EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

HIGHER CERTIFICATE IN STATISTICS, 2016

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 8 printed pages.
This front cover is page 1.
Question 1 starts on page 2.

There are 4 questions altogether in the paper.

1. In order to start playing a board game, a player throws an irregular die until he obtains a six. The random variable X is the number of throws the player makes up to and including the first six. The probability mass function of X is given by

$$P(X = x) = \theta(1-\theta)^{x-1} \quad \text{for } x = 1, 2, 3, \ldots$$

where θ is the probability that the irregular die shows a six.

(i) Show that $P(X > k) = (1-\theta)^k$ for $k = 1, 2, 3, \ldots$.

(4)

200 players are observed whilst trying to start the game. It is found that 24 obtain a six on their first throw, 48 obtain their first six on their second throw and the remainder require more than 2 throws before obtaining a six.

(ii) Show that the likelihood of these data can be written as

$$L(\theta) = \theta^{72}(1-\theta)^{304}$$

and find the maximum likelihood estimate $\hat{\theta}$ of θ.

(9)

(iii) Find an approximate 95% confidence interval for θ.

(7)

2. (a) The random variable X has moment generating function (mgf) $M_X(t)$. Let $R(t) = \log M_X(t)$.

(i) Find $R'(t)$ and $R''(t)$ in terms of the first two derivatives of $M_X(t)$.

(2)

(ii) Show that $R'(0) = E(X)$ and $R''(0) = \text{Var}(X)$.

(4)

(b) The random variable Y has a Poisson distribution with probability mass function (pmf)

$$P(Y = y) = \frac{e^{-\lambda}\lambda^{y}}{y!}, \quad y = 0, 1, 2, \ldots,$$

where $\lambda > 0$.

(i) Show that the mgf of Y is $M_Y(t) = e^{\lambda(e^t - 1)}$.

(3)

(ii) Use the mgf to show that $E(Y) = \lambda$. Given also that $E(Y^2) = \lambda + \lambda^2$ and $E(Y^3) = \lambda + 3\lambda^2 + \lambda^3$, show that $E[(Y - E(Y))^3] = \lambda$.

(5)

(iii) $Y_1, Y_2, \ldots, Y_n$ are independent identically distributed Poisson random variables each with pmf given above. Use their mgfs to find the distribution of their sum $S = \sum_{i=1}^{n} Y_i$. State any properties of mgfs which you use in your solution.

(6)

3. In a dice game, a bag contains four red dice, three white dice and two blue dice. A random sample of three dice is drawn without replacement. The random variable X is the number of white dice in the sample and the random variable Y is the number of blue dice in the sample. The joint probability mass function of X and Y is shown in the table.

		X			
		0	1	2	3
Y	0	$4k$	$18k$	$12k$	k
	1	$12k$	$24k$	$6k$	0
	2	$4k$	$3k$	0	0

(i) Find the value of k.

(2)

(ii) Explain, from the rules of the dice game, why $P(X=2, Y=2)=0$ and confirm the value of $P(X=2, Y=0)$ from the rules of the game.

(5)

(iii) Find the conditional probability mass function of X given that $Y = 0$. Show that $E(X \mid Y=0)=\frac{9}{7}$ and find $\mathrm{Var}(X \mid Y=0)$.

(6)

(iv) Show that $\mathrm{Cov}(X, Y)=-\frac{1}{6}$ and explain briefly why you would expect this quantity to have a negative value.

(7)

4. The random sample $X_1, X_2, ..., X_n$ comes from a $\mathrm{N}(\mu, \sigma^2)$ distribution.

(i) Show that

$$E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \frac{(n-1)\sigma^2}{n}$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $n > 1$. Use this result to obtain an unbiased estimator for σ^2.

[You may assume $E(\bar{X}) = \mu$ and $\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$ without proof.]

(9)

A general estimator of the form $S_a^2 = a\sum_{i=1}^{n}(X_i - \bar{X})^2$ is proposed for σ^2, where a is an arbitrary constant.

(ii) Show that the bias of S_a^2 is $(an - a - 1)\sigma^2$.

(4)

(iii) The mean square error (MSE) of an estimator is defined to be the variance of the estimator plus the square of its bias. Given that

$$\mathrm{Var}\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right) = 2(n-1)\sigma^4,$$

find the MSE of S_a^2 and show that the value of a which minimises this MSE is $\frac{1}{n+1}$.

(7)

BLANK PAGE

BLANK PAGE

BLANK PAGE