EXAMINATION

12 April 2005 (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.

1 Calculate the sample mean and standard deviation of the following claim amounts (£):

534	671	581	620	401	340	980	845	550	690

[3]

2 Suppose A, B and C are events with $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$, $P(C) = \frac{1}{3}$, $P(A \cup B) = \frac{3}{4}$, $P(A \cap C) = \frac{1}{6}$, $P(B \cap C) = \frac{1}{6}$ and $P(A \cap B \cap C) = \frac{1}{12}$.

(a) Determine whether or not the events A and B are independent.

(b) Calculate the probability $P(A \cup B \cup C)$.

[4]

3 Claim sizes in a certain insurance situation are modelled by a distribution with moment generating function $M(t)$ given by

$$M(t) = (1 - 10t)^{-2}.$$

Show that $E[X^2] = 600$ and find the value of $E[X^3]$. [3]

4 Consider a random sample of size 16 taken from a normal distribution with mean $\mu = 25$ and variance $\sigma^2 = 4$. Let the sample mean be denoted $\overline{X}$.

State the distribution of $\overline{X}$ and hence calculate the probability that $\overline{X}$ assumes a value greater than 26. [3]

5 Consider a random sample of size 21 taken from a normal distribution with mean $\mu = 25$ and variance $\sigma^2 = 4$. Let the sample variance be denoted S^2.

State the distribution of the statistic $5S^2$ and hence find the variance of the statistic S^2.

[3]

6 In a survey conducted by a mail order company a random sample of 200 customers yielded 172 who indicated that they were highly satisfied with the delivery time of their orders.

Calculate an approximate 95% confidence interval for the proportion of the company's customers who are highly satisfied with delivery times. [3]

7 For a group of policies the total number of claims arising in a year is modelled as a Poisson variable with mean 10. Each claim amount, in units of £100, is independently modelled as a gamma variable with parameters $\alpha = 4$ and $\lambda = 1/5$.

Calculate the mean and standard deviation of the total claim amount. [5]

8 The distribution of claim size under a certain class of policy is modelled as a normal random variable, and previous years' records indicate that the standard deviation is £120.

(i) Calculate the width of a 95% confidence interval for the mean claim size if a sample of size 100 is available. [2]

(ii) Determine the minimum sample size required to ensure that a 95% confidence interval for the mean claim size is of width at most "±£10". [2]

(iii) Comment briefly on the comparison of the confidence intervals in (i) and (ii) with respect to widths and sample sizes used. [1]

[Total 5]

9 Let $X_1, \ldots, X_n$ denote a large random sample from a distribution with unknown population mean μ and known standard deviation 3. The null hypothesis H_0: $\mu = 1$ is to be tested against the alternative hypothesis H_1: $\mu > 1$, using a test based on the sample mean with a critical region of the form $\bar{X} > k$, for a constant k.

It is required that the probability of rejecting H_0 when $\mu = 0.8$ should be approximately 0.05, and the probability of not rejecting H_0 when $\mu = 1.2$ should be approximately 0.1.

(i) Show that the test requires

$$\Phi\left(\frac{k-0.8}{3/\sqrt{n}}\right) \approx 0.95 \text{ and } \Phi\left(\frac{k-1.2}{3/\sqrt{n}}\right) \approx 0.10$$

where Φ is the standard normal distribution function. [4]

(ii) The values for the sample size n and the critical value k which satisfy the requirements of part (i) are $n = 482$ and $k = 1.025$ (you are **not** asked to verify these values).

Calculate the approximate level of significance of the test, and comment on the value. [3]

[Total 7]

10 A model used for claim amounts (X, in units of £10,000) in certain circumstances has the following probability density function, $f(x)$, and cumulative distribution function, $F(x)$:

$$f(x) = \frac{5(10^5)}{(10+x)^6}, \ x > 0 \ ; \ F(x) = 1 - \left(\frac{10}{10+x}\right)^5 .$$

You are given the information that the distribution of X has mean 2.5 units (£25,000) and standard deviation 3.23 units (£32,300).

(i) Describe briefly the nature of a model for claim sizes for which the standard deviation can be greater than the mean. [2]

(ii) (a) Show that we can obtain a simulated observation of X by calculating

$$x = 10\left[(1-r)^{-0.2} - 1\right]$$

where r is an observation of a random variable which is uniformly distributed on (0,1).

(b) Explain why we can just as well use the formula

$$x = 10\left[r^{-0.2} - 1\right]$$

to obtain a simulated observation of X.

(c) Calculate the missing values for the simulated claim amounts in the table below (which has been obtained using the method in (ii)(b) above):

r	*Claim (£)*
0.7423	6,141
0.0291	10,2872
0.2770	29,272
0.5895	11,148
0.1131	54,635
0.9897	207
0.6875	7,782
0.8525	3,243
0.0016	?
0.5154	?

