EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

HIGHER CERTIFICATE IN STATISTICS, 2013

MODULE 5 : Further probability and inference

Time allowed: One and a half hours

Candidates should answer **THREE** *questions.*

Each question carries 20 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 4 questions altogether in the paper.

1. (a) Random variables X_1 and X_2 have a bivariate Normal distribution with $E(X_1) = 10$, $E(X_2) = 9$, $\text{Var}(X_1) = 16$, $\text{Var}(X_2) = 9$ and $\text{Corr}(X_1, X_2) = -0.5$.

(i) Find $E(X_1 - X_2)$ and $\text{Var}(X_1 - X_2)$.

(4)

(ii) Using your answer to part (i) and the table of the Normal cumulative distribution function, find $P(X_1 > X_2)$.

(3)

(b) An internet consulting business has developed a score, S, for visitors to a shopping website which indicates their propensity to make purchases. Visitors with a score $S = 1$ have a 70% chance of making no purchases and a 30% chance of making one purchase; visitors with a score $S = 2$ have a 50% chance of making no purchases, a 40% chance of making one purchase and a 10% chance of making two purchases; visitors with a score $S = 3$ have a 30% chance of making no purchases, a 50% chance of making one purchase and a 20% chance of making two purchases. It is found that 70% of visitors have score $S = 1$, 20% have $S = 2$ and 10% have $S = 3$. Let Y ($Y = 1, 2, 3$) be the number of purchases made by a randomly selected visitor to the website.

(i) Construct a table showing the joint distribution of S and Y.

(6)

(ii) Find the marginal distribution of Y.

(1)

(iii) Find $\text{Cov}(S, Y)$.

(4)

(iv) A visitor to the website has made no purchases. What is the probability that this visitor had score $S = 3$?

(2)

2. The time, T, to an event has the exponential distribution with mean λ^{-1} ($\lambda > 0$) i.e. $f(t) = \lambda e^{-\lambda t}$ for $t > 0$. The random variable X takes the value 0 if $T \leq 1$ and the value 1 if $T > 1$.

(i) Using integration, evaluate $P(T \leq t \mid X = 1)$ for $t > 1$ and hence show that the conditional distribution of T given $X = 1$ has probability density function

$$\lambda e^{-\lambda(t-1)} \quad \text{for } t > 1.$$

(7)

(ii) Show that the moment generating function of the distribution found in part (i) is

$$m(s) = \frac{\lambda e^{s}}{\lambda - s} \quad \text{for } s < \lambda.$$

(3)

(iii) Using the moment generating function found in part (ii), find the mean and variance of the conditional distribution of T given $X = 1$.

(6)

(iv) Find the probability density function of the conditional distribution of T given $X = 0$.

(4)

3. The independent discrete random variables $X_1, X_2, \ldots, X_n$ each have the probability distribution

$$P(X = x) = \frac{p}{2^x} + \frac{2(1-p)}{3^x} \quad \text{for } x = 1, 2, 3, \ldots,$$

where p $(0 < p < 1)$ is an unknown parameter.

(i) Write down the likelihood function of p.

(1)

(ii) Show that the first derivative of the log likelihood is

$$\left\{ \sum_{i=1}^{n} \frac{\frac{1}{2^{x_i}} - \frac{2}{3^{x_i}}}{\frac{p}{2^{x_i}} + \frac{2(1-p)}{3^{x_i}}} \right\}$$

(2)

(iii) Find the second derivative of the log likelihood and show that it must be strictly less than zero for all values of p between 0 and 1.

(4)

(iv) For a particular set of data from the distribution above, the first and second derivatives of the log likelihood have been evaluated at several values of p and are shown in the table below. By drawing appropriate diagrams and explaining your reasoning carefully, find the maximum likelihood estimate of p and an approximate 90% confidence interval for p. (You may assume that regularity conditions hold.)

p	0.10	0.12	0.14	0.16
First derivative of log likelihood	10.80	4.18	–1.92	–7.58
Second derivative of log likelihood	–345.3	–317.7	–293.5	–273.1

(13)

4. Random variables X_1, X_2 and X_3 are independent observations from a distribution with probability density function $f(x)$ where

$$f(x) = \begin{cases} 1 & \text{for } \theta - 0.5 < x < \theta + 0.5 \\ 0 & \text{otherwise} \end{cases}$$

where θ is an unknown parameter.

(i) Find the method of moments estimator of θ.

(4)

(ii) Show that this estimator is unbiased and find its variance.

(6)

Let Y be the median of X_1, X_2 and X_3. It can be shown that the probability density function of Y is $g(y)$ where

$$g(y) = 1.5 - 6(y - \theta)^2 \quad \text{for } \theta - 0.5 < y < \theta + 0.5.$$

(iii) Show that Y is an unbiased estimator of θ.

[Hint. First evaluate $E(Y - \theta)$, using the change of variable technique when evaluating the integral.]

(4)

(iv) By evaluating $E(Y - \theta)^2$, or otherwise, find Var(Y) and the efficiency of Y relative to the method of moments estimator.

(6)

BLANK PAGE

BLANK PAGE

BLANK PAGE