MATH11400 Statistics 1 2010–11

Homepage http://www.stats.bris.ac.uk/%7Emapjg/Teach/Stats1/

Problem Sheet 8

1. Use the **R** commands `help(TDist)` and `help(Chisquare)` to find out how to compute the probability density function, distribution function and the inverse distribution function for the t and the χ^2 families of distributions.

   (a) Plot the probability density function for the t_r distribution over the interval $(-4, 4)$ for $r = 1, 5, 10$ and 15 degrees of freedom, and compare it with the probability density function for the $N(0, 1)$ distribution.

   (b) Plot the probability density function for the χ^2_r distribution over the interval $(0, 20)$ for $r = 5, 10$ and 15 degrees of freedom.

*2. Let $X_1, \dots, X_n$ be a random sample of size n from the $N(\mu, \sigma^2)$ distribution. Denote the sample variance by $S^2 = \sum_{i=1}^n (X_i - \overline{X})^2/(n-1)$ and denote the maximum likelihood estimator of σ^2 by $\hat{\sigma}^2_{mle} = \sum_{i=1}^n (X_i - \overline{X})^2/n$.

   (a) State the distribution of $\sum_{i=1}^n (X_i - \overline{X})^2/\sigma^2$ and its mean and variance, and thus find the mean and variance of both S^2 and $\hat{\sigma}^2_{mle}$.

   (b) Let $\hat{\sigma}^2$ denote an estimator for σ^2. The bias of $\hat{\sigma}^2$ is defined as $\mathrm{E}(\hat{\sigma}^2 - \sigma^2)$, while the mean square error is defined as $\mathrm{E}[(\hat{\sigma}^2 - \sigma^2)^2]$. Note that it can be easier to calcuate the mean square error from its representation as $\mathrm{E}[(\hat{\sigma}^2 - \sigma^2)^2] = \mathrm{Var}(\hat{\sigma}^2) + [\text{bias}(\hat{\sigma}^2)]^2$.
   Use the results of part (a) to compare the performance of S^2 and $\hat{\sigma}^2_{mle}$ as estimators of σ^2 in terms of their bias (i.e. their average error) and their mean square error (i.e. their average squared error).

*3. Suppose that $X_1, X_2, \dots, X_n$ and $Y_1, Y_2, \dots, Y_n$ are all i.i.d. Normal$(0, \theta^2)$ where θ is unknown.

   (a) Write down the distribution of the random variable $T_i = X_i^2 + Y_i^2$ for each i. Hence find the maximum likelihood estimate of θ based on observations $t_1, t_2, \dots, t_n$ of $T_1, T_2, \dots, T_n$.

   (b) Find also the maximum likelihood estimate of θ based on observations $x_1, x_2, \dots, x_n$ and $y_1, y_2, \dots, y_n$ of $X_1, X_2, \dots, X_n$ and $Y_1, Y_2, \dots, Y_n$, respectively.

   (c) Compare both of these estimates to the method of moments estimator found in problem sheet 3, question 2.

*4. In an experiment to determine the fuel consumption of a new model of car, a driver was employed to drive nine new cars, each for 100km. The fuel consumption in litres for each of the nine 100km drives is displayed in the table below. The data is contained in the Statistics 1 data set `fuel`.

12.09 11.18 9.97 10.50 9.92 9.97 11.84 10.93 10.70.

Stating clearly any assumptions you make, find a 90% confidence interval for the mean fuel consumption per 100km for cars of this type. Explore the data (e.g. with a stem and leaf plot) to confirm that your assumptions are appropriate.

5. Neurobiological arguments suggest that learning to play an instrument may improve the spatial-temporal reasoning of pre-school children. A study measured the spatial-temporal reasoning of 34 preschool children before and after six months of piano lessons. The changes in the reasoning scores of the children are displayed in the table below. You may assume the data represents the observed values of a simple random sample of size $n = 34$ from a population with unknown mean μ and unknown variance σ^2. The data is contained in the Statistics 1 data set `piano`.

2	5	7	−2	2	7	4	1	0	7	3	4	3	4	9	4	5
2	9	6	0	3	6	−1	3	4	6	7	−1	7	−3	3	4	4

Construct a 95% confidence interval for the population mean improvement in reasoning scores, stating clearly any assumptions you make. You should summarize and display the distribution of the data, say in a histogram, and hence check that your assumptions are appropriate.

Note: Given a single data set `xdata`, containing n values with sample mean $\bar{x}$ and sample variance s^2, the **R** command `t.test(xdata,conf.level=0.95)` will produce output that includes a 95% confidence interval based on the end points

$$c_L = \bar{x} - t_{n-1;\alpha/2}\, s/\sqrt{n} \qquad \text{and} \qquad c_U = \bar{x} + t_{n-1;\alpha/2}\, s/\sqrt{n}.$$

with $\alpha = 0.05$. The confidence level can, of course, be changed as desired. You should answer the questions above by going through the relevant working yourself, but you may wish to use this command to check your answers in cases where it is appropriate.