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

GRADUATE DIPLOMA IN STATISTICS, 1996

Statistical Theory and Methods II

Time Allowed: Three Hours

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

All questions carry equal marks.

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* $\log_e$.

1. (a) Define the *mean square error* of an estimator and show why there is not usually an estimator with uniformly least mean square error.

(b) A random sample $X_1, X_2, ..., X_n$ is obtained from a distribution with probability density function $f(x) = 3\alpha^3 / (\alpha + x)^4$, $x > 0$, where α (>0) is an unknown parameter.

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

(ii) Show that $\hat{\alpha}$ is an unbiased estimator of α and find its variance.

(iii) Find the Cramer-Rao lower bound for the variance of unbiased estimators of α and deduce the efficiency of $\hat{\alpha}$.

2. What is meant by the *invariance property* of maximum likelihood estimators.

A random sample $T_1, T_2, , ... , T_n$ comes from a Uniform distribution on the range $[\theta_1, \theta_2]$, where $\theta_1 < \theta_2$ and the values of θ_1 and θ_2 are unknown. It is required to estimate $\delta = \theta_2 - \theta_1$.

(i) Show that the maximum likelihood estimator of δ is $\hat{\delta} = X - Y$, where $X = \max_i (T_i)$ and $Y = \min_i (T_i)$.

(ii) Evaluate $P(X \le x)$ for $\theta_1 < x < \theta_2$ and $P(Y \ge y)$ for $\theta_1 < y < \theta_2$. Hence, or otherwise, find the probability density functions of X and Y.

(iii) Evaluate the expected value of $\hat{\delta}$ and show that $\hat{\delta}$ is a biased estimator of δ.

Turn over

3. Explain carefully the relationship between *statistical tests* and *confidence intervals* in classical statistical inference.

The weights of tubs filled with margarine on a production line are independent and have a Normal distribution with mean μ and variance σ^2, where μ is unknown and σ^2 is known. Prior to the weighing device developing a fault, n tubs are weighed and their weights accurately recorded as $X_1, X_2, \ldots, X_n$. After developing the fault, the error in the device is proportional to the amount weighed, so that a tub of true weight w is recorded as having weight ρw, where ρ (>0) is unknown (i.e. the mean of the recorded weight is $\rho\mu$ and the variance is $\rho^2\sigma^2$). The weights of n tubs weighed after the fault has occurred are $Y_1, Y_2, \ldots, Y_n$.

(i) Find the distributions of $\overline{X}$ (= $\Sigma X_i/n$) and $\overline{Y}$ (=$\Sigma Y_i/n$).

(ii) Show that

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

is a pivotal quantity for ρ.

(iii) Deduce a 95% confidence interval for ρ. (It can be assumed that the probability that $\overline{X} - 1.96\sqrt{2\sigma^2/n}$ is negative is so small that this possibility may be ignored.)

4. In a certain library, the distribution of the numbers of times books are borrowed over a 6-month period is Geometric, parameter p, where p $(0 < p < 1)$ is unknown. The Librarian wants to test at the 10% level the hypothesis that $p = 0.2$ against the hypothesis that $p < 0.2$. A random sample of 200 books is selected and the numbers of times the books have been borrowed in the previous 6 months found to be $N_1, N_2, \ldots, N_{200}$.

(i) Show that there is a uniformly most powerful test and find its form. (Any general result that you use must be stated clearly.)

(ii) Using a generating function, or otherwise, find the mean and variance of N_1.

(iii) Use the Central Limit Theorem to find the uniformly most powerful test at approximately the 10% level.

[Note: the Geometric distribution, parameter p, has probability distribution:

$$P(X = k) = p(1-p)^k \text{ for } k = 0, 1, 2, \ldots .]$$

5. The following 3-stage sampling scheme has been proposed for determining whether a large batch of components should be accepted or rejected. At Stage 1, 40 components are sampled at random; the batch is accepted if there are no defectives, rejected if there is more than one defective and otherwise the Stage 2 sample is taken. At Stage 2, a further 40 components are sampled at random; the batch is accepted if there are no defectives, rejected if there are more than 2 defectives and otherwise the Stage 3 sample is taken. At Stage 3, 40 more components are sampled at random; the batch is accepted if the total number of defectives in all samples is less than 4 and is rejected otherwise. The proportion of defectives in the batch is denoted by p and λ = 40p. Use the Poisson approximation to the Binomial distribution to:

(i) find the probability distribution of the number of components sampled per batch;

(ii) evaluate the expected number of components sampled per batch;

(iii) show that the probability that the batch is accepted is approximately equal to

$$e^{-\lambda}(1+\lambda e^{-\lambda}+\lambda^2 e^{-2\lambda}(2+3\lambda)/2);$$

(iv) show that the probability of sampling 120 components in a batch is maximised when p is approximately 0.03.

6. Show that, under regularity conditions, the Bayes estimator of a parameter is the mean of the posterior distribution when the loss function is quadratic. (You may assume that the Bayes estimator minimises the posterior expected loss.)

The numbers of goals scored by a certain football team in different games can be assumed to be independent and to follow a Poisson distribution, parameter λ (> 0). The prior distribution of λ is Gamma, with parameters ν = 3, k = 3 and in the last 10 matches, the team has scored 12 goals.

(i) Find the posterior distribution of λ.

(ii) Find the posterior probability that λ is less than 0.711.

(iii) Assuming quadratic loss, estimate the probability that the team will fail to score in its next game.

[If X has the Gamma distribution with parameters k and ν, then the probability density function of X is: $f(x)=\nu(\nu x)^{k-1}e^{-\nu x}/\Gamma(k)$ for x > 0, where Γ() is the gamma function, and, if k is an integer, $2\nu X$ has the χ^2_{2k} distribution.]

Turn over

7. A particular device can be either faulty, in which case its time till failure has the exponential distribution with mean μ = 10, or satisfactory, in which case its time till failure is exponential with μ = 1000. Before accepting the device, you can, if you wish, test it. The cost for testing for planned time t is 5+t, irrespective of whether the device fails before time t. If you decide not to test the device, there is, of course, no testing cost. If you test the device and it does not fail, or if you decide not to test, you will accept the device and there will be a cost of 25 if it subsequently turns out to be faulty.

   (i) Let δ_t be the decision to test for planned time t (> 0) and let $R_{10}(t)$ and $R_{1000}(t)$ be the values of the risk function when μ = 10 and 1000 respectively. Let δ_0 be the decision not to test and let $R_{10}(0)$ and $R_{1000}(0)$ be the corresponding values of the risk function when μ = 10 and 1000 respectively. Evaluate $R_{10}(0)$, $R_{1000}(0)$ and, for $t > 0$, $R_{10}(t)$ and $R_{1000}(t)$.

   (iii) Find the value of t for which $R_{10}(t)$ is minimised and confirm that this corresponds to the minimax decision.

   (iv) Find the Bayes Solution when the prior probability of the device being faulty is 0.1.

8. Describe the various criteria that are used for deciding between estimators, indicating the importance of each of the criteria and the way in which they are used.