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

GRADUATE DIPLOMA, 1999

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 silent, cordless, non-programmable electronic calculators.

Where a calculator is used the **method** *of calculation should be stated in full.*

Note that $\binom{n}{r}$ *is the same as* ${}^nC_r$ *and that* ln *stands for* $\log_e$.

This examination paper consists of 8 printed pages. This front cover is page 1. The reverse of the front cover, which is intentionally left blank, is page 2. Question 1 starts on page 3.

There are 8 questions altogether in the paper.

1. Explain what is meant by a *method of moments estimator*.

(3)

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution with probability density function

$$f(x) = \frac{5\theta^5}{(x+\theta)^6}, \qquad x > 0,$$

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

(i) Find $\hat{\theta}$, the method of moments estimator of θ.

(4)

(ii) Verify that $\hat{\theta}$ is an unbiased estimator of θ and find its variance. Say, with reasons, whether $\hat{\theta}$ is consistent for θ.

(5)

(iii) Find the Cramér-Rao lower bound for the variance of unbiased estimators of θ, and deduce the efficiency of $\hat{\theta}$.

(8)

2. A botanist has the theory that the number of different plant species, X, growing in a unit area of a certain region has a distribution with probability mass function

$$P(X = k) = \frac{-1}{\ln(1-\alpha)} \frac{\alpha^k}{k}, \qquad k = 1, 2, \ldots ,$$

where $0 < \alpha < 1$ is an unknown parameter. In a random sample of 150 such areas, the sample mean of the number of different plant species in a unit area was found to be 4.0.

(i) Assuming that the appropriate regularity conditions are satisfied, find an equation satisfied by $\hat{\alpha}$, the maximum likelihood estimate of α.

(6)

(ii) Without performing any computations, and assuming that an initial estimate of $\hat{\alpha}$ is available, describe how the Newton-Raphson method may be used to construct an iterative algorithm for finding the value of $\hat{\alpha}$ numerically.

(6)

(iii) Given that $\hat{\alpha}$ is approximately equal to 0.90, and assuming the asymptotic efficiency of the maximum likelihood estimator, find an approximate 99% confidence interval for α.

(8)

3. An experiment is conducted to compare m fertiliser treatments for strawberries. Each of these fertilisers is applied to n plants and the weights of the crops, in kilograms, are recorded. Let X_{ij} denote the weight of the crop from the jth plant that receives the ith fertiliser for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$, and suppose that the X_{ij} are independent random variables such that X_{ij} has a Normal distribution with mean μ_i and known variance σ^2. It is required to test the null hypothesis that all the μ_i are equal against the alternative hypothesis that they are not all equal.

(i) Show that the generalised likelihood ratio test depends on the data only through the value of

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\left\{\left(X_{ij}-\overline{X}\right)^2-\left(X_{ij}-\overline{X}_i\right)^2\right\}$$

where $\overline{X}_i = \sum_{j=1}^{n} X_{ij}\big/n$ and $\overline{X} = \sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}\big/(mn)$.

(9)

(ii) By using the relation $X_{ij} - \overline{X}_i = (X_{ij} - \overline{X}) - (\overline{X}_i - \overline{X})$, or otherwise, show that the test in part (i) has critical region depending on the value of $\sum_{i=1}^{m}\left(\overline{X}_i-\overline{X}\right)^2$.

(5)

(iii) Explain how χ^2 tables may be used to construct a test of size α based on the value of $\sum_{i=1}^{m}\left(\overline{X}_i-\overline{X}\right)^2$, and carry out the test at the 5% level when $m = 3$, $n = 7$, $\sum_{i=1}^{3}\left(\overline{X}_i-\overline{X}\right)^2 = 112$ and $\sigma^2 = 30$.

(6)

4. The probability that a certain flower seed germinates is p. The prior distribution of p is uniform over the interval (0,1). A gardener sows a set of 45 such seeds and finds that 25 of them germinate. It can be assumed that the seeds germinate independently of one another.

