

Bayesian Thinking: Fundamentals, Regression and Multilevel Modeling

Jim Albert and Monika Hu

1/1/23

General Overview

This webinar series is divided into two parts.

- Part 1 (January 9) Bayesian Fundamentals / Bayesian Regression
- Part 2 (January 11) Bayesian Regression / Multilevel Modeling

Structure

- Each part will consist of three 50-minute presentations, divided by two 10 minute breaks for questions.
- We encourage you to submit questions during the presentations.
- All presentations and R code will be posted on our Github site.

Our Backgrounds

- Jim Albert has taught a Bayesian graduate course at Bowling Green State University for many years.
- Monika Hu has taught a Bayesian class at the undergraduate level at Vassar College.
- Recently we coauthored the text *Probability and Bayesian Modeling*.
- Web version of the book available at

<http://bitly.com/ProbBayes>

All of the webinar files (markdown, pdf, data files) can be found at
<https://github.com/bayesball/JSM2020>

Any Questions?



- Can ask questions by use of the Zoom chat window.
- We will try to address all questions during the Webinar or afterwards

Computation

- Our text focuses on the use of JAGS for simulating from general Bayesian models.
- Here we are going to focus on the use of Stan which implements Hamiltonian MCMC sampling.
- Stan is well-supported and is especially efficient in fitting multilevel models.

Using Stan

- Stan modeling language
- Interfaces with popular computing environments
- Package `rstan` is the R interface to Stan
- Write a script defining the Bayesian model

Higher Level Interface

- Packages `brms` and `rstanarm` provide R formula type interfaces to Stan.
- We will illustrate the use of `brms` for the regression and multilevel modeling examples.

Example: Sleeping Patterns of Students

- Recent StatCrunch survey of high school students
- Each student asked “What is an average hours of sleep for you per night?”
- Interested in the outcome “average is at least 8 hours”
- Observed data y_1, \dots, y_n where $y_i = 1$ (student averages at least 8 hours of sleep) or $y_i = 0$

A Bayesian Model

- (Sampling) y_1, \dots, y_N are independent Bernoulli(p)
- p is the proportion of all students who average at least 8 hours of sleep
- (Prior) p is random, assign it a prior density $g(p)$
- Prior represents one's subjective beliefs about the location of p

Choice of Prior?

- Convenient to let p have a beta density

$$g(p) \propto p^{\alpha-1}(1-p)^{\beta-1}, 0 < p < 1$$

- Choose shape parameters α and β to reflect beliefs about p

Specifying a Beta Prior

- Hard to specify values of the shape parameters directly.
- Indirectly specify shape parameters by specifying quantiles of p
- Specify a median (best guess at p)
- Specify a 90th percentile (indicates sureness of your guess)
- Find values of α , β that match values of median and 90th percentile

Shiny App

- I wrote a Shiny app to help one specify a subjective beta prior for a proportion

https://bayesball.shinyapps.io/ChooseBetaPrior_3/

- Use a slider to specify two percentiles
- Graph shows the matching beta prior

Predictive density

- Bayesian model specifies joint density of (p, y) :

$$f(p, y) = g(p)f(y|p)$$

- (Prior) predictive density is marginal density of y

$$f(y) = \int g(p)f(y|p)dp$$

- This represents what one predicts in a future sample of a particular size.

Choosing a Prior

- I think relatively few students average 8 or more hours of sleep
- My best guess at p is 0.15
- Pretty sure (with probability 0.90) that p is smaller than 0.25
- This information is matched up with a beta prior with $\alpha = 4.42$ and $\beta = 23.51$

Checking My Prior

- This prior says that my 90% interval estimate is

$$P(0.062 < p < 0.283) = 0.90$$

- Suppose I think about a future sample of 50 where Y is the number of students who average 8+ hours of sleep. My prior implies

$$P(1 \leq Y \leq 14) = 0.917$$

- If these bounds don't seem right, adjust your statements about the median and 90th percentile

Constructing a Beta(a, b) Prior From Two Quantiles

Choose Median and 90th Percentile of Prior:

0 0.15 -- 0.25 1

Choose Probability Content for Middle Interval:

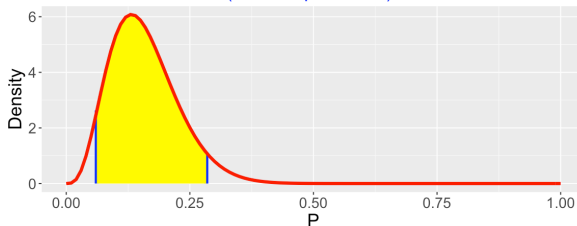
0.5 0.9

Choose Future Sample Size N:

