ST 740: Bayesian Basics

Alyson Wilson

Department of Statistics
North Carolina State University

August 26, 2013

Bayes Theorem

A good way to remember Bayes Theorem:

The posterior is proportional to the prior times the likelihood.

Jessica Utts Video

Be prepared to discuss in class on Monday.

- Describe experiment, data, likelihood
- What priors were used?
- What impact did the prior have on the analysis?

Example

Calculating the Posterior Distribution

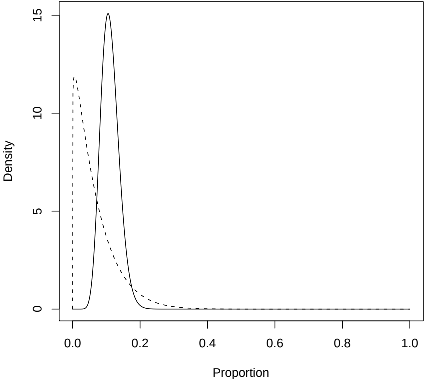
$$\begin{aligned} Y &\sim \text{Binomial}(n = 120, p) \\ p &\sim \text{Beta}(1.08, 14.96) \\ f(y \mid p) &\propto p^{14}(1-p)^{106} \\ \pi(p) &\propto p^{1.08-1}(1-p)^{14.96-1} \end{aligned}$$

$$\begin{aligned} \pi(p \mid y) &\propto f(y \mid p)\pi(p) \\ &\propto p^{14}(1-p)^{106} p^{1.08-1}(1-p)^{14.96-1} \\ &\propto p^{15.08-1}(1-p)^{120.96-1} \end{aligned}$$

We recognize that the posterior distribution for p is Beta(15.08, 120.96).

Posterior Inference

How do we summarize the posterior distribution?



Remember Intro Stats: posterior mean, posterior median, posterior standard deviation, posterior quantiles

Posterior Inference

$$
\begin{aligned}
\mathrm{E}[p \mid y] &= \frac{\alpha}{\alpha+\beta} \\
\mathrm{Var}[p \mid y] &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \\
\mathrm{Mode}[p \mid y] &= \frac{\alpha-1}{\alpha+\beta-2}
\end{aligned}
$$

where α and β are the parameters of the posterior beta distribution. You can look up quantiles of the beta distribution using the R function *qbeta*.

Posterior Inference

Intervals

We can also calculate intervals using the posterior distribution. Because we are treating the parameter as a random variable, we can make probability statements about the intervals. In particular, "Given our data, there is a 95% chance that the parameter is in this interval."

- Central credible interval: Find an interval with equal probability in each tail. For a 90% interval, use the 5th and 95th quantiles of the posterior distribution.
- Highest posterior density (HPD) interval: Find the shortest interval that contains the specified amount of the posterior distribution.

For this example the 95% central credible interval is (0.064, 0.169), and the 95% HPD interval is (0.061, 0.164).

Why Bayesian?

- Philosophy
- Pragmatism

- "After a point Bayesians decided to stop arguing so much about the philosophy of statistics and to get on with the process of doing coherent statistics." D. V. Lindley (1999)
- You will find that in many of the things that you read about Bayesian statistics, the authors spend a lot of effort justifying why the Bayesian approach is correct and what the classical approach is flawed. There is a long and interesting history behind this.

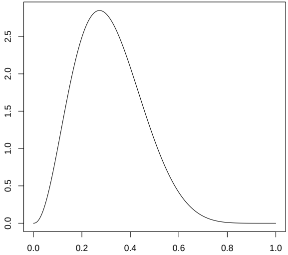
Example

We are interested in learning about the sleeping habits of American college students. In particular, what proportion of college students get at least eight hours of sleep on a typical weeknight?

Prior Information

We believe that

- it is equally likely that the proportion of students getting at least eight hours of on a typical weeknight is greater than or less than 0.3
- there is only a 10% chance that the proportion is greater than 0.5



$\text{Beta}(3.4, 7.4)$

`parametersolver(c(0.3,0.5),c(0.5,0.9),c(1,1))`

Data

A sample of 27 students is taken, and 11 record that they had at least eight hours of sleep on the previous night. We model this data as Binomial($n = 27, p$).

Example

Bayes Theorem

We apply *Bayes Theorem* to use the data we've collected to update our "pre-data" information about the parameter to "posterior" information about the parameter.

$$\pi(\theta \mid \mathbf{y}) = \frac{f(\mathbf{y} \mid \theta)\pi(\theta)}{f(\mathbf{y})}$$

where

- $\pi(\theta \mid \mathbf{y})$ is the posterior distribution for the parameter (what we know about the parameter after combining our prior knowledge and the data)
- $f(\mathbf{y} \mid \theta)$ is the likelihood function (remember that the posterior distribution is a function of θ)
- $\pi(\theta)$ is the prior distribution for the parameter
- $f(\mathbf{y})$ is the marginal distribution for the data

Example

Bayes Theorem

We apply *Bayes Theorem* to use the data we've collected to update our "pre-data" information about the parameter to "posterior" information about the parameter.

