EXAMINATION

30 March 2006 (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 12 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 The stem and leaf plot below gives the surrender values (to the nearest €1,000) of 40 endowment policies issued in France and recently purchased by a dealer in such policies in Paris. The stem unit is €10,000 and the leaf unit is €1,000.

5	3
5	6
6	02
6	5779
7	122344
7	556677899
8	1123444
8	567778
9	024
9	6

Determine the median surrender value for this batch of policies. [2]

2 In a certain large population 45% of people have blood group A. A random sample of 300 individuals is chosen from this population.

Calculate an approximate value for the probability that more than 115 of the sample have blood group A. [3]

3 A random sample of size 10 is taken from a normal distribution with mean $\mu = 20$ and variance $\sigma^2 = 1$.

Find the probability that the sample variance exceeds 1, that is find $P(S^2 > 1)$. [3]

4 In a one-way analysis of variance, in which samples of 10 claim amounts (£) from each of three different policy types are being compared, the following means were calculated:

$$\bar{y}_{1\bullet} = 276.7\,, \quad \bar{y}_{2\bullet} = 254.6\,, \quad \bar{y}_{3\bullet} = 296.3$$

with residual sum of squares SS_R given by

$$SS_R = \sum_{i=1}^{3}\sum_{j=1}^{10}(y_{ij} - \bar{y}_{i\bullet})^2 = 15{,}508.6$$

Calculate estimates for each of the parameters in the usual mathematical model, that is, calculate $\hat{\mu}, \hat{\tau}_1, \hat{\tau}_2, \hat{\tau}_3,$ and $\hat{\sigma}^2$. [4]

5 A large portfolio of policies is such that a proportion p ($0 < p < 1$) incurred claims during the last calendar year. An investigator examines a randomly selected group of 25 policies from the portfolio.

(i) Use a Poisson approximation to the binomial distribution to calculate an approximate value for the probability that there are at most 4 policies with claims in the two cases where (a) $p = 0.1$ and (b) $p = 0.2$. [3]

(ii) Comment briefly on the above approximations, given that the exact values of the probabilities in part (i), using the binomial distribution, are 0.9020 and 0.4207 respectively. [2]

[Total 5]

6 One variable of interest, T, in the description of a physical process can be modelled as $T = XY$ where X and Y are random variables such that $X \sim N(200, 100)$ and Y depends on X in such a way that $Y|X = x \sim N(x, 1)$.

Simulate two observations of T, using the following pairs of random numbers (observations of a uniform (0, 1) random variable), explaining your method and calculations clearly:

Random numbers

0.5714 , 0.8238

0.3192 , 0.6844

[6]

7 Let $(X_1, X_2, \ldots, X_n)$ be a random sample from a uniform distribution on the interval $(-\theta, \theta)$, where θ is an unknown positive number.

A particular sample of size 5 gives values 0.87, −0.43, 0.12, −0.92, and 0.58.

(i) Draw a rough graph of the likelihood function $L(\theta)$ against θ for this sample. [3]

(ii) State the value of the maximum likelihood estimate of θ. [2]

[Total 5]

8 The events that lead to potential claims on a policy arise as a Poisson process at a rate of 0.8 per year. However the policy is limited such that only the first three claims in any one year are paid.

(i) Determine the probabilities of 0, 1, 2 and 3 claims being paid in a particular year. [2]

(ii) The amounts (in units of £100) for the claims paid follow a gamma distribution with parameters $\alpha = 2$ and $\lambda = 1$.

Calculate the expectation of the sum of the amounts for the claims paid in a particular year. [3]

(iii) Calculate the expectation of the sum of the amounts for the claims paid in a particular year, given that there is at least one claim paid in the year. [2]

[Total 7]

9 The total claim amount on a portfolio, S, is modelled as having a compound distribution

$$S = X_1 + X_2 + \ldots + X_N$$

where N is the number of claims and has a Poisson distribution with mean λ, X_i is the amount of the i^{th} claim, and the X_i's are independent and identically distributed and independent of N. Let $M_X(t)$ denote the moment generating function of X_i.

(i) Show, using a conditional expectation argument, that the cumulant generating function of S, $C_S(t)$, is given by

$$C_S(t) = \lambda\{M_X(t) - 1\}.$$

Note: You may quote the moment generating function of a Poisson random variable from the book of Formulae and Tables. [4]

(ii) Calculate the variance of S in the case where $\lambda = 20$ and X has mean 20 and variance 10. [2]

[Total 6]

10 A marketing consultant was commissioned to conduct a questionnaire survey of the clients of a financial company. The total number of respondents was 650, of whom 220 had investments above a specified threshold.

(i) Each respondent who had investments above the threshold was asked about the percentage of these investments that was held in the form of a certain type of trust. The respondents answered by ticking appropriate boxes and the results led to the following frequency distribution.

percentage	*< 10*	*10–25*	*25–50*	*> 50*
frequency	22	76	73	49

(a) Present these data graphically using a carefully drawn histogram.

