INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

28 September 2017 (am)

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. *You have 15 minutes of planning and reading time before the start of this examination. You may make separate notes or write on the exam paper but not in your answer booklet. Calculators are not to be used during the reading time. You will then have three hours to complete the paper.*

4. *Mark allocations are shown in brackets.*

5. *Attempt all 10 questions, beginning your answer to each question on a new page.*

6. *Candidates should show calculations where this is appropriate.*

Graph paper is NOT 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 number of cans of fizzy drinks consumed by teenagers each day is the subject of an empirical study. The following data have been recorded.

cans per day	0	1	2	3	4	5
number of teenagers	25	30	26	20	14	10

Assume that no teenager drinks more than five cans per day.

(i) Calculate the mean, median and mode for this sample. [3]

(ii) Comment on the symmetry of the observed data, using your answer to part (i) and without making any further calculations. [2]
[Total 5]

2 Assume that n independent random variables $X_1, X_2, \ldots, X_n$ are observed, which all have the discrete distribution given by

x	1	2	3
$P\left[X_i = x\right]$	0.6	0.3	0.1

and define the mean as $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

(i) Determine the exact distribution of $\overline{X}$ for $n = 2$. [3]

(ii) Determine the approximate distribution of $\overline{X}$ for $n = 50$ including all relevant parameters, stating any assumptions you make. [5]
[Total 8]

3 Let $X_1, X_2, \ldots, X_n$ be a random sample of independent random variables from a $N(\mu, \sigma^2)$ distribution.

(i) Comment on the shape of the sampling distribution of the sample variance S^2 with respect to the sample size n. [2]

(ii) Determine the variance of S^2 based on its sampling distribution. [2]
[Total 4]

4 An airline is analysing the punctuality of its scheduled flights. It measures departures and arrivals and classifies them as early, on time or late. From its records no flights depart early and 85% depart on time. Arrivals are early 10% of the time and late 20% of the time.

(i) Determine the probability that a flight arrives on time. [1]

Further examination shows that:

- none of the flights that depart late are early arrivals.
- 10% of the flights that depart on time are late arrivals.

(ii) Determine the probability that a flight will arrive early if its departure is on time. [3]

(iii) Show that the probability that a flight both departs on time and arrives on time is 0.665. [2]

(iv) Determine the probability that a flight will arrive on time if it departs late. [3]

(v) Determine the probability that if a flight arrives late it departed on time. [2]

[Total 11]

5 In an election between two candidates A and B in a large district, a sample poll of 100 voters chosen at random, indicated that 55% were in favour of candidate A.

(i) Calculate a 95% confidence interval for the proportion of all voters in favour of candidate A based on the above sample. [2]

A candidate is elected if they win more than 50% of the votes. We want a test in which the alternative hypothesis is that support for candidate A is such that she will win the election.

(ii) (a) Write down the hypotheses for this test in terms of a suitable parameter. [1]

(b) Explain whether or not the confidence interval in part (i) can be used to test the hypothesis in part (ii)(a) at the 5% level of significance. [1]

It has been reported in the news that a new poll estimates support for candidate A at 52%, with a margin of error of no more than ±2% with confidence 95%.

(iii) Determine the minimum size of the sample of voters that was taken in this new poll. [2]

[Total 6]

6 Some tea experts claim that the taste of a cup of tea does not change according to whether the tea or the milk is added first to the cup. To test the hypothesis that people cannot tell the difference, an actuary organises a tasting experiment where an individual is asked to taste 10 randomly presented cups of tea; five of these cups were prepared with tea added first, and the other five with milk added first. The individual does not know how many cups of tea where prepared in either way.

Under the null hypothesis that people cannot tell the difference, the probability of correctly recognising the preparation method purely by chance is considered to be 0.5. The actuary adopts the following decision rule for testing the hypothesis:

> *Reject* the null hypothesis if seven or more cups are correctly identified; conclude that the individual can tell the difference.
>
> *Do not reject* the null hypothesis if less than seven cups are correctly identified; conclude that the individual cannot tell the difference.

(i) Determine the probability of a type I error for this test. [3]

Based on past experience we can quantify the probability p that an individual recognises the tea preparation method correctly.

(ii) Determine the probability of a type II error assuming that the true value of p is:

(a) $p = 0.7$. [2]

(b) $p = 0.8$. [2]

(iii) Comment on the power of the test, based on your answers in part (ii). [3]
[Total 10]

7 The annual number of claims an insurance company incurs, N, is believed to follow a Poisson distribution with mean λ. The value of each claim X_i, $i = 1, 2, \ldots$ follows a known distribution with mean μ and variance σ^2. The value of each claim is independent of the value of any other claim and of the number of claims. Let $S = X_1 + X_2 + \ldots + X_N$ denote the total claims in any given year.

