wanted to compute the probability that the next survey participant would have at least 10 subscriptions. Denote the next participant's number of subscriptions as Y*. They assumed Y* ~ Poisson(6.4) and got P(Y*>9) = 0.114 (see the plot below).




(a) Would you expect the probability to be higher or lower than 0.114 if it was computed using the posterior predictive distribution? Why?

It could depend on the prior, but assuming it is uninformative, because of uncertainty about the true value of the mean parameter, the PPD will be wider than the Poisson distribution giving more probability above 9.

(b) Would you expect the probability computed using the posterior predictive distribution to be more similar to 0.114 if the sample size increased or decreased? Why?

As the sample size increases the posterior of the true mean will be 6.4 with no uncertainty and so the PPD will be exactly Poisson(6.4) and the PPD prob will be 0.114.

(c) Describe how to compute probabilities from the posterior predictive distribution using R.

We would use Monte Carlo sampling to get a bunch of samples for Y*, and then compute the proportion of the MC samples that are at least 10.

(5) Say $Y|\theta$ ~ Binomial(n, $\theta$) and $\theta$ ~ beta(a,b). The derivation of the posterior in the notes was

$$p(\theta|Y) \propto \mathrm{f}(\mathrm{Y}|\theta)\pi(\theta) \propto \theta^{(Y+a)-1}(1-\theta)^{(n-Y+b)-1}$$

and then we concluded that the posterior was Beta(Y+a,n-Y+b). Explain why we never had to compute the marginal distribution of the data, m(y).

The marginal m(y) as a function of theta is a constant, so we can just say it's proportional to the term that include theta.

(6) Assume $Y_1|\theta$ ~ Binomial($n_1$,θ) independent of $Y_2|\theta$ ~ Binomial($n_2$,θ), and the prior is θ ~ Beta(a,b). Derive the posterior distribution of $\theta|Y_1,Y_2$.

Since Y1 and Y2 are independent the likelihood is f(Y1,Y2|theta) = f(Y1|theta)f(Y2|theta). Thus

$$
\begin{aligned}
p(\theta|\mathrm{Y1},\mathrm{Y2}) &\propto f(\mathrm{Y1},\mathrm{Y2}|\theta)\pi(\theta)\\
&\propto f(\mathrm{Y1}|\theta)f(\mathrm{Y2}|\theta)\pi(\theta)\\
&\propto [\theta^{\mathrm{Y1}}(1-\theta)^{\mathrm{n1-Y1}}][\theta^{\mathrm{Y2}}(1-\theta)^{\mathrm{n2-Y2}}][\theta^{a-1}(1-\theta)^{b-1}]\\
&\propto \theta^{(\mathrm{Y1+Y1}+a)-1}(1-\theta)^{(\mathrm{n1-Y1+n2-Y2}+b)-1}
\end{aligned}
$$

Therefore, the posterior is Beta(Y1+Y2+a, n1-Y1+n2-Y2+b), or Beta(Y+A, n+b) where Y=Y1+Y2 and n=n1+n2.

(7) You are tasked with writing the R package to conduct a Bayesian analysis of a proportion.

(a) What are the inputs to your function and what (if any) are the default values?

The model is Likelihood Y ~ Binomial (n,theta) and theta ~ beta(a,b). So we would have to have inputs Y, n, a and b. Y and n have no defaults, but maybe a=b=1 is a reasonable default.

(b) What are the outputs?

A table with the posterior mean, mode and 95% credible interval. Other options are a plot of the prior and posterior PDFs and the probability above a threshold so the users could test say Ho: theta<0.5 versus Ha: theta>0.5.