(b) Calculate the mean percentage for the full set of 220 such respondents, assuming that the frequencies in each category are uniformly spread over the corresponding range. [5]

(ii) Calculate a 95% confidence interval for the percentage of such investors who would have investments above the threshold. [4]

The same respondents with investments referred to in part (i) were also asked to specify their satisfaction with the current return received from their full portfolio of investment. This was in the form of a four-point qualitative scale: very satisfied, quite satisfied, a little disappointed, very disappointed. The following two-way table of frequencies was obtained.

	percentage in type of trust			
	<10	*10–25*	*25–50*	*>50*
very satisfied	1	6	7	6
quite satisfied	8	29	36	27
a little disappointed	10	37	28	15
very disappointed	3	4	2	1

In order to investigate whether there is any relationship between the percentage in such trusts and satisfaction with current return, a χ^2 test is to be performed.

(iii) Calculate the expected frequencies for the above table under an appropriate hypothesis (which should be stated) and comment on why it would be inappropriate to carry out a χ^2 test directly with these data. [3]

(iv) Combining the "very satisfied" and "quite satisfied" categories together and the "a little disappointed" and "very disappointed" categories together results in the following reduced two-way table.

	percentage in type of trust			
	<10	*10–25*	*25–50*	*>50*
satisfied	9	35	43	33
disappointed	13	41	30	16

Perform the required χ^2 test at the 5% level using this reduced table and comment on your conclusion. [7]

[Total 19]

11 An actuary has been advised to use the following positively-skewed claim size distribution as a model for a particular type of claim, with claim sizes measured in units of £100,

$$f(x;\theta) = \frac{x^2}{2\theta^3}\exp\left(-\frac{x}{\theta}\right) \quad : \quad 0 < x < \infty,\ \theta > 0$$

with moments given by $E[X] = 3\theta$, $E[X^2] = 12\theta^2$ and $E[X^3] = 60\theta^3$.

(i) Determine the variance of this distribution and calculate the coefficient of skewness. [4]

(ii) Let $X_1, X_2, \ldots, X_n$ be a random sample of n claim sizes for such claims.

Show that the maximum likelihood estimator (MLE) of θ is given by $\hat{\theta} = \dfrac{\bar{X}}{3}$ and show that it is unbiased for θ. [5]

(iii) A sample of $n = 50$ claim sizes yields $\Sigma x_i = 313.6$ and $\Sigma x_i^2 = 2{,}675.68$.

(a) Calculate the MLE $\hat{\theta}$.

(b) Calculate the sample variance and comment briefly on its comparison with the variance of the distribution evaluated at $\hat{\theta}$.

(c) Given that the sample coefficient of skewness is 1.149, comment briefly on its comparison with the coefficient of skewness of the distribution.

[4]

(iv) (a) Write down a large-sample approximate 95% confidence interval for the mean of the distribution in terms of the sample mean $\bar{x}$ and the sample variance s^2. Hence obtain an approximate 95% confidence interval for θ and evaluate this for the data in part (iii) above.

(b) Evaluate the variance of the distribution at both the lower and upper limits of this confidence interval and comment briefly with reference to your answer in part (iii)(b) above.

[5]
[Total 18]

12 In an experiment to compare the effects of vaccines of differing strengths intended to give protection to children against a particular condition, twelve batches of vaccine were tested in twelve equal-sized groups of children. The percentages of children who subsequently remained healthy after exposure to the condition, named the *PRH* values, were recorded. The strength of each batch of vaccine was measured by an independent test and recorded as the *SV* value.

The recorded values are:

Batch:	1	2	3	4	5	6	7	8	9	10	11	12
PRH (y):	16	68	23	35	42	41	46	48	52	50	54	53
SV (x):	0.9	1.6	2.3	2.7	3.0	3.3	3.7	3.8	4.1	4.2	4.3	4.5

$$\sum x = 38.4\,;\ \sum y = 528\,;\ \sum x^2 = 137.16\,;\ \sum y^2 = 25{,}428\,;\ \sum xy = 1{,}778.4$$

(i) Draw a rough plot of the data to show the relationship between the *SV* and *PRH* values. [2]

It is evident that one of the observations is "out of line" and so may have an undue effect on any regression analysis. You are asked to investigate this as follows.

(ii) (a) Calculate the total, regression, and error sums of squares for a least-squares linear regression analysis for predicting *PRH* values from *SV* values using all 12 data observations.

(b) Determine the coefficient of determination R^2.

(c) Determine the equation of the fitted regression line.

(d) Examine whether or not there is evidence, at the 5% level of testing, to enable one to conclude that the slope of the underlying regression equation is non-zero.

[11]

The details of the regression analysis after removing the data for batch 2 are given in the box below.

Regression equation: $y = 3.76 + 11.4\,x$

	Coef	*Stdev*	*t-ratio*	*p-val*
Intercept	3.757	3.092	1.22	0.255
x	11.377	0.8838	12.9	0.000

Analysis of Variance

Source	*df*	*SS*	*MS*	*F*	*p-val*
Regression	1	1486.9	1486.9	165.69	0.000
Error	9	80.8	8.98		
Total	10	1567.6			

(iii) (a) Comment on the main differences in the results of the regression analysis resulting from removing the data for batch 2.

(b) Calculate a 95% confidence interval for the expected (mean) *PRH* value for a batch of vaccine with *SV* value 3.5.

[9]

[Total 22]

END OF PAPER