EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

GRADUATE DIPLOMA, 2005

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 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 7 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. (i) The events $E_1, E_2, \ldots, E_n$ partition the sample space, S. Another event A in S has probability $P(A) > 0$. Write down the law of total probability, which expresses $P(A)$ in terms of conditional and unconditional probabilities involving the events $E_1, E_2, \ldots, E_n$. Write down Bayes' Theorem for probabilities of the form $P(E_j \mid A)$.

(4)

(ii) A department store has several fitting rooms where customers may try on clothes before buying them. The store's policy is to allow a customer to take up to four items of clothing into the fitting room to try on. Experienced shop assistants judge that customers who use the fitting rooms are equally likely to wish to try on 1, 2, 3 or 4 items of clothing. The total time (minutes) spent in the fitting room by a customer who tries on x items of clothing ($x = 1, 2, 3, 4$) is a random variable that has an exponential distribution with expected value $3x$.

(a) Find the cumulative distribution function of the time (minutes) that a customer spends in the fitting room. Use it to evaluate the probability that a customer spends more than 5 minutes in the fitting room.

(8)

(b) Find the expected value and variance of the length of time that a customer spends in the fitting room.

(8)

Hint. You may use the following results without proof.

$$E(Y) = E\{E(Y \mid X)\}$$

$$\mathrm{Var}(Y) = E\{\mathrm{Var}(Y \mid X)\} + \mathrm{Var}\{E(Y \mid X)\}$$

2. Suppose $0 < \theta < 1$ and that X, Y are independent binomial random variables such that

$$P(X = x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \quad x = 0, 1, \ldots, n,$$

$$P(Y = y) = \binom{m}{y} \theta^y (1-\theta)^{m-y}, \quad y = 0, 1, \ldots, m.$$

(i) Show that the random variable $X + Y$ also follows a binomial distribution.

(9)

(ii) Suppose that z is an integer in the range 0 to $n + m$. Find $P(X = x \mid X + Y = z)$ for $x = 0, 1, \ldots, z$.

(6)

(iii) A network consists of two sub-networks, the first consisting of 20 components and the second consisting of 30 components. Each component has probability 0.1 of failing within one year, independently of all the other components. After one year, it is found that exactly six of the components have failed. Find the conditional probability that exactly three components in the first sub-network have failed.

(5)

3. The continuous random variables X and Y have the joint probability density function

$$f(x, y) = \begin{cases} 12x^2, & 0 < x < y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

(i) Show that, for all non-negative integers r and s,

$$E(X^r Y^s) = \frac{12}{(r+3)(r+s+4)}.$$

Hence find the marginal expected values and variances of X and Y, and the correlation between X and Y.

(14)

(ii) For any value z, such that $0 \le z \le 1$, find $P(Y - X > z)$. Hence find the cumulative distribution function and probability density function of the random variable $Z = Y - X$.

(6)

4. (i) Suppose that X and Y are independent gamma random variables with a common scale parameter θ, i.e. their respective probability density functions are

$$f_X(x) = \frac{\theta^\alpha x^{\alpha-1} e^{-\theta x}}{\Gamma(\alpha)}, \quad x > 0,$$

$$f_Y(y) = \frac{\theta^\beta y^{\beta-1} e^{-\theta y}}{\Gamma(\beta)}, \quad y > 0.$$

Here α, β and θ are all positive constants and Γ denotes the gamma function.

Define new random variables U and V as follows:

$$U = \frac{X}{X+Y}, \qquad V = X + Y.$$

Show that V has a gamma distribution, also with scale parameter θ, that U has a beta distribution and that U and V are independent.

(16)

(ii) A particular job consists of two consecutive tasks, whose duration times are independent and identically distributed exponential random variables. Use the result of part (i) to deduce the distribution of the proportion of total duration time that is spent on the first task.

(4)

5. (i) The continuous random variable X follows the chi-squared distribution with one degree of freedom. This means that X has probability density function

$$f_X(x) = \frac{1}{\sqrt{2\pi x}} \exp\left\{-\frac{x}{2}\right\}, \qquad x > 0.$$

Show that X has moment generating function (mgf)

$$M_X(t) = \frac{1}{\sqrt{(1-2t)}}, \qquad t < \tfrac{1}{2}.$$

Hence find the expected value and variance of X.

