EXAMINATION

30 April 2009 (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. *Mark allocations are shown in brackets.*
4. *Attempt all 13 questions, beginning your answer to each question on a separate sheet.*
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 A random sample of 12 claim amounts (in units of £1,000) on a general insurance portfolio is given by:

14.9	12.4	19.4	3.1	17.6	21.5	15.3	20.1	18.8	11.4	46.2	16.2

For these data: $\Sigma x = 216.9$, $\Sigma x^2 = 5{,}052.13$, sample mean $\bar{x} = £18{,}075$
sample median = £16,900, sample standard deviation $s = £10{,}143$.

Calculate the sample mean, median, and standard deviation of the sample (of size 10) which remains after we remove the claim amounts 3.1 and 46.2 from the original sample (you should show intermediate working and/or give justifications for your answers). [6]

2 Consider three events A, B, and C for which A and C are independent, and B and C are mutually exclusive. You are given the probabilities $P(A) = 0.3$, $P(B) = 0.5$, $P(C) = 0.2$ and $P(A \cap B) = 0.1$.

Find the probability that none of A, B, or C occurs. [3]

3 The random variable X has probability density function

$$f(x) = k(1-x)(1+x), \quad 0 < x < 1,$$

where k is a positive constant.

(i) Show that $k = 1.5$. [2]

(ii) Calculate the probability $P(X > 0.25)$. [2]
[Total 4]

4 Let the random variable Y denote the size (in units of £1,000) of the loss per claim sustained in a particular line of insurance. Suppose that Y follows a chi-square distribution with 2 degrees of freedom. Two such claims are randomly chosen and their corresponding losses are assumed to be independent of each other.

(i) Determine the mean and the variance of the total loss from the two claims. [2]

(ii) Find the value of k such that there is a probability of 0.95 that the total loss from the two claims exceeds k. [2]
[Total 4]

5 The human resources department of a large insurance company currently estimates that 82% of new employees recruited by their call centres will still be employed by the company after one year. A recent extension to the call centre business led to 280 new employees being recruited.

Calculate an approximate value for the probability that at least 240 of these new employees will still be employed by the company after one year. [3]

6 The variables $X_1, X_2, \ldots, X_{40}$ give the size (in units of £100) of each of 40 claims in a random sample of claims arising from damage to cars by vandals. The size of each claim is assumed to follow a gamma distribution with parameters $\alpha = 4$ and $\lambda = 0.5$ and each is independent of all others. Let $\bar{X} = \frac{1}{40}\sum_{i=1}^{40} X_i$ be the random variable giving the mean size of such a sample.

(i) State the approximate sampling distribution of $\bar{X}$ and determine its parameters. [2]

(ii) Determine approximately the median of $\bar{X}$. [1]
[Total 3]

7 A survey is undertaken to investigate the frequency of motor accidents at a certain intersection.

It is assumed that, independently for each week, the number of accidents follows a Poisson distribution with mean λ.

(i) In a single week of observation two accidents occur. Determine a 95% confidence interval for λ, using tables of "Probabilities for the Poisson distribution". [3]

(ii) In an observation period of 30 weeks an average of 2.4 accidents is recorded. Determine a 95% confidence interval for λ, using a normal approximation. [3]

(iii) Comment on your answers in parts (i) and (ii) above. [1]
[Total 7]

8 A random sample of 25 recent claim amounts in a general insurance context is taken from a population that you may assume is normally distributed. In units of £1,000, the sample mean is $\bar{x} = 9.416$ and the sample standard deviation is $s = 2.105$.

Calculate a 95% one-sided upper confidence limit (that is, the upper limit k of a confidence interval of the form $(0,k)$) for the standard deviation of the claim amounts in the population. [5]

9 An analysis of variance investigation with samples of size eight for each of four treatments results in the following ANOVA table.

Source of variation	*d.f.*	*SS*	*MSS*
Between treatments	3	6716	2239
Residual	28	3362	120
Total	31	10078	

(i) Calculate the observed F statistic, specify an interval in which the resulting P-value lies, and state your conclusion clearly. [3]

(ii) The four treatment means are:

$$\bar{y}_1 = 85.0,\ \bar{y}_2 = 66.5,\ \bar{y}_3 = 59.0,\ \text{and}\ \bar{y}_4 = 95.5\,.$$

(a) Calculate the least significant difference between pairs of means using a 5% level.

(b) List the means in order, illustrate the non-significant pairs using suitable underlining, and comment briefly.

[3]
[Total 6]

10 For a group of policies the probability distribution of the total number of claims, N, arising during a period of one year is given by

$$P(N = 0) = 0.70,\quad P(N = 1) = 0.15,\quad P(N = 2) = 0.10,\quad P(N = 3) = 0.05.$$

Each claim amount, X (in units of £1,000), follows a gamma distribution with parameters $\alpha = 2$ and $\lambda = 0.1$ independently of each other claim amount and of the number of claims.

