EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

GRADUATE DIPLOMA, 2005

Statistical Theory and Methods II

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.

1. Let $X_1, X_2, ..., X_n$ be a random sample from a population with probability density function

$$f(x) = \sqrt{\frac{2}{\pi\theta}} \exp\left(-\frac{x^2}{2\theta}\right), \qquad x > 0.$$

(i) Show that $E(X^2) = \theta$.

(4)

(ii) Find the maximum likelihood estimator (MLE), $\hat{\theta}$, of θ.

(5)

(iii) Show that $\hat{\theta}$ is an unbiased estimator of θ and that the Cramér-Rao lower bound is attained. [You may assume that $\text{Var}(X^2) = 2\theta^2$.]

(6)

(iv) Suppose now that $\phi = \sqrt{\theta}$ is the parameter of interest. Without undertaking further calculations, write down the MLE of ϕ and explain briefly why it is a biased estimator of ϕ.

(5)

2. Let $X_1, X_2, ..., X_n$ be a random sample from a population that is uniformly distributed on the interval $(0, \theta)$, where the parameter θ is positive.

(i) Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Show that $\tilde{\theta} = 2\bar{X}$ is the method of moments estimator of θ and that it is unbiased.

(5)

(ii) Consider the random sample 0.2, 0.3, 1.0, 0.1 from the above population. Evaluate $\tilde{\theta}$ and comment on the usefulness, or otherwise, of the estimate.

(4)

(iii) Let $Y = \max X_i$. Show that the probability density function of Y is

$$g(y) = \frac{ny^{n-1}}{\theta^n}, \qquad 0 < y < \theta.$$

(5)

(iv) An estimator $\hat{\theta} = cY$ is to be used to estimate θ, where the multiplier c is to be chosen. Show that the mean square error of $\hat{\theta}$ is minimised when $c = \dfrac{n+2}{n+1}$.

(6)

3. Let $X_1, X_2, ..., X_n$ be a random sample from a population with probability density function $f(x;\theta)$, where θ is a parameter. Let $S = S(X_1, X_2, ..., X_n)$ be a sufficient statistic for θ.

(i) What can be said about the conditional distribution of $X_1, X_2, ..., X_n$ given that $S = s$?

(4)

(ii) State the factorisation theorem for sufficient statistics.

(5)

(iii) Suppose now that

$$f(x;\theta) = \frac{x^{\theta-1}e^{-x}}{\Gamma(\theta)}, \qquad x > 0,$$

where $\Gamma(.)$ is the gamma function and $\theta > 0$ is a positive parameter. Show that $S = \sum_{i=1}^{n} \log X_i$ is a sufficient statistic for θ.

(5)

(iv) In the situation given in (iii) it is required to test the null hypothesis H_0: $\theta = 1$ against the alternative hypothesis H_1: $\theta = 2$. Use the Neyman-Pearson method to give the form of the most powerful test of a given size, in terms of S.

(6)

4. A random sample of 100 observations is drawn from a continuous distribution whose median, θ, is of interest.

(i) Describe the sign test, and the corresponding Normal approximation, for testing hypotheses about θ.

(5)

(ii) It is required to test the null hypothesis H_0: $\theta = 17$ against the alternative hypothesis H_1: $\theta \neq 17$ using a 5% significance level. Using a Normal approximation, find the critical region for the sign test.

(5)

(iii) Explain what is meant by a non-parametric confidence interval.

(5)

(iv) Find an approximate 95% non-parametric confidence interval for θ of the form $(X_{(a)}, X_{(b)})$, where $X_{(i)}$ denotes the ith order statistic, $i = 1, 2, \ldots, 100$.

(5)

5. Let X_1, X_2, …, X_n be a random sample from a Bernoulli distribution with success probability θ, so that $P(X = 1) = \theta$ and $P(X = 0) = 1 - \theta$. Suppose that θ has the following prior probability density function (pdf):

$$g(\theta) = 6\theta(1-\theta), \qquad 0 < \theta < 1.$$

(i) You are given that the beta distribution with positive parameters a and b has pdf

$$f(y) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1}(1-y)^{b-1}, \qquad 0 < y < 1.$$

Show that the posterior distribution for θ is beta with parameters $2 + \sum_{i=1}^{n} X_i$ and $n + 2 - \sum_{i=1}^{n} X_i$.

(5)

(ii) Using a squared error loss function, show that the Bayes estimator of θ is

$$\frac{2 + \sum_{i=1}^{n} X_i}{4 + n}.$$

(8)

(iii) Find the bias of this estimator and calculate its mean square error.

(7)

6. A random sample of observations $x_1, x_2, \ldots, x_n$ is available from a distribution with probability density function (pdf)

$$f(x \mid \theta) = 2\theta x \exp(-\theta x^2), \qquad x > 0,$$

where θ has a prior gamma distribution with mean $\frac{a}{b}$ and variance $\frac{a}{b^2}$, i.e. its pdf is of the form

$$g(\theta) \propto \theta^{a-1} \exp(-b\theta), \qquad \theta > 0.$$

(i) Show that the gamma distribution is a conjugate prior.

(5)

(ii) Write down the posterior mean and posterior standard deviation of θ.

(4)

(iii) Suppose that $n = 48$, $a = b = 1$ and $\sum_{i=1}^{n} x_i^2 = 48.0$. Sketch the prior and posterior pdfs of θ. [You may assume that the gamma distribution with parameters a and b is approximately Normal when a is large.]

(8)

(iv) Using a Normal approximation, calculate a Bayesian 95% posterior interval for θ.

(3)

7. (a) Data from a distribution indexed by a real parameter θ are available. It is required to test a simple null hypothesis against a simple alternative hypothesis. Two possible procedures are to use a fixed sample size for a Neyman-Pearson test and to use a sequential probability ratio test. Explain how tests of these types may be constructed and discuss briefly their relative strengths and weaknesses.

(10)

(b) A random sample of individuals is taken from a population. The reaction time of each individual to a stimulus is measured immediately before and one hour after a drug treatment. It is of interest to test whether the drug treatment has had an effect on the reaction times. One statistician proposes using a paired t test. Another statistician recommends using a Wilcoxon signed-rank test. Compare and contrast these two tests.

(10)

8. A random sample of size n has yielded a sample mean of zero and a sample standard deviation of one. A further observation is taken, yielding a value y.

(i) Show that the sample mean and sample variance of the augmented sample are $\dfrac{y}{n+1}$ and $1-\dfrac{1}{n}+\dfrac{y^2}{n+1}$ respectively.

(6)

(ii) Let μ be the mean of the population and let t denote the usual t statistic for testing $H_0:\ \mu=\mu_0$ against $H_1:\ \mu\neq\mu_0$. Evaluate t for the augmented sample of size $n+1$ and show that $|t|\to 1$ as $|y|\to\infty$.

(7)

(iii) Use the results obtained above to discuss the effect of an outlier on the one-sample t test.

(7)