# Chapter 3

# Bayes Inference for proportions (I)

## 3.1 Introduction

Binary data is categorical data such that each observation is either 1 or 0, with 1 meaning that the observation has a characteristic that interests us and 0 it does not. For example, in a survey we may ask individuals whether they know how to write or not. The possible responses are YES or NO. The unknown of interest is the proportion of the population that knows how to write, $p$. We call this a simple proportion.

Simple proportions are very often the object of statistical inference. For example, proportion of patients that respond to a treatment, proportion of people in the population with AIDS, proportion of adults in the USA that approve the last bill passed by Congress. Surveys and polls are often used to collect data that allows us to estimate these simple proportions. They are interpreted as the probabilities that an observation satisfies the characteristic of interest in the investigation. In the Bayesian context, the estimation of a single proportion $p$ is often done with a Beta prior and a product of Bernoullis likelihood or a Binomial likelihood function and a Beta prior. This is in contrast with the case we saw in Chapter 1, discrete prior and product of Bernoullis likelihood.

We continue in this chapter with the sleep problem of Chapter 1.

## 3.2 The Beta Density

Since the proportion is a continuous parameter, an alternative approach to a discrete prior like that seen in an earlier chapter is to construct a prior density $g(p)$ on the interval $(0, 1)$ that represents the person's initial beliefs. Suppose the investigator believes that the proportion is equally likely to be smaller or larger than 0.3. Moreover, she is 90% confident that $p$ is less than 0.5. A convenient family of densities for a proportion is the Beta density (see the handout with all the distributions that I gave you in class).

$$g(p) = \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1} \qquad 0 < p < 1 \tag{3.1}$$

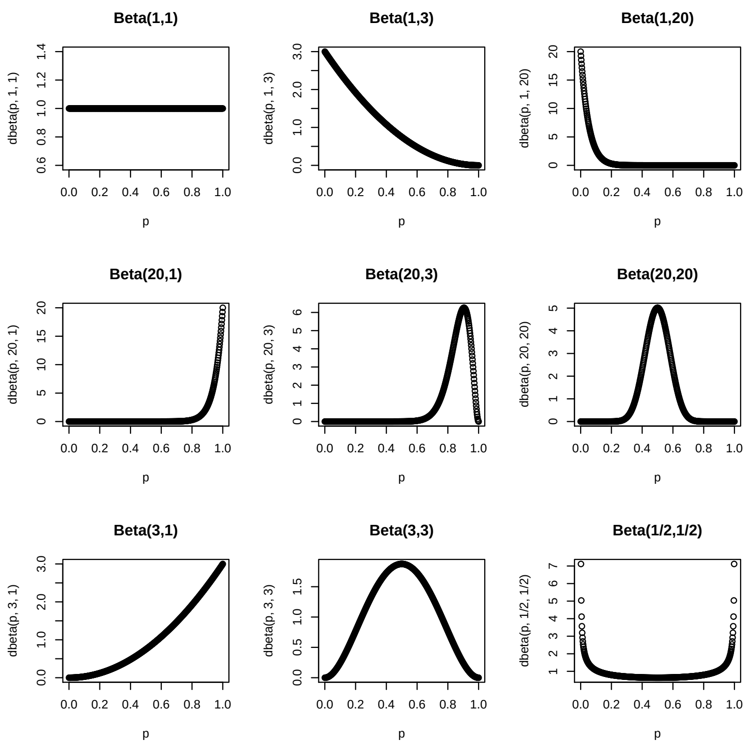
where $a$ and $b$ are positive numbers and $B(a, b) = \int_0^1 p^{a-1}(a-p)^{b-1}dp = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

It can be proved that

$$E(p) = \frac{a}{a+b} \qquad\qquad Var(X) = \frac{ab}{(a+b)^2(a+b+1)} \tag{3.2}$$

The parameters of the prior distribution are often referred to as hyperparameters. The Beta prior distribution is indexed by two *hyperparameters*, which means we can specify a particular prior distribution by fixing two features of the distribution, for example its mean and variance. For now,assume that we can select reasonable values of $a$ and $b$. Appropriate methods for working with unkwnown hyperparameters in certain problems are described later in the course.

