STAT 345-01: Nonparametric Statistics

Problem Set 3

Assigned 2018 September 11
Due 2018 September 18

Show your work on all problems! Be sure to give credit to any collaborators, or outside sources used in solving the problems. Note that if using an outside source to do a calculation, you should use it as a reference for the method, and actually carry out the calculation yourself; it's not sufficient to quote the results of a calculation contained in an outside source.

Please hand in parts one and two separately.

1 Part One

1.1 Conover Problems on Quantile Tests

Exercise 3.2.1

Exercise 3.2.2

Exercise 3.2.4

Problem 3.2.1

1.2 Conover Problems on The Sign Test

Exercise 3.4.2 [Give a p-value.]

Exercise 3.4.4 [Give a p-value.]

Exercise 3.4.6 [Give a p-value.]

2 Part Two: Comparison Between Quantile and t Tests

Please turn in some sort of transcript of your python session, along with answers to the questions posed. If you want to submit electronically, please send either a pdf (with a white background so it can be printed) or a plain ASCII file. No Word documents!

(a) Consider a sample of size $n = 100$ and suppose we wish to determine whether there's evidence that the population median is positive. We wish to define a quantile test with null hypothesis H_1: $x_{0.5} = 0$ and alternative hypothesis H_1: $x_{0.5} > 0$. If we want to define a test with confidence level as close as possible to 0.05, but not below, what is the rejection region? What is the actual confidence level α?

(b) Suppose the sample was drawn from a normal distribution with mean θ and standard deviation 1. Create a vector `theta` of 121 values of θ from 0 to 0.6 (using `np.linspace()`) and a corresponding vector `ptheta` containing, for each θ value, $P(X \leq 0|\theta)$ where $X \sim N(\theta, 1)$. This can be done with `stats.norm(loc=theta).cdf(0)` or `stats.norm.cdf(0,loc=theta)`. (Numpy can vectorize over the parameters of distributions.)

(c) Using the vector of probabilities from part (b), construct a vector containing the power $\gamma(\theta)$ at each θ value. This can be done by using the vectorized distribution `stats.norm(loc=theta)` and then evaluating whatever `cdf()` or `sf()` method you did in part (a) to get the actual confidence level. You'll plot θ versus $\gamma(\theta)$ a couple of steps down.

(d) Now consider a t-test of the hypthesis H_0: $\theta = 0$ versus H_1: $\theta > 0$. Use the method from problem set 2 to empirically estimate the power $\gamma(\theta)$ for the t test, using $N = 10^4$ Monte Carlo trials (since we saw last time that 10^3 trials gives a bit too much Monte Carlo noise). Note that the test in this problem is one-tailed and that one was two-tailed; don't forget to adjust your expressions as appropriate.

(e) On the same set of axes, plot the power curves $\gamma(\theta)$ obtained in part (c) for the quantile test and part (d) for the t-test. Label both of your curves (you can use the `legend()` command to make a legend box), and make sure they're different styles (e.g., by using `ls='-'` for one curve and `ls='--'` for the other) and not just different colors (so they're distinguishable when printed out in black and white). Which test is more powerful?

(f) Repeat steps (b)-(e), supposing the sampling distribution is a Laplace distribution with location parameter θ and scale parameter $1/\sqrt{2}$, as defined by `scipy.stats` (so the distribution is defined by `scipy.stats.laplace(loc=theta,scale=np.sqrt(0.5))`). Now which test is more powerful?