EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

GRADUATE DIPLOMA, 2010

MODULE 2 : Statistical Inference

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* ${}^{n}C_{r}$.

This examination paper consists of 6 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.

© RSS 2010

1. The times in seconds T_1, T_2, …, T_n between messages arriving at a node in a telecommunications system constitute a random sample from an exponential distribution with probability density function $f(t) = \mu^{-1}e^{-t/\mu}$ ($t > 0$), where μ (> 0) is an unknown parameter.

(i) Show that $\sum_{i=1}^{n} T_i$ is a sufficient statistic for μ. (3)

(ii) Show that $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} T_i$ is an unbiased estimator of μ and find its variance. (5)

(iii) Is $\hat{\mu}_n$ a consistent estimator of μ? Justify your answer. (2)

(iv) Show that $E\left(\sqrt{T_i}\right) = \tfrac{1}{2}\sqrt{\pi\mu}$ and hence show that $\tilde{\mu} = \frac{4}{\pi}\sqrt{T_1T_2}$ is an unbiased estimator of μ. (6)

(v) Find the relative efficiency of $\tilde{\mu}$ compared to $\hat{\mu}_2$. (4)

2. The number of eggs laid by a breeding female of a certain species of sea bird has a Poisson distribution with mean λ (> 0) and is independent for different birds. An ornithologist wants to estimate λ and examines a number of possible nests. She can only be sure that the nest belongs to the correct species if there is at least one egg in it. The number of eggs in the ith nest identified as belonging to this species is denoted by X_i; by the end of the day, she has identified n (> 0) such nests. (Note that $X_i > 0$ for $i = 1, 2, …, n$.)

(i) Obtain the value of $P(X_i = k)$ for $k = 1, 2, 3, …$.

(3)

(ii) Find an equation for determining $\hat{\lambda}$, the maximum likelihood estimator of λ based on $X_1, X_2, …, X_n$.

(6)

(iii) Find the asymptotic variance of $\hat{\lambda}$.

(6)

(iv) Describe a numerical method for determining $\hat{\lambda}$. Demonstrate one iteration of the method when $n = 30$, $\sum X_i = 50$ and the initial estimate of λ is 2.0.

(5)

3. (a) Suppose that $X_1, X_2, \ldots, X_n$ constitute a random sample from a distribution with probability density $f(x \mid \theta)$, where θ is a real parameter with prior density $\pi(\theta)$. The loss when θ is estimated by $\hat{\theta}$ (a function of $X_1, X_2, \ldots, X_n$) is $\ell(\hat{\theta}; \theta)$.

(i) Define the *Bayes risk* of $\hat{\theta}$ and the *Bayes estimator* of θ.

(3)

(ii) Show that the Bayes estimator can be found by minimising the posterior expected loss for given $(x_1, x_2, \ldots, x_n)$.

(4)

(iii) Suppose now that $\ell(\hat{\theta}; \theta) = \left|\hat{\theta} - \theta\right|$, the absolute value loss function. Show that the Bayes estimator is the median of the posterior distribution of θ.

(5)

(b) An interval estimator (m_1, m_2) (where $m_1 < m_2$) is required for the parameter μ of a continuous distribution. The posterior distribution of μ is N(100, 1) and the loss associated with an estimator (m_1, m_2) is $m_2 - m_1$ if $m_1 \le \mu \le m_2$ and $5 + m_2 - m_1$ if $\mu < m_1$ or $\mu > m_2$. Show that the Bayes rule is to choose $m_1 \approx 98.825$ and $m_2 \approx 101.175$. [You may assume that $\phi(1.175) \approx 0.2$, where $\phi(.)$ is the probability density function of the standard Normal distribution.]

(8)

4. (a) Explain what is meant by the *most powerful test at level α* between two simple hypotheses.

(3)

(b) The strength Y of a product has the Weibull distribution with probability density $f(y) = \theta\phi y^{\phi-1}\exp(-\theta y^{\phi})$ for $y > 0$, where θ (> 0) is unknown and ϕ (> 0) is known. The strengths of a random sample of n products are $Y_1, Y_2, \ldots, Y_n$.

(i) Find the form of the most powerful test of the null hypothesis $\theta = 0.5$ against the alternative hypothesis $\theta = 1.0$.

(6)

(ii) Show that Y_i^{ϕ} has an exponential distribution.

(3)

(iii) Use χ^2 tables to find the critical region of the most powerful test at the 5% level when $n = 20$. [You may assume the result that if $X_1, X_2, \ldots, X_m$ are independent, each with an exponential distribution, mean μ, then $2\mu^{-1}\sum X_i$ has the χ^2_{2m} distribution.]