10 50

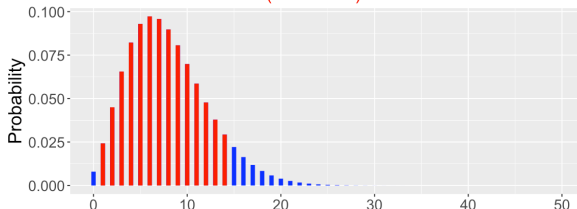
Beta(a, b) Prior with $a = 4.42$, $b = 23.51$

$\text{Prob}(0.062 < p < 0.283) = 0.9$



Predictive Distribution for Sample of Size 50

$\text{Prob}(1 \leq Y \leq 14) = 0.917$



Updating Beliefs

- Sample $N = 44$ students, observe $Y = 7$ who average more than 8 hours of sleep.
- $[Y \mid p]$ is $\text{binomial}(N, p)$
- Posterior density is product of likelihood and prior

$$g(p|y) \propto p^Y (1-p)^{N-Y} \times p^{\alpha-1} (1-p)^{\beta-1}$$

- Here we get

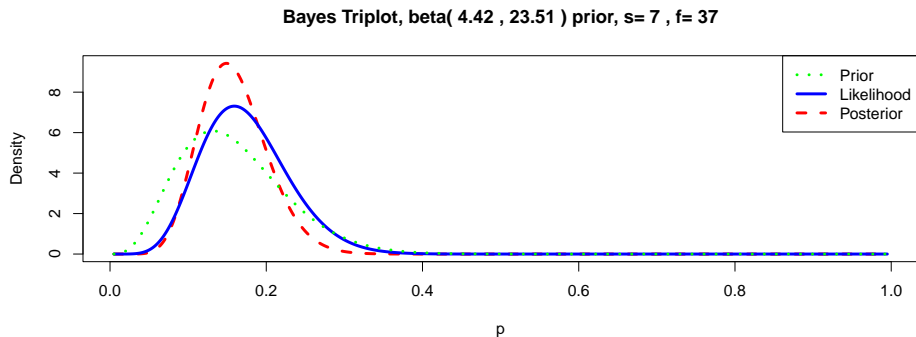
$$g(p|y) \propto p^{\alpha+Y-1} (1-p)^{\beta+N-Y-1}, 0 < p < 1$$

- Posterior is also beta with parameters $\alpha + Y$ and $\beta + N - Y$

Example

- Prior is $\text{Beta}(4.42, 23.51)$
- Observe $Y = 7$ in a sample of $N = 44$
- Posterior is $\text{Beta}(4.42 + 7, 23.51 + 37) = \text{Beta}(11.42, 60.51)$

Bayesian Triplot - Show Prior, Likelihood and Posterior

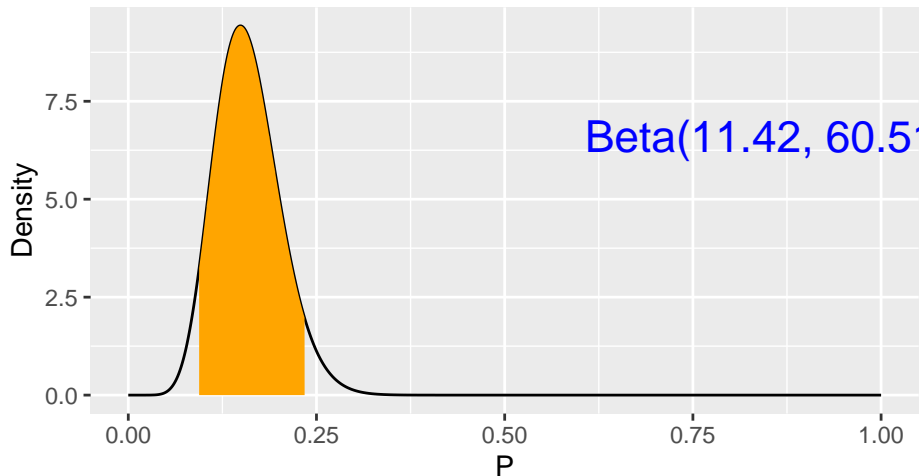


Bayesian Inferences

- All inferences about p are summaries of this posterior density
- For example, a 90% interval estimate is an interval that covers 90% of the posterior probability

90% Interval Estimate

$$P(0.094 < P < 0.234) = 0.9$$



Simulation-Based Inference

- Here we can find the exact posterior distribution
- But in most situations, we cannot, but it is possible to simulate from the posterior
- Simulate many draws from the posterior and implement inference by summarizing the simulated sample

Simulation for Our Example

- Simulate 1000 draws from the beta posterior.

```
p_sim <- rbeta(1000, 11.42, 60.51)
```

- Find 90% interval estimate by computing quantiles of the simulated draws.

