EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

GRADUATE DIPLOMA, 2009

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 10 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. The hypergeometric distribution has probability mass function

$$P(X = x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \qquad x = 0, 1, \ldots, M,$$

where M, n and N are positive integers, $n \geq M$ and $N \geq M + n$. You are given that it has mean $\frac{Mn}{N}$ and variance $\frac{Mn(N-M)(N-n)}{N^2(N-1)}$.

A standard pack of 52 playing cards consists of four suits (clubs, diamonds, hearts and spades) each containing 13 cards numbered 2, 3, 4, …, 10, Jack, Queen, King, Ace (their face values).

A bridge hand of 13 cards is dealt from a well-shuffled pack of 52 cards. Cards of face value Jack, Queen, King or Ace in any of the 4 suits are called "honours", and have associated "points" (Jack = 1, Queen = 2, King = 3, Ace = 4). Let X_A, X_K, X_Q and X_J respectively denote the numbers of Aces, Kings, Queens and Jacks in the hand, and let Z be the total points of all the honours in the hand.

(i) Write down Z in terms of X_A, X_K, X_Q and X_J.

(1)

(ii) Explain why the joint probability distribution of X_A and X_K is

$$P(X_A = x, X_K = y) = \frac{\binom{4}{x}\binom{4}{y}\binom{44}{13-x-y}}{\binom{52}{13}}, \qquad x, y = 0, 1, 2, 3, 4.$$

Show that the marginal distribution of X_A is of hypergeometric form given by

$$P(X_A = x) = \frac{\binom{4}{x}\binom{48}{13-x}}{\binom{52}{13}}, \qquad x = 0, 1, 2, 3, 4.$$

Explain why X_A, X_K, X_Q and X_J all have the same marginal distribution, write down their common expectation, and hence obtain $E(Z)$.

(8)

Question 1 is continued on the next page

(iii) You may assume that, given that $X_A = x$, the conditional distribution of X_K is hypergeometric such that

$$P\left(X_K = y \middle| X_A = x\right) = \frac{\binom{4}{y}\binom{44}{13-x-y}}{\binom{48}{13-x}}, \qquad y = 0, 1, 2, 3, 4.$$

Write down $E(X_K \mid X_A = x)$ and deduce that

$$E\left(X_A X_K\right) = \frac{52}{48} - \frac{E(X_A^{\;2})}{12}.$$

Obtain $\mathrm{Var}(X_A)$ and show that $E(X_A^2) = 29/17$. Show further that $E(X_A X_K) = 16/17$, deduce the value of $\mathrm{Cov}(X_A, X_K)$ and, by using symmetry or otherwise, show that $\mathrm{Var}(Z) = 290/17$.

(8)

(iv) Obtain a Normal approximation for $P(Z > 20)$, and suggest with a reason whether the true probability is likely to be greater or less than the Normal approximate value.

(3)

2. (i) (a) The independent events A and B are defined on the sample space of a given random experiment. If $\overline{A}$ and $\overline{B}$ respectively denote the complements of A and B, show that $\overline{A}$ and $\overline{B}$ are also independent.

(4)

(b) The events A, B and C are defined to be *pairwise independent* if A and B are independent, A and C are independent and B and C are independent. Four cards each bear the numbers 0 or 1 in each of three positions, as shown.

	Position		
	1	**2**	**3**
Card 1	1	0	0
Card 2	0	1	0
Card 3	0	0	1
Card 4	1	1	1

One of the four cards is chosen at random and the events A, B and C are defined as follows.

A: the card chosen has a 1 in the first position

B: the card chosen has a 1 in the second position

C: the card chosen has a 1 in the third position

Establish whether A, B and C are pairwise independent.

Establish whether A, B and C are mutually independent.

(4)

(ii) In answering a question on a multiple-choice test, a student either knows the answer or guesses, independently for each question. Let p be the probability that the student knows the answer, and assume that a student who guesses picks an answer at random from the m multiple-choice answers given.

(a) A student answers a question correctly. Show that the conditional probability that this student knows the answer to this question is

$$\frac{mp}{1+(m-1)p}.$$ (5)

(b) If $p = 1/2$, how many multiple-choice answers for a question should the examiner include in order to be at least 90% certain that a student who answers correctly actually knows the answer? (3)

(c) A test consists of 10 multiple-choice questions in which $m = 3$. A student knows the answers to 4 of the questions and guesses the rest. Let X denote the number of questions this student gets right. Write down the distribution of X. What is the probability that the student gets at most 5 right? What is the probability that the student gets at least 8 right? (4)

3. (i) The random variable X follows a gamma distribution with shape parameter α and scale parameter λ, so that the probability density function (pdf) of X is

$$f_X(x) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, \qquad x > 0,\ \alpha > 0,\ \lambda > 0.$$

Show that the moment generating function (mgf) of X is

$$M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-\alpha}, \qquad |t| < \lambda.$$

Hence or otherwise obtain $E(X)$ and $\mathrm{Var}(X)$.

