

Quant I - Problem Set III

Distributions, Confidence Intervals, and Simulation

Kenneth Benoit

Assigned: Wednesday, October 27th, 2010

Due: Wednesday, November 10th, 2010

1. Suppose you are playing the board game Risk, where attacking and defending is done by rolling dice. Your opponent attacks you, and (s)he rolls two dice returning a five and a six. You are allowed to roll two dice as well. In order to resist this attack successfully, you have to roll one six and at least a five with the other die. The probability of doing so in a single try is $\frac{1}{6} * \frac{1}{3} = \frac{1}{18}$. Assume your opponent is both mean and lucky and attacks you ten times, each time rolling a five and a six. What is the probability to successfully resist at least three times out of ten?

Please do this in R using one of the d,p,q,r-distribution functions. Interpret the result in one sentence. (4 points)

2. Verzani p. 158, Q 5.15

Please do this in R using one of the d,p,q,r-distribution functions. Interpret the result in one sentence. (4 points)

3. Verzani p. 164, Q 5.32

- (a) Please do this in R using one of the d,p,q,r-distribution functions, making use of the central limit theorem. Interpret the result in one or two sentences. Hints: a) Rather than using the information that the total number of tickets should be less than 75 over 21 days directly, think about what this information implies about the mean over the 21 days. b) You can use the condition ≤ 74.99 instead of < 75 . (6 points)

- (b) Additional question, optional for *extra credit*:
Simulate 10,000 21-days-cycles in R (any way you like) to calculate the probability asked for in the main question. (*optional 5 points*)

- (c) Additional question (**obligatory**):
In R, calculate a 95%-confidence interval for the mean of the sampling distribution, using code that reflects the single steps in constructing the confidence interval. Briefly interpret. (5 points)

4. The dataset `GBRleaders97.RData` contains two variables (for a sub-sample of $N = 500$) from a British election survey from 1997. Data come in form of a dataframe called `d`. After loading the dataset, you can see the variable names by typing `names(d)` and access them using the `d$. . .` notation. Both variables represent answers to the question “I’d like to ask you how much you like or dislike some political leaders”, where answer categories are on a scale from 0 (strongly dislike) to 10 (strongly like). The variables refer to John Major (*major*) and Tony Blair (*blair*).
- Calculate a 95% confidence interval for the median of the *major* variable. Use a bootstrap with 1000 replications. You can do this in R any way you like, but using a loop may be the easiest variant. (5 points)
 - It is interesting to know how polarized the public view of the two leaders was. For this purpose, it is helpful to measure how strongly the “fans” of a certain leader liked him compared to how strongly the opponents disliked him. One possible measure of this is the ratio of the third quartile and the first quartile ($\frac{p_{75}}{p_{25}}$) of the like-dislike variable.
 - Calculate this ratio for each of the variables (for the given sample). On which leader did the respondents in the sample hold more polarized views? (3 points)
 - Use a bootstrap with 1000 replications in order to calculate (for each leader separately) a 95% confidence interval for this ratio. Again, you can do this in R any way you like, but using a loop may be the easiest variant. (Hint: One way to do this is to first draw a sample from the data, and then calculate the quantity of interest and store it. Repeat this a 1000 times.) Do the confidence intervals overlap? Judged on the basis of this result, did polarization of public opinion with regard to the two politicians differ? (7 points)
5. On September 14th, 2010, the Irish Times reported: “Bord Bia said research carried out by it and released yesterday to coincide with the opening of National Organic Week showed 33 per cent of Irish grocery shoppers purchased an organic product in the past week. This was a 2 per cent increase on figures for 2008 when this figure was 31 per cent even though the volume of organic food sales has remained steady over the year.”

The sample sizes of these surveys were not reported. Let us for now assume that in each survey 1000 respondents were randomly selected.

- (a) Using R code that shows how you arrive at the solution, calculate confidence intervals for each of the two proportions as discussed in Verzani Section 7.2. (You can check your result with `prop.test()` if you want. The results will be the same only approximately, since `prop.test()` does not rely on the normal approximation.) Do the confidence intervals overlap? Briefly interpret what this result tells you about Irish consumer behaviour. (5 points)
- (b) Re-run your calculations for the previous question assuming that the sample sizes were 7500 each. Does the conclusion change? (2 points)

6. Verzani p. 195, Q 7.19

Construct a confidence interval using the t-distribution for the 4-card observed sample. Show your R code for each step. Hint: To answer this you need to consider the size of your sample, not of the 750 groups and 3000 criminals. (4 points)

Total: 45 points (+5 optional)