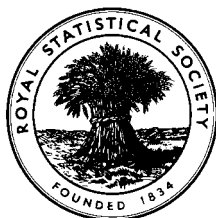


**EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY**  
*(formerly the Examinations of the Institute of Statisticians)*



**GRADUATE DIPLOMA, 2000**

**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 10 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) Explain what is meant by a *likelihood function*. (2)
- (b) Let  $X_1, X_2, \dots, X_n$  be a random sample of size  $n$  from a uniform distribution over the interval  $(0, \theta)$ , where  $\theta > 0$  is an unknown parameter.
- (i) Show that the method of moments estimator of  $\theta$  is  $\hat{\theta}_1 = 2\bar{X}$ , where  $\bar{X}$  denotes the sample mean. Verify that  $\hat{\theta}_1$  is an unbiased estimator of  $\theta$  and find its variance. (5)
- (ii) Write down the likelihood function and explain carefully why the maximum likelihood estimator of  $\theta$  is  $\hat{\theta}_2 = \max_i(X_i)$ , the maximum of the  $n$  random variables. (4)
- (iii) Let  $Y = \max_i(X_i)$ . Find the distribution function of  $Y$  and deduce its probability density function. Hence show that the mean of  $Y$  is  $n\theta/(n+1)$ . (6)
- (iv) Write down an unbiased estimator,  $\tilde{\theta}$ , of  $\theta$  based on  $Y$ . Given that the variance of  $Y$  is  $n\theta^2/\{(n+1)^2(n+2)\}$ , find the efficiency of  $\tilde{\theta}$  relative to  $\hat{\theta}_1$ . (3)

2. Explain what is meant by a *sequential probability ratio test* and give the no-overshoot approximations to the stopping boundaries. (6)

Let  $x_1, x_2, \dots$  be a sequence of independent observations from the distribution with probability mass function

$$P(X = k) = \frac{1}{k \ln(\theta)} \left( \frac{\theta - 1}{\theta} \right)^k, \quad k = 1, 2, \dots,$$

where  $\theta > 1$  is an unknown parameter. It is required to test the null hypothesis that  $\theta = 2$  against the alternative that  $\theta = 4$ .

- (i) Construct a sequential probability ratio test for which the Type I and II errors are both approximately 0.01. (7)
- (ii) Show how a graph may be used to help carry out the test. (3)
- (iii) Find the approximate expected sample size when  $\theta = 2$ . (4)

3. Describe the merits and limitations of *maximum likelihood estimation*. (6)

The deviations, in pounds, from the nominal weight of one pound of bags of carrots have a Normal distribution with mean zero and variance  $\sigma^2$ , and are independent. Let  $X_1, X_2, \dots, X_n$  denote the deviations for a random sample of  $n$  one-pound bags of carrots.

- (i) Write down the likelihood function of the standard deviation  $\sigma$  and find  $\hat{\sigma}$ , the maximum likelihood estimator of  $\sigma$ . (4)

- (ii) Assuming the asymptotic efficiency of the maximum likelihood estimator, find the asymptotic variance of  $\hat{\sigma}$ , and hence construct an approximate 90% confidence interval for  $\sigma$ . (4)

- (iii) Show that  $\sum_{i=1}^n X_i^2 / \sigma^2$  is a pivotal quantity. Using this property, construct an exact 90% confidence interval for  $\sigma$  when  $n = 12$  and  $\sum_{i=1}^{12} X_i^2 = 0.46$ . (6)

4. Some biology students were interested in analysing the amount of time that bees spend in flower patches gathering nectar. Suppose that the times, in seconds, spent by bees at low-density and high-density flower patches have exponential distributions with respective parameters  $\lambda_1 > 0$  and  $\lambda_2 > 0$ . Random samples of visiting times were obtained for the two types of flower patches. These were  $X_1, X_2, \dots, X_m$  and  $Y_1, Y_2, \dots, Y_n$  corresponding to  $\lambda_1$  and  $\lambda_2$ , respectively.

(i) Show that the generalised likelihood ratio test for testing the null hypothesis that  $\lambda_1 = \lambda_2$  against the alternative that  $\lambda_1 \neq \lambda_2$  depends on the data only through the value of  $\bar{Y}/\bar{X}$ , where  $\bar{X}$  and  $\bar{Y}$  denote the sample mean visiting times.

(9)

(ii) Show that  $2\lambda_1 m \bar{X}$  and  $2\lambda_2 n \bar{Y}$  have  $\chi^2$  distributions, and hence that  $\bar{Y}/\bar{X}$  has an  $F$  distribution under the null hypothesis.

(6)

(iii) Explain how  $F$  tables may be used to construct a test of size  $\alpha$  based on the value of  $\bar{Y}/\bar{X}$ , and carry out the test at the 5% level when  $m = 37$ ,  $n = 39$ ,  $\bar{X} = 229$  and  $\bar{Y} = 208$ .