(9)

(ii) Suppose that $X_1, X_2, ..., X_n$ are independent random variables, each following the chi-squared distribution with one degree of freedom. Find the mgf of

$$Z = \frac{1}{\sqrt{2n}}(X_1 + ... + X_n) - \sqrt{\frac{n}{2}}.$$

Find the limiting form of this mgf as $n \to \infty$. By recognising the limiting mgf, state the limiting distribution of Z.

(11)

6. A random sample of size n is drawn from the uniform distribution over the interval $-\theta$ to θ, where $\theta > 0$. The ordered values in the sample are denoted $U_{(1)} \le U_{(2)} \le \ldots \le U_{(n)}$.

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

(6)

(ii) Show that the sample range, $R = U_{(n)} - U_{(1)}$, has probability density function

$$f(r) = \frac{n(n-1)r^{n-2}(2\theta - r)}{(2\theta)^n}, \qquad 0 < r < 2\theta.$$

(8)

(iii) Find the expected value of R, and comment on the possible use of $\frac{1}{2}R$ as an estimator of θ.

(6)

Turn over

7. (i) Take the following values as a random sample from a U(0,1) distribution, i.e. a uniform distribution on the range 0 to 1.

0.1423 0.3372 0.6772 0.9192

Use these values to generate four random variates from each of the following distributions, explaining carefully the method you use in each case.

(a) The standard Normal distribution. (6)

(b) The Normal distribution with expected value –2 and variance 0.81. (3)

(c) The chi-squared distribution with one degree of freedom. (3)

(ii) A busy taxi rank is situated close to a city railway station. A customer is defined to be an individual or a group of individuals requiring one taxi. The number of taxis that arrive at this rank in any minute is a Poisson random variable with mean 2. The number of customers who arrive at the rank in any minute is also a Poisson random variable with mean 2. It may be assumed that the number of taxis that arrive and the number of customers that arrive at the rank in the same minute are independent. For both taxis and customers, the numbers of arrivals in disjoint time periods are independent, and pick-ups are instantaneous.

At 3.00 p.m. one afternoon, there are no taxis and no customers at this taxi rank. Use the following U(0,1) variates to simulate the number of **taxis** that arrive at the rank in each of the next five one-minute intervals.

0.553 0.817 0.356 0.955 0.201

Use the following U(0,1) variates to simulate the number of **customers** that arrive at the rank in the same five one-minute intervals.

0.783 0.207 0.178 0.408 0.644

Now simulate the whole process of arrivals at and departures from the taxi rank, starting at 3.00 p.m. You should record the number of taxis and the number of customers standing at the taxi rank at each of the following times: 3.01, 3.02, 3.03, 3.04, 3.05. You may assume that a taxi that arrives at the taxi rank picks up a waiting customer (if there is one) or waits for the next customer to arrive (if there is no customer at the rank). A customer who arrives at the taxi rank takes a waiting taxi (if there is one) or waits for the next taxi to arrive (if there is no taxi at the rank). (8)

8. (i) Consider a two-state Markov chain with transition matrix

$$\mathbf{P} = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix},$$

where $0 < \alpha < 1$ and $0 < \beta < 1$.

(a) Show that $\mathbf{P} = \mathbf{CDC}^{-1}$, where

$$\mathbf{C} = \begin{bmatrix} 1 & -\alpha \\ 1 & \beta \end{bmatrix}$$

and $$\mathbf{D} = \begin{bmatrix} 1 & 0 \\ 0 & 1-\alpha-\beta \end{bmatrix}.$$

(4)

(b) Hence prove that the n-step transition matrix of this Markov chain can be written in the form

$$\mathbf{C}\begin{bmatrix} 1 & 0 \\ 0 & \lambda^n \end{bmatrix}\mathbf{C}^{-1},$$

where $\lambda = 1 - \alpha - \beta$. Find the limiting value of this transition matrix as $n \to \infty$.

(8)

(ii) In a certain region, a day with no rain is followed by a day when it rains with probability 0.2. A day when it rains is followed by a day with no rain with probability 0.9.

Assuming the Markov property holds, write down the transition matrix of a Markov chain model of this process.

Suppose that, on a day-long visit to this region, you observe that there is no rain. Find the approximate probability that, if you make a further day-long visit to the region, many days later, there will be rain. Show that your answer would not change if there were rain during your first visit. Interpret this answer.

(8)