Practical 3: Simulations for hypothesis tests

November 9, 2011

1. Suppose you have 4 independent observations from a normal distribution with unknown mean and unknown variance. You estimate the SD from the data, and use this to compute a Student t statistic. You report it to someone, who mistakenly believes that the variance is known, and thus performs a Z test for the hypothesis that the mean of the distribution is 0. What will the impact be on the statistical conclusions? What if the error goes in the other direction (that is, you compute from a known variance, but someone believes you have delivered a T statistic)? Carry out some simulations to demonstrate the error that will be made.

2. You have a small number (2,3,5, or 10) of independent samples from an unknown distribution with expectation 1. You perform a t test at level 0.01 for the null hypothesis that the expectation is 1. However, the distribution that you are sampling from is not normal.

   (a) Simulate this experiment 1000 times with each of a number of different distributions: Poisson with parameter 1, exponential, uniform on $[0, 2]$, Pareto (with density $2(1 + x)^{-3}$ on $0 < x < \infty$), and perhaps some others. Under which circumstances does the test "work" (in the sense that the probability of rejecting a true null hypothesis is close to 0.01)?

   (b) Compare the distribution of your simulated statistic to Student t distribution using a Q-Q plot.

   (c) Try to formulate a guess about which properties of a distribution are required for the t test to be accurate, or for it to be conservative.

   (d) Take some lists of simulated data as above, from different distributions but with the same mean, and formally test the hypothesis that they have the same mean, first using the t test, then using the rank sum test. (Let each sample have size 10 or 20.) Do the same for distributions with different means. Looking at the results of 1000 experiments, which distributions are best distinguished by these different tests? (That is, which pairs of distributions give the rank sum test the most power? Which ones are indistinguishable by this test?)

3. How would you design a distribution that is not normal, but such that the χ^2 test has very little power to distinguish it from a normal distribution? What about a Poisson distribution? Test your ideas with simulations.