EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

HIGHER CERTIFICATE IN STATISTICS, 2011

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

This examination paper consists of 5 printed pages, **each printed on one side only**.
This front cover is page 1.
Question 1 starts on page 2.

There are 4 questions altogether in the paper.

1. The continuous random variables X and Y are jointly distributed with joint probability density function

$$f(x, y) = kxy \quad (0 \le x \le 1,\ 0 \le y \le 2 - 2x)$$

and zero elsewhere, where k is a constant.

(i) Sketch the region where the joint density is non-zero.

(2)

(ii) Use integration to show that $k = 6$.

(5)

(iii) Find the marginal probability density function of X and use it to show that $P\left(X \le \tfrac{1}{2}\right) = \tfrac{11}{16}$.

(5)

(iv) Find $f(y \mid x)$, the conditional probability density function of Y given x, and use it to evaluate $P\left(Y \le \tfrac{1}{2} \mid X = \tfrac{1}{2}\right)$.

(4)

(v) Evaluate $P\left(Y \le \tfrac{1}{2} \mid X \le \tfrac{1}{2}\right)$.

(4)

Turn over

2. (a) The random variables X and Y are jointly distributed with a bivariate Normal distribution.

(i) Sketch a typical scatter plot of data from this distribution.

(2)

(ii) Define the five parameters usually used to specify this distribution.

(2)

(iii) State the names of the marginal distributions of X and Y.

(2)

(iv) What is the form of the conditional mean of Y given $X = x$, considered as a function of x?

(2)

(b) Suppose that the random variables V and W have $\mathrm{N}(\mu_1, \sigma_1^2)$, $\mathrm{N}(\mu_2, \sigma_2^2)$ distributions respectively and that V and W are independent.

(i) Given that the moment generating function of a $\mathrm{N}(\mu, \sigma^2)$ random variable is $\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$, find the distribution of $S = V + W$.

(4)

(ii) State the distribution of $U = V - W$.

(1)

(iii) Find $E(SU)$ and $\mathrm{Cov}(S, U)$.

(4)

(iv) State the name of the joint distribution of S and U and give the values of the parameters of this distribution in terms of μ_1, μ_2, σ_1^2 and σ_2^2.

(3)

3. (a) Explain what is meant by the *likelihood function*, and why it may be useful in estimating the value of a parameter.

(5)

(b) The continuous random variable X has probability density function

$$f(x) = \frac{\theta^3 x^2 e^{-\theta x}}{2} \qquad (x > 0)$$

where $\theta > 0$ is an unknown parameter. A random sample of values $X_1, X_2, \ldots, X_n$ is available from this distribution.

(i) Show that the maximum likelihood estimator of θ is $\hat{\theta} = \frac{3}{\overline{X}}$, where $\overline{X}$ is the sample mean of $X_1, X_2, \ldots, X_n$.

(5)

(ii) Find the approximate distribution of $\hat{\theta}$ when n is large and use this result to find an approximate 95% confidence interval for θ when $n = 200$ and $\overline{X} = 6.0$.

(5)

(iii) Show that $\hat{\theta}$ is a biased estimator of θ when $n = 1$.

(5)

4. The discrete random variable X has probability distribution given by

$$P(X=k)=(k+1)\,p^2(1-p)^k \qquad (k=0,1,2,3,\ldots)$$

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

(i) Show that the probability generating function of X is given by

$$\pi(t)=\frac{p^2}{\left(1-t(1-p)\right)^2} \qquad (\text{for } |t|<(1-p)^{-1}).$$

[You may use the results that, for $|a|<1$, $\sum_{k=0}^{\infty} a^k = \frac{1}{1-a}$ and $\sum_{k=0}^{\infty} k a^k = \frac{a}{(1-a)^2}$.]

(5)

(ii) Use the probability generating function to find the mean and variance of X.

(6)

(iii) A random sample X_1, X_2, …, X_n is available from this distribution. Find the method of moments estimator of p.

(3)

(iv) Let Y_i take the value 1 if $X_i = 0$, and the value 0 otherwise, for $i = 1, 2, \ldots, n$. State the distribution of $U=\sum_{i=1}^{n} Y_i$.

(2)

(v) Find an unbiased estimator of p^2 based on U and show that this estimator is consistent. [You do not need to prove any results you use concerning the moments of standard distributions.]

(4)