Chapter 4

The Beta-Binomial Conjugate Model and its Applications

4.1 Introduction

We are going to work with binary data again in this chapter. We have again a random sample of Bernoulli random variables $Y_1, Y_2, ..., Y_n$ where each $Y_i = 0 \quad or 1$, with 1 representing that the observation has the trait we are studying: likes Hollywood movies, lives with parents, temperature dropped after tornado, program compiled successfully, etc. The unknown of interest is the proportion p of successes (i.e., of 1's) in the population.

If we lack information distinguishing the n observations, we may treat their values as being exchangeable. Thus, the information contained in $Y_1, Y_2, ..., Y_n$ is equivalent to saying that there are $s = \sum_{i=1}^{n} Y_i$ observations with the characteristic of interest (i.e., with value 1). For example, if a data set is: 0,0,1,0,1,0,0,0,1,1,1, then we can say that there are $n = 11$ observations in the sample and that there are $s = 5$ observations that have the characteristic we are studying. The amount of information we obtain by looking at the string of numbers or by knowing the s is exactly the same. That is why summarizing the joint distribution or likelihood function with $p^s(1-p)^(n-s)$ is equivalent to seeing the whole string of numbers. We then say that $s = \sum_{i=1}^{n} Y_i$ is a sufficient statistic.

This quantity, s, is the total number of successes in the sample. It has its own distribution, a Binomial distribution, with parameter (n, p). A sum of n Bernoulli distributions with parameter p is a Binomial with parameters (n, p)

In this chapter, we study the Binomial distribution as a likelihood function for the same inference problem we studied in the last chapter. For a random sample of size n, we will say that the random variable s (which will vary with the sample), is a Binomial(n, p).

Combined with a Beta prior, a Binomial likelihood and a Beta prior give a Beta-binomial model.

4.2 The Beta-Binomial model

We will use the notes in section 3.1 of Hoff's book as notes for the theoretical discussion of the model. There is not much difference between this model and the "product of Bernoulii's" - Beta prior. So the results done with one will be the same as those done with the other. The Beta-Binomial model is used more often, though.

4.3 Crime victimization Survey Data

The crime victimization survey of the Bureau of Justice Statistics contains data on the number of females who were hurt by domestic violence in 1988 and 1994. In 1988, of the 4905 women with valid responses, 4795 (97.8%) were not hurt and 110 (2.2%) were. Armed with this prior information, we observe in 1994 that 4815 (98.2%) of the women were not hurt that year and 90 (1.8 %) were. Does this mean that female victimization in domestic violence is decreasing?

It must be pointed out that the total number of cases studied was 8643. Thus, 3738 of the respondents entered either invalid answers or no answers at all when surveyed.

Additional information in the survey reveals that 14.5 percent of the women hurt the first year, were also hurt the second year, while only 1.5 percent of those not hurt the first year were hurt the second year. Does this mean that victimization is chronic? That is, do women who get hurt once tend to be hurt more than once? We will answer this question in a later chapter.

We can also find in the survey several demographic and financial situation variables, as well as drug and alcohol use information for each of the households. With those in hand, could we determine what are the most important determinants of the fact that women get hurt in domestic violence and of the chronicity of victimization? Later in the course we will use logistic models to analyze these data.

How can we analyze these data using Bayesian methods to answer the question we have?

The simplest question to answer is whether female victimization in domestic violence is decreasing or not. In this chapter, we will answer only the first question. But first, we will introduce the Beta-Binomial pair for Bayesian analysis.

4.4 Posterior distribution of the proportion of victimized women in 1994. Uninformative and Informative priors

We will answer this question in two ways. In the first way, we will assume that we don't have any strong opinions about the p and just use a uniform prior $U(0, 1)$. This is equivalent to using a $Beta(1, 1)$. With the second approach, we will let the information from 1988 be our prior information and we will use it to construct a more informative prior for p.

Using the Binomial model for the data twice, and two types of prior, Chapter 3, Section 3.1 in Hoff's book contains the Bayesian theory relevant to this model, and we will spend one lecture discussing that theory as well. That Section 3.1 is required reading.

4.4.1 Ignore 1988 and just use the data of 1994. Uninformative prior

We use a uniform prior U(0, 1) = Beta($a = 1, b = 1$). Under this prior, our posterior distribution is

$$p(\theta \mid y_1, y_2, ..., y_{4905}) = \text{Beta}(0 + 90, 1 + 4815) \tag{4.1}$$

This is a Beta density with parameters $a = 91$ and $b = 4816$.

