

STA 360/602, Case Study

Rebecca C. Steorts

March 26, 2019

You should turn in your work in a well thought out and summarized paper/report giving your results as well as any plots you were asked to find. Be sure to include figures and label them appropriately in your report. Include two appendices. Include your R source code in Appendix A. In Appendix B include the work to back up the findings in your paper.

When you turn in your assignment, e-mail me your group's R code so that I can verify it runs (without errors).

1. In the evaluation of drugs for possible clinical application, studies are routinely performed on rodents. For a particular study drawn from the statistical literature, suppose the immediate aim is to estimate θ , the probability of tumor in a population of female laboratory rats of type 'F344' that receive a zero dose of the drug (control group). The data show that 4 out of 14 rats developed endometrial stromal polyps (a type of tumor). It is natural to assume a binomial model for the number of tumors.
 - (a) For convenience, we select a $\text{Beta}(a, b)$ prior. What is the posterior distribution of θ ?
 - (b) Assume that your prior distribution is $\text{Beta}(5, 6)$. Plot the prior, posterior, and likelihood in R (as a function of θ). Label your x -axis as θ and y -axis as density. Provide an appropriate title and a proper legend to identify each distribution.
 - (c) Find the posterior mean and variance and interpret the values you find in terms of the problem. Also, give a 95 percent credible interval (equal tailed) for θ and interpret your interval.

- (d) Now suppose we want to use a prior from historical data (this is what is done in practice). The historical data is given in `tumor.txt`. Using this historical information on 70 groups of rats, what is commonly done is to assume that for the sake of updating, we can use this information to estimate the prior mean and prior variance of our Beta distribution.

Calculate the prior mean and prior variance for the data provided. The experimenters do not assume that the groups (70 of them) can be combined, so the prior mean is

$$\mu = \frac{1}{N} \sum_{j=1}^N \frac{y_j}{n_j}.$$

The prior variance can be found by

$$\sigma^2 = \frac{1}{N-1} \sum_{j=1}^N \left(\frac{y_j}{n_j} - \mu \right)^2.$$

Using this information, you can easily calculate these two quantities in R using `mean` and `var` if applied properly.

- (e) Now solve for a and b using the `BBsolve` function in R. If you have trouble with this remember you can get help using `?BBsolve`. We find that the observed sample mean and variance for the 70 groups are 0.136 and 0.0107. The resulting $(a, b) = (1.4, 8.6)$. Make sure that you reach these values before continuing on in your analysis.
- (f) Using this new prior on θ and the prior from part (b), plot the two priors, two posteriors and the likelihood (all as a function of θ). Save your plot to a pdf file that opens properly. Label your x -axis as θ and y -axis as density. Provide an appropriate title and a proper legend to identify each distribution.

In creating your plots, do the following: the posteriors should be coded to have line type 1, the priors line type 2, and the likelihood line type 3. This will allow you to distinguish the three different types of densities on your graph. Also, code your plots so that your posterior and prior using the historical data are in red, your posterior and prior using the `Gamma(5,6)` are in blue, and the

likelihood is in green. Finally, be sure to incorporate all this information into your legend. This will allow you to easily visualize your plot and draw conclusions.

- (g) Interpret the results of your plot in part (g). What behavior in terms of the posterior distribution do you notice when you use the same prior (but different prior parameters)? Is this optimal?
- (h) Find the posterior mean and variance using the updated prior from part (f). Compare the posterior mean to the crude proportion, $4/14$. What do you find, and is this expected?
- (i) Also, give a 95 percent credible interval for θ . Compare this to the 95 percent credible interval you found for θ under the $\text{Beta}(5, 6)$ prior. What do you find?