The curves in Figure 1 show several *Beta*$(a, b)$ densities. The horizontal axis is p, the population proportion. We will call $p$ the model. For larger $a, b$ the density becomes narrower. For $a$ larger than $b$, more of the probability is concentrated on models to the right of 0.5.; for $b$ larger than $a$ the converse holds.




{fig:betas} Figure 3.1: Examples of Beta($a, b$ densities

These curves in Figure **??** were obtained with the following R code:

```
 p=seq(0,1,length=500)
par(mfrow=c(3,3))
plot(p,dbeta(p,1,1), main="Beta(1,1)")
 plot(p,dbeta(p,1,3), main="Beta(1,3)")
 plot(p,dbeta(p,1,20), main="Beta(1,20)")
 plot(p,dbeta(p,20,1), main="Beta(20,1)")
 plot(p,dbeta(p,20,3), main="Beta(20,3)")
 plot(p,dbeta(p,20,20), main="Beta(20,20)")
 plot(p,dbeta(p,3,1), main="Beta(3,1)")
 plot(p,dbeta(p,3,3), main="Beta(3,3)")
 plot(p,dbeta(p,1/2,1/2), main="Beta(1/2,1/2)")
```

Beta densities are particular types of prior (and posterior) distributions. There is a different density for each pair of $a, b$. An important special case is Beta(1,1). When $a = b = 1$, the Beta density is a constant over all $p$ values between 0 and 1. It represents a type of indifference over these values. Any interval of $p$ values between 0 and 1 has the same probability as any other interval of the same width.

When $a, b$ are large, the Beta density approaches a Normal density.

In Bayesian analysis, we usually say that the kernel of the Beta density is proportional to

$$p^{a-1}(1-p)^{b-1} \qquad 0 < p < 1 \tag{3.3}$$

In our example, we believe that the median and 90th percentiles are given, respectively, by .3 and .5, and this can be matched, by trial and error, with a beta density with $a = 3.4$ and $b = 7.4$.

## 3.3 Updating rules for Beta Densities

Suppose you observe $s$ successes and $f$ failures in the binary data you collected. How can we find the posterior probability distribution? As you very well know, answering this question requires assessing your prior probabilities of $p$ and using Bayes' rule to find the posterior probabilities. Bayes' rule says to multiply prior probabilities (or densities) by likelihoods. When your prior density is the $Beta(a, b)$, there is a very convenient updating formula.

As seen above, the mathematical form of the Beta density is $p^{a-1}(1-p)^{b-1}$

The likelihood is the probability of observing $s$ successes and $f$ failures, assuming $p$. In calculating the likelihood, whenever a success appears, multiply by $p$, and whenever a failure occurs, multiply by $1 - p$. This assumes that the individual observations are independent, assuming $p$. So the likelihood has $s$ factors of $p$ and $f$ factors of $1 - p$:

$$likelihood = (p)(p)\ldots\ldots\ldots(p)(1-p)(1-p)\ldots\ldots\ldots(1-p) = p^s(1-p)^f \tag{3.4}$$

Bayes's rule says to multiply the prior probabilities [ in this case multiply the density $p^{a-1}(1-p)^{b-1}$ by the likelihood $p^s(1-p)^f$ ] to obtain the posterior probabilities for p, given by

$$g(p \mid D) \propto p^{a-1}(1-p)^{b-1}p^s(a-p)^f = p^{a+s-1}(1-p)^{b+f-1} \tag{3.5}$$

The right-hand side is the updated beta density, which is a Beta$(a + s, b + f\ )$. So when your prior is a Beta density, you can use this updating property to find your posterior:

Thus, to summarize: combining a beta prior with the likelihood function, one can show that the posterior density is also of the beta form with updated parameters $a + s$ and $b + f$. (This is an example of a conjugate analysis where the prior and posterior densities have the same functional form.) Thus the posterior mean of $p$, which may be interpreted as the posterior probability of success for a future draw from the population, is now

$$E(p) = \frac{a+s}{a+s+b+f} \tag{3.6}$$

which always lies between the sample proportion $\frac{s}{n}$ and the prior mean $\frac{a}{a+b}$.

