

INSTITUTE AND FACULTY OF ACTUARIES



EXAMINATION

30 April 2014 (pm)

Subject CT3 – Probability and Mathematical Statistics Core Technical

Time allowed: Three hours

INSTRUCTIONS TO THE CANDIDATE

1. *Enter all the candidate and examination details as requested on the front of your answer booklet.*
2. *You must not start writing your answers in the booklet until instructed to do so by the supervisor.*
3. *Mark allocations are shown in brackets.*
4. *Attempt all 10 questions, beginning your answer to each question on a new page.*
5. *Candidates should show calculations where this is appropriate.*

Graph paper is required for this paper.

AT THE END OF THE EXAMINATION

Hand in BOTH your answer booklet, with any additional sheets firmly attached, and this question paper.

In addition to this paper you should have available the 2002 edition of the Formulae and Tables and your own electronic calculator from the approved list.

- 1** The following sample shows the durations x_i in minutes for 20 journeys from Edinburgh to Glasgow:

51 53 54 55 59 59 60 60 60 69 71 72 74 90 97 104 107 108 115 167

with $\sum_{i=1}^{20} x_i = 1,585$ and $\sum_{i=1}^{20} x_i^2 = 142,127$.

- (i) Calculate the mean and the median of this sample. [2]
- (ii) Calculate the standard deviation of this sample. [2]
- [Total 4]

- 2** A set of data has mean 62 and standard deviation 6.

Derive a linear transformation for these data that will result in the new data having mean 50 and standard deviation 12. [3]

- 3** Sixty per cent of new drivers in a particular country have had additional driving education. During their first year of driving, new drivers who have not had additional driving education have a probability 0.09 of having an accident, while new drivers who have had additional driving education have a probability 0.05 of having an accident.

- (a) Calculate the probability that a new driver does not have an accident during their first year of driving.
- (b) Calculate the probability that a new driver has had additional driving education, given that the driver had no accidents in the first year. [5]

- 4** Let X be a random variable with probability density function:

$$f(x) = \begin{cases} \frac{1}{2}e^x & ; x \leq 0 \\ \frac{1}{2}e^{-x} & ; x > 0 \end{cases}$$

- (i) Show that the moment generating function of X is given by:

$$M_X(t) = (1 - t^2)^{-1},$$

for $|t| < 1$. [3]

- (ii) Hence find the mean and the variance of X using the moment generating function in part (i). [3]
- [Total 6]

5 Consider ten independent random variables X_1, \dots, X_{10} which are identically distributed with an exponential distribution with expectation 4.

- (i) Specify the approximate distribution of $X = \sum_{i=1}^{10} X_i$, including all parameters, using the central limit theorem. [2]
- (ii) Calculate the approximate value of the probability $P[X < 40]$ using the result in part (i). [1]
- (iii) Calculate the exact probability $P[X < 40]$. [3]
- (iv) Comment on the answers in parts (ii) and (iii). [1]

[Total 7]

6 In an opinion poll, a sample of 100 people from a large town were asked which candidate they would vote for in a forthcoming national election with the following results:

Candidate	A	B	C
Supporters	32	47	21

- (i) Determine the approximate probability that candidate B will get more than 50% of the vote. [3]

A second opinion poll of 150 people was conducted in a different town with the following results:

Candidate	A	B	C
Supporters	57	56	37

- (ii) Use an appropriate test to decide whether the two towns have significantly different voting intentions. [7]

[Total 10]

7 Let X and Y be two continuous random variables.

- (i) Prove that $E[E[Y|X]] = E[Y]$. [3]

Suppose the number of claims, N , on a policy follows a Poisson distribution with mean μ , and the amount of the i^{th} claim, X_i , follows a Gamma distribution with parameters α and λ . Let S denote the total value of claims on a policy in a given year.

- (ii) Derive the mean of S using the result in part (i). [2]

Suppose $\mu = 0.15$, $\alpha = 100$, and $\lambda = 0.1$.

- (iii) Calculate the variance of S . [3]

[Total 8]

- 8 Let X_1, X_2, \dots, X_n be a random sample from a distribution with parameter θ and density function:

$$f(x) = \begin{cases} \frac{2x}{\theta^2} & ; \quad 0 \leq x \leq \theta \\ 0 & ; \quad x < 0 \text{ or } x > \theta \end{cases} .$$

Suppose that $\underline{x} = (x_1, x_2, \dots, x_n)$ is a realisation of X_1, X_2, \dots, X_n .

- (i) (a) Derive the likelihood function $L(\theta; \underline{x})$ and produce a rough sketch of its graph.
- (b) Use the graph produced in part (i)(a) to explain why the maximum likelihood estimate of θ is given by $x_{(n)} = \max\{x_1, x_2, \dots, x_n\}$.

[4]

Let $X_{(n)} = \max\{X_1, X_2, \dots, X_n\}$ be the estimator of θ , that is the random variable corresponding to $x_{(n)}$.

- (ii) (a) Show that the cumulative distribution function of the estimator $X_{(n)}$ is given by:

$$F_{X_{(n)}}(x) = \left(\frac{x}{\theta}\right)^{2n}$$

for $0 \leq x \leq \theta$.