(5)

(ii) Let $\bar{X}$ be the mean of a random sample of size n from the above distribution. Obtain the moment generating function of $\bar{X}$ and deduce the form of the distribution of $\bar{X}$, making your reasoning clear. Write down $E(\bar{X})$ and $\mathrm{Var}(\bar{X})$.

(5)

(iii) You are given that, for any positive integer k, a χ^2 distribution on k degrees of freedom is of gamma form with shape parameter $k/2$ and scale parameter $1/2$. Use this fact and earlier results to show that if 2α is an integer then $2n\lambda\bar{X}$ is distributed as χ^2 on $2n\alpha$ degrees of freedom.

(4)

(iv) A random sample of size 20 is taken from a gamma distribution with $\alpha = 1/2$ and unknown scale parameter λ. Write down the mean μ of this distribution. You are given that a 90% confidence interval for μ, based on the above results and using the "Statistical tables for use in examinations", is given in terms of the sample mean $\bar{x}$ as $(0.637\bar{x}, 1.843\bar{x})$.

Use the Central Limit Theorem to obtain an approximate 90% confidence interval for μ based on the Normal distribution. Say with justification which confidence interval you prefer.

(6)

Turn over

4. The random variable Y has the geometric probability mass function given by

$$p_Y(y) = \frac{1}{4}\left(\frac{3}{4}\right)^y, \qquad y = 0, 1, 2, \ldots .$$

The random variable X has the distribution with probability mass function

$$p_X(x) = \frac{\Gamma(x+0.5)\,0.75^x}{2\Gamma(x+1)\Gamma(0.5)}, \qquad x = 0, 1, 2, \ldots .$$

[You may use without proof the results $\Gamma(x + 1) = x\,\Gamma(x)$, $\Gamma(1) = 1$ and $\Gamma(½) = \sqrt{\pi}$.]

(i) Given a supply of pseudorandom numbers $u_1, u_2, \ldots$ from the continuous uniform distribution on the range 0 to 1 (i.e. from the U(0, 1) distribution), show how the inverse transform method for discrete distributions may be used to generate pseudorandom realisations $y_1, y_2, \ldots$ of Y.

(3)

(ii) Calculate the ratios $\dfrac{p_X(0)}{p_Y(0)}$ and $\dfrac{p_X(1)}{p_Y(1)}$.

Show that $\dfrac{p_X(x)}{p_Y(x)} < \dfrac{p_X(x-1)}{p_Y(x-1)}$ for $x \geq 2$.

Deduce that $\max\left[\dfrac{p_X(x)}{p_Y(x)}\right] = 2$.

(5)

(iii) It is required to generate pseudorandom realisations $x_1, x_2, \ldots$ of X. Apply the results of part (ii) to show how this may be done using the discrete form of the acceptance-rejection method based on a supply of pseudorandom numbers $y_1, y_2, \ldots$ from the probability mass function $p_Y(y)$ and an additional supply of U(0, 1) pseudorandom numbers $v_1, v_2, \ldots$. Also show that on average four U(0, 1) numbers are needed to generate each realisation of X.

(5)

(iv) The pseudorandom numbers $u_1 = 0.612$, $u_2 = 0.173$, $u_3 = 0.864$, $u_4 = 0.305$ may be used to generate realisations y_1, y_2, y_3, y_4 of Y. Additional pseudorandom numbers $v_1 = 0.374$, $v_2 = 0.287$, $v_3 = 0.569$, $v_4 = 0.087$ are available to be used as required in applying the argument of part (iii). Use as many of both sets of numbers as you need to generate a realisation x_1 of X.

(7)

Turn over

5. Urn A contains x white balls and $n-x$ black balls, urn B contains x black balls and $n-x$ white balls. One ball is chosen at random from urn A and, at the same time and independently, one ball is chosen at random from urn B. The ball chosen from urn A is then put into urn B and the ball chosen from urn B is put into urn A.

(i) Let Y be the random variable denoting the number of white balls in urn A after this operation. Show that

(a) $$P(Y=x)=\frac{2x(n-x)}{n^2}, \quad x = 0, 1, \ldots, n,$$

(b) $$P(Y=x+1)=\frac{(n-x)^2}{n^2}, \quad x = 0, 1, \ldots, n-1,$$

(c) $$P(Y=x-1)=\frac{x^2}{n^2}, \quad x = 1, \ldots, n.$$

(7)

(ii) (a) This operation of choosing a ball from each urn and transferring it to the other urn is repeated several times. Explain why the resulting process can be represented as a discrete-time Markov chain and identify the state space.