[5]
[Total 7]

11 Twenty insects were used in an experiment to examine the effect on their activity level, y, of 3 standard preparations of a chemical. The insects were randomly assigned, 4 to receive each of the preparations and 8 to remain untreated as controls. Their activity levels were metered from vibrations in a test chamber and were as follows:

	Activity levels (y)								*Totals*
Control	43	40	65	51	33	39	54	62	387
Preparation A	73	55	61	65					254
Preparation B	84	63	51	72					270
Preparation C	46	91	84	71					292

For these data $\sum y = 1,203$, $\sum y^2 = 77,249$.

(i) Conduct an analysis of variance test to establish whether the data indicate significant differences amongst the results for the four treatments. [7]

(ii) (a) Complete the following table of residuals for the data and analysis in part (i) above:

Control	?	?	16.6	2.6	–15.4	–9.4	5.6	13.6
Preparation A	9.5	–8.5	–2.5	1.5				
Preparation B	16.5	?	–16.5	4.5				
Preparation C	?	?	?	–2				

(b) Make a rough plot of the residuals against the treatment means.

(c) State the assumptions underlying the analysis of variance test conducted in part (i).

(d) Comment on how well the data conform to these assumptions in the light of the residual plot. [8]

(iii) It is suggested that any differences can be explained in terms of a difference between controls on the one hand and treated groups on the other.

Comment on any evidence for this and state how you would formally test for this effect (but do not carry out the test). [4]

[Total 19]

12 (i) A random variable Y has a Poisson distribution with parameter θ *but* there is a restriction that zero counts cannot occur. The distribution of Y in this case is referred to as the zero-truncated Poisson distribution.

(a) Show that the probability function of Y is given by

$$p(y) = \frac{\theta^y e^{-\theta}}{y!(1-e^{-\theta})} \quad (y = 1, 2, \ldots).$$

(b) Show that $E[Y] = \theta/(1-e^{-\theta})$. [4]

(ii) (a) Let $y_1, \ldots, y_n$ denote a random sample from the zero-truncated Poisson distribution.

Show that the maximum likelihood estimate of θ may be determined by the solution to the following equation:

$$\bar{y} - \theta - \frac{\theta e^{-\theta}}{1-e^{-\theta}} = 0,$$

and deduce that the maximum likelihood estimate is the same as the method of moments estimate.

(b) Obtain an expression for the Cramer-Rao lower bound (*CR*lb) for the variance of an unbiased estimator of θ.

[9]

(iii) The following table gives the numbers of occupants in 2,423 cars observed on a road junction during a certain time period on a weekday morning.

Number of occupants	1	2	3	4	5	6
Frequency of cars	1,486	694	195	37	10	1

The above data were modelled by a zero-truncated Poisson distribution as given in (i).

The maximum likelihood estimate of θ is $\hat{\theta} = 0.8925$ and the Cramer-Rao lower bound on variance at $\hat{\theta} = 0.8925$ is 5.711574×10^{-4} (you do not need to verify these results.)

(a) Obtain the expected frequencies for the fitted model, and use a χ^2 goodness-of-fit test to show that the model is appropriate for the data.

(b) Calculate an approximate 95% confidence interval for θ and hence calculate a 95% confidence interval for the mean of the zero-truncated Poisson distribution.

[9]
[Total 22]

13 As part of an investigation into health service funding a working party was concerned with the issue of whether mortality rates could be used to predict sickness rates. Data on standardised mortality rates and standardised sickness rates were collected for a sample of 10 regions and are shown in the table below:

Region	*Mortality rate m (per 10,000)*	*Sickness rate s (per 1,000)*
1	125.2	206.8
2	119.3	213.8
3	125.3	197.2
4	111.7	200.6
5	117.3	189.1
6	100.7	183.6
7	108.8	181.2
8	102.0	168.2
9	104.7	165.2
10	121.1	228.5

Data summaries:

$\Sigma m = 1136.1$, $\Sigma m^2 = 129{,}853.03$, $\Sigma s = 1934.2$, $\Sigma s^2 = 377{,}700.62$, $\Sigma ms = 221{,}022.58$

(i) Calculate the correlation coefficient between the mortality rates and the sickness rates and determine the probability-value for testing whether the underlying correlation coefficient is zero against the alternative that it is positive. [4]

(ii) Noting the issue under investigation, draw an appropriate scatterplot for these data and comment on the relationship between the two rates. [3]

(iii) Determine the fitted linear regression of sickness rate on mortality rate and test whether the underlying slope coefficient can be considered to be as large as 2.0. [5]

(iv) For a region with mortality rate 115.0, estimate the expected sickness rate and calculate 95% confidence limits for this expected rate. [4]

[Total 16]

END OF PAPER