$$\pi(\theta \mid \mathbf{y}) = \frac{f(\mathbf{y} \mid \theta)\pi(\theta)}{f(\mathbf{y})}$$

where

- $\pi(\theta \mid \mathbf{y})$ is the posterior distribution for the parameter (what we know about the parameter after combining our prior knowledge and the data)
- $f(\mathbf{y} \mid \theta)$ is the likelihood function (remember that the posterior distribution is a function of θ)
- $\pi(\theta)$ is the prior distribution for the parameter
- $f(\mathbf{y})$ is the marginal distribution for the data

Example

Bayes Theorem

We apply *Bayes Theorem* to use the data we've collected to update our "pre-data" information about the parameter to "posterior" information about the parameter.

$$\pi(\theta \mid \mathbf{y}) = \frac{f(\mathbf{y} \mid \theta)\pi(\theta)}{f(\mathbf{y})}$$

where

- $\pi(\theta \mid \mathbf{y})$ is the posterior distribution for the parameter (what we know about the parameter after combining our prior knowledge and the data)
- $f(\mathbf{y} \mid \theta)$ is the likelihood function (remember that the posterior distribution is a function of θ)
- $\pi(\theta)$ is the prior distribution for the parameter
- $f(\mathbf{y})$ is the marginal distribution for the data

Example

Bayes Theorem

We apply *Bayes Theorem* to use the data we've collected to update our "pre-data" information about the parameter to "posterior" information about the parameter.

$$\pi(\theta \mid \mathbf{y}) = \frac{f(\mathbf{y} \mid \theta)\pi(\theta)}{f(\mathbf{y})}$$

where

- $\pi(\theta \mid \mathbf{y})$ is the posterior distribution for the parameter (what we know about the parameter after combining our prior knowledge and the data)
- $f(\mathbf{y} \mid \theta)$ is the likelihood function (remember that the posterior distribution is a function of θ)
- $\pi(\theta)$ is the prior distribution for the parameter
- $f(\mathbf{y})$ is the marginal distribution for the data

Example

Bayes Theorem

We apply *Bayes Theorem* to use the data we've collected to update our "pre-data" information about the parameter to "posterior" information about the parameter.

$$\pi(\theta \mid \mathbf{y}) = \frac{f(\mathbf{y} \mid \theta)\pi(\theta)}{f(\mathbf{y})}$$

where

- $\pi(\theta \mid \mathbf{y})$ is the posterior distribution for the parameter (what we know about the parameter after combining our prior knowledge and the data)
- $f(\mathbf{y} \mid \theta)$ is the likelihood function (remember that the posterior distribution is a function of θ)
- $\pi(\theta)$ is the prior distribution for the parameter
- $f(\mathbf{y})$ is the marginal distribution for the data

Example

Calculating the Posterior Distribution

$$\begin{aligned} Y &\sim \text{Binomial}(n = 27, p) \\ p &\sim \text{Beta}(3.4, 7.4) \\ f(y \mid p) &\propto p^{11}(1-p)^{16} \\ \pi(p) &\propto p^{3.4-1}(1-p)^{7.4-1} \end{aligned}$$

$$\begin{aligned} \pi(p \mid y) &\propto f(y \mid p)\pi(p) \\ &\propto p^{11}(1-p)^{16}p^{3.4-1}(1-p)^{7.4-1} \\ &\propto p^{14.4-1}(1-p)^{23.4-1} \end{aligned}$$

We recognize that the posterior distribution for p is Beta(14.4, 23.4).

Example

Calculating the Posterior Distribution

$$\begin{aligned} Y &\sim \text{Binomial}(n = 27, p) \\ p &\sim \text{Beta}(3.4, 7.4) \\ f(y \mid p) &\propto p^{11}(1-p)^{16} \\ \pi(p) &\propto p^{3.4-1}(1-p)^{7.4-1} \end{aligned}$$

$$\begin{aligned} \pi(p \mid y) &\propto f(y \mid p)\pi(p) \\ &\propto p^{11}(1-p)^{16} p^{3.4-1}(1-p)^{7.4-1} \\ &\propto p^{14.4-1}(1-p)^{23.4-1} \end{aligned}$$

We recognize that the posterior distribution for p is Beta(14.4, 23.4).

Example

Calculating the Posterior Distribution

$$
\begin{aligned}
Y &\sim \text{Binomial}(n = 27, p) \\
p &\sim \text{Beta}(3.4, 7.4) \\
f(y \mid p) &\propto p^{11}(1-p)^{16} \\
\pi(p) &\propto p^{3.4-1}(1-p)^{7.4-1}
\end{aligned}
$$

$$
\begin{aligned}
\pi(p \mid y) &\propto f(y \mid p)\pi(p) \\
&\propto p^{11}(1-p)^{16} p^{3.4-1}(1-p)^{7.4-1} \\
&\propto p^{14.4-1}(1-p)^{23.4-1}
\end{aligned}
$$

We recognize that the posterior distribution for p is Beta(14.4, 23.4).

