EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

HIGHER CERTIFICATE IN STATISTICS, 2007

(Modular format)

MODULE 2 : Probability models

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 4 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. 100 men are surveyed as to whether they play cricket, tennis or golf. It is found that

- 10 play none of these sports
- 5 play all three of these sports
- 88 play cricket or tennis or both
- 78 play cricket or golf or both
- 30 play golf and tennis but not cricket
- 38 play golf
- 74 play tennis.

Using a Venn diagram, or otherwise, find the following.

(i) The number of the men who play at least one of these sports. (2)

(ii) The number of the men who play exactly one of these sports. (8)

(iii) The number of the men who play exactly two of these sports. (4)

(iv) Of those who do not play golf, the proportion who play cricket. (3)

(v) The mean number of sports played by these men. (3)

2. The random variable X has the exponential probability density function (pdf) given by

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \quad \lambda > 0.$$

(i) Show that $E(X) = \dfrac{1}{\lambda}$ and find the standard deviation of X.

(8)

(ii) Show that, for any $c > 0$, $P(X > \text{c}) = \exp(-\lambda c)$.

Hence show that, for any $x > c$, $P(X > x \mid X > c) = \exp(-\lambda(x-c))$. Deduce the conditional pdf of X given that $X > c$, and comment briefly.

(6)

(iii) A random sample has been selected from a distribution that is thought to be exponential. The values obtained, arranged in ascending order, are 0.1, 0.1, 0.2, 0.4, 1.1, 2.3, 2.5, 3.4, 4.3, 5.6. [You are given that the sum and sum of squares of these values are 20.0 and 74.38 respectively.] Calculate the sample mean and the sample standard deviation and say with a reason whether you think the exponential model is suitable for the distribution underlying this sample.

(6)

3. A coin has probability p of showing heads and probability $1 - p$ of showing tails when it is tossed, independently each time.

(i) (a) Let X be the random variable denoting the number of times the coin shows heads when it is tossed n times. Show that

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

making clear all the steps of your reasoning. Under what conditions can the distribution of X be approximated by a Normal distribution?

(6)

(b) A student uses the Normal approximation to approximate $P(X \le 3)$ when $n = 20$ and $p = 0.2$. Calculate the answer he should obtain, use tables of the exact distribution of X to compute the percentage error in the answer, and comment briefly.

(6)

(ii) For integer $x \ge 1$, let N be the random variable denoting the number of tosses of the coin needed to obtain x heads. Show from first principles that

$$P(N = n) = \binom{n-1}{x-1} p^x (1-p)^{n-x}, \quad n = x, x+1, x+2, ...$$

Evaluate this probability for the case $p = 0.2$, $x = 3$ and $n = 20$, and compare your result with the exact $P(X = 3)$ for the binomial distribution with the same values of p, x and n.

(8)

4. (i) Let $A_1, A_2, \ldots, A_k$ be mutually exclusive and exhaustive events, and let B be an event with $P(B) > 0$. Obtain an expression for $P(A_i | B)$ from first principles in terms of probabilities of the form $P(B | A_j)$ and $P(A_j)$.

(5)

The number of flaws, X, in a standard length of yarn is assumed to be Poisson distributed with probability mass function

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots,$$

where λ is a positive parameter. A textile manufacturer buys yarn from suppliers P, Q and R in the long-run proportions 1:2:3. It is known from experience that the numbers of flaws in lengths of yarn from these suppliers are independently Poisson distributed with respective parameter values $\lambda_P = 3$, $\lambda_Q = 2$ and $\lambda_R = 1$.

(ii) An unlabelled length of yarn is found to have 2 flaws. Is it more likely to have come from supplier Q or supplier R?

(10)

(iii) A second unlabelled length of yarn, known to be from the same supplier as the first, is also found to have 2 flaws. Are the two lengths of yarn more likely to have come from supplier Q or supplier R? Comment briefly on this result in comparison with that of part (ii).

(5)