979

OLLSCOIL NA hÉIREANN, GAILLIMH

NATIONAL UNIVERSITY OF IRELAND, GALWAY

SEMESTER II (SUMMER) EXAMINATIONS, 2004/2005

FOURTH SCIENCE EXAMINATION

Module Code:	**MA486**
Module:	**STATISTICAL INFERENCE**
External Examiner	Dr. T. C. Bailey
Internal Examiner	Prof. J. P. Hinde

Instructions:

Duration: **Two** Hours.

Answer any *Three* of the four questions.

Relevant tables are supplied.

Question 1 is on the next page

1. (a) Consider a random sample $X_1, X_2, \ldots, X_n$ from a probability density (mass) function $f(x;\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of parameters. Explain the following terms:

- T is a *sufficient statistic* for $\boldsymbol{\theta}$;
- S is a *minimal sufficient* statistic for $\boldsymbol{\theta}$;
- *simple* and *composite* hypotheses for $\boldsymbol{\theta}$.

Why are (minimal) sufficient statistics so important in statistical inference? (8)

(b) For a random sample $X_1, X_2, \ldots, X_n$ from a Poisson distribution with mean λ, show that $\sum_{i=1}^{n} X_i$ is a minimal sufficient statistic for λ. (5)

Note: the probability mass function for a Poisson(λ) distribution is

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad x = 0, 1, 2 \ldots$$

Verify that $T = \sum_{i=1}^{n} X_i$ is sufficient by obtaining the conditional distribution of the observations given $T = t$. (7)

(c) Suppose now that X_1, X_2 is a random sample of size 2 from a Poisson(λ) distribution.

i) Show that $U = 2^{X_1}$ is an unbiased estimator of e^{λ} with variance $e^{2\lambda}[e^{\lambda} - 1]$.

ii) Using the fact that $T = X_1 + X_2$ is minimal sufficient, obtain an unbiased estimator of λ with smaller variance than U.

iii) Find the variance of your new estimator and show that is does indeed have smaller variance than U. (14)

Note: if $X \sim$ Poisson(λ) then

$$\mathrm{E}\left[s^X\right] = e^{\lambda(s-1)} \quad \text{for all } s \in \mathbb{R}$$

[34]

Question 2 is on the next page

2. (a) State and prove the Neyman-Pearson lemma on testing a simple null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against a simple alternative hypothesis $H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_1$. (13)

(b) For a random sample of size n from an exponential distribution with parameter θ and probability density function

$$f(x;\theta) = \theta e^{-\theta x}, \quad x > 0, \quad \theta > 0$$

show that the uniformly most powerful size-α test of

$$H_0 : \theta = \theta_0 \quad \text{against} \quad H_1 : \theta > \theta_0$$

involves rejecting H_0 for small values of $\sum_i x_i$. Begin by considering a simple alternative hypothesis. (8)

Explain briefly how the critical region is determined to give the correct size for the test – use the fact that if $X_1, X_2, \ldots, X_n$ are independent random variables from an exponential distribution with parameter θ, then (4)

$$\sum_{i=1}^{n} X_i \sim \Gamma(n, \theta).$$

If $n = 1$, show that the power function of the size-α test is

$$\pi(\theta) = 1 - (1 - \alpha)^{\theta/\theta_0}.$$

Sketch this power function and indicate on your plot what the power function for a sample of size $n = 10$ would look like (do not try to derive this merely show where you would expect it to be). (9)

[34]

Question 3 is on the next page

3. (a) Given a random sample of observations $x_1, x_2, \ldots, x_n$ from a single parameter probability density function $f(x;\theta)$, suppose that we wish to test the hypothesis $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$. Give the form of the test statistics for the following:

i) the log likelihood ratio test $(-2\log_e LR)$;

ii) the Wald test;

iii) the score test.

