EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

GRADUATE DIPLOMA, 2010

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 6 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.

© RSS 2010

1. Let X be a random variable with probability mass function

$$P(X = n) = \frac{K}{n(n+2)}, \quad n = 1, 2, 3, \ldots$$

for some constant K.

(i) By noting that $\frac{1}{n(n+2)} = \frac{1}{2}\left(\frac{1}{n} - \frac{1}{n+2}\right)$, or otherwise, show that $K = \frac{4}{3}$.

(6)

(ii) Calculate the probability that the value of X is an odd number.

(6)

(iii) Suppose Y is independent of X and has the same distribution as X. Find the probability that

(a) $X + Y$ is even,

(4)

(b) X is even, given that $X + Y$ is even.

(4)

2. [In this question, all random variables take non-negative integer values only, and you may quote standard properties of probability generating functions (pgfs) without proof.]

Let N be a random variable with pgf $g(z)$, and let $X_1, X_2, X_3, \ldots$ be independent random variables all having the same distribution whose pgf is $f(z)$.

(i) Define $S_N = X_1 + X_2 + ... + X_N$ for $N \geq 1$, and $S_0 = 0$. Show that the pgf of S_N is the composition $g\left(f(z)\right)$.

(6)

(ii) Suppose that N has the Poisson distribution with mean $\lambda > 0$, while each X_i takes only the values 0 and 1 with respective probabilities $1 - p$ and p $(0 < p < 1)$. Find the corresponding pgfs, and deduce the distribution of S_N.

(7)

(iii) Suppose instead that, for $m = 0, 1, 2, \ldots, p_m$ denotes the probability that, in any given road accident, m casualties will need hospital treatment. Assume that the numbers of casualties in different accidents are independent, and that the number of road accidents relevant to a particular hospital on a Sunday afternoon has the Poisson distribution with mean 1. Show that the probability that this hospital will have to treat at most two road accident casualties next Sunday afternoon is

$$\left(1 + p_1 + p_2 + \frac{p_1^2}{2}\right)\exp\left(p_0 - 1\right).$$

(7)

3. Suppose that X and Y are independent random variables with the same probability density function (pdf) $f(x)$. Write down, without proof, a formula for the pdf of $X + Y$.

(2)

Suppose that $f(x) = x/2$ for $0 < x < 2$ (and $f(x) = 0$ elsewhere).

(i) Find the pdf of $W = X + Y$ for $0 < w < 2$ and for $2 < w < 4$.

(12)

(ii) Find the pdf of $V = (X - 1)^2$.

(6)

4. A sequence of independent Bernoulli trials, in which the probability of success is p, with $0 < p < 1$, is carried out.

Let X denote the number of trials up to and including the first success. Find the distribution of X and obtain its expected value and variance.

(8)

A further sequence of x trials is carried out, where x is the observed value of the random variable X. Let Y denote the number of successes in these trials. Show that Y has expected value 1 and find its variance.

(8)

[You may use the result $\text{Var}(Y) = E(\text{Var}(Y \mid X)) + \text{Var}(E(Y \mid X))$.]

Explain briefly why the result $E(Y) = 1$ should cause no surprise, and why it might be anticipated that the variance of Y decreases as p increases.

(4)

5. Suppose $\{X_n\}$ ($n = 1, 2, 3, \ldots$) are independent random variables, all having the same continuous distribution. When observations are made sequentially, a *record* occurs whenever a value exceeds all previous values (note that the first value is itself a record). Let A_n denote the event that X_n is a record, and let S_n denote the number of records in the sequence $\{X_1, X_2, \ldots, X_n\}$.

(i) Show that the probability of A_n is $1/n$.

(2)

(ii) Explain why the events A_n and A_m are independent when $m \neq n$.

(4)

(iii) Find the mean and variance of S_n.

(6)

(iv) Use the central limit theorem, with a continuity correction, to estimate the probability of at least 10 records in the first 100 values of $\{X_n\}$.

(8)

[You may use the results $\sum_{r=1}^{n} \frac{1}{r} \approx \log n$ and $\sum_{r=1}^{n} \frac{1}{r^2} \approx \frac{\pi^2}{6}$.]

6. Suppose that U has the uniform distribution over the interval (0, 1).

Write $V = \exp(U^2)$, $W = V + K(U - 0.5)$ where K is a constant, and $\theta = \int_0^1 \exp\left(t^2\right) dt$.

(i) Show that $E(W) = E(V) = \theta$.

(5)

(ii) Show that the covariance of U and V is $(e - 1 - \theta)/2$.

(5)

(iii) Deduce the value of K that minimises the variance of W.

(5)

(iv) The mean of 1000 simulated values of V yields a preliminary estimate of θ as approximately 1.46. Assume that you have an unlimited supply of values from independent random variables all of which have the same distribution as U. Describe an efficient simulation method, based on (i) and (iii), that would be expected to lead to the estimation of the value of θ to a high degree of precision.

(5)

7. (i) Show that, if U has the uniform distribution on the interval (0, 1), then $-2\log U$ has the exponential distribution with mean 2.

(5)

Suppose that X and Y are independent standard Normal random variables. Let (R, Θ) be the corresponding polar coordinates, i.e.

$$R = \sqrt{X^2 + Y^2} \quad \text{and} \quad \cos\Theta = X/R,\ \sin\Theta = Y/R.$$

(ii) Show that the joint density of (R, Θ) is $g\left(r, \theta\right) = \frac{r}{2\pi}\exp\left(-\frac{r^2}{2}\right)$ over $0 < r < \infty$, $0 < \theta < 2\pi$. Deduce the density function of R.

(10)

(iii) Hence or otherwise show that R^2 has the exponential distribution with mean 2.

(5)

8. Let X denote the smallest and Y the largest of four independent random variables, each having the continuous uniform distribution over the interval $(0, \theta)$ where $\theta > 0$.

(i) For given $0 < x < y < \theta$, show that $P\left(X > x,\ Y < y\right) = \dfrac{\left(y - x\right)^4}{\theta^4}$. Deduce the joint density of X and Y.

(4)

(ii) Write $U = Y - X$ and $V = Y + X$. Find the joint density of U and V. Sketch the region over which this density is non-zero.

(6)

(iii) Deduce the marginal densities of U and V.

(4)

(iv) You are given that $E(X) = \theta/5$, $\text{Var}(X) = 2\theta^2/75$, $E(Y) = 4\theta/5$, $\text{Var}(Y) = 2\theta^2/75$, $\text{Var}(U) = \theta^2/25$ and $\text{Var}(V) = \theta^2/15$. Find constants c_1, c_2, c_3, c_4 such that $E(c_1X) = E(c_2Y) = E(c_3U) = E(c_4V) = \theta$. Which of the four quantities c_1X, c_2Y, c_3U, c_4V has the smallest variance?

(6)