(i) Find the posterior distribution of p.

(6)

(ii) Find the mode of the posterior distribution. Under what circumstances might this value be used as a Bayes estimate of p?

(4)

(iii) Assuming a quadratic loss function and <u>stating clearly any result that you use</u>, find the Bayes estimate of p.

(4)

(iv) Assuming that the posterior distribution of p can be approximated by a Normal distribution, find an approximate 95% Bayesian confidence interval for p.

(6)

[The beta distribution with parameters $\alpha > 0$ and $\beta > 0$ has probability density function

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \qquad 0 < x < 1,$$

where $\Gamma(.)$ denotes the gamma function.]

5. Explain what is meant by a *confidence set* in classical inference.

(4)

Suppose that X_1, X_2,…, X_n is a random sample from a uniform distribution over the interval $(0, \theta)$, where $\theta > 0$ is an unknown parameter.

(i) Let $Y = \max_i(X_i)$ be the maximum of the n random variables. Find the distribution function of Y and deduce that its probability density function is

$$f_Y(y) = ny^{n-1}/\theta^n$$

for $0 < y < \theta$.

(5)

(ii) Show that Y/θ is a pivotal quantity.

(4)

(iii) Show that, for $0 < \alpha < 1$, $100(1-\alpha)\%$ confidence intervals for θ are of the form $Y/R_2 < \theta < Y/R_1$, where $0 < R_1 < R_2$ and R_1 and R_2 satisfy a condition which you should state.

(3)

(iv) Show that the shortest $100(1-\alpha)\%$ confidence interval for θ of the form given in part (iii) is of length $Y(\alpha^{-1/n} - 1)$.

(4)

6. Under what circumstances is it appropriate to use a *Wilcoxon signed ranks test*? Describe carefully, but without any calculation, how the figures in Table XVII of the "Abridged Tables for use by Examination Candidates", which relate to the Wilcoxon signed ranks test, may be calculated.

(10)

The consumption rates of nectar in cubic centimetres per hour for a random sample of eight honey-eating birds at sunrise and sunset are as follows.

Sunrise	0.9	1.6	1.4	1.2	1.6	1.1	0.8	1.0
Sunset	0.8	1.1	1.2	1.3	1.1	1.0	0.7	0.8

Carry out a two-sided Wilcoxon signed ranks test to test whether the distributions of the sunrise and sunset consumption rates are equal, and report your conclusions.

(10)

7. Explain what is meant by the *size* of a test. State the *Neyman-Pearson Lemma*.
(4)

The final weight of a certain organism, X, has a lognormal distribution with probability density function

$$f(x) = \frac{1}{\theta x\sqrt{2\pi}} \quad e^{-\frac{1}{2}\left\{\frac{\ln(x)}{\theta}\right\}^2}, \qquad x > 0,$$

where $\theta > 0$ is an unknown parameter. A random sample of n such organisms is selected and their final weights are found to be $X_1, X_2, \ldots, X_n$.

(i) Show that the most powerful test of the null hypothesis that $\theta = \theta_0$ against the alternative that $\theta = \theta_1$, where $\theta_1 > \theta_0$, has critical region depending on the value of $\sum_{i=1}^{n} \{\ln(X_i)\}^2$.
(6)

(ii) Find the distribution function and probability density function of $Y = \ln(X)/\theta$. Hence show that $\sum_{i=1}^{n} \{\ln(X_i)\}^2 / \theta^2$ has a chi-squared distribution with n degrees of freedom.
(5)

(iii) Find the critical region of the most powerful test with size 0.05 when $n = 25$ and $\theta_0 = 1$.
(2)

(iv) Evaluate the power of the test found in part (iii) when $\theta_1 = \sqrt{3}$.
(3)

8. Give an account of the uses of the *Central Limit Theorem* in sampling theory, the construction of statistical tables, estimation and testing. Explain how a statistician can ensure the reasonableness of an approximation or assumption that is made when appealing to the Central Limit Theorem.
(20)

BLANK PAGE