



EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

GRADUATE DIPLOMA, 2015

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 nC_r .

This examination paper consists of 8 printed pages.

This front cover is page 1.

Question 1 starts on page 2.

There are 8 questions altogether in the paper.

1. Independent failure times T_1, T_2, \dots, T_n come from a continuous distribution with probability density function

$$f(t) = \frac{1}{2\beta} t^{-\frac{1}{2}} \exp\left(-\frac{1}{\beta} t^{\frac{1}{2}}\right) \quad \text{for } t > 0,$$

where $\beta (> 0)$ is an unknown parameter.

- (i) Find the method of moments estimator, $\tilde{\beta}$, of β . [Hint: note that the gamma function can be written $\Gamma(\alpha) = \int_0^\infty v^\alpha x^{\alpha-1} e^{-vx} dx$ and that $\Gamma(\alpha) = (\alpha-1)!$ if α is a positive integer.] (6)
- (ii) Find the maximum likelihood estimator, $\hat{\beta}$, of β . (6)
- (iii) Show that $\hat{\beta}$ is an unbiased estimator of β . (4)
- (iv) Find the Cramér-Rao lower bound for the variance of unbiased estimators of β and show that $\hat{\beta}$ is an efficient estimator of β . (4)

2. A single observation X is taken from the distribution with probability density function

$$f(x) = 2e^{x-\lambda} (1+e^{x-\lambda})^{-3} \quad \text{for } -\infty < x < \infty,$$

where λ is an unknown parameter.

- (i) Show that the cumulative distribution function (c.d.f.) of X is
- $$F(x) = 1 - (1 + e^{x-\lambda})^{-2} \quad \text{for } -\infty < x < \infty. \quad (2)$$
- (ii) Find the c.d.f. of $Y = X - \lambda$ and say why Y is a pivotal quantity. (4)
- (iii) Find a 90% confidence interval for λ in terms of X based on the central 90% region of the distribution of Y (i.e. using the interval between the 5th and 95th percentiles of Y). (8)
- (iv) It is required to test the null hypothesis $\lambda = 1$ against the alternative hypothesis $\lambda > 1$ at the 5% level using a critical region of the form $X > k$ for some value of k . What value must be chosen for k and what is the power of this test at $\lambda = 5$?

3. Given a model involving a parameter θ , suppose that the likelihood obtained from a set of data is given by $L(\theta)$ but no simple expression can be found for the maximum likelihood estimate $\hat{\theta}$. Describe the Newton-Raphson method for finding the value of $\hat{\theta}$ numerically. (6)

(4)

X_1, X_2, \dots, X_n is a random sample from a population with probability density function

$$f(x) = \frac{3x^2}{\alpha^3} \exp\left(-\frac{x^3}{\alpha^3}\right) \quad \text{for } x > 0,$$

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

- (i) Show that $E(X^3) = \alpha^3$. (4)
- (ii) Find $\hat{\alpha}$, the maximum likelihood estimator of α . (6)
- (iii) Using an asymptotic result, find the approximate distribution of $\hat{\alpha}$ when n is large. (3)
- (iv) Hence calculate an approximate 95% confidence interval for α when $n = 200$ and $\sum X_i^3 = 1600$. (3)

4. The proportion of visits to a website during a day which result in a sale is a random variable X with probability density function

$$f(x) = \phi(1-x)^{\phi-1} \quad \text{for } 0 < x < 1,$$

where $\phi (> 0)$ is an unknown parameter. The proportions X_1, X_2, \dots, X_n for a random sample of days have been observed. It is required to test the null hypothesis $H_0: \phi = 0.5$ against the alternative hypothesis $H_1: \phi = 1$.

- (i) Find the form of the most powerful test, expressed in terms of $\frac{Y}{n}$, where $Y = \sum \log(1 - X_i)$.

(5)

- (ii) Use integration to show that, under the null hypothesis, $E(\log(1 - X_i)) = -2$ and $E[(\log(1 - X_i))^2] = 8$.

(6)

- (iii) Using the central limit theorem and the results given in part (ii), show that, under the null hypothesis, when n is large $\frac{Y}{n}$ is approximately Normally distributed with mean -2 and variance $\frac{4}{n}$.

(3)

- (iv) Find, in terms of $\frac{Y}{n}$, the most powerful test with approximate size 0.05 when n is large.

(3)

- (v) Show that there is not a uniformly most powerful test of $H_0: \phi = 0.5$ against $H_1: \phi \neq 0.5$.

