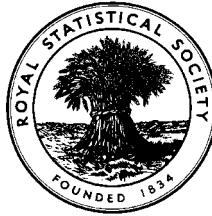


EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY
(formerly the Examinations of the Institute of Statisticians)



GRADUATE DIPLOMA IN STATISTICS, 1998

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. The reading on a voltage meter connected to a test circuit is uniformly distributed over the interval $(\theta, \theta + 1)$, where θ is the true but unknown voltage of the circuit. A random sample of size n , X_1, X_2, \dots, X_n , of these readings is available.
 - (i) Show that $\hat{\theta} = \bar{X} - \frac{1}{2}$ where \bar{X} denotes the sample mean, is an unbiased estimator of θ and find its variance.
 - (ii) Let $Y = \max_i(X_i)$ be the maximum of the n readings. Find the distribution function of Y and deduce its probability density function. Hence show that the mean and variance of Y are $\theta + n/(n+1)$ and $n/\{(n+1)^2(n+2)\}$, respectively.
 - (iii) Write down an unbiased estimator of θ based on Y and find its efficiency relative to $\hat{\theta}$. Show that both estimates are consistent for θ .

2. Explain what is meant by the *power* of a test and describe how the power may be used to help determine the most appropriate sample size.

Let X_1, X_2, \dots, X_n be a random sample from the gamma distribution with probability density function

$$f(x) = \frac{\nu^k}{\Gamma(k)} x^{k-1} e^{-\nu x}, \quad x > 0,$$

where $k > 0$ is known and $\nu > 0$ is unknown.

[If X is a gamma random variable with parameters $k > 0$ and $\nu > 0$, it has moment generating function $\{\nu/(\nu-t)\}^k$ for $t < \nu$.]

- (i) Find the form of the most powerful test of the null hypothesis that $\nu = \nu_0$ against the alternative hypothesis that $\nu = \nu_1$, where $\nu_1 > \nu_0$.
- (ii) Show that $\sum_{i=1}^n X_i$ has a gamma distribution with parameters nk and ν .
- (iii) Find the critical region of the most powerful test at the $100\alpha\%$ level, $0 < \alpha < 1$, when $k = \frac{1}{n}$.
- (iv) Show that the power of the test found in part (iii) is $1 - (1 - \alpha)^{\frac{\nu_1}{\nu_0}}$.

Turn over

3. Explain what is meant by a *pivotal quantity*, and show how such a quantity may be used to construct a confidence set for a parameter.

A single observation on a random variable X is taken from a distribution with probability density function

$$f(x) = \lambda x^{\lambda-1}, 0 < x < 1,$$

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

- (i) Show that $Y = -\ln(X)$ has an exponential distribution with parameter λ .
 - (ii) Show that $Y\lambda$ is a pivotal quantity.
 - (iii) Find a 95% confidence interval for λ .
4. A clinician has to decide whether or not to recommend a new drug in the treatment of a rare disease. There is a standard drug which is successful in about 40% of the cases treated, but it is hoped that the new drug will do better, although it is more expensive and can have occasional unfortunate side-effects. The clinician agrees to treat a sequence of patients with the new drug and to apply a sequential probability ratio test to the results. If the new drug has a success rate of 70%, then the test should accept it with probability 0.98, but if the success rate is only 35%, then the test should accept it with probability only 0.01.
- (i) Construct a sequential probability ratio test with approximately these error probabilities.
 - (ii) Find the approximate expected sample size when the true success rate for the new drug is 70%.
 - (iii) Draw a graph that allows the test to be carried out easily. Suppose that S denotes a 'success' and F a 'failure'. Demonstrate the use of the graph assuming that the responses for the first 16 patients are as follows:

$F, S, S, F, F, S, F, S, S, S, S, S, S, S, S, F.$

5. Explain what is meant by a *conjugate family of distributions*.

The number of telephone calls a man receives in a week has a Poisson distribution with mean θ . Let X_1, X_2, \dots, X_n denote the numbers of telephone calls that he receives during a random sample of n weeks, and suppose that the prior distribution of θ is gamma with parameters $\nu = 1/3$ and $k = 3$.

- (i) Find the posterior distribution of θ .
- (ii) Assuming a squared-error loss function and stating clearly any result that you use, find the Bayes estimator of θ .
- (iii) Find a 95% Bayesian confidence interval for θ if $n = 4$ and $\sum X_i = 26$.
- (iv) Show that a Bayes estimate of the probability that the man will receive no telephone calls next week, in the case $n = 4$ and $\sum X_i = 26$, is $(13/16)^{29}$, assuming a squared-error loss function.

[If Y is a gamma random variable with parameters $k > 0$ and $\nu > 0$, it has probability density function

$$f(y) = \frac{\nu^k y^{k-1}}{\Gamma(k)} e^{-\nu y}, \quad y > 0,$$

where $\Gamma(\cdot)$ denotes the gamma function, moment generating function $\{\nu/(v-t)\}^k$ for $t < \nu$, and, if k is a positive integer, $2\nu Y$ has χ_{2k}^2 distribution.]

6. Apples are packed in containers for marketing, and each container holds three apples. Although the apples are checked visually for damage before packing, there is a constant probability p that any apple will have developed a soft spot by the time it is sold. An inspector checks 216 containers on the shelves of supermarkets, and the distribution of the number of soft apples is as follows:

Number of soft apples	0	1	2	3	Total
Frequency	110	85	20	1	216

Assume that the numbers of soft apples in the different containers are independent.

- (i) Show that, if the number of soft apples in a container has a binomial distribution with parameter p , then the likelihood function is proportional to $p^{128} (1-p)^{520}$.
- (ii) Use a goodness-of-fit test to test whether the number of soft apples in a container has a binomial distribution.
- (iii) Find an approximate 90% confidence interval for p .

Turn over

7. Give an account of the classical, Bayesian and likelihood approaches to interval estimation.

8. Let x_1, x_2, \dots, x_n be a random sample from a continuous distribution with distribution function F . Describe the *Kolmogorov-Smirnov test* for the null hypothesis that $F(x) = F_0(x)$, where F_0 is some specified distribution function, and comment on the merits of the test.

Let X denote the distance in feet between flaws on a used computer tape. A random sample of ten observations on X are as follows:

18, 6, 1, 32, 116, 23, 12, 58, 101, 68.

Carry out a Kolmogorov-Smirnov test to test the null hypothesis that the above distances have an exponential distribution with mean 40 feet.