Calculate the expected value and the standard deviation of the total of the claim amounts for a period of one year. [5]

11 The number of claims, X, which arise in a year on each policy of a particular class is to be modelled as a Poisson random variable with mean λ. Let $\underline{X} = (X_1, X_2, \ldots, X_n)$ be a random sample from the distribution of X, and let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

(i) (a) Use moment generating functions to show that $\sum_{i=1}^{n} X_i$ has a Poisson distribution with mean $n\lambda$.

(b) State, with a brief reason, whether or not the variable $2X_1 + 5$ has a Poisson distribution.

(c) State, with a brief reason, whether or not $\bar{X}$ has a Poisson distribution in the case that $n = 2$.

(d) State the approximate distribution of $\bar{X}$ in the case that n is large.

[8]

An actuary is interested in the level of claims being experienced and wants in particular to test the hypotheses

$$H_0: \lambda = 1 \;\; v \;\; H_1: \lambda > 1 .$$

He decides to use a random sample of size $n = 100$ and the best (most powerful) available test. You may assume that this test rejects H_0 for $\bar{x} > k$, for some constant k.

(ii) (a) Show that the value of k for the test with level of significance 0.01 is $k = 1.2326$.

(b) Calculate the power of the test in part (ii)(a) in the case $\lambda = 1.2$ and then in the case $\lambda = 1.5$.

(c) Comment briefly on the values of the power of the test obtained in part (ii)(b).

[9]
[Total 17]

12 In a genetic plant breeding experiment a total of 1,500 plants were categorised into one of four classes (labelled *A*, *B*, *C* and *D*) with the following results:

class:	*A*	*B*	*C*	*D*
frequency:	1071	62	68	299

A genetic model specifies that the probability that an individual plant belongs to each class is given by:

class:	*A*	*B*	*C*	*D*
probability:	$\frac{1}{4}(2+\theta)$	$\frac{1}{4}(1-\theta)$	$\frac{1}{4}(1-\theta)$	$\frac{1}{4}\theta$

where θ is an unknown parameter such that $0 < \theta < 1$.

(i) (a) Write down the likelihood for these data and determine the log-likelihood.

(b) Show that the maximum likelihood estimate (MLE) of θ is a solution of the quadratic equation

$$750\theta^2 - 256\theta - 299 = 0$$

and hence that the MLE is given by $\hat{\theta} = 0.825$. [7]

(ii) (a) Determine the second derivative of the log-likelihood and use this, evaluated at $\hat{\theta} = 0.825$, to obtain an approximation for the Cramer-Rao lower bound for this situation.

(b) Hence calculate an approximate 95% confidence interval for θ, using the asymptotic distribution of the MLE.

[5]

(iii) An extension of the genetic model suggests that the value of θ should be equal to 0.775.

(a) Carry out an appropriate χ^2 test to investigate the extent to which the current data support the extended model with this value of θ (you should calculate and comment on the *P*-value).

(b) Comment briefly on how this relates to your approximate confidence interval in part (ii)(b).

[6]

[Total 18]

13 The following table gives the scores (out of 100) that 10 students obtained on a midterm test (x) and the final examination (y) in a course in statistics.

Midterm x	65	62	50	82	80	68	88	67	90	92
Final y	44	49	54	59	66	67	71	81	89	98

For these data you are given: $S_{xx} = 1{,}760.4$, $S_{yy} = 2{,}737.6$, $S_{xy} = 1{,}529.8$

(i) (a) Draw a scatterplot of the data and comment briefly on the relationship between the score in the final examination and that in the midterm test.

(b) The equation of the line of best fit is given by $y = 3.146 + 0.869x$. Perform a suitable test involving the slope parameter β, to test the null hypothesis H_0: $\beta = 0$ against H_1: $\beta > 0$.

(c) Calculate a 95% confidence interval for the mean final examination score for a midterm score of 75.

(d) Consider now that we require a 95% confidence interval for an individual predicted final examination score for a midterm score of 75.

State (giving reasons) whether this interval will be narrower or wider than the one calculated in part (i)(c) above. (You are not asked to calculate the interval.)

[13]

The lecturer of this course decides to assess the linear relationship between the score in the final examination and that in the midterm test, by using the sample correlation coefficient r.

The hypothesis H_0: $\rho = 0$ (where ρ denotes the population correlation coefficient) can be tested against H_1: $\rho > 0$, by using the result that under H_0 the sampling distribution of the statistic $\dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ is the t_{n-2} distribution (where n is the size of the sample).

(ii) (a) Show algebraically, that is without referring to the specific data given here, that in general the above statistic and the statistic involving β that you used in (i)(b) produce equivalent tests.

(b) Calculate the value of r for the given data and hence verify numerically the result of part (ii)(a) above.

[6]
[Total 19]

END OF PAPER