(3)

5. The amount of oil, in suitable units, recoverable from a test well has a distribution with probability density function given by

$$f(x) = \frac{\theta 2^\theta}{x^{\theta+1}} \quad \text{for } x > 2$$

and is zero otherwise, where $\theta (> 0)$ is an unknown parameter. The amounts of oil recoverable from a random sample of tests are X_1, X_2, \dots, X_n . It is required to test the null hypothesis $\theta = 2$ against the alternative hypothesis $\theta \neq 2$.

- (i) Show that the maximum likelihood estimator $\hat{\theta}$ of θ satisfies the equation

$$\sum \log(X_i) = \frac{n}{\hat{\theta}} + n \log 2. \tag{5}$$

- (ii) Hence show that

$$\hat{\theta} = \frac{n}{\sum \log(0.5 X_i)}. \tag{2}$$

- (iii) Using the result in part (i), show that the generalised likelihood ratio test statistic λ satisfies

$$\log \lambda = n \log\left(\frac{2}{\hat{\theta}}\right) + n - \frac{2n}{\hat{\theta}}. \tag{8}$$

- (iv) Suppose now that n is large and that the size of the test is to be 0.05. Show that the acceptance region is

$$\log \hat{\theta} + \frac{2}{\hat{\theta}} \leq \frac{1.92}{n} + 1 + \log 2. \tag{5}$$

6. A sample of n independent measurements is drawn from a symmetric distribution, so that the mean and median of the distribution, m , are equal. Describe how a Wilcoxon signed-rank test can be used to test the null hypothesis $m = m_0$ against the alternative hypothesis $m \neq m_0$ at the 5% level, where m_0 is a given value. Include in your answer a discussion of the use of tables and of a large-sample formula.

(7)

The ages, in years, of six randomly selected purchasers of a particular product are 20, 41, 24, 61, 32 and 48. The distribution of ages can be assumed to be symmetric.

- (i) Use a Wilcoxon signed-rank test to test the null hypothesis $m = 26$ against the alternative hypothesis $m \neq 26$ at the 10% level.

(6)

- (ii) Explain how a 90% confidence interval for m can be obtained from this test and investigate whether the following values are in this interval:

(a) 23.9;

(b) 50.9.

(7)

7. X_1, X_2, \dots, X_n is a random sample from the Poisson distribution with unknown mean $\lambda (> 0)$.

(i) State the mean and variance of $\sum X_i$ and hence show that

$$E\left(\left(\sum X_i\right)^2\right) = n\lambda + n^2\lambda^2. \quad (3)$$

(ii) It is required to estimate $\theta = \lambda^2$ and the following estimator has been proposed:

$$\hat{\theta} = \left(\frac{\sum X_i}{n}\right)^2.$$

Show that the jack-knife estimator of θ based on $\hat{\theta}$ is

$$\hat{\theta}_J = \frac{\left(\sum X_i\right)^2 - \sum X_i^2}{n(n-1)}. \quad (6)$$

(iii) Show that $\hat{\theta}_J$ is an unbiased estimator of θ . (3)

Suppose now that the prior distribution of λ is gamma, with parameters $k (> 0)$ and $\nu (> 0)$. [You may use the results that the gamma distribution with parameters k and ν has probability density function $f(y) = \frac{\nu^k y^{k-1} e^{-\nu y}}{\Gamma(k)}$ and has mean $\frac{k}{\nu}$ and variance $\frac{k}{\nu^2}$.]

(iv) Find the posterior distribution of λ . (5)

(v) Assuming quadratic loss, find the Bayes estimator of θ . (3)

8. (a) Describe how computer Monte Carlo simulation can be used to
- (i) compare estimators, (4)
 - (ii) draw inferences in Bayesian analysis. (4)
- (b) Each item on a production line is given a quick test which has two possible results: s_1 (appears satisfactory) and s_2 (appears unsatisfactory). However, the test is itself prone to error so that if the item is satisfactory, $P(s_1) = 0.9$ and $P(s_2) = 0.1$, while if the item is unsatisfactory, $P(s_1) = 0.4$ and $P(s_2) = 0.6$. After each item is inspected, it is either sold or scrapped. If a satisfactory item is sold, there is a net loss of -2 units (i.e. a profit of 2 units), while if an unsatisfactory item is sold there is a penalty resulting in a net loss of 10 units. Any item that is scrapped results in a net loss of 1 unit.
- (i) List the four decision rules for deciding whether each item should be scrapped. (2)
 - (ii) Calculate the risk table. (8)
 - (iii) State, with reasons, which is the minimax rule. (2)