(4)

(iv) Use χ^2 tables and linear interpolation to evaluate the approximate power of the test found in part (iii).

(4)

5. A new measuring device is being tested under 3 different conditions, and it is assumed that its errors have a Normal distribution with mean 0 and variance σ_i^2 under condition i (i = 1, 2, 3). Let X_{ij} be the error for the jth measurement under the ith condition ($j = 1, 2, \ldots, n_i$; i = 1, 2, 3). It can be assumed that all measurements are independent. It is required to test the null hypothesis that $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$ against the alternative that the variances are not all equal.

(i) Find the maximum likelihood estimators of σ_1^2, σ_2^2 and σ_3^2.

(6)

(ii) Derive the generalised likelihood ratio test in as simple a form as possible.

(9)

(iii) Carry out the test at approximately the 5% level in the case $n_1 = 50$, $n_2 = n_3 = 40$, $\sum_{j=1}^{50} X_{1j}^2 = 5.5$, $\sum_{j=1}^{40} X_{2j}^2 = 3.6$ and $\sum_{j=1}^{40} X_{3j}^2 = 5.2$.

(5)

6. (a) Explain what is meant by *nonparametric inference* and discuss its advantages and disadvantages compared to parametric inference.

(7)

(b) A random sample of 12 observations is available from a continuous distribution whose (unknown) median is m.

(i) Find the critical region of the sign test of the null hypothesis $m = 20$ against the alternative hypothesis $m \neq 20$, when the significance level is as near as possible to 5% but is no more than 5%.

(7)

(ii) Use the sign test to find an approximate 95% confidence interval for m.

(6)

7. Independent random samples $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ have been taken from Normal distributions, both with known variance σ^2 (> 0) but with means μ ($\gg \sigma$) and $\alpha\mu$ respectively, where μ and α (> 0) are unknown.

(i) By considering the distribution of $\bar{X} - (\bar{Y}/\alpha)$, find a pivotal quantity for α. State its distribution and explain why it is pivotal for α.

(5)

(ii) Show that the maximum likelihood estimator of α is $\hat{\alpha} = \bar{Y}/\bar{X}$. [You need only consider the first partial derivatives of the log likelihood.]

(6)

(iii) Use your answers to parts (i) and (ii) to show that an approximate 95% confidence interval for α is given by $\dfrac{\bar{Y}}{\bar{X}} \pm 1.96\,\dfrac{\sigma}{\bar{X}}\sqrt{\dfrac{1}{n}\left(1+\left(\dfrac{\bar{Y}}{\bar{X}}\right)^2\right)}$.

(3)

(iv) Explain how a 95% bootstrap confidence interval for α based on the maximum likelihood estimator can be constructed.

(6)

8. (a) A random sample of data has been obtained from a distribution with unknown parameter η, and the null hypothesis $\eta = \eta_0$ is to be tested against the alternative $\eta = \eta_1$, where $\eta_0 \neq \eta_1$. Explain what are meant by the *prior* and *posterior odds* of the null hypothesis and by the *Bayes factor*. Show how these three quantities are related.

(5)

(b) Let $X_1, X_2, \ldots, X_{20}$ be a random sample from a geometric distribution with $P(X_i = k) = \theta(1-\theta)^k$ for $k = 0, 1, 2, \ldots$, where θ ($0 < \theta < 1$) is an unknown parameter. It is found that $\sum X_i = 40$.

(i) It is required to test the null hypothesis $\theta = 0.5$ against the alternative hypothesis $\theta = 0.25$, the prior probability of the null hypothesis being 0.75. Evaluate the Bayes factor and the posterior odds of the null hypothesis.

(5)

(ii) Suppose now that it is required to test the null hypothesis $\theta = 0.5$ against the alternative hypothesis $\theta \neq 0.5$, where the prior distribution of θ under the alternative hypothesis is

$$\pi(\theta) = 12\theta^2(1-\theta), \quad 0 < \theta < 1 .$$

Show that the Bayes factor is equal to

$$\frac{1}{12}\left(\frac{1}{2}\right)^{60}\frac{64!}{22!41!}.$$

[Hint: the beta distribution, parameters α (> 0) and β (> 0), has probability density function $\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1}$ for $0 \le y \le 1$.]

(5)

(iii) Show how the Bayes factor evaluated in part (ii) is related to $P(W = 22)$, where W has the binomial distribution with $n = 64$ and $p = 0.5$. By using the Normal approximation to the binomial distribution, find the approximate value of the Bayes factor.

(5)