EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

GRADUATE DIPLOMA, 2008

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. The times, in hours, between accidents that are reported to an ambulance service have an exponential distribution with mean θ. A random sample of these times T_1, T_2, …, T_n is available.

(i) Show that $\hat{\theta} = \Sigma T_i / n$ is an unbiased estimator of θ. Show that Var(T_i) = θ^2 and hence find $\mathrm{Var}(\hat{\theta})$.

(5)

(ii) Find the Cramér-Rao lower bound for the variance of unbiased estimators of θ. What implications does this result have for the estimator $\hat{\theta}$?

(5)

(iii) $X = \min(T_i)$ is the minimum of the observed times. By evaluating $P(X > x)$, or otherwise, find the probability density function of X and hence identify the distribution of X.

(5)

(iv) Write down an unbiased estimator of θ based on X and find its efficiency relative to $\hat{\theta}$. Giving a reason, say which estimator is to be preferred.

(5)

[In part (iv), you may use without proof the results that $E(X) = \theta / n$ and Var(X) = θ^2 / n^2.]

2. The random variable Y has probability density function

$$f(y) = \frac{(y+1)e^{-y/\theta}}{\theta(\theta+1)} \quad (y > 0)$$

where θ (> 0) is an unknown parameter. You are given that $E(Y) = \theta(2\theta+1)/(\theta+1)$ and Var(Y) = $\theta^2(2\theta^2+4\theta+1)/(\theta+1)^2$.

A random sample $y_1, y_2, \ldots, y_{25}$ from this distribution is available and it is desired to test the null hypothesis $\theta = 1$ against the alternative $\theta = 2$.

(i) Find the form of the critical region for the most powerful test.

(6)

(ii) Use the central limit theorem to obtain, approximately, the critical region for the most powerful test at the 5% level.

(5)

(iii) Find the approximate power of this test at $\theta = 2$.

(5)

(iv) Say, with reasons, whether this test is uniformly most powerful for the null hypothesis $\theta = 1$ against the alternative $\theta \neq 1$.

(4)

Turn over

3. (a) Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a distribution that has an unknown parameter θ that can take values in the set Ω. Describe the *generalised likelihood ratio test* of the null hypothesis $\theta \in \omega \subset \Omega$ against the general alternative. State the asymptotic distribution, under the null hypothesis, of $-2\log\Lambda$, where Λ is the test statistic.

(5)

(b) A manufacturer of computers finds that X_i, the number of claims under guarantee for computers of model i, has a probability distribution satisfying

$$P(X_i = 0) = p_i, \quad P(X_i = 1) = p_i(1-p_i) \quad \text{and} \quad P(X_i > 1) = (1-p_i)^2$$

where $0 < p_i < 1$ (i = 1, 2). The manufacturer wishes to compare the distributions for the two models. Suppose that, for model i, out of n_i computers sold for which the guarantee period has expired, there have been no claims under the guarantee for m_i computers, a single claim for a further r_i computers, and more than one claim for the rest (i = 1, 2).

(i) Find the form of the generalised likelihood ratio test for testing the null hypothesis $p_1 = p_2$ against the alternative $p_1 \neq p_2$.

(8)

(ii) Records show that $n_1 = 4000$, $m_1 = 3500$, $r_1 = 400$, $n_2 = 6000$, $m_2 = 5000$ and $r_2 = 700$. Carry out the test and report your conclusions.

(7)

4. (a) Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a distribution that has an unknown parameter θ. What is meant by a *pivotal quantity* and by a 95% *confidence interval*? Show how a pivotal quantity may be used to find a confidence interval.

(7)

(b) A random sample $Y_1, Y_2, \ldots, Y_{10}$ is drawn from the uniform distribution between 0 and α, where α (> 0) is an unknown parameter. The largest of the ten observations is denoted by X.

(i) Find $P(X \le x)$, where $0 < x < \alpha$.

(3)

(ii) Show that $W = X/\alpha$ is a pivotal quantity and derive its probability density function.