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

	5%	95%
	0.08962417	0.23496112

Prediction

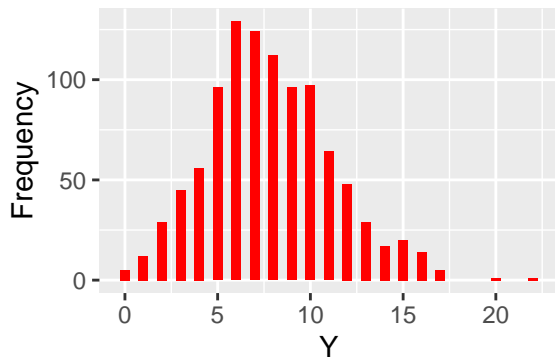
- Suppose one wishes to predict the number y^* of 8+ hours of sleep in future sample of 50 students
- Interested in posterior predictive (PP) density.

$$f(y^*|y) = \int f(y^*|p)g(p|y)dp$$

- Simulate draws from PP density by (1) simulating p from posterior and (2) simulating $y|p$ from the sampling density

```
p_sim <- rbeta(1000, 11.42, 60.51)
ys <- rbinom(1000, size = 50, prob = p_sim)
```

Prediction in Our Example



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

5%	95%
3	14

Using Stan

- Write a Stan script defining the Bayesian model.

```
data {  
  int<lower=0> N;  
  int<lower=0,upper=1> y[N];  
}  
parameters {  
  real<lower=0,upper=1> theta;  
}  
model {  
  theta ~ beta(4.42, 23.51);  
  for (i in 1:N) {  
    y[i] ~ bernoulli(theta);  
  }  
}
```

Enter data by a list

```
library(readr)
d <- read_csv("Happiness_vs_Sleep_Exercise.csv")
d$y <- ifelse(d$Exercise >= 8, 1, 0)
my_data <- list(N = 44, y = d$y)
```

Run Stan Using the rstan package

- Inputs are Stan model file and the data list.

```
library(rstan)
fit_bern <- stan(file = "bern_beta.stan",
                 data = my_data,
                 refresh = 0)
```

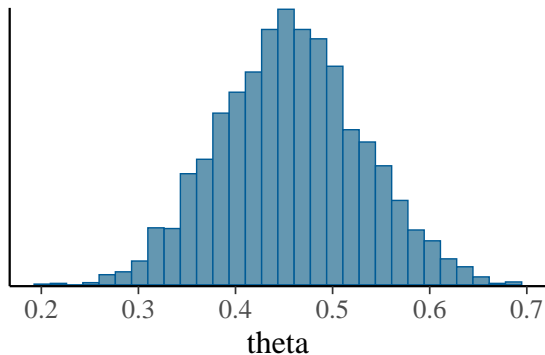
Extract the posterior draws

```
draws <- as.data.frame(fit_bern)
head(draws)
```

	theta	lp__
1	0.3782732	-32.29627
2	0.4473198	-31.71851
3	0.4563823	-31.71064
4	0.4260967	-31.79723
5	0.4466790	-31.71964
6	0.5602757	-32.70604

Histogram of the simulated draws of p

```
library(bayesplot)
mcmc_hist(draws, pars='theta')
```



Posterior summaries

```
summary(fit_bern)
```

```
$summary
```

	mean	se_mean	sd	2.5%	97.5%
theta	0.4570535	0.002025998	0.07316045	0.3166979	0.4065714
lp__	-32.2227561	0.016884303	0.71583646	-34.3140683	-32.3971406

	75%	97.5%	n_eff	Rhat
theta	0.5053301	0.6043663	1303.991	1.0006571
lp__	-31.7594573	-31.7111289	1797.469	0.9997978

```
$c_summary
```

```
, , chains = chain:1
```

```
stats
```

parameter	mean	sd	2.5%	25%	75%	97.5%
theta	0.4557443	0.07304661	0.3194328	0.4050488	0.4949488	0.6100488

Wrap-Up: Some Attractive Features of Bayes

- One recipe (Bayes' rule) for implementing inference
- Conditional inference
- Allows input of prior opinion

More Attractive Features

- Intuitive conclusions
- The probability that p is in $(0.23, 0.45)$ is 90 percent.
- If you have hypothesis $p \leq 0.7$, you can compute the probability a hypothesis is true.
- Prediction and inference: in both cases you are learning about unobserved quantities given observations

Some More Attractive Features of Bayes

- Flexibility in modeling
- Advances in Bayesian computation
- Attractive way to implement multilevel modeling
- Can handle sparse data (say, many 0's in categorical response data)

Some Issues with Bayes

- “QUESTION: What if I use the wrong prior?”
- “QUESTION: Aren't I introducing errors by simulating from the posterior?”