Joe Brew - PHC 7000 - HW 7

Readings

Greenland, S. Bayesian perspectives for epidemiological research: I. Foundations and basic models. International Journal of Epidemiology 2006; 35:765-775.Bayesian Perspectives..Greenland.pdf

Dunson, DB. Commentary: Practical Advantages of Bayesian Analysis of Epidemiologic Data. American Journal of Epidemiology 2001; 153 (12) : 1222-1226.Advantages of Bayesian Analysis ..Dunson.pdf

Homework Questions

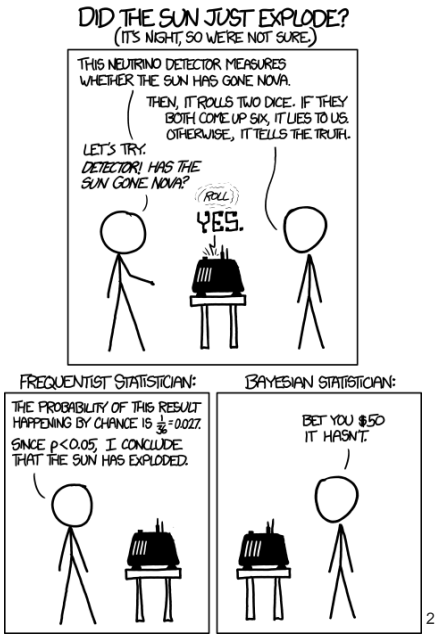
Discuss two similarities and two differences between frequentist and Bayesian.

Similiarities: Both deal with a posterior "likelihood" function (though frequentists wouldn't call it that). Both attempt to estimate an effect.

Differences: Whereas frequentists use probability only to assess the likelihood of a sample representing the population from whence it came, bayesians use probability both for that purpose as well as to model other kinds of uncertainty.[1] A second difference is that frequentist statistics assume a "truth" about which we try to estimate a statistic; bayesian statistics never quite get to 100% truth - rather, they're always on their way.

Perhaps a better explanation of the differences can be found in this comic strip:

[1] Almost directly scooped from http://www.quora.com/What-is-the-difference-between-Bayesian-and-frequentist-statisticians, which though a discussion forum, is a lot more clear about bayesian vs. frequentist statistics than most of the internet.



[2]

How is prior determined? (2-3 sentences)

A prior is a (subjective) probability distribution of effect/events *before* the beginning of our study. It does not need to actually come before the study, but it does need to be apart from the study. A prior can be determined by literature review, best guess, or some more quantitative approach, but in any case, it's the least rigorous part of bayesian statistics. (That said, a subjectively assigned prior is no worse than frequentists' blissful ignorance of prior probability).

What are some advantages and limitations of Bayesian analysis?

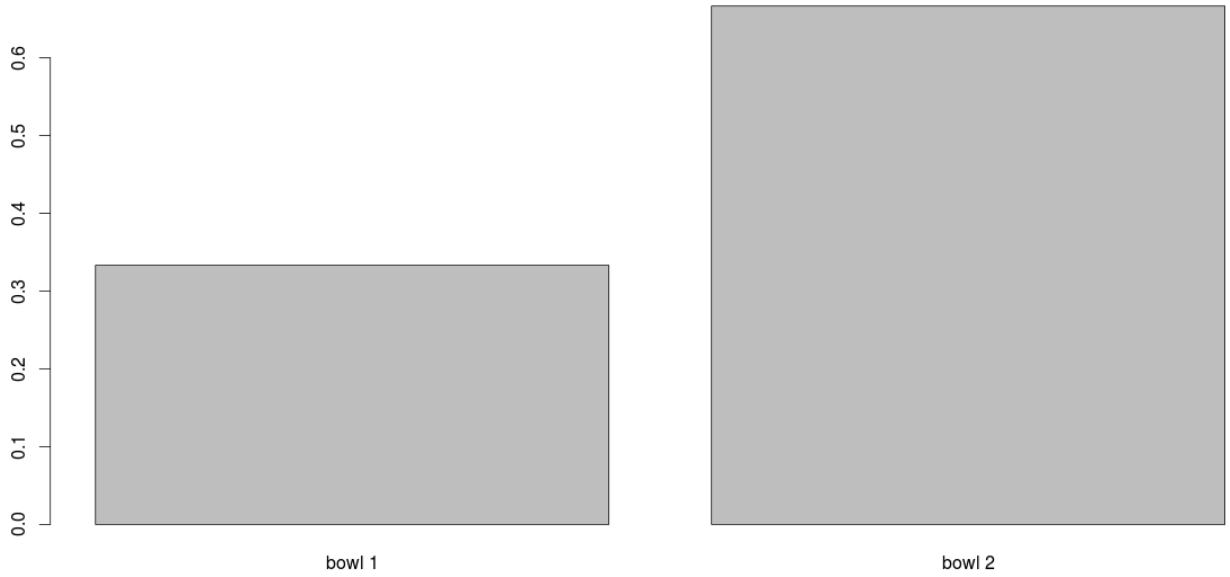
The main advantage of bayesian analysis is that it explicitly addresses the issue of likelihood/probability before a study begins, thereby reducing the likelihood of type II error (which, even by a frequentist's admission, should occur about 95% of the time). If a prior distribution is introduced clearly and on good grounding (ie, it makes sense), the results of a bayesian analysis can be intuitive and useful.

The main disadvantage is the perception of bias regarding the assignment of prior. A second disadvantage is the difficulty in transmitting results (given that the field of epidemiology, for better or for worse, is still dominated by frequentists).

[2] http://xkcd.com/1132/

Problem: Suppose there are two full bowls of round balls. Bowl 1 has 10 red balls and 10 blue balls, while Bowl 2 has 20 of each. You pick a bowl at random, and then pick a ball at random. Let's assume there is no reason to believe that you treat one bowl differently from another, likewise for the balls. The ball turns out to be a blue one. How probable is it that you picked it out of Bowl 1?

66% (see next page for process)




```
df <- data.frame(bowl = c(rep(1, 20), rep(2, 40)),
                 col = c(rep('red', 10), rep('blue', 10),
                         rep('red', 20), rep('blue', 20)))

picked_blue <- df[which(df$col == "blue"),]

tab <- table(picked_blue$bowl)
ptab <- prop.table(tab)
barplot(ptab, names.arg = paste("bowl", names(ptab)))
```