Intro to Bayesian Methods: Part II

Rebecca C. Steorts
Bayesian Methods and Modern Statistics: STA 360/601

Lecture 2

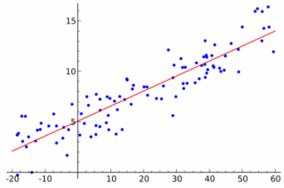
Last Time

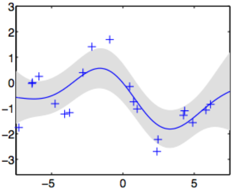
- Why should we learn about Bayesian concepts?
- Natural if thinking about unknown parameters as random.
- They naturally give a full distribution when we perform an update.
- Today: An example about students and sleep.

Bayesian Motivation

Parameters

$$P(X|\theta) \quad = \quad \text{Probability}[\text{data}|\text{pattern}]$$





Inference idea

$$\text{data} \quad = \quad \text{underlying pattern} \quad + \quad \text{independent noise}$$

[Picture: Peter Orbanz, Columbia University]

Review of the Beta-Binomial

$$X|\theta \sim \text{Binomial}(n, \theta) \tag{1}$$
$$\theta \sim \text{Beta}(a, b) \tag{2}$$

Find $\pi(\theta \mid X)$

Review of the Beta-Binomial

$$X|\theta \sim \text{Binomial}(n, \theta) \quad (1)$$
$$\theta \sim \text{Beta}(a, b) \quad (2)$$

Find $\pi(\theta \mid X)$

$$\pi(\theta|x) \quad \propto \quad p(x|\theta)p(\theta)$$

Review of the Beta-Binomial

$$X|\theta \sim \text{Binomial}(n, \theta) \quad (1)$$
$$\theta \sim \text{Beta}(a, b) \quad (2)$$

Find $\pi(\theta \mid X)$

$$\begin{aligned} \pi(\theta|x) &\propto p(x|\theta)p(\theta) \\ &\propto \binom{n}{x}\theta^x(1-\theta)^{n-x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1} \end{aligned}$$

Review of the Beta-Binomial

$$X|\theta \sim \mathsf{Binomial}(n,\theta) \tag{1}$$
$$\theta \sim \mathsf{Beta}(a,b) \tag{2}$$

Find $\pi(\theta \mid X)$

$$\begin{aligned}
\pi(\theta|x) &\propto p(x|\theta)p(\theta)\\
&\propto \binom{n}{x}\theta^x(1-\theta)^{n-x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\\
&\propto \theta^x(1-\theta)^{n-x}\theta^{a-1}(1-\theta)^{b-1}
\end{aligned}$$

Review of the Beta-Binomial

$$
\begin{align}
X|\theta &\sim \textsf{Binomial}(n, \theta) \tag{1}\\
\theta &\sim \textsf{Beta}(a, b) \tag{2}
\end{align}
$$

Find $\pi(\theta \mid X)$

$$
\begin{aligned}
\pi(\theta|x) &\propto p(x|\theta)p(\theta) \\
&\propto \binom{n}{x}\theta^x(1-\theta)^{n-x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1} \\
&\propto \theta^x(1-\theta)^{n-x}\theta^{a-1}(1-\theta)^{b-1} \\
&\propto \theta^{x+a-1}(1-\theta)^{n-x+b-1} \implies
\end{aligned}
$$

Review of the Beta-Binomial

$$X|\theta \sim \text{Binomial}(n, \theta) \tag{1}$$
$$\theta \sim \text{Beta}(a, b) \tag{2}$$

Find $\pi(\theta \mid X)$

$$\begin{aligned}
\pi(\theta|x) &\propto p(x|\theta)p(\theta) \\
&\propto \binom{n}{x}\theta^x(1-\theta)^{n-x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1} \\
&\propto \theta^x(1-\theta)^{n-x}\theta^{a-1}(1-\theta)^{b-1} \\
&\propto \theta^{x+a-1}(1-\theta)^{n-x+b-1} \implies
\end{aligned}$$

$$\theta|x \sim \text{Beta}(x+a, n-x+b).$$

How Much Do You Sleep

We are interested in a population of American college students and the proportion of the population that sleep at least eight hours a night, which we denote by θ.

