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

GRADUATE DIPLOMA, 2001

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. A random sample of 30 observations has been taken from an unknown continuous distribution and it is required to draw inferences about the median, η, of the distribution.

(i) Describe the *sign test*, and the corresponding Normal approximation, for testing hypotheses about η.

(8)

(ii) Using a Normal approximation, find the critical region for the sign test when the null hypothesis is $\eta = 10$, the alternative hypothesis is $\eta \neq 10$ and the significance level is 5%.

(4)

(iii) Explain what is meant by a *nonparametric confidence interval*.

(3)

(iv) Show that an approximate 95% nonparametric confidence interval for η is given by $\left[X_{(10)}, X_{(21)}\right)$, where $X_{(i)}$ denotes the ith order statistic for $i = 1, 2, \ldots, 30$.

(5)

2. Let $X_1, X_2, \ldots, X_n$ be the survival times of a random sample of n subjects given a particular medical treatment and suppose that these have an exponential distribution with mean ν.

(i) Show that $\hat{\nu} = \overline{X}$, where $\overline{X}$ denotes the sample mean, is an unbiased estimator of ν and find its variance.

(4)

(ii) Find the Cramér-Rao lower bound for the variance of unbiased estimators of ν and say whether $\hat{\nu}$ is efficient.

(7)

(iii) Let $Y = \min_i(X_i)$. Find $P(Y > y)$ and deduce the probability density function of Y. Hence show that the respective mean and variance of Y are ν / n and ν^2 / n^2.

(5)

(iv) Write down an unbiased estimator of ν based on Y and find its efficiency relative to $\hat{\nu}$. State, with reasons, which of these two estimators is consistent for ν.

(4)

3. A random sample of n flowers is taken from a colony and the numbers X, Y and Z of the three genotypes AA, Aa and aa are observed, where $X + Y + Z = n$. Under the hypothesis of random cross-fertilisation, each flower has probabilities θ^2, $2\theta(1-\theta)$ and $(1-\theta)^2$ of belonging to the respective genotypes, where $0 < \theta < 1$ is an unknown parameter.

(i) Show that the maximum likelihood estimator of θ is $\hat{\theta} = (2X + Y)/(2n)$.

(7)

(ii) Consider the test statistic $T = 2X + Y$. Given that T has a binomial distribution with parameters $2n$ and θ, obtain a critical region of approximate size α based on T for testing the null hypothesis that $\theta = \theta_0$ against the alternative that $\theta = \theta_1$, where $\theta_1 < \theta_0$ and $0 < \alpha < 1$.

(2)

(iii) Show that the test in part (ii) is the most powerful of size α.

(3)

(iv) Deduce approximately how large n must be to ensure that the power is at least 0.9 when $\alpha = 0.05$, $\theta_0 = 0.4$ and $\theta_1 = 0.3$.

(8)

4. Explain what is meant by a *risk function* and an *admissible decision rule* in the context of decision theory.

(4)

Let $X_1, X_2, \ldots, X_n$ be a random sample from a Normal distribution with mean θ and known variance σ^2, and suppose that θ has a Normal prior distribution with mean u_0 and variance v_0.

(i) Show that the posterior distribution of θ is Normal, and find the mean and variance of this distribution.

(7)

(ii) Assuming a quadratic loss function and stating clearly any result that you use, write down the Bayes estimator of θ. Also give a 95% Bayesian confidence interval for θ.

(4)

(iii) Find the risk function for the Bayes estimator for the case $u_0 = 0$, and show that it is smaller than that for the estimator $\bar{X}$ if $\theta^2 < 2v_0 + \dfrac{\sigma^2}{n}$.

(5)

5. Suppose that the household incomes in a certain country have the Pareto distribution with probability density function

$$f(x) = \frac{\theta \nu^{\theta}}{x^{\theta+1}} \ , \qquad \nu \le x < \infty \ ,$$

where $\theta > 0$ is unknown and $\nu > 0$ is known. Let $X_1, X_2, \ldots, X_n$ denote the incomes for a random sample of n such households. It is required to test the null hypothesis that $\theta = 1$ against the alternative that $\theta \neq 1$.

(i) Show that the maximum likelihood estimator of θ is

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n} \ln\left(X_i / \nu\right)} \ .$$

(7)

(ii) Show that the generalised likelihood ratio test statistic, $\lambda(x)$, satisfies

$$\ln\{\lambda(x)\} = n - n\ln\left(\hat{\theta}\right) - \frac{n}{\hat{\theta}} \ .$$

(6)

(iii) Show that the test accepts the null hypothesis if

$$k_1 < \sum_{i=1}^{n} \ln\left(X_i\right) < k_2 \ ,$$

and state how the values of k_1 and k_2 may be determined.

(7)

6. The lengths, in millimetres, of cuckoos' eggs found in hedge sparrows' nests may be modelled by a random variable X with probability density function

$$f(x) = \frac{\theta x^{\theta-1}}{25^{\theta}} , \qquad 0 \le x \le 25 ,$$

where $\theta > 0$ is an unknown parameter. It is required to test the null hypothesis that $\theta = 3$ against the alternative that $\theta = 6$.

(i) Construct a sequential probability ratio test for which the Type I and Type II errors are approximately 0.05 and 0.10, respectively.

(7)

(ii) Show that

$$E\{\ln(X)\} = \ln(25) - \frac{1}{\theta} ,$$

and hence show that the expectation of the sample size of the test in part (i) when $\theta = 6$ is approximately 12.3.

(8)

(iii) Use a table to carry out the test in part (i) when the first 14 cuckoos' eggs have the following lengths.

22.0	23.9	20.9	23.8	25.0	24.0	21.7	23.8
22.8	23.1	23.1	23.5	23.0	23.0		

(5)

7. Explain what is meant by a *pivotal quantity*.

(4)

(a) Suppose that X is a continuous random variable with distribution function $F(x) = P(X \le x)$ which is strictly increasing in x, at least for $0 < F(x) < 1$. Show that $Y = F(X)$ has a uniform distribution over the interval (0,1).

(3)

(b) Let X have the logistic distribution with probability density function

$$f(x) = \frac{e^{x-\theta}}{\left(1+e^{x-\theta}\right)^2} , \qquad -\infty < x < \infty ,$$

where $-\infty < \theta < \infty$ is an unknown parameter.

(i) Show that $X - \theta$ is a pivotal quantity and hence, given a single observation on X, construct an exact $100(1-\alpha)\%$ confidence interval for θ. Evaluate the interval when $\alpha = 0.05$ and $X = 10$.

(10)

(ii) Given a random sample of size n from the above distribution, $X_1, X_2, \ldots, X_n$, and that $E(X) = \theta$ and $\text{Var}(X) = \pi^2/3$, briefly explain how you would use the central limit theorem to construct an approximate 95% confidence interval for θ.

(3)

8. Explain what is meant by *simulation*, and describe how it may be used in estimation and model validation. Your answer should include, for example, discussion of how the accuracy of asymptotic results may be assessed, how different inference procedures may be compared and how the assumptions of a model may be checked.

(20)