(3)

(b) Write down the transition probability matrix for the case $n = 4$. Verify that the equilibrium distribution is given by

$$(\pi_0, \pi_1, \pi_2, \pi_3, \pi_4) = \left(\frac{1}{70}, \frac{16}{70}, \frac{36}{70}, \frac{16}{70}, \frac{1}{70}\right).$$

(10)

6. (i) (a) The random variable X is distributed Normally with mean μ and variance σ^2. Show that the moment generating function (mgf) of X is given by

$$M_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right).$$

[You may use without proof the result $\int_{-\infty}^{\infty} \frac{\exp\left(-z^2/2\right)}{\sqrt{2\pi}} dz = 1$.]

(4)

(b) The random variable $Y = \exp(X)$, so that Y has a lognormal distribution. By using the result of part (a), or otherwise, obtain $E(Y)$ and $\text{Var}(Y)$.

(4)

(ii) The sizes (in £ sterling) of claims $Y_1, Y_2, Y_3, \ldots$ in the house insurance portfolio of an insurance company are believed to follow a distribution of the above lognormal form, such that log(claim size) is Normally distributed with mean 7 and variance 4. It is assumed that claims are independent and that the number N of claims in a year is a Poisson random variable with mean 100.

(a) Let $T = Y_1 + Y_2 + \ldots + Y_N$ be the total claim amount in a year. Use the results

$$E(T) = E(Y)\,E(N)$$

and

$$\text{Var}(T) = E(N)\,\text{Var}(Y) + \{E(Y)\}^2\,\text{Var}(N)$$

to obtain $\text{E}(T)$ and $\text{Var}(T)$.

(5)

(b) Use a Normal approximation based on the results of part (ii)(a) to find the necessary reserve c which is sufficient with probability 0.999 to meet the annual total claims on this portfolio. Why might you hesitate in adopting this figure?

(7)

7. (i) A random sample $X_1, X_2, \ldots, X_n$ is drawn from a continuous distribution with cumulative distribution function $F(x)$ and probability density function (pdf) $f(x)$. Write down the formula for the joint pdf of the rth and sth order statistics, $X_{(r)}$ and $X_{(s)}$, of this random sample.

(2)

(ii) Here and in the rest of this question let $n = 4$ and suppose that the distribution sampled is continuous uniform on the range 0 to 1, so that the pdf is given by $f(x) = 1$ for $0 \le x \le 1$.

(a) Find $F(x)$.

(1)

(b) Writing U for $X_{(2)}$ and V for $X_{(3)}$, show that the joint pdf of U and V is given by

$$f_{U,V}(u,v) = 24u(1-v), \qquad 0 \le u \le v \le 1.$$

Deduce the marginal pdfs of U and V, $f_U(u)$ and $f_V(v)$.

(5)

(iii) (a) Show that $E(U) = 2/5$, find $E(U^2)$ and deduce $\mathrm{Var}(U)$.

(4)

(b) Given that $E(V) = 3/5$ and $E(V^2) = 2/5$, find $\mathrm{Var}(V)$.

(1)

(c) Show that $E(UV) = 4/15$ and deduce $\mathrm{Cov}(U, V)$.

(4)

(iv) Express the median M of this sample in terms of U and V, write down $E(M)$ and find $\mathrm{Var}(M)$.

(3)

Turn over

8. (i) If X has a continuous uniform distribution on the interval $(-\pi/2, \pi/2)$, find the probability density function (pdf) of $Y = \tan X$ and name the distribution of Y.

(8)

(ii) Let Z be a standard Normal random variable, and let $Y = Z_1^2 + Z_2^2 + \ldots + Z_k^2$ where $Z_1, Z_2, \ldots, Z_k$ are independent standard Normal random variables, all also independent of Z. Given that the pdf of Y is

$$f_y(y) = \frac{2^{-k/2} y^{(k/2)-1} e^{-k/2}}{\Gamma(k/2)} \qquad \text{for } y > 0,$$

write down the joint pdf of Y and Z. If $W = \dfrac{Z}{\sqrt{Y/k}}$, obtain the joint pdf of Y and W. Hence show that the marginal pdf of W is given by

$$f_w(w) = \frac{\Gamma\left(\dfrac{k+1}{2}\right)}{\Gamma(k/2)\sqrt{\pi k}} \cdot \frac{1}{\left(1+\dfrac{w^2}{k}\right)^{(k+1)/2}} \qquad \text{for } -\infty < w < \infty,$$

and name the distribution of W. State also the name of the distribution of W^2.

(12)