(i) Write down an expression for the moment generating function of S in terms of the moment generating function of X_i. [1]

(ii) Derive formulae for the mean and variance of S using your answer to part (i). [5]

[Total 6]

8 The two random variables X_1 and X_2 are independent from each other and follow a uniform $U(-\theta, \theta)$ distribution, where $\theta > 0$ is a parameter.

Let $\hat{\theta}_1 = 3Z$ denote a possible estimator of θ, where $Z = \max(X_1, X_2)$.

(i) Show that the probability density function of Z is given by $f_Z(z) = \frac{z+\theta}{2\theta^2}$, by first deriving its cumulative distribution function. [4]

(ii) Show that $E(Z) = \frac{\theta}{3}$. [1]

(iii) (a) Derive the bias of $\hat{\theta}_1$. [1]

(b) Derive an expression for the mean squared error (MSE) of $\hat{\theta}_1$ in terms of the unknown parameter θ. [4]

Let $\hat{\theta}_2 = 2Z$ denote a different estimator of θ, where again $Z = \max(X_1, X_2)$.

(iv) (a) Show that $\text{bias}\left(\hat{\theta}_2\right) = -\frac{\theta}{3}$. [2]

(b) Show that $\text{MSE}\left(\hat{\theta}_2\right) = \theta^2$. [2]

(v) Comment on how good the two estimators are, based on your answers in parts (iii) and (iv). [2]

[Total 16]

9 The reaction time of drivers under the influence of alcohol is the subject of an empirical study with an emphasis on the relationship between regular physical exercise and reaction time. For this study drivers are categorised into four groups according to the time spent exercising per week: group A (no regular exercising), group B (0–2 hours), group C (2–4 hours) and group D (more than 4 hours per week).

Five drivers are randomly chosen from each group and their reaction time in seconds is measured during a simulated driving test 60 minutes after ingesting a specific amount of alcohol. The measured reaction times are in the following table.

	Reaction time, x_i					$\sum x_i$	$\sum x_i^2$
Group A	0.95	1.40	1.00	1.35	1.30	6.00	7.375
Group B	1.25	1.35	0.90	1.20	1.35	6.05	7.458
Group C	1.10	1.30	1.10	0.95	1.30	5.75	6.703
Group D	1.05	0.80	1.25	1.05	1.05	5.20	5.510

In a first approach the mean reaction time is studied without considering the time spent on exercising.

(i) Calculate a 95% confidence interval for the mean reaction time based on the sample of all twenty drivers, stating any assumptions you make. [6]

It is estimated that the average reaction time of drivers not under the influence of alcohol is 0.9 seconds.

(ii) Test the null hypothesis that the mean reaction time under the influence of alcohol is equal to 0.9 seconds against the alternative hypothesis that the reaction time is different from 0.9 seconds. The test decision should be based on the sample of all twenty drivers. Use a significance level of 5%. [2]

To investigate the impact of regular exercising on the reaction time under the influence of alcohol an analyst suggests carrying out an ANOVA test based on the grouping of drivers and the data in the table above.

(iii) State the mathematical model underlying the one-way analysis of variance together with all assumptions. [2]

(iv) Carry out an analysis of variance to test the null hypothesis that the time spent on exercising has no effect on the average reaction time under the influence of alcohol. [9]

[Total 19]

10 A company leases animals, which have been trained to perform certain tasks, for use in the movie industry. The table below gives the number of tasks that each of nine monkeys in a random sample can perform, along with the number of years the monkeys have been working with the company.

Name	*Hellion*	*Freeway*	*SuSu*	*Henri*	*Jo*	*Peepers*	*Cleo*	*Jeep*	*Maggie*
Years	10	8	6.5	6	5	1.5	0.5	0.5	0.4
Tasks	28	24	28	28	27	23	15	6	23

The random variable Y_i denotes the number of years and T_i the number of tasks for each monkey $i = 1, \ldots, 9$.

$$\sum y_i = 38.4,\ \sum y_i^2 = 270.16,\ \sum y_i t_i = 1011.2,\ \sum t_i = 202,\ \sum t_i^2 = 4976$$

(i) Explain the roles of response and explanatory variables in a linear regression. [2]

(ii) Determine the correlation coefficient between Y and T. [4]

(iii) Perform a statistical test using Fisher's transformation to determine whether the population correlation coefficient is significantly different from zero. [6]

(iv) Determine the parameters of a linear regression, including writing down the equation. [3]
[Total 15]

END OF PAPER