Hypothesis Testing Practical

Daniel Lawson

October 21, 2014

0. Comparison between distributions can be made visually using Quantile-Quantile plots. Standardize the sums as in and plot the empirical quantiles against the Norm(0, 1) quantiles.
   (a) Look at the help for normrnd,unifrnd,exprnd and poissrnd to find out how to generate data from each.
   (b) Try qqplot(data1) for plotting them against the quantiles of a Normal distribution.
   (c) Use qqplot(data1,data2) for comparing empirical quantiles of two datasets. Do this for each pair of distributions and for different parameter values within each.

1. To determine the effectiveness of a new drug on the level of haemoglobin in the blood of anemic patients, 10 randomly selected patients who underwent this treatment were sampled. The table below shows the level of haemoglobin in the patients blood before and after the treatment.

Patient	1	2	3	4	5	6	7	8	9	10
Before	11.2	9.4	9.9	9.3	8.9	8.2	10.5	8.8	10.3	9.8
After	12.9	10.8	10.3	10.9	8.5	8.9	10.4	8.5	11.2	10.1

   (a) Try to get a sense of the data by using appropriate plots to investigate their characteristics. Compare the mean and standard deviation of the level of haemoglobin before and after the treatment.
   (b) Carry out both the two-sample t test and the paired sample t test at a significance level of 0.05 to determine if the drug is effective in increasing the level of haemoglobin in the blood. Check your results with the output of ttest and ttest2 in Matlab. Given the setting of the experiment, which test is more appropriate? Why?
   (c) Perform two different non-parametric tests – one for paired samples and one for non-paired samples – and compare the results to the ones you had from the t test. Recall the use of signtest and signrank in Matlab.

2. Responding to anecdotal reports in air squadrons that military pilots father more girls than boys, [Sny61] tabulated the sex of pilots offspring for three kinds of flight duty during the month of conception. Is there any evidence of an association between fathers activity and offspring gender? Use a 5% level of significance. Check your results with the output of chisq.test() in R.

	Female offspring	Male offspring
Flying Fighters	51	38
Flying Transports	14	16
Not flying	38	46

3. You have a small number (2, 3, 5, or 10,20) 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]. 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?)

   Extra credit: Repeat this process with a multimodal distribution of two normals with sd 1 and means (0,2), with relative weights 0.25 and 0.75.

4. Consider two distributions: $X \sim x$ with mean 0 and $Y \sim y$ with mean δ. Both have standard deviation σ. Consider that we have N samples from each, with N large.

   (a) How big does δ have to be in order to be likely to tell the populations apart? (Assume a 1 tailed test.) Use the central limit theorem, and $p = 0.05$ rejection of the null under the null, 0.95 rejection of the null under the alternative.

(b) Generate samples of $N = 20^2$ and 50^2 with the appropriate δs and confirm that you do reject the null with the appropriate frequency.

(c) Now imagine that we have $N = 1,000,000$ samples. Lets say that we reject the null with a small p-value. What does that mean about δ?

(d) If these large samples were obtained from an internet survey 10 years ago (X), compared to an internet survey now (Y), what can we deduce about the population of internet users? How about the population at large?

5. Suppose you have 4 independent observations from a normal distribution with unknown mean and unknown variance. You estimate the mean and SD from the data, and use them 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.

6. The Cauchy distribution is peaked around 0 like the normal, but has a 'heavier tail'. You can simulate from it saving the following code into cauchyrnd.m:

```
function r = cauchyrnd(mu,c,varargin)
r = c.*tan(pi*(rand(varargin{:})-0.5))+mu;
```

Where μ is the mean and c is a measure of variation. (The final argument is for the size of data returned). Access it via e.g.

```
cauchyrnd(0,2)
cauchyrnd(0,2,n,m)
```

where the second form returns an n by m matrix of values.

(a) Following the examples in Question 3, simulate $n = 2, 3, 5, 10, 100, 1000, 5000$ Cauchy Random variables with position $\mu = 0$ and $c = 2$. Analyse the mean of these variables with the standard t test procedure.

(b) Monitor the $0.01, 0.05, 0.1$ quantiles.

(c) Do some standard exploritory plots.

(d) Try the other appropriate tests we have covered.

(e) Discuss why the procedure is not working.

[sny61] Snyder, R.G. (1961) The sex ratio of offspring of pilots of high performance military aircraft. Human Biology, 33(1):1-10.