



EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

GRADUATE DIPLOMA, 2015

MODULE 1 : Probability distributions

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 nC_r .

This examination paper consists of 12 printed pages.
This front cover is page 1.
Question 1 starts on page 2.

There are 8 questions altogether in the paper.

1. The events E_1, E_2, E_3, \dots form a mutually exclusive and exhaustive partition of the sample space S . The event A in S has probability $P(A) > 0$.

(i) 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, E_3, \dots .
(1)

(ii) Given the Law of Total Probability, prove *Bayes' Theorem* in the form

$$P(E_j | A) = \frac{P(A | E_j)P(E_j)}{\sum_i P(A | E_i)P(E_i)}, \quad j = 1, 2, 3, \dots .$$

(4)

It may be assumed from experience that 1 in 1000 fetuses has the genetic disorder trisomy 21 ("Down's Syndrome"). A screening test carried out early in pregnancy classifies a foetus as either having trisomy 21 (a positive result) or not having trisomy 21 (a negative result). The test gives a positive result for 90% of fetuses that have trisomy 21 but also gives a positive result for 5% of fetuses that do not have trisomy 21.

(iii) What proportion of all fetuses have a positive test result for trisomy 21?
(4)

(iv) What proportion of fetuses that have a positive test result do not have trisomy 21?
(3)

(v) In a sample of 1000 fetuses that independently have positive test results, find an approximation for the probability that at least 990 do not have trisomy 21.
(8)

2. A school has 25 members of staff, of whom 7 are classed as Support, 14 as Teaching and 4 as Management staff. The staff agree that they should set up a five-person social committee and, after long discussion, decide that the only fair way of choosing who will serve on the committee is by simple random sampling from the entire staff. Let the random variables X , Y and Z , respectively, be the numbers of Support, Teaching and Management staff who are chosen to serve on the social committee.

- (i) Explain why the joint probability distribution of X and Y is

$$P(X = x, Y = y) = \frac{\binom{7}{x} \binom{14}{y} \binom{4}{5-x-y}}{\binom{25}{5}}$$

for appropriate values of x and y , which you should specify.

(3)

- (ii) Find the probability that the committee will consist of 1 member of Support staff, 3 of Teaching staff and 1 of Management staff.

(3)

- (iii) Marginally, the random variable X has a hypergeometric distribution. Without doing any algebra, explain why. Write down the form of $P(X = x)$.

(3)

- (iv) Find the probability that there is at least one member of Management staff on the social committee.

(4)

- (v) The random variable with probability distribution $\frac{\binom{M}{t} \binom{N-M}{n-t}}{\binom{N}{n}}$ has expected

value $n \frac{M}{N}$ and variance $n \frac{M}{N} \left(1 - \frac{M}{N}\right) \left(\frac{N-n}{N-1}\right)$, for appropriate values of M , N , n and t . Use these results to obtain the expected values and variances of X , Y and Z .

(3)

- (vi) Write down $E(X + Y + Z)$ and $\text{Var}(X + Y + Z)$. Explain why

$$\text{Var}(X + Y + Z) \neq \text{Var}(X) + \text{Var}(Y) + \text{Var}(Z).$$

(4)

3. A large group of players is playing the game of "chuck a luck" with 3 fair six-sided dice. Each player bets a stake of 1 dollar on one of the numbers 1, 2, 3, 4, 5 or 6. If the player's number comes up on k dice ($k = 1, 2$ or 3), then the player gets his stake back and wins a further k dollars. Otherwise, the player loses his stake (which can be thought of as winning -1 dollars).

(i) Show that, on average, an individual player loses $\frac{17}{216}$ dollars on a play of this game. (5)

(ii) Find the variance of an individual player's winnings on a play of this game. (3)

(iii) Three friends, A, B and C, are part of this group of players. Suppose they each choose a number to bet on without discussing their choices with one another. Find the expected value and variance of their total winnings, T , on a play of the game. (3)

(iv) Suppose instead that A, B and C arrange to bet on three different numbers. The probability distribution of T in this case is shown below.

t	-3	-1	0	1	2	3
$P(T = t)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{19}{72}$	$\frac{1}{12}$	$\frac{1}{36}$

(a) Explain why $P(T = -3) = \frac{1}{8}$. (2)

(b) Explain why $P(T = 2) = \frac{1}{12}$. (3)

(c) Find the expected value and variance of their total winnings on one play of the game. Compare these values with the corresponding values obtained in part (iii) and comment. (4)

4. (a) (i) The number, X , of vehicles passing a fixed point on a motorway in a period of 1 minute has a Poisson distribution with expected value λ . A proportion θ (where $0 < \theta < 1$) of all the vehicles that pass this fixed point will leave the motorway at the next intersection; the other vehicles will remain on the motorway. Let Y be the number of vehicles that pass the fixed point in a period of 1 minute and leave the motorway at the next intersection, and let Z be the number of vehicles that pass the fixed point in the same period and remain on the motorway. Explain why, for $y = 0, 1, 2, \dots$ and $z = 0, 1, 2, \dots$,

$$P(Y = y, Z = z) = P(Y = y | X = y + z)P(X = y + z)$$

and use this result to prove that

$$P(Y = y, Z = z) = \frac{e^{-\lambda}}{y!z!} \lambda^{y+z} \theta^y (1-\theta)^z. \quad (6)$$

- (ii) Show that the marginal probability distribution of Y is

$$P(Y = y) = \frac{e^{-\lambda\theta}}{y!} (\lambda\theta)^y$$

for $y = 0, 1, 2, \dots$. Identify this distribution.