We will summarize the posterior exactly, obtaining an exact confidence interval, since we know the form of the posterior density. We do that with the following code.

```
  theta=seq(0,0.05,length=1000)
 post.density=dbeta(theta,91,4816)
 plot(theta,post.density, type="l",main="posterior prob of female vict-U(0,1) prior")
 post.interval=qbeta(c(0.025,0.975), 91,4816)
post.interval
 abline(v=post.interval)

 post.mean=91/(91+4816)   # the posterior mean
```

We will present the plot of this posterior together with the plot of the posterior obtained in the section with the informative prior.

An alternative method of summarization of a posterior distribution is based on simulation. In this case, we can simulate a large number of values from the beta posterior density and summarize the simulated output. Using the random beta command rbeta, we simulate 1000 random proportion values from the $Beta(91, 4816)$.

```
ps=rbeta(1000,91,4816)   # draw Beta(91,4816) random numbers
hist(ps,xlab="p",main="posterior for theta, simulated")    # histogram of random numbers
quantile(ps,c(0.025,0.975))    # posterior interval
mean(ps)  # posterior mean
```

Note that the summaries of the posterior density for theta based on simulation are approximately equal to the exact values based on calculation from the beta distribution.

4.4.2 Taking into account 1988 and using the data of 1994. Informative prior

I am going to approximate the prior reasoning as follows: the average prior p will be the observed proportion of women that were victimized (110/4905=0.02242610). Of course, with so much missing data and so much risk of respondents not being honest, we don't want our prior to be too strong. Thus, we make the standard deviation a little large.

The analysis now is more complicated because we must assess an informative Beta prior. Tentatively, and pending more serious assessment of the prior, we will go with a Beta(1,30).

With this Beta(1,30) and $s = 90$, $f = 4815$, we know that the posterior distribution will be
$Beta(1 + 90, 30 + 4815) = Beta(91, 4845)$

We can now summarize the posterior distribution with a posterior confidence interval and the mean, the same way we did with the non-informative prior.

```
  theta=seq(0,0.05,length=1000)
 post.density1=dbeta(theta,91,4845)
 plot(theta,post.density1, type="l",main="posterior prob of female Beta(91,4845) prior")
 post.interval1=qbeta(c(0.025,0.975), 91,4845)
post.interval1  # to see the posterior interval
 abline(v=post.interval1)

 post.mean1=91/(91+4845)   # the posterior mean
 post.mean1   # to see the mean
```

4.4.3 Comparing results with the non-informative prior and the informative prior

With the uninformative, uniform prior, we have obtained a posterior mean of 0.01854494 and a posterior 95% interval $(0.01495989, 0.02250143)$. With the more informative prior, we have obtained a mean of 0.01843 and a posterior interval $(0.01487183, 0.02236948)$. There is almost no difference in the information provided by the two densities.

A plot of the the two posterior density functions, gives almost identical distributions. It is hard to tell one from the other. To obtain the two plots together, we use the following commands

```
theta=seq(0,0.05,length=1000)
post.density=dbeta(theta,91,4816)
post.density1=dbeta(theta,91,4845)

plot(theta,post.density,type="l",main="posterior prob of female vict-U(0,1) and Beta(91,4845) prior")
lines(theta,post.density1,col="red")
```

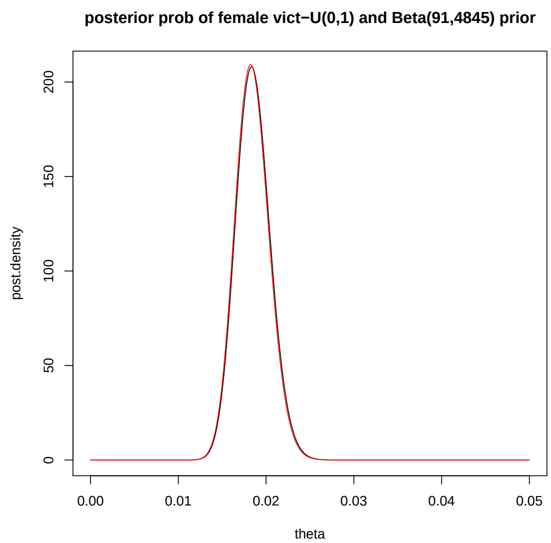

Figure 4.1: Posterior distribution of proportion of victimized women in 1994-black curve is when using a Beta(1,1) noninformative prior; red curve is with prior Beta(1,30) prior {fig:vict1}

Our question though was whether there has been a decrease in victimization.

With the more informative prior, the prior mean was $1/31 = 0.0322$ and the 95% prior interval is (0.0008435710.115703308). The posterior mean then was 0.0184 and the 95%. interval is $(0.01487183, 0.02236948)$Thus, we would reach the conclusion that victimization has decreased. We are less uncertain about the value of p after observing the data. The data has had a big impact on our prior beliefs.