Prior Data

- *The Gamecock*, at the USC printed an internet article "College Students Don't Get Enough Sleep" (2004).
  - Most students spend six hours sleeping each night.
- 2003: University of Notre Dame's paper, *Fresh Writing*.
  - The article reported took random sample of 100 students:
  - "approximately 70% reported to receiving only five to six hours of sleep on the weekdays,
  - 28% receiving seven to eight,
  - and only 2% receiving the healthy nine hours for teenagers."

- Have a random sample of 27 students is taken from UF.
- 11 students record that they sleep at least eight hours each night.
- Based on this information, we are interested in estimating θ.

- From USC and UND, believe it's probably true that most college students get less than eight hours of sleep.
- Want our prior to assign most of the probability to values of $\theta < 0.5$.
- From the information given, we decide that our best guess for θ is 0.3, although we think it is very possible that θ could be any value in $[0, 0.5]$.

- Given this information, we believe that the median of θ is 0.3 and the 90th percentile is 0.5.
- Knowing this allows us to estimate the unknown values of a and b.
- How do we actually calculate a and b?

- Given this information, we believe that the median of θ is 0.3 and the 90th percentile is 0.5.
- Knowing this allows us to estimate the unknown values of a and b.
- How do we actually calculate a and b?

We would need to solve the following equations:

$$\int_0^{0.3} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1}\, d\theta = 0.5$$

$$\int_0^{0.5} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1}\, d\theta = 0.9$$

In non-calculus language, this means the 0.5 quantile (50th percentile) = 0.3. The 0.9 quantile (90th percentile) = 0.5.

The equations are written as percentiles above!

- We can easily solve this numerically in `R` using a numerical solver `BBsolve`.
- The documentation for this package is not great, so beware.

Since you won't have used this command before, you'll need to install the package `BB` and load the library.

Here is the code in R to find a and b.

```
## install the BBsolve package
install.packages("BB", repos="http://cran.r-project.org")
library(BB)

## using percentiles
myfn <- function(shape){
test <- pbeta(q = c(0.3, 0.5), shape1 = shape[1],
 shape2 = shape[2]) - c(0.5, 0.9)
return(test)
}
BBsolve(c(1,1), myfn)


## using quantiles
fn = function(x){qbeta(c(0.5,0.9),x[1],x[2])-c(0.3,0.5)}
BBsolve(c(1,1),fn)
```

Using our calculations from the Beta-Binomial our model is

$$
\begin{aligned}
X \mid \theta &\sim \text{Binomial}(27, \theta) \\
\theta &\sim \text{Beta}(3.3, 7.2) \\
\theta \mid x &\sim \text{Beta}(x + 3.3, 27 - x + 7.2) \\
\theta \mid 11 &\sim \text{Beta}(14.3, 23.2)
\end{aligned}
$$

```
th = seq(0,1,length=500)
a = 3.3
b = 7.2
n = 27
x = 11
prior = dbeta(th,a,b)
like = dbeta(th,x+1,n-x+1)
post = dbeta(th,x+a,n-x+b)
pdf("sleep.pdf",width=7,height=5)
plot(th,post,type="l",ylab="Density",lty=2,lwd=3,
xlab = expression(theta))
lines(th,like,lty=1,lwd=3)
lines(th,prior,lty=3,lwd=3)
legend(0.7,4,c("Prior","Likelihood","Posterior"),
lty=c(3,1,2),lwd=c(3,3,3))
dev.off()
```

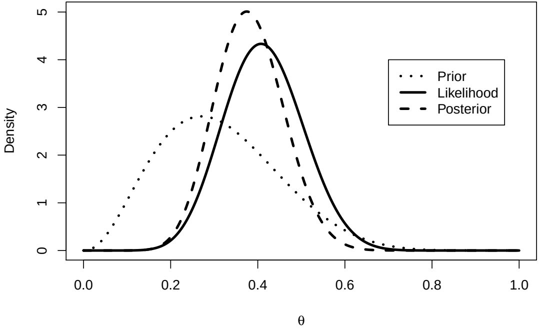


Figure 1: Likelihood $p(X|\theta)$, Prior $p(\theta)$, and Posterior Distribution $p(\theta|X)$

What questions might we want to ask?

- What is the posterior mean and variance? (Homework 1).
- Confidence intervals
- If new data comes in, how do we predict whether these students are getting less than 8 hours of sleep?

Above questions we'll cover in Module 3!

Coming up next

Make sure you're feeling very familiar with R. See the Intro to Bayes Lab (with solutions).

Module 2: Decision theory: how do we think about loss, risk (general decision making)?

Wednesday: Lab with decision theory.