(4)

- (iii) Deduce the marginal probability distribution of Z , $P(Z = z)$. Show that Y and Z are independent random variables.

(4)

- (b) The number, X_t , of vehicles passing a fixed point on a motorway in any time period of length t minutes is a Poisson random variable with expected value λt . The numbers of vehicles that arrive in non-overlapping time intervals can be assumed to be independent random variables.

Suppose that a vehicle passes the fixed point and let the continuous random variable T be the length of time (in minutes) that elapses until the next vehicle passes the same point. Explain why, for $t > 0$,

$$P(T \leq t) = 1 - P(X_t = 0).$$

Use this result to obtain the cumulative distribution function and probability density function of T . Identify this distribution.

(6)

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

$$f(u) = \frac{(m+n-1)!}{(m-1)!(n-1)!} u^{m-1} (1-u)^{n-1}, \quad 0 < u < 1,$$

where m and n are positive integers. Derive $E(U)$ and $\text{Var}(U)$.

(6)

[You may use without proof the result that, for non-negative integers r and s ,

$$\int_0^1 u^r (1-u)^s du = \frac{r!s!}{(r+s+1)!} .]$$

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

$$f_{XY}(x, y) = \begin{cases} 60x^2y, & 0 < x < 1, 0 < y < 1, 0 < x + y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Derive the marginal probability density functions of X and Y , and use part (i) to deduce their expected values and variances. Find and interpret the correlation between X and Y .

(14)

6. The continuous random variables X and Y are independent, and each has the uniform distribution on the interval 0 to 1. The random variables U and V are defined as follows.

$$U = (-2\log X)^{\frac{1}{2}} \sin(2\pi Y) \quad V = (-2\log X)^{\frac{1}{2}} \cos(2\pi Y)$$

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

$$f(u, v) = \frac{1}{2\pi} \exp\left\{-\frac{1}{2}(u^2 + v^2)\right\}, \quad -\infty < u < \infty, \quad -\infty < v < \infty. \quad (11)$$

[You may use without proof the result that, if $g(x) = \tan^{-1}(x)$, then $g'(x) = \frac{1}{1+x^2}$.]

- (ii) Deduce that U and V are independent and write down their marginal probability density functions. Identify the marginal distributions of U and V . (5)

- (iii) Describe how this result would enable you to simulate pseudo-random variates from

(a) the standard Normal distribution,

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

(4)

7. (i) The random variable X has the Normal distribution with probability density function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \quad -\infty < x < \infty.$$

Show that X has moment generating function

$$M_X(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right). \quad (7)$$

- (ii) For constants a and b , show that the moment generating function of $Y = aX + b$ is $e^{bt} M_X(at)$. Use this result to obtain the moment generating function of

$$Z = \frac{X - \mu}{\sigma}.$$

Deduce the distribution of Z .

(5)

- (iii) The kurtosis of X is

$$K(X) = \frac{E([X - \mu]^4)}{\sigma^4} - 3.$$

Writing $K(X) = E(Z^4) - 3$, use the moment generating function of Z to find the value of $K(X)$.

(8)

8. (i) Let X be any continuous random variable, and let $F(x)$ denote its cumulative distribution function. Suppose that U is a continuous random variable with the uniform distribution on the interval 0 to 1, and define the new random variable Y by $Y = F^{-1}(U)$, where $F^{-1}(\cdot)$ is the inverse function of $F(\cdot)$. By considering the cumulative distribution function of Y , show that Y has the same distribution as X .

(5)

(ii) Briefly describe a method of simulating pseudo-random variates from a continuous probability distribution, based on the result of part (i).

(3)

(iii) The continuous random variable X has the probability density function

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

(a) Explain why the method described in part (ii) could not easily be used to obtain pseudo-random variates from this distribution.

(2)

(b) Suppose that you have a method of obtaining pseudo-random variates from the uniform distribution on the interval 0 to 1. Describe in detail a rejection method for obtaining pseudo-random variates from the distribution $f(x)$.

(7)

(c) Using your rejection method, how many uniform pseudo-random variates will you require, on average, in order to obtain one acceptable simulated value from the distribution $f(x)$?

(3)

BLANK PAGE

BLANK PAGE

BLANK PAGE