(5)

[The  $\chi^2$  distribution with  $k$  degrees of freedom has moment generating function  $(1-2t)^{-k/2}$  for  $t < 1/2$  and the exponential distribution with parameter  $\lambda$  has moment generating function  $\lambda/(\lambda-t)$  for  $t < \lambda$ .]

5. In an experiment on a random sample of 250 insects, it was found that 180 were killed following treatment by a certain insecticide.

(i) Let  $p$  denote the proportion of all insects that would be killed using this insecticide. Using classical methods, find an approximate 95% confidence interval for  $p$ .

(8)

(ii) Find the posterior distribution of  $p$ , assuming that the prior distribution of  $p$  is uniform over the interval  $(0, 1)$ .

(3)

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

(4)

(iv) Four more insects are treated by the insecticide. Assuming that their responses to the insecticide are independent, find the probability, in terms of  $p$ , that exactly three of them are killed. Using the Bayesian approach, show that the predicted probability that exactly three are killed is approximately 0.41.

(5)

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

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

where  $\Gamma(\cdot)$  denotes the gamma function; it has mean  $\alpha/(\alpha + \beta)$  and variance  $\alpha\beta/\{(\alpha + \beta)^2(\alpha + \beta + 1)\}$ .]

6. Under what circumstances is it appropriate to use a *Mann-Whitney U test*? Describe carefully how the figures in Table XIV of the “Abridged Tables for use by Examination Candidates”, which relate to the *Mann-Whitney U test*, may be calculated.

(10)

The weight gains in grams of female rats under two diets were recorded. Of 19 female rats, 12 were randomly assigned to a high protein diet, with the remaining seven being assigned to a low protein diet. The weight gains were as follows:

High protein 134 146 104 119 124 161 107 83 113 129 97 123

Low protein 70 118 101 85 107 132 94

Using the Abridged Tables, carry out a two-sided *Mann-Whitney U test* to test whether the distributions of weight gains under the two diets are equal, give a range within which the  $p$ -value lies and report your conclusions.

(10)

7. Explain what is meant by a *decision rule* and a *minimax decision rule* in the context of decision theory. (4)

Suppose that  $X$  is a random variable with probability mass function

$$P(X = i) = (1 - p)p^i, \quad i = 0, 1, \dots,$$

where  $0 < p < 1$ . After observing the value of  $X$ , one of two actions  $A$  and  $B$  must be taken. If action  $A$  is taken, there is a loss of 4 if  $p > 0.7$  or  $-8$  if  $p \leq 0.7$ . If action  $B$  is taken, there is a loss of 0 if  $p > 0.7$  or 2 if  $p \leq 0.7$ . Under decision rule  $\delta_k$ , action  $A$  is taken if  $X < k$ , otherwise action  $B$  is taken.

- (i) Find the minimax decision rule among the decision rules  $\delta_0, \delta_1, \dots$ . (7)
- (ii) Show that the Bayes risk of  $\delta_k$ ,  $B(k)$ , is given by

$$B(k) = -4.4 - \frac{4}{k+1} + \frac{14(0.7)^{k+1}}{k+1}, \quad k = 0, 1, \dots,$$

if the prior distribution of  $p$  is uniform over the interval  $(0, 1)$ . (5)

- (iii) By showing that

$$B(k+1) - B(k) = \frac{1}{(k+1)(k+2)} \left\{ 4 - 14(0.7)^{k+1} (0.3k + 1.3) \right\},$$

or otherwise, show that  $\delta_6$  minimises the Bayes risk among the decision rules  $\delta_0, \delta_1, \dots$ . (4)

8. Explain what is meant by the *Neyman-Pearson* approach to testing one simple hypothesis against another simple hypothesis. (4)

Suppose that  $X_1, X_2, \dots, X_n$  are independent random variables such that  $X_i$  has a Poisson distribution with mean  $\lambda v^i$ , where  $\lambda > 0$  is known and  $v > 0$  is unknown. It is required to test the null hypothesis that  $v = 1$  against the alternative hypothesis that  $v = 2$ .

- (i) Show that the most powerful test has critical region depending on the value of  $\sum_{i=1}^n iX_i$ . (5)
- (ii) Briefly explain how you would use the central limit theorem to obtain a test with approximate significance level  $\alpha$ . (5)
- (iii) Suppose that  $\lambda = \frac{1}{3}$ ,  $n = 2$  and that the test rejects the null hypothesis if  $X_1 + 2X_2 > 2$ . Find the exact significance level and power of the test. (6)

[You are given that  $\sum_{i=1}^n i^2 = \frac{1}{6}n(n+1)(2n+1)$ .]