- (b) Hence, derive the probability density function of the estimator $X_{(n)}$.
- (c) Determine the expected value $E(X_{(n)})$ and the variance $V(X_{(n)})$.
- (d) Show that the estimator $\frac{2n+1}{2n} X_{(n)}$ is an unbiased estimator of θ .

[9]

- (iii) (a) Derive the mean square error of the estimator given in part (ii)(d).
- (b) Comment on the consistency of this estimator.

[5]

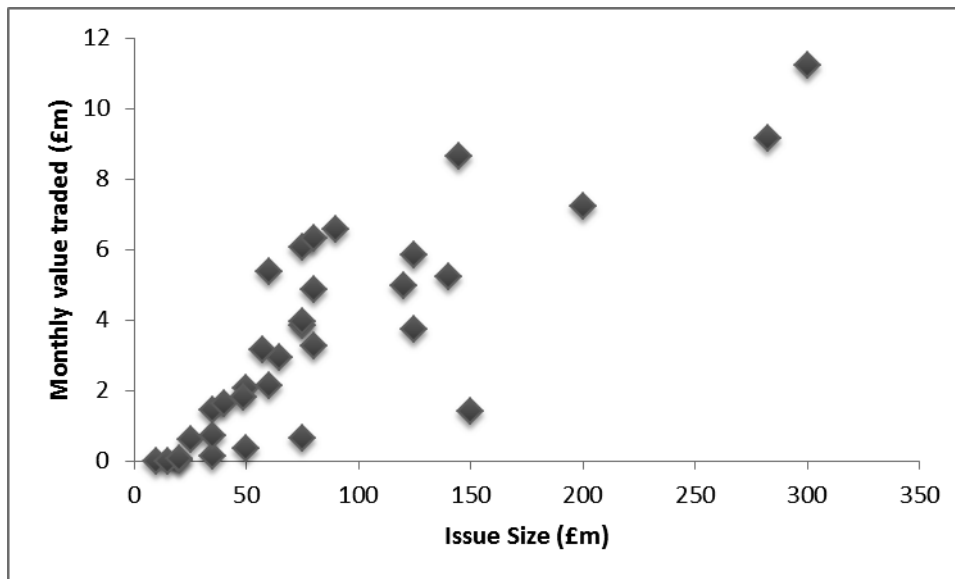
[Total 18]

- 9 The weekly amount spent on childcare for one child is believed to depend on the age of the child. We denote by X the random variable describing the cost per child for a randomly selected child of age one year, Y being the cost for a three year old child, and Z the cost for a five year old child. It is assumed that X , Y , and Z are normally distributed and that childcare costs are independent between children. Random samples of children of different ages are taken and the weekly childcare costs are recorded during the year 2012. A summary of the data is given in the following table:

Random variable	X	Y	Z
Age of child	1	3	5
Average cost per week per child	200	170	155
Sample standard deviation	30	30	20
Sample size	25	25	25

- (i) Calculate the overall average weekly cost of childcare per child for the children in these samples. [1]
- (ii) Calculate a 95% confidence interval for the expected childcare cost for a child aged one year. [2]
- (iii) Calculate a 95% confidence interval for the expected childcare cost for a child aged five years. [2]
- (iv) Calculate a 95% confidence interval for the ratio of the variances of X and Z . [3]
- (v) Perform a test at 5% significance level for the null hypothesis that the variances of X and Z are equal based on your answer to part (iv). [1]
- (vi) Calculate an approximate 95% confidence interval for the difference between the average weekly childcare cost per child for children aged one and for children aged five. Justify any assumptions that you make and explain any approximate values you use. [5]
- (vii) Perform a test to decide if there is a difference between the expected weekly childcare cost per child spent for children aged one and for children aged five based on your answer to part (vi). [1]
- (viii) Perform an analysis of variance to decide if the age of a child has an impact on the weekly amount spent on childcare. [7]
- [Total 22]

- 10 An analyst is instructed to investigate the relationship between the size of a bond issue and its trading volumes (value traded). The data for 33 bonds are plotted in the following chart.



- (i) Comment on the relationship between issue size and value traded. [2]

The analyst denotes issue size by s and monthly value traded by v . He calculates the following from the data:

$$\sum s_i = 2,843.7, \sum s_i^2 = 397,499.8, \sum v_i = 115.34, \sum v_i^2 = 689.37, \sum s_i v_i = 15,417.75$$

- (ii) (a) Determine the correlation coefficient between s and v .
 (b) Perform a statistical test to determine if the correlation coefficient is significantly different from 0. [7]
- (iii) Determine the parameters of a linear regression of v on s and state the fitted model equation. [3]
- (iv) State the outcome of a statistical test to determine whether the slope parameter in part (iii) differs significantly from zero, justifying your answer. [2]

A colleague suggests that the central part of the data, with issue sizes between £50m and £150m, seem to have a greater spread of value traded and without the bonds in the upper and lower tails the linear relationship would be much weaker.

- (v) Comment on the colleague's observation. [3]
 [Total 17]

END OF PAPER