Highest Posterior Density Intervals
A Brief Diversion

Let Θ be the support of θ. (The support of a distribution is the smallest closed interval/set whose complement has probability zero. Think "places where the parameter has positive probability.")

The 100(1-α)% HPD interval for θ is the subset C of Θ of the form

$$C = \{\theta \in \Theta : \pi(\theta \mid \mathbf{y}) \geq k(\alpha)\}$$

where $k(\alpha)$ is the largest constant such that

$$P(C \mid x) \geq 1 - \alpha.$$

In words, the density for every point in the interval is greater than for every point outside the interval.

Package `TeachingDemos`, functions `hpd` and `emp.hpd`.

```
hpd(qbeta,shape1=14.4,shape2=23.4)
```

Example

Predictive Distributions

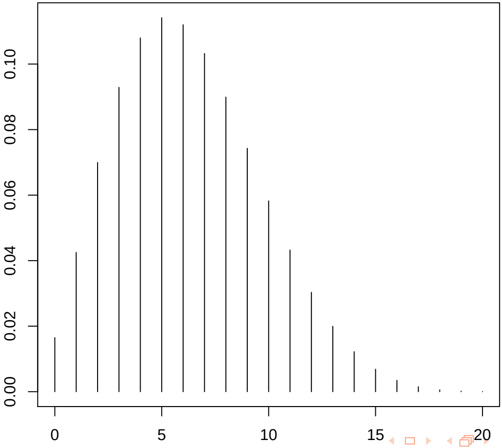
Suppose that after we assess our prior but before we talk to any students, we want to make a prediction about how many of the next 20 students we talk to will have gotten eight hours of sleep on the previous night. Our *prior predictive distribution* is

$$
\begin{aligned}
f(\tilde{y}) &= \int_0^1 f(\tilde{y} \mid p)\pi(p)dp \\
&= \int_0^1 \binom{20}{\tilde{y}} p^{\tilde{y}}(1-p)^{20-\tilde{y}} \frac{p^{3.4-1}(1-p)^{7.4-1}dp}{\mathrm{B}(3.4, 7.4)} \\
&= \frac{\binom{20}{\tilde{y}}}{\mathrm{B}(3.4, 7.4)} \int_0^1 p^{\tilde{y}+3.4-1}(1-p)^{20-\tilde{y}+7.4-1}dp \\
&= \binom{20}{\tilde{y}} \frac{\mathrm{B}(\tilde{y}+3.4, 20-\tilde{y}+7.4)}{\mathrm{B}(3.4, 7.4)}
\end{aligned}
$$

Example

Predictive Distributions

This is our prediction, accounting for our prior uncertainty, of the number of students who will have gotten eight hours of sleep.



Example

Predictive Distributions

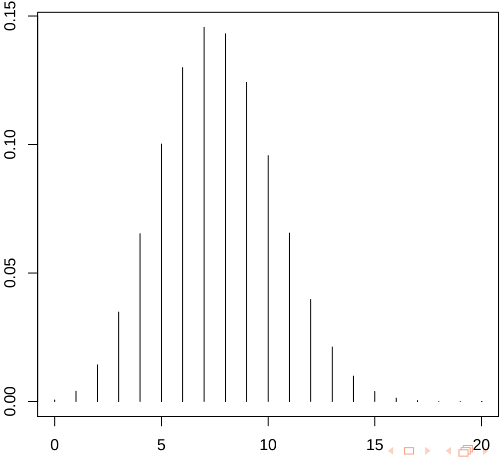
After we observed the data, our *posterior predictive distribution* about how many of the next 20 students we talk to will have gotten eight hours of sleep on the previous night is

$$
\begin{aligned}
f(\tilde{y} \mid y) &= \int_0^1 f(\tilde{y} \mid p)\pi(p \mid y)dp \\
&= \int_0^1 \binom{20}{\tilde{y}} p^{\tilde{y}}(1-p)^{20-\tilde{y}} \frac{p^{14.4-1}(1-p)^{23.4-1}dp}{B(14.4, 23.4)} \\
&= \frac{\binom{20}{\tilde{y}}}{B(14.4, 23.4)} \int_0^1 p^{\tilde{y}+14.4-1}(1-p)^{20-\tilde{y}+23.4-1}dp \\
&= \binom{20}{\tilde{y}} \frac{B(\tilde{y}+14.4, 20-\tilde{y}+23.4)}{B(14.4, 23.4)}
\end{aligned}
$$

Example

Predictive Distributions

This is our prediction, accounting for our posterior uncertainty, of the number of students who will have gotten eight hours of sleep.



Example

Predictive Distributions

We predict that there is (about) a 5% chance that we will see 4 or fewer students getting eight hours of sleep, a 50% chance that we will see 8 or fewer students, a 90% chance that we will see 12 or fewer.

Posterior Distribution

Suppose we talk to five more students. What is the distribution of the expected number of students who get at least eight hours of sleep per night?

This is a question about a function of the parameter p. We want to know the distribution of $\phi = 5p$, which is easy to calculate.