Explain the idea behind each test, using a diagram of the log-likelihood function for illustration, and describe how the tests are applied in practice, i.e. when do we reject H_0 and what reference distribution do we use? (9)

Find the form of each of these test statistics when the observations are from an exponential distribution with parameter θ and probability density function

$$f(x;\theta) = \theta e^{-\theta x}, \quad x > 0, \quad \theta > 0$$

(8)

(b) In a genetic breeding experiment using guinea pigs, 20 piglets of 60 offspring from a particular cross were red, 10 were black and 30 were white. Genetic theory suggests that these numbers should be in the proportions

$$\theta\,,\ \frac{(1-\theta)}{2}\,,\ \frac{(1-\theta)}{2}$$

where θ is an unknown parameter.

Write down the likelihood function using the multinomial distribution.

Use the likelihood ratio test to test whether the data are consistent with the genetic model at the 0.05 significance level, giving a clear statement of the null and alternative hypotheses.

Now use a likelihood ratio test, at the 0.05 level, to test

$$H_0 : \theta = \frac{1}{3} \quad \text{against} \quad H_1 : \theta \neq \frac{1}{3}.$$

Explain your conclusions from the two tests. (17)

Note: the k-category multinomial distribution is given by

$$\frac{n!}{n_1!n_2!\cdots n_k!}p_1^{n_1}p_2^{n_2}\cdots p_k^{n_k}, \quad 0 \leq n_i \leq n,\ \sum_{i=1}^{k} n_i = n,\ \sum_{i=1}^{k} p_i = 1$$

[34]

Question 4 is on the next page

4. (a) Over the period September 1996 to June 2001 Tiger Woods played 112 tournaments and won 32 of these. Writing the outcome for Woods as $x_i = 1$ for a win in the ith tournament and $x_i = 0$ otherwise, if we assume that he had a constant probability π of winning each tournament, explain why a simple model for the outcome x_i of each of n tournaments leads to the following log-likelihood for π

$$\ell(\pi) = \sum_{i=1}^{n} x_i \log(\pi) + \sum_{i=1}^{n}(1 - x_i)\log(1 - \pi)$$ (3)

i) Find general expressions for the maximum likelihood estimate of π, the observed information $I(\pi)$ and the expected information $i(\pi)$. (6)

ii) Evaluate these quantities for the observed outcomes of 32 wins in 112 tournaments. (4)

(b) Examination of the patterns of Tiger's wins suggests that something changed at around the 64th tournament – he won only 8 of the first 64 tournaments, but then went on to win 24 of the next 48 tournaments. Extend the model used above to have one success probability, π_1, for tournaments 1 to m and a changed probability $\theta\pi_1$ for the remaining $n - m$ tournaments.

i) Show that the log-likelihood is now

$$\begin{aligned}\ell(\pi_1, \theta) = &\sum_{i=1}^{m} x_i \log(\pi_1) + \sum_{i=1}^{m}(1 - x_i)\log(1 - \pi_1) \\ &+ \sum_{i=m+1}^{n} x_i \log(\theta\pi_1) + \sum_{i=m+1}^{n} (1 - x_i)\log(1 - \theta\pi_1)\end{aligned}$$

and that by writing $t_1 = \sum_{i=1}^{m} x_i$ and $t_2 = \sum_{i=m+1}^{n} x_i$ this can be expressed as

$$\ell(\pi_1, \theta) = t_1 \log(\pi_1) + (m - t_1)\log(1 - \pi_1) + t_2 \log(\theta\pi_1) + (n - m - t_2)\log(1 - \theta\pi_1)$$ (4)

ii) Find expressions for the joint maximum likelihood estimates of π_1 and θ, in terms of t_1 and t_2. (6)

iii) For the observed data on Tiger Woods, perform a likelihood-ratio test of the constant success probability model, i.e. test $H_0 : \theta = 1$ against $H_1 : \theta \neq 1$. (6)

iv) Obtain the profile log-likelihood $\ell(\pi_1, \hat{\theta}(\pi_1))$ for π_1 and show that, up to an additive constant, it only depends upon the data from the first m tournaments. What does this mean? (5)

[34]