Astronomy A3/A4H

Statistical Astronomy I: Example Sheet 4

(In questions involving Hypothesis tests, if the significance level is not explicitly stated then you should choose the level yourself)

1. The distribution of (natural) log distance of galaxies in a survey is found to be normal with mean μ and variance σ^2. Derive the pdf, $p(r)$, of the galaxy distance, R, and determine the expected value and variance of R.

2. The distribution of X = log temperature for a stellar population is modelled to be Gaussian in form. X is measured for a sample of 16 stars with the following results (in suitably scaled units)

$$\sum x_i = 51.2 \qquad \sum x_i^2 = 243.19$$

   Test the hypothesis that μ, the population mean log temperature, is equal to 4.0 against the alternative hypothesis that $\mu < 4.0$:-

   (a) assuming that σ, the population standard deviation, is known to be 1.9

   (b) when σ is not known *a priori* and must be estimated from the sample data.

   Suggest why (a) is the better hypothesis test *if* σ is known.

3. The apparent magnitudes of two samples of galaxies observed in two different clusters are given below.

cluster A	13.1	14.0	14.9	13.7	15.5	13.2	13.1	13.3	13.4	13.9
cluster B	13.7	13.1	14.6	13.6	13.4	13.2	13.3	13.8		

   Ignoring apparent magnitude errors and assuming both samples to be drawn from the same universal gaussian LF, construct a hypothesis test to decide if the clusters are equidistant based on these sample data.

4. By comparing the expected and observed frequencies under the null hypothesis, carry out a χ^2 goodness of fit test for the data of question 6 and 7 on example sheet 3, testing the goodness of fit to a binomial and Poisson model respectively.

5. A particular cluster of galaxies is modelled as consisting of two subgroups, within each of which the redshift distribution follows a normal pdf. Subgroup A has mean redshift 3000 kms^{-1} and standard deviation 140 kms^{-1}, while subgroup B has mean redshift 3500 kms^{-1} and standard deviation 200 kms^{-1}. Suppose a galaxy with redshift z is selected at random from the cluster. The following decision rule is adopted for assigning membership of the galaxy to subgroup A or B.

   galaxy belongs to subgroup A if $z \leq 3200$ kms^{-1}.
   galaxy belongs to subgroup B if $z > 3200$ kms^{-1}.

By considering the classification of group membership as a test of the null hypothesis H_1 : galaxy belongs to subgroup A, against the alternative hypothesis H_2 : galaxy belongs to subgroup B, determine the probability of a type I and type II error – and hence the power of the test – with this critical region.

Suppose the critical region is now changed to:-

galaxy belongs to subgroup A if $z \leq 3300$ kms^{-1}.
galaxy belongs to subgroup B if $z > 3300$ kms^{-1}.

How does this change the values of P(I), P(II) and the power of the test?

6. Let the discrete random variable r have the binomial distribution, defined in question 5 of examples sheet 3. If n is large and p is not too close to either 0 or 1 then the variable:-

$$z = \frac{r-np}{\sqrt{np(1-p)}}$$

is approximately normally distributed with mean zero and variance 1.

A coin is tossed n times and the number, r, of heads is noted, and the results used to test the hypothesis that the coin is 'fair' (i.e. $p = 0.5$). It is decided thhat the null hypothesis of a fair coin should be rejected if more than $r_{\rm crit}$ heads are obtained in the n tosses.

Write down integral expression for the probability of a type I and type II error, given the alternative hypothesis that $p = 0.6$.

How many tosses are required to ensure that the both P(I) and P(II) are no more than 5% with these null and alternative hypotheses?

What is the corresponding value of $r_{\rm crit}$ in this case?

7. Observations are made of galaxies in two clusters in order to determine the dispersion of the Tully-Fisher distance indicator relation. In the first cluster distances were estimated to 11 Sc galaxies and a sample variance of 13.67 was obtained. In the second cluster distances were estimated to 9 Sb galaxies and a sample variance of 9.47 was obtained. On the assumption that the Tully-Fisher residuals are normally distributed, test the hypothesis that the TF relations for Sc and Sb galaxies have identical dispersion.

Dr. Martin Hendry
Room 607, Kelvin Building
Ext 5685; email martin@astro.gla.ac.uk