EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

GRADUATE DIPLOMA, 2002

Statistical Theory and Methods I

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 silent, cordless, non-programmable electronic calculators.

Where a calculator is used the ***method*** *of calculation should be stated in full.*

Note that $\binom{n}{r}$ *is the same as* ${}^nC_r$ *and that* ln *stands for* $\log_e$.

This examination paper consists of 10 printed pages. This front cover is page 1. The reverse of the front cover, which is intentionally left blank, is page 2. Question 1 starts on page 3.

There are 8 questions altogether in the paper.

1. (i) If the random variable T can take only non-negative values, then its survivor function, $S(t)$, is defined by

$$S(t) = P(T > t), \qquad t \geq 0.$$

When T is a continuous random variable, that takes only non-negative values, prove that

$$\int_0^\infty S(t)\,dt = E(T).$$

[Hint. Change the order of integration in a double integral.]

(5)

(ii) A random variable X has cumulative distribution function $F(x)$ given by

$$F(x) = \begin{cases} 0 & x < 1, \\ 1 - xe^{-(x-1)} & x \geq 1. \end{cases}$$

Using the result proved in part (i), show that $E(X) = 3$.

[You may use the fact that

$$\int_0^\infty u^{m-1} e^{-u}\,du = \Gamma(m) = (m-1)!$$

when m is a positive integer.

You may also use this result in part (iii).]

(6)

(iii) Now let $Y = X^2$, where X is the random variable defined in part (ii). Derive the cumulative distribution function and survivor function of Y. *Hence, using the result proved in part (i)*, find $E(X^2)$ and the variance of X.

(9)

2. (a) The continuous random variable U follows a beta distribution with probability density function

$$\frac{(m+n-1)!}{(m-1)!(n-1)!}u^{m-1}(1-u)^{n-1}, \qquad 0 \le u \le 1,$$

where m and n are positive integers. Find the expected value and variance of U in terms of m and n.

(8)

(b) The continuous random variables X and Y have joint probability density function

$$f(x,y) = 12x^2, \qquad 0 \le x \le y \le 1.$$

Derive the marginal probability density functions of X and Y. Using the results of part (a), or otherwise, find their expected values and variances. Find the correlation between X and Y.

(12)

3. The continuous random variables X and Y independently follow the uniform distribution on the interval 0 to 1. The random variables U and V are defined by

$$U = (-2\ln X)^{1/2}\sin(2\pi Y), \qquad V = (-2\ln X)^{1/2}\cos(2\pi Y).$$

(i) Show that $X = \exp\left[-\frac{1}{2}\left(U^2 + V^2\right)\right]$ and $Y = \frac{1}{2\pi}\tan^{-1}\left(\frac{U}{V}\right)$.

(4)

(ii) Show that the joint probability density function of U and V is

$$f(u,v) = \frac{1}{2\pi}\exp\left[-\left(u^2+v^2\right)/2\right], \qquad -\infty < u < \infty,\ -\infty < v < \infty.$$

[Hint. If $g(t) = \tan^{-1}(t)$, then $g'(t) = \frac{1}{1+t^2}$.]

(7)

(iii) Explain why U and V are independent, and derive the marginal probability density functions of U and V.

(5)

(iv) Describe how this result would enable you to simulate from

(a) the standard Normal distribution,

(b) the chi-squared distribution with 2 degrees of freedom.

(4)

4. (i) The continuous random variable X follows the exponential distribution with probability density function

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

where $\theta > 0$. Show that X has moment generating function

$$M_X(t) = \frac{\theta}{\theta - t}, \qquad t < \theta.$$

Using this result, find the expected value and variance of X.

(9)

(ii) Suppose that $X_1, X_2, ..., X_n$ are independently distributed, each following the exponential distribution described in part (i). Find the moment generating function of

$$Z = \frac{\theta}{\sqrt{n}}(X_1 + ... + X_n) - \sqrt{n}.$$

Find the limiting form of this moment generating function as $n \to \infty$.

[Hint. Consider taking the limit of the logarithm of the moment generating function.]

Using this result, name the limiting distribution of Z.

(11)

5. A random sample of size n is drawn from the uniform distribution on the interval 0 to 1. The sample values, in increasing order of size, are $U_{(1)}, U_{(2)}, \ldots, U_{(n)}$.

(i) Derive the cumulative distribution function and probability density function of $U_{(1)}$.

(5)

(ii) Determine the joint probability density function of $U_{(1)}$ and $U_{(2)}$.

(4)

(iii) Hence show that $U_{(2)} - U_{(1)}$ has the same probability density function as $U_{(1)}$.

(8)

(iv) When $n = 10$, find the probability that $U_{(2)} - U_{(1)}$ is less than 0.1.

(3)

6. A light aircraft has been lost during a flight and is known to have come down in one of three geographical regions, 1, 2 or 3. Those planning to search for it believe that the probabilities of the aircraft being in regions 1, 2, 3 are θ_1, θ_2, θ_3 respectively (where $\theta_1 + \theta_2 + \theta_3 = 1$). An aerial search of one of the regions on one occasion is called a "sortie". If a sortie is made over region i, and the aircraft is in region i but has not previously been discovered, there is a probability α (where $0 < \alpha < 1$) that it will be found, irrespective of the number of previous sorties over region i.

(i) The first sortie is made over region 1.

(a) Show that the probability that the aircraft is not found in this sortie is $(1-\alpha\theta_1)$.

(2)

(b) Given that the aircraft is not found in the first sortie, find the posterior probability that the aircraft actually came down in region i ($i = 1, 2, 3$).

(5)

(ii) Suppose that, after all three regions have been searched once in turn, the aircraft still has not been found. Find the posterior probability that it actually came down in region i ($i = 1, 2, 3$).

(4)

(iii) Sorties are to be flown in regions 1, 2 and 3, consecutively and in that order, until the missing aircraft is found. Let the random variable X be the total number of sorties required in order to find the aircraft. *Given* that the aircraft actually came down in region i ($i = 1, 2, 3$), show that the *conditional* expected value of X is $\dfrac{3}{\alpha} + (i-3)$.

Hence determine the (unconditional) expected value of X.

(9)

7. (i) The continuous random variable X has probability density function $f(x)$ and cumulative distribution function $F(x)$. Explain carefully how the inversion method (i.e. the inverse c.d.f. method) of simulation could be used to simulate observations from this distribution. What restrictions on $F(x)$ are required in order to make this method of simulation work?

(4)

(ii) The following numbers are a random sample of real numbers from the uniform distribution on the interval 0 to 1:

0.167 0.236 0.778 0.968.

Use these values to generate four random variates from each of the following distributions:

(a) $f_X(x) = \exp(-x), \qquad x \geq 0.$

(4)

(b) $f_X(x) = 4x(1-x^2), \qquad 0 \leq x \leq 1.$

(7)

(c) $P(X = x) = \dfrac{e^{-2}2^x}{x!}, \qquad x = 0, 1, \ldots .$

(5)

Turn over

8. If a certain team loses one of its matches, then it has probability 0.5 of losing the next match and probability 0.4 of drawing it. If the team draws a match, then it has probability 0.3 of losing the next match and probability 0.4 of drawing it. If the team wins a match, then it has probability 0.2 of losing the next match and probability 0.4 of drawing it.

(i) Model this as a Markov Chain, and write down its transition matrix.

(5)

(ii) If the team loses its first game of the season, find the probability that it wins its third game.

(5)

(iii) Find the stationary distribution of this model. The team is awarded 0 points when it loses, 1 point when it draws and 3 points when it wins. Find the expected number of points per game awarded to this team in the long run.

(10)