EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY
(formerly the Examinations of the Institute of Statisticians)

GRADUATE DIPLOMA, 2007

Statistical Theory and Methods II

Time Allowed: Three Hours

Candidates should answer **FIVE** *questions.*

All questions carry equal marks.
The number of marks allotted for each part-question is shown in brackets.

Graph paper and Official tables are provided.

Candidates may use calculators in accordance with the regulations published in the Society's "Guide to Examinations" (document Ex1).

The notation log *denotes logarithm to base* ***e***.
Logarithms to any other base are explicitly identified, e.g. $\log_{10}$.

Note also that $\binom{n}{r}$ *is the same as* ${}^nC_r$.

This examination paper consists of 5 printed pages, **each printed on one side only**.
This front cover is page 1.
Question 1 starts on page 2.

There are 8 questions altogether in the paper.

1. (a) Let $X_1, X_2, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ population with sample mean $\bar{X}$, and let $Y_1, Y_2, \ldots, Y_n$ be an independent random sample from a $N(\rho\mu, \rho^2\sigma^2)$ population with sample mean $\bar{Y}$. Here σ is known but μ and ρ are unknown parameters. Assuming that $\rho \neq 0$, show that

$$\frac{\bar{X} - (\bar{Y}/\rho)}{\sqrt{2\sigma^2/n}}$$

is a pivotal quantity. (6)

(b) Let $U_1, U_2, \ldots, U_n$ be a random sample from a population with probability density function

$$f(u) = \theta^{-1} u^{(1-\theta)/\theta}, \qquad 0 < u < 1,$$

where $\theta > 0$ is an unknown parameter.

(i) Show that $-\theta^{-1}\log U_i$ has an exponential distribution. Show also that the mean and variance of this distribution are both equal to 1. (8)

(ii) Hence explain briefly why $V = -\theta^{-1}\sum_{i=1}^{n}\log U_i$ is a pivotal quantity. (2)

(iii) Use the central limit theorem to obtain an approximate 95% confidence interval for θ based on V when n is large. (4)

2. A random sample of size n is taken from a population with probability density function $f(x, \theta)$ with unknown parameter θ.

(i) A sufficient statistic T is available. What can be said about the distribution of the data given T? State the Neyman-Fisher factorisation theorem. (5)

(ii) Show that the maximum likelihood estimator of θ is a function of T. (5)

(iii) Give an example where the method of moments estimator of θ is not a function of a sufficient statistic, stating clearly any standard distributional results you have used. (4)

(iv) Suppose now that

$$f(x, \theta) = \frac{1}{x\theta\sqrt{2\pi}} \exp\{-(\log x)^2 / 2\theta^2\}, \qquad x > 0,$$

where $\theta > 0$. Use the factorisation theorem to obtain a sufficient statistic for θ. (6)

3. Let X_1, X_2, …, X_n be a random sample from a population with probability mass function

$$p(x) = \theta(1-\theta)^x, \quad x = 0, 1, 2, \ldots ,$$

where $0 < \theta < 1$ is an unknown parameter.

(i) Obtain the maximum likelihood estimator of θ and show that it is the same as the method of moments estimator of θ.

[You are given that $\sum_{x=1}^{\infty} x a^{x-1} = (1-a)^{-2}$ when $0 < a < 1$.] (6)

(ii) Show that the Cramér-Rao lower bound for the variance of unbiased estimators of θ is $\dfrac{\theta^2(1-\theta)}{n}$. (5)

(iii) Obtain a large sample approximate 95% confidence interval for θ. (5)

(iv) A researcher comments that such an interval is unhelpful because the confidence interval may extend below the value 0. Show that, in this example, such a comment has no validity, provided the sample size is at least four. (4)

4. Suppose $T > 0$ is an estimator of an unknown parameter $\theta > 0$, and that the loss function

$$L(T, \theta) = \frac{T}{\theta} + \frac{\theta}{T} - 2$$

is to be used.

