

MATHEMATICS E-156, SPRING 2014
MATHEMATICAL FOUNDATIONS OF STATISTICAL SOFTWARE
Module #10 (Bayesian Methods)

Last modified: April 16, 2014

Reading from Chihara and Hesterberg

- Chapter 10, but you can skip the messy formulas at the start of section 10.4 and 10.5.
- Appendix B.12 on the beta distribution. The proof of Theorem B.20 is in the math notes.

Proof of the Week

- Suppose that random variable X has a beta distribution, i.e. its probability density function is

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, 0 \leq x \leq 1$$

Prove the following:

—

$$\int_0^1 f(x) dx = 1.$$

—

$$E[X] = \frac{\alpha}{\alpha + \beta}$$

—

$$\text{Var}[X] = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

(This is done on pp. 390-391, though partially left as an exercise)

R scripts

- Script 10A-BayesDiscrete.R
 - Topic 1 - Binomial data with a discrete prior distribution
 - Topic 2 - the posterior distribution does not depend on how we incorporate the data
 - Topic 3 - there is no need to make the prior mass function sum to 1
- Script 10B-BayesBeta.R
 - Topic 1 - looking at some beta distributions
 - Topic 2 - mean and variance of a beta distribution
 - Topic 3 - using a continuous Bayesian prior to estimate a proportion
- Script 10C-BayesNormal.R
 - Topic 1 - if the samples have a normal distribution, use a normal prior
 - Topic 2 - incorporating the data sequentially leads to the same posterior
- Script 10D-BayesMultiarm.R
 - Topic 1 - trying to identify the best version of a Web site
 - Topic 2 - getting the same results using a binomial simulation

Mathematical notes

1. Bayes' theorem

- (a) State and prove Bayes' theorem for two events A and B , neither of which has probability zero.
- (b) Specialize to the case where X are some data (one or more random variables) from a discrete probability distribution that is specified by one or more parameters collectively represented by θ . The event A is “the parameter value is θ_j ”; the event B is “the observed data are X .”
- (c) State Bayes' theorem for the case where you need to determine a density function $p(\theta|x)$. Parameter θ has a continuous probability distribution with pdf $\pi(\theta)$, while the data X has a continuous distribution whose density function $f(x|\theta)$ depends on the value of the parameter. (No proof required.)

2. Proof of the week

Suppose that random variable X has a beta distribution, i.e. its probability density function is

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, 0 \leq x \leq 1$$

Prove that

$$\int_0^1 f(x) dx = 1.$$

3. Expectation and variance of the beta distribution

Knowing that $\int_0^1 \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} dx = 1$, prove that

•

$$E[X] = \frac{\alpha}{\alpha + \beta}$$

•

$$\text{Var}[X] = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

- A beta distribution with expectation μ , variance σ^2 has parameters

$$\alpha = \frac{\mu^2 - \mu^3}{\sigma^2} - \mu, \beta = \frac{\alpha}{\mu} - \alpha$$

4. Suppose that the parameter θ for a binomial distribution has the prior distribution $\theta \sim \text{Beta}(\alpha, \beta)$ and the data are given by the binomial distribution $X \sim \text{Binom}(n, \theta)$. Prove that the posterior distribution $\theta|X \sim \text{Beta}(\alpha + x, \beta + n - x)$.

Reminder:
$$\int_0^1 \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1} d\theta = 1.$$

Section problems

1. Exercises 3 and 4 on page 323. Partial answers to 3 are on page 323. Editing script 10A will solve this problem pretty quickly.
2. Exercise 5 on page 324. Answers on page 323. Note that part (a) is frequentist, not Bayesian. You can get a good approximate answer in one line by using `qbinom()`. I cannot figure out how to get exactly the same confidence interval as on page 323, even by going to page 192 (which we have not covered yet) and using `prop.test`.
3. Exercise 15 on page 326. Parts (a) and (b) are not really R problems, but after you solve them, you can have R plot graphs of the prior and posterior distributions for θ .

Homework assignment This assignment should be submitted as a single R script. Include enough comments so that it is clear what you are doing and where each problem begins. You can upload it to the dropbox on the Class 10 page of the Web site.

It is OK to paste and edit lines from the scripts on the course Web site. It is not OK to paste lines from your classmates' solutions!

1. You are playing a computer game where you can battle the AI by having a hero from your army fight one-on-one against a randomly chosen unit from the opposing army. Your probability of winning depends on what type of opposing unit is chosen, as follows:
 - If it is a dragon, your probability of winning is 0.1.
 - If it is an orc, your probability of winning is 0.3.
 - If it is a goblin, your probability of winning is 0.6.
 - If it is an elf, your probability of winning is 0.7.
 - If it is a fairy, your probability of winning is 0.8.

You know that the opposing army consists of one dragon, two orcs, three goblins, three elves, and one fairy, and for your Bayesian prior you assume that each of these ten units is equally likely to be chosen.

- (a) You fight five battles, winning two and losing three. Find the posterior distribution for the probability of selecting each type of unit.
- (b) You fight ten more battles, winning three and losing seven. What is now your estimate of the probability that the AI selects a dragon? An orc? A fairy?
- (c) Show that if you group together the result of the fifteen battles, you get the answers to (b) in a single step.

Feel free to reuse code from script 10A.

2. Page 324, Exercise 6. For part (a), see the comment on the second section problem.
3. Page 324, Exercise 8. Feel free to reuse code from script 10C.
4. Page 326, Exercise 16. This is another cute conjugate-prior problem like the third section problem. Add part (f):
You estimate that the prior distribution has mean 8 and variance 4, and you assume that it is a gamma distribution because otherwise you cannot calculate the posterior density. For the same observed values as in part (d), find the posterior density. Plot the prior and posterior densities, and mark, using vertical lines of different colors, the 95% credible interval for each.

5. Among players aged 25 or less, here are the top six hitters in major league baseball for 2013:

- Freddie Freeman, 176 hits in 551 at bats.
- Paul Goldschmidt, 182 hits in 602 at bats.
- Eric Hosmer, 188 hits in 623 at bats.
- Salvador Perez, 145 hits in 496 at bats.
- Jean Segura, 173 hits in 588 at bats.
- Mike Trout, 190 hits in 589 at bats.

You are a Yankees general manager whose motto is "Great hitters are bought, not made," and you have 50 million dollars to spend on hiring one of these folks when he becomes a free agent. Using the "Bayes multi-arm" approach, identify which (if any) of these players has a probability of less than 5 percent of being the top hitter in the group and is therefore not worth considering. After doing the problem by Bayesian methods, repeat the analysis by simply simulating 10000 seasons, using the 2013 batting average as the parameter for a binomial distribution. You should reach the same conclusion.