# M. Sc. Examination by course unit 2015

# MTH731U: Computational Statistics

## Duration: 3 hours

## Date and time: 12th May 2015, 14:30–17:30

---

**Apart from this page, you are not permitted to read the contents of this question paper until instructed to do so by an invigilator.**

**You should attempt ALL questions. Marks awarded are shown next to the questions.**

**Calculators ARE permitted in this examination. Please state on your answer book the name and type of machine used.**
**The New Cambridge Statistical Tables are provided.**

Complete all rough workings in the answer book and **cross through any work that is not to be assessed**.

Possession of unauthorised material at any time when under examination conditions is an assessment offence and can lead to expulsion from QMUL. Check now to ensure you do not have any notes, mobile phones, smartwatches or unauthorised electronic devices on your person. If you do, raise your hand and give them to an invigilator immediately. It is also an offence to have any writing of any kind on your person, including on your body. If you are found to have hidden unauthorised material elsewhere, including toilets and cloakrooms it shall be treated as being found in your possession. Unauthorised material found on your mobile phone or other electronic device will be considered the same as being in possession of paper notes. A mobile phone that causes a disruption in the exam is also an assessment offence.

**Exam papers must not be removed from the examination room.**

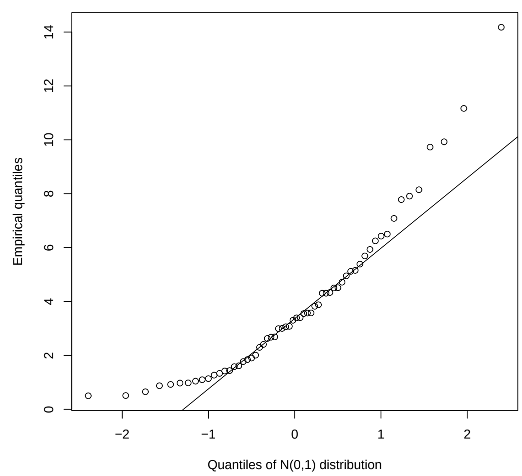
**Examiner(s): B. M. Parker**

---

© Queen Mary, University of London (2015) 

**Question 1.**

(a) Consider the following quantile-quantile plot. Explain how this plot helps us to assess whether a sample comes from a particular distribution. **[5]**




(b) Consider the following distributions:

(i) The normal distribution with mean $\mu = 1$ and variance $\sigma^2 = 2.5$.

(ii) The chi-squared distribution with 4 degrees of freedom.

(iii) Student's-t distribution with 4 degrees of freedom.

For a sample of size 60, which of these distributions would you expect to give a quantile-quantile plot similar to that shown above? Justify your answer. **[5]**

**Question 2.** We wish to test at the 10% level of significance if the computer generated sample

$$1.85, \quad 2.03, \quad 1.00, \quad 7.40, \quad 1.45$$

of size $n = 5$ comes from an exponential distribution with parameter $\lambda = 0.5$, using the Kolmogorov-Smirnov one sample test.

(a) Carry out this test, stating clearly your hypothesis and conclusions. **[12]**

(b) Why might it be preferable to carry out this test at the 10% level of significance rather than using, for example, the 1% level of significance? **[3]**

(c) If $Y$ is uniformly distributed on the interval (0,1), show that the distribution of $X = -2\log(1-Y)$ is exponentially distributed with parameter $\lambda = 0.5$. **[5]**

© **Queen Mary, University of London (2015)**

**Question 3.** Consider the following data:

$$1.4, \quad 2.7, \quad 2.8, \quad 3.2, \quad 5.9, \quad 6.8.$$

(a) Find the histogram estimator $\hat{f}_H$ and sketch its graph. Assume that the intervals are of equal width 1 with integer endpoints. [**8**]

(b) Comment on the bias and variance of the histogram estimator. How does this vary with the interval width? [**3**]

(c) Rosenblatt's estimator is given by

$$\hat{f}_{n,h}(y) = \frac{\#\{i : y_i - h < y < y_i + h\}}{2nh}$$

For the same data, compute Rosenblatt's estimator for a bandwidth of $h = 1$ and sketch its graph. [**9**]

(d) Comment on how changing the bandwidth would affect the graph. [**3**]

(e) Define the rectangular kernel, and prove that it is a kernel estimator for any given sample $y_1, \ldots, y_n$. [**7**]

**Question 4.** For $n = 4$ psychology students, the level of aptitude for a particular task was measured before and after lunch. A continuous variable was measured. The results are shown below, where the first value in each pair was measured before, and the second value after lunch:

$$(1.76, 3.68), \quad (3.95, 3.90), \quad (4.52, 4.56), \quad (3.50, 4.10).$$

(a) Perform a permutation test for matched pairs to test at the 10% level of significance whether there is a treatment effect. [**8**]

(b) Name a parametric test which could generally be used for this problem. State why a non-parametric test might be more appropriate here. [**3**]

(c) Describe briefly how one would carry out a permutation test using a computer. You do not have to write any code, but suggest the form of an algorithm that a program would carry out. [**4**]

© Queen Mary, University of London (2015) 

**Question 5.** Let $y_1, \ldots, y_n$ be realizations of independent and identically distributed random variables $Y_1, \ldots, Y_n$ with variance $\sigma^2$. Show that when using

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

to estimate $\sigma^2$, the jackknife estimate of bias for $\hat{\sigma}^2$ is equal to

$$\widehat{bias}_{\text{jack}} = -\frac{1}{n} s^2,$$

where $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$. [**10**]

**Question 6.**

(a) The coefficient of skewness of a random variable $Y$ with mean $\mu = E(Y)$ and standard deviation $\sigma = \sqrt{Var(Y)}$ is defined by

$$\gamma = \frac{E[(Y - \mu)^3]}{\sigma^3}.$$

For a random sample $y_1, \ldots, y_n$ of realizations of $Y$, find the plug-in estimate of $\gamma$. [**4**]

(b) For a random sample $y_1, \ldots, y_n$ of realizations of $Y$, describe briefly how we might use the plug-in estimate with a bootstrap procedure to estimate the bias of the estimator for $\gamma$. [**5**]

(c) Given a data set $y_1, \ldots, y_n$ of $n$ distinct values, show that the number of distinct nonparametric bootstrap samples is ${}^{2n-1}C_n = \binom{2n-1}{n}$. [**6**]

---

**End of Paper.**

© **Queen Mary, University of London (2015)**