Since the prior, likelihood and posterior are all in the beta family, we can use the `R` command `dbeta` to compute values of prior, likelihood and posterior. These three densities are displayed using the `R` commands `plot` and `lines` in the same graph in Figure **??** This figure is helpful in seeing that the posterior density in this case compromises between the initial prior beliefs and the information in the data.

```
p=seq(0,1,length=500)
a=3.4
b= 7.4
s= 11
f = 16
prior=dbeta(p,a,b)
like=dbeta(p, s+1, f+1)
post=dbeta(p, a+s, b+f)
plot(p, post, type="l", ylab="Density", lty=2, lwd=3)
lines(p, like, lty=1, lwd=3)
lines(p, prior, lty=3, lwd=3)
legend(.7, 4, c("Prior", "Likelihood", "Posterior"), lty=c(3,1,2), lwd=c(3,3,3))
```

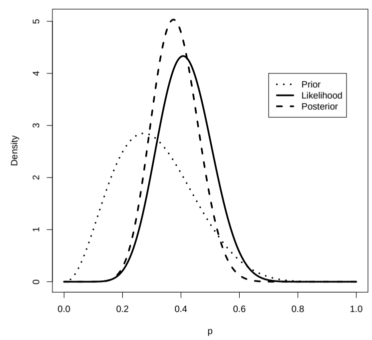



Figure 3.2: A beta prior density $g(p)$, the likelihood function $L(D \mid p)$, and the posterior density $g(p \mid data)$ for learning about a proportion $p$. {fig:sleep3}

## 3.4 Summarizing the Posterior distribution using the mathematical form of the posterior density

We will use the example we are covering in this chapter to illustrate different ways of summarizing the beta posterior distribution to make inferences about the proportion $p$.

The beta cdf and inverse cdf functions `pbeta` and `qbeta` are useful in computing probabilities and constructing interval estimates for $p$. Is it likely that $p$ is larger than 0.5? This is answered by computing the posterior probability $P(p > 0.5|D)$, which is given by the R command

```
1 - pbeta(0.5, a+s, b+f)
```

This probability is small, 0.0684257, so it is unlikely that more than half of the population have the characteristic.

A 90 percent interval estimate for $p$ is found by computing the 5th and 95th percentiles of the beta density

```
qbeta(c(0.05,0.95), a+s, b+f)
```

We are confident that the proportion of interest is between 0.2562364 and 0.5129274.

These summaries are exact because they are based on `R` functions for the beta posterior density.

## 3.5 Summarizing the Posterior distribution using basic simulation from the known posterior distribution

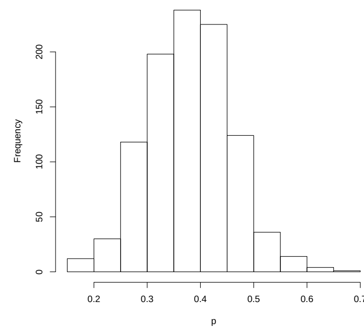
An alternative method of summarization of a posterior density is based on simulation. In this case, we can simulate a large number of values from the beta posterior density and summarize the simulated output. Using the random beta command `rbeta`, we simulate 1000 random proportion values from the $beta(a + s, b + f)$ posterior by the command

```
ps = rbeta(1000, a+s, b+f)
```

and display the posterior as a histogram of the simulated values in Fig..

```
hist(ps, xlab="p", main=" " )
```

The probability that the proportion is larger than 0.5 is estimated by the proportion of simulated values in this range, using the following command.




{fig:sleep4} Figure 3.3: Simulated posterior distribution, g(p —D).

```
sum(ps>=0.5)/1000
```

This gave me 0.055, but may give you something slightly different because it is based on random numbers and each of us may get different random numbers.

A 90 percent interval estimate can be estimated by the 5th and 95th sample quantiles of the simulated sample using the following commands:

```
quantile(ps, c(0.05, 0.95))
```

This gave me 0.2537476 and 0.5020906. You probably will get slightly different numbers.

Note that these summaries of the posterior density for $p$ based on simulation are approximately equal to the exact values based on calculations from the beta distribution.

## 3.6 Predictive probability

A pleasant characteristic of Beta densities generally is that predictive probabilities are given by a simple formula:

Predictive probability of success = $\frac{a}{a+b}$ Notice that this is the mean of the density of population models.

Thus the posterior predictive probability of succes is

$$\frac{a+x}{a+b+s+f}$$

Suppose instead of the next observation, one is interested in the next two or more observations. How many successes will there be?

## 3.7 Required additional reading

Section 3.1 in Hoff's book.