Gx2050

OLLSCOIL NA hÉIREANN, GAILLIMH

NATIONAL UNIVERSITY OF IRELAND, GALWAY

SEMESTER II (SUMMER) EXAMINATIONS, 2003/2004

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.

Requirements:

Statistical Tables

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 $\boldsymbol{\theta}$;
- C is an *ancillary statistic* for $\boldsymbol{\theta}$.

(6)

(b) Consider a random sample $X_1, X_2, \ldots, X_n$ from the Bernoulli distribution with success probability θ, $\text{Bern}(\theta) \equiv \text{Bin}(1,\theta)$. Show that $\sum_{i=1}^n X_i$ is a minimal sufficient statistic for θ. (5)

Note: the probability mass function for a $\text{Bern}(\theta)$ distribution is

$$P(X = x) = \theta^x (1-\theta)^{1-x} \quad x = 0, 1.$$

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

Show that $Y = 2^{X_1}$ is an unbiased estimator of $1 + \theta$.

Considering the minimal sufficient statistic T, show that

$$P(Y = y \mid T = t) = \begin{cases} \dfrac{n-t}{n} & y = 1 \\ \dfrac{t}{n} & y = 2 \end{cases}$$

Hence, obtain an unbiased estimator of $1 + \theta$ with smaller variance than Y and compare its variance to

i) that of Y;

ii) the Cramér-Rao lower bound.

Note: You may assume that the variance of the Bernoulli distribution is $\theta(1-\theta)$. (16)

[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 $H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_1$. (16)

(b) For a random sample $X_1, X_2, \ldots, X_n$ of size n from a normal distribution $N(\mu, \sigma^2)$ what is the form of the best critical region for testing

$$H_0 : \mu = \mu_0,\ \sigma = \sigma_0 \quad \text{against} \quad H_1 : \mu = \mu_1,\ \sigma = \sigma_1$$

where $\mu_1 > \mu_0$ and $\sigma_1 > \sigma_0$? (6)

By taking $\sigma_1 = \sigma_0$, or otherwise, show that the uniformly most powerful size-α test of

$$H_0 : \mu = \mu_0 \quad \text{against} \quad H_1 : \mu > \mu_0,$$

with σ assumed known, involves rejecting H_0 for large values of $\overline{X}$. (6)

Obtain an explicit expression for the critical region of a size-α test and derive and sketch the power function. (6)

[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 these test statistics when the observations are from a Poisson distribution with mean θ.

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

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

(8)

(b) In a genetic breeding experiment using guinea pigs, of 56 offspring from a particular cross there were 32 red, 8 black and 16 white piglets. Genetic theory suggests that these numbers should be in the proportions

$$\frac{2(1+\theta)}{4}, \frac{(1-2\theta)}{4}, \frac{1}{4}$$

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

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

Note: the k-category multinomial distribution is given by (17)

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

[34]

Question 4 is on the next page

4. Assume that the number of winter moth larvae on oak trees in a forest has a Poisson distribution with parameter λ. In a study of these a biologist counted the number of larvae $y_1, y_2, \ldots, y_n$ on a random selection of n small oak trees.

   (i) Obtain the log-likelihood function of λ and hence the maximum likelihood estimate $\widehat{\lambda}$. (4)

   (ii) Calculate the observed information $I(\lambda)$ and the expected information $i(\lambda)$, and hence obtain an approximate standard error for $\widehat{\lambda}$. (6)

   Growing tired of this extensive method of data collection the biologist decided to just record whether any moth larvae were present or not on a further random selection of m trees. However, being tired and rather careless his chance of making correct observations was affected. Writing X as the binary random variable for whether moth larvae were recorded as present ($x = 1$) or not ($x = 0$), assume that

$$P(X = x) = \left(1 - \omega e^{-\lambda}\right)^x \left(\omega e^{-\lambda}\right)^{1-x} = \begin{cases} \omega e^{-\lambda} & x = 0 \\ 1 - \omega e^{-\lambda} & x = 1 \end{cases}$$

   where ω is an additional parameter reflecting the accuracy of the biologist's observations ($\omega = 1$ corresponds to perfect observation).

   This further sample of trees led to observations $x_1, x_2, \ldots, x_m$.

   (a) Show that the log-likelihood function of λ and ω based on both sets of measurements can be written as

$$\ell(\lambda, \omega) = c - n\lambda + \sum_{i=1}^{n} y_i \log_e(\lambda) + z \log_e\left(1 - \omega e^{-\lambda}\right) + (m - z) \log_e\left(\omega e^{-\lambda}\right)$$

   where c is a constant not depending on the parameters and $z = \sum_{j=1}^{m} x_j$. (4)

   (b) Find the joint maximum likelihood estimates of λ and ω. Comment on the answers. (6)

   (c) Calculate the observed information matrix $I(\lambda, \omega)$ and obtain the expected information matrix. (6)

   (d) Obtain the profile log-likelihood $\ell(\lambda, \widehat{\omega}(\lambda))$ for λ and show that, up to an additive constant, this only depends upon the data from the first experiment. (4)

   (e) Suppose that the first part of the data collection examined 100 trees and counted a total of 120 moth larvae. Perform a likelihood ratio test of $H_0 : \lambda = 1$ against $H_1 : \lambda \neq 1$. (4)

[34]

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

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