4.5 Mixtures of Binomial Distributions as likelihood

4.5.1 Tossing coins

Suppose that there is a Beta(4,4) prior distribution on the probability θ that a coin will yield a "head" when spun in a specified manner. The coin is independently spun 10 times and heads appear less than 3 times. You are not told how many heads were seen, only that the number is less than 3 times. Calculate your exact posterior, posterior mean, variance, and 95% posterior interval.

Solution

Prior: $p(\theta) \propto \theta^3(1-\theta)^3$

Likelihood: $p(D \mid \theta) = \binom{10}{0}\theta^0(1-\theta)^{10} + \binom{10}{1}\theta(1-\theta)^9 + \binom{10}{2}\theta^2(1-\theta)^8$

Then the exact posterior density (up to a proportionality constant) for θ is $p(\theta \mid D) \propto \theta^3(1-\theta)^{13} + 10\theta^4(1-\theta)^{12} + 45\theta^5(1-\theta)^1 1$

In order to calculate the posterior mean, variance and 95% interval, we adopt the simulation-based inference approach. By realizing that the posterior results from a *mixture* of three Beta distributions, we have to find the weights, π_1, π_2 and π_3 such that

$$p(\theta \mid D) = \pi_1 \text{ Beta}(4,14) + \pi_2 \text{ Beta}(5,13) + \pi_3 \text{ Beta}(6,12) \propto \theta^3(1-\theta)^{13} + 10\theta^4(1-\theta)^{12} + 45\theta^5(1-\theta)^{11}$$

Since

$$\text{Beta}(\alpha,\beta) = c\theta^{(\alpha-1)}(1-\theta)^{(\beta-1)}$$

with

$$c = \frac{1}{\text{Beta}(\alpha, \beta)}$$

is the normalizing constant,

$\pi_1 = \frac{1/c_1}{\sum \pi_i}$, $\pi_2 = \frac{1/c_2}{\sum \pi_i}$, $\pi_3 = \frac{1/c_3}{\sum \pi_i}$ are the weights we look for.

In fact,

$$\pi_1 \text{ Beta}(4, 14) + \pi_2 \text{ Beta}(5, 13) + \pi_3 \text{ Beta}(6, 12) = \frac{1}{(\sum \pi_i)(\theta^3(1-\theta)^{13} + 10\theta^4(1-\theta)^{12} + 45\theta^5(1-\theta)^{11}})$$

To draw $p(\theta \mid y)$ and calculate the posterior mean, variance and 95% confidence interval, give the following commands in R. Notice that the mathematical beta function $\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

```
pi.sum = sum(1*beta(4,14)+10*beta(5,13)+45*beta(6,12)) # these betas are the math betas

########### Find pi1, pi2, pi3 #############

pi.1= 1*beta(4,14)/pi.sum     ; pi.1   # first we create object pi.1, then we view its value
pi.2 = 10*beta(5,13)/pi.sum   ; pi.2    # etc...
pi.3 = 45*beta(6,12)/pi.sum    ; pi.3

#############Simulated Summary of the Posterior ####################

############ Draw random numbers from components of posterior mixture ######
########  This will help obtain a simulated solution ###################

m1=rbeta(10156,4,14)   # draw random numbers from beta densities (4,14)
m2=rbeta(31250,5,13)   #   etc.
m3=rbeta(58594,6,12)

######## Compute approximate posterior mean, variance and interval ###############
##### We pool together all the random numbers with the R function c() #######

approx.mean=mean(c(m1,m2,m3)) ; approx.mean  #Take the mean of all random numbers
post.var=var(c(m1,m2,m3))  ; post.var  # Take variance of all random numbers
post.interval = quantile( c(m1,m2,m3),c(.025,.975)) ; post.interval  # compute posterior 95 percent int

############  Some exact summaries of the Beta posterior #################

var.beta=function(a,b) {
(a*b)/((a+b)^2 *(a+b+1))
}
exact.mean= pi.1*4/18+pi.2*5/18+pi.3*6/18  # this is the exact mean from exact posterior Beta density
exact.var=(pi.1)^2 *var.beta(4,14)+ (pi.2)^2 * var.beta(5,13) + (pi.3^2)*var.beta(6,12)


################  Plot the exact posterior ###########################

 theta=seq(0,1,.001)
 post.dens=theta^3*(1-theta)^(13)+10*theta^4*(1-theta)^(12)+45*theta^5*(1-theta)^(11)
 plot(theta, post.dens,ylim=c(0,max(post.dens)),type="l",xlab="theta",ylab="density", xaxs="i",yaxs="i"
 abline(v=approx.mean,lty=2,lwd=2)
```

```
abline(v=post.interval,lty=3)
```

4.6 Additional required reading

Section 3.1 pf Hoff's book