Mathematics
Higher level
Paper 3 – statistics and probability

Thursday 21 November 2019 (afternoon)

1 hour

Instructions to candidates

- Do not open this examination paper until instructed to do so.
- Answer all the questions.
- Unless otherwise stated in the question, all numerical answers should be given exactly or correct to three significant figures.
- A graphic display calculator is required for this paper.
- A clean copy of the **mathematics HL and further mathematics HL formula booklet** is required for this paper.
- The maximum mark for this examination paper is **[50 marks]**.

8819 – 7210

Please start each question on a new page. Full marks are not necessarily awarded for a correct answer with no working. Answers must be supported by working and/or explanations. In particular, solutions found from a graphic display calculator should be supported by suitable working. For example, if graphs are used to find a solution, you should sketch these as part of your answer. Where an answer is incorrect, some marks may be given for a correct method, provided this is shown by written working. You are therefore advised to show all working.

1. [Maximum mark: 7]

Peter, the Principal of a college, believes that there is an association between the score in a Mathematics test, X, and the time taken to run $500\,\text{m}$, Y seconds, of his students.
The following paired data are collected.

Mathematics test score X	70	75	76	66	60	61
Time taken to run $500\,\text{m}$ Y	100	105	95	109	89	101

It can be assumed that (X, Y) follow a bivariate normal distribution with product moment correlation coefficient ρ.

(a) (i) State suitable hypotheses H_0 and H_1 to test Peter's claim, using a two-tailed test.

(ii) Carry out a suitable test at the 5% significance level. With reference to the p-value, state your conclusion in the context of Peter's claim. [5]

(b) Peter uses the regression line of y on x as $y = 0.248x + 83.0$ and calculates that a student with a Mathematics test score of 73 will have a running time of 101 seconds. Comment on the validity of his calculation. [2]

2. [Maximum mark: 15]

(a) Three independent random variables X_1, X_2, X_3 are taken from a distribution with mean μ and variance σ^2. Three estimators are proposed for μ.

$$T_1 = \frac{X_1 + X_2 + X_3}{3},\ T_2 = \frac{X_1 + 2X_2 + 3X_3}{3},\ T_3 = \frac{X_1 + 2X_2}{3}$$

(i) Show that one of these estimators for μ is biased and show that the other two are unbiased.

(ii) For the two unbiased estimators, determine, with a reason, which one is more efficient. [9]

(b) Consider the random variable Y, which follows a negative binomial distribution $Y \sim \mathrm{NB}(4, p)$. A random sample is taken from this distribution and the mean is denoted by $\bar{Y}$.

(i) Find $\mathrm{E}(\bar{Y})$.

(ii) Hence suggest an unbiased estimator for $\dfrac{1}{p}$ in terms of $\bar{Y}$. [2]

(This question continues on the following page)

(Question 2 continued)

(c) A discrete random variable W has a probability distribution given by the following table.

w	1	2
$\mathrm{P}(W = w)$	0.5	0.5

(i) Calculate $\mathrm{E}(W)$.

(ii) Calculate $\mathrm{E}\left(\frac{1}{W}\right)$.

(iii) Hence explain why your estimator for $\frac{1}{p}$ in (b)(ii) does not directly suggest an unbiased estimator for p. [4]

3. [Maximum mark: 14]

(a) State the central limit theorem as applied to a random sample of size n, taken from a distribution with mean μ and variance σ^2. [2]

A random variable X has a distribution with mean μ and variance 4. A random sample of size 100 is to be taken from the distribution of X.

(b) Jack takes a random sample of size 100 and calculates that $\bar{x} = 60.2$. Find an approximate 90% confidence interval for μ. [2]

(c) Josie takes a different random sample of size 100 to test the null hypothesis that $\mu = 60$ against the alternative hypothesis that $\mu > 60$ at the 5% level.

(i) Find the critical region for Josie's test, giving your answer correct to two decimal places.

(ii) Write down the probability that Josie makes a Type I error.

(iii) Given that the probability that Josie makes a Type II error is 0.25, find the value of μ, giving your answer correct to three significant figures. [10]

4. [Maximum mark: 14]

Consider the random variable X, which follows a negative binomial distribution $X \sim \mathrm{NB}(r, p)$. The probability generating function for X is given by

$$G_X(t) = \frac{p^r t^r}{(1-qt)^r}\text{, where } q = 1 - p.$$

(a) Use this probability generating function to find and simplify $\mathrm{E}(X)$. [5]

Consider another independent random variable Y, where $Y \sim \mathrm{NB}(s, p)$.
Let $W = X + Y$.

(b) (i) Find the probability generating function for W.

(ii) Hence identify the distribution that W follows and state its parameters.

(iii) Given that $r = 2$ and $s = 3$, calculate $\mathrm{P}(X = 3 \mid W = 7)$. [9]