(i) Sketch the loss function as a function of T when $\theta = 1$. (2)

(ii) Find the loss when the observed value of T is $r\theta$ for some positive value $r \neq 1$. Find another value of T that has the same loss. Briefly give a practical interpretation of this result. (6)

Consider the case when a random sample of size two, (X_1, X_2), is taken from an exponential distribution with mean θ, and T takes the form $c\bar{X}$, where $\bar{X}$ is the sample mean and c is a positive constant to be chosen.

(iii) Explain why the density of $X_1 + X_2$ is $y\exp(-y/\theta)/\theta^2$, for $y > 0$. Hence show that $(X_1 + X_2)^{-1}$ has mean θ^{-1}. (6)

(iv) Find the expected loss of T as a function of c and obtain the value of c that minimises the expected loss. (6)

5. Let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution with mean θ. The null hypothesis H_0: $\theta = 1$ is to be tested against the alternative hypothesis H_1: $\theta = 2$.

(i) Show that the Neyman-Pearson approach leads to rejection of H_0 in favour of H_1 when

$$\sum_{i=1}^{n} X_i \geq k$$

for some suitable k. (6)

(ii) It is required that the Type I and Type II error probabilities should each be no more than 0.01. By using the Normal approximation to the Poisson distribution, estimate the smallest value of n that will satisfy this requirement. (8)

(iii) Suppose now that H_0 is to be tested against the composite alternative hypothesis H_1: $\theta > 1$. Deduce the uniformly most powerful test with size approximately 0.01. Draw a rough sketch of the power function of this test when n is as specified in part (ii). (6)

6. Observations are available sequentially from a distribution with unknown parameter θ. The null hypothesis H_0: $\theta = \theta_0$ is to be tested against the alternative hypothesis H_1: $\theta = \theta_1$ with Type I and Type II error probabilities of approximately α and β respectively.

(i) Give instructions, in terms of α and β, as to how to perform the sequential probability ratio test (SPRT), and state, without proof, Wald's formulae for the (approximate) mean sample sizes needed, under each hypothesis. (6)

(ii) Without doing any calculations, discuss briefly the sample sizes needed for the above SPRT compared to the sample size of the corresponding most powerful fixed sample size test when the true value of θ lies strictly between θ_0 and θ_1. (4)

Now suppose that the observations come from a $N(0, \sigma^2)$ population and we wish to test H_0: $\sigma = 1$ against H_1: $\sigma = 2$ with the Type I and Type II error probabilities approximately equal to 0.05.

(iii) (a) Obtain the SPRT. Calculate the approximate mean sample sizes needed under each hypothesis. (7)

(b) Suppose that the first three observations are 1.6, –0.9 and –2.5. Carry out the steps in the SPRT for these data. (3)

Turn over

7. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a N(μ, 1) population.

(i) Suppose the prior distribution for μ is N(a, 1). Show that the posterior distribution for μ is

$$\mathrm{N}\left(\frac{\sum_{i=1}^{n} X_i + a}{n+1},\ \frac{1}{n+1} \right).$$

(6)

(ii) Suppose now that the prior distribution for μ has probability density function

$$\pi(\mu) = \frac{1}{2\sqrt{2\pi}} \exp\left\{ -\frac{(\mu-2)^2}{2} \right\} + \frac{1}{2\sqrt{2\pi}} \exp\left\{ -\frac{(\mu+2)^2}{2} \right\}.$$

(a) Give a rough sketch of this prior density function for values of μ between –4 and +4.

(4)

(b) Find the posterior distribution for μ. When n = 99 and $\sum_{i=1}^{99} X_i = 88$, sketch the posterior density on a separate graph, but using the same scale and axes as in part (a). Summarise the main differences between the prior and posterior distributions.

(10)

8. (a) Describe the Kolmogorov-Smirnov and χ^2 goodness-of-fit tests for a continuous univariate distribution (i) when the distribution is fully specified, and (ii) when the distribution depends on an unknown parameter. Compare and contrast the two tests in each situation.

(12)

(b) Data are available from two independent random samples and the equality of the two location parameters is to be tested. Outline one Normal-based parametric approach and one rank-based approach. Briefly compare the former with the latter.

(8)