(4)

(iii) Deduce the shortest 95% confidence interval for α based on W, when the observed value of X is 6.4.

(6)

Turn over

5. Manufacturers S and T both make widgets, but the probability that one of S's widgets is faulty is 0.3 whilst the corresponding probability for T is 0.7. A customer receives a batch of widgets but does not know whether it was from S or T. He decides to choose two widgets at random from the batch, examine each carefully to see whether it is faulty, and then take one of two actions: a_1, accept the batch; or a_2, reject the batch. The losses associated with these actions are as follows.

		a_1	a_2
Manufacturer	S	0	8
	T	16	4

(i) List the possible pure decision functions.

(6)

(ii) Evaluate the risk function of each decision and, by plotting a graph or otherwise, determine which of the decisions are admissible.

(8)

(iii) Find the minimax mixed decision function.

(6)

6. (a) Explain how in Bayesian inference the posterior distribution of a parameter θ can be used to give (i) a point estimate of θ, (ii) an interval estimate for θ.

(4)

(b) The numbers of flaws in one-km lengths of fibres can be assumed to be independent and to follow a Poisson distribution with mean λ (> 0). The prior distribution of λ is gamma (see below), with parameters $v = 12$, $k = 2$. When 10 randomly selected one-km lengths are inspected, 8 flaws are found in total.

(i) Find the posterior distribution of λ.

(6)

(ii) Find the approximate posterior probability that λ is less than 0.646.

(5)

(iii) Assuming quadratic loss, find the Bayes estimate of the probability that a new randomly selected one-km length of fibre will contain no faults.

(5)

[If X has the gamma distribution with parameters v and k, then its probability density function is

$$f(x) = v^k x^{k-1} e^{-vx} / \Gamma(k) \quad \text{for } x > 0,$$

where $\Gamma(k)$ is the gamma function. When k is a positive integer, then $\Gamma(k) = (k-1)!$ and $2vX$ has the χ^2_{2k} distribution.]

7. (a) What is meant by a *sequential* test? What are the benefits and drawbacks of sequential tests, compared to tests based on fixed sample sizes?

(5)

(b) In an accelerated life experiment, the times to failure, in hours, of a certain type of device have probability density function

$$f(x) = v^2 x e^{-vx}$$

for $x > 0$. It is required to test the null hypothesis $v = 0.2$ against the alternative $v = 0.1$, using a sequential probability ratio test. The significance level is to be 0.05 and the power should equal 0.8.

(i) Show that the mean time to failure is $2/v$. [You may use without proof the result that $\int_0^\infty y^k e^{-y/\theta} dy = k!\theta^{k+1}$, when k is a positive integer.]

(2)

(ii) Devise a sequential probability ratio test with approximately the above characteristics.

(6)

(iii) Sketch a chart that could be used in practice to help carry out the test.

(3)

(iv) Find the approximate expected sample size for your test if $v = 0.2$.

(4)

8. (a) It is required to test whether a sample of values could be assumed to have come from a specified distribution. Why might the Kolmogorov-Smirnov one-sample test be preferred to the χ^2 goodness-of-fit test when the size of the sample is small?

(4)

(b) A measurement, X, has probability density function given by

$$f(x) = \frac{\lambda x^{\lambda-1}}{\theta^\lambda} \exp(-(x/\theta)^\lambda) \quad (x > 0)$$

where λ and θ are positive parameters. Show that the $100p$ percentile of this distribution is $\theta(-\log(1-p))^{1/\lambda}$. Hence use the respective estimates 4.425 and 5.575 of the median and upper quartile to deduce that the values of λ and θ can be estimated as approximately 3 and 5 respectively.

(8)

Carry out a Kolmogorov-Smirnov test of the hypothesis that the measurements below come from this distribution.

5.1, 6.2, 3.4, 2.2, 4.7, 3.3, 1.6, 4.6, 5.0, 4.3

[You are given that the critical value for the test statistic at the 5% significance level is 0.41.]

(8)