INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

15 April 2011 (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 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 The numbers of claims which have arisen in the last twelve years on 60 motor policies (continuously in force over this period) are shown (sorted) below:

0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	3	3	3
3	3	3	3	3	3	3	4	4	4	4	4	5	5	5	5	6	6	6	7

Derive:

(i) The sample median, mode and mean of the number of claims. [3]
(ii) The sample inter-quartile range of the number of claims. [2]
(iii) The sample standard deviation of the number of claims. [3]
[Total 8]

2 A random sample of size $n = 36$ has sample standard deviation $s = 7$.

Calculate, approximately, the probability that the mean of this sample is greater than 44.5 when the mean of the population is $\mu = 42$. [3]

3 In a large population, 35% of voters intend to vote for party A at the next election. A random sample of 200 voters is selected from this population and asked which party they will vote for.

Calculate, approximately, the probability that 80 or more of the people in this sample intend to vote for party A. [4]

4 Let N be the random variable that describes the number of claims that an insurer receives per month for one of its claim portfolios. We assume that N has a Poisson distribution with $E[N] = 50$. The amount X_i of each claim in the portfolio is normally distributed with mean $\mu = 1{,}000$ and variance $\sigma^2 = 200^2$. The total amount of all claims received during one month is

$$S = \sum_{i=1}^{N} X_i$$

with $S = 0$ for $N = 0$. We assume that $N, X_1, X_2, \ldots$ are all independent of each other.

(i) Specify the type of the distribution of S. [1]

(ii) Calculate the mean and standard deviation of S. [3]
[Total 4]

5 Let X_1, X_2, X_3, X_4, and X_5 be independent random variables, such that X_i ~ gamma with parameters i and λ for $i = 1, 2, 3, 4, 5$. Let $S = 2\lambda\sum_{i=1}^{5} X_i$.

(i) Derive the mean and variance of S using standard results for the mean and variance of linear combinations of random variables. [3]

(ii) Show that S has a chi-square distribution using moment generating functions and state the degrees of freedom of this distribution. [4]

(iii) Verify the values found in part (i) using the results of part (ii). . [1]

[Total 8]

6 Consider two random variables X and Y, for which the variances satisfy $V[X] = 5V[Y]$ and the covariance $Cov[X,Y]$ satisfies $Cov[X,Y] = V[Y]$.

Let $S = X + Y$ and $D = X - Y$.

(i) Show that the covariance between S and D satisfies $Cov[S,D] = 4V[Y]$. [3]

(ii) Calculate the correlation coefficient between S and D. [3]

[Total 6]

7 An insurance company distinguishes between three types of fraudulent claims:

Type 1: legitimate claims that are slightly exaggerated
Type 2: legitimate claims that are strongly exaggerated
Type 3: false claims

Every fraudulent claim is characterised as exactly one of the three types. Assume that the probability of a newly submitted claim being a fraudulent claim of type 1 is 0.1. For type 2 this probability is 0.02, and for type 3 it is 0.003.

(i) Calculate the probability that a newly submitted claim is not fraudulent. [1]

The insurer uses a statistical software package to identify suspicious claims. If a claim is fraudulent of type 1, it is identified as suspicious by the software with probability 0.5. For a type 2 claim this probability is 0.7, and for type 3 it is 0.9.

Of all newly submitted claims, 20% are identified by the software as suspicious.

(ii) Calculate the probability that a claim that has been identified by the software as suspicious is:

(a) a fraudulent claim of type 1,
(b) a fraudulent claim of any type.

[5]

(iii) Calculate the probability that a claim which has NOT been identified as suspicious by the software is in fact fraudulent. [3]
[Total 9]

8 Two medications, labelled A and B, were being investigated using a group of twelve patients each of whom was approximately at the same stage of suffering from a severe cough. The patients were divided randomly into two groups of six and medication A was administered to each patient in the first group while medication B was administered to each patient in the second group. Over the next three days the total number of coughs was recorded for each patient, with the following results:

A:	321	585	468	619	447	532
B:	478	381	596	552	358	426

$$\Sigma x_A = 2,972, \quad \Sigma x_A^2 = 1,530,284, \quad \Sigma x_B = 2,791, \quad \Sigma x_B^2 = 1,343,205$$

(i) Apply an appropriate t-test to determine whether these two medications differ in their effectiveness for the relief of coughing, assuming that the two samples are independent and come from normal populations with equal variances. [7]

(ii) In relation to the test performed in part (i):

(a) Comment on the required assumption of independence.

(b) Present the data graphically and hence comment on the required assumption of normality.

(c) Apply an appropriate F-test to comment on the required assumption of equal variances.

[6]

Suppose that the investigators had used a total of eighteen such patients divided into three groups of six and that a placebo (an inactive substance) was administered to each patient in the third group, labelled C. Suppose that the resulting data for medications A and B were as above together with the following results for the placebo group.

C:	691	827	785	531	603	714

$$\Sigma x_C = 4{,}151\,, \quad \Sigma x_C^2 = 2{,}933{,}001$$

(iii) Perform an analysis of variance to test the hypothesis that there is no difference among the three groups as regards coughing. [8]

(iv) Comment briefly on any difference among the three groups. [1]

[Total 22]

9 Claims on a certain type of policy are such that the claim amounts are approximately normally distributed.

(i) A sample of 101 such claim amounts (in £) yields a sample mean of £416 and sample standard deviation of £72. For this type of policy:

(a) Obtain a 95% confidence interval for the mean of the claim amounts.

(b) Obtain a 95% confidence interval for the standard deviation of the claim amounts.

[8]

The company makes various alterations to its policy conditions and thinks that these changes may result in a change in the mean, but not the standard deviation, of the claim amounts. It wants to take a random sample of claims in order to estimate the new mean amount with a 95% confidence interval equal to

sample mean ± £10.

(ii) Determine how large a sample must be taken, using the following as an estimate of the standard deviation:

(a) The sample standard deviation from part (i).

(b) The upper limit of the confidence interval for the standard deviation from part (i)(b).

[6]

(iii) Comment briefly on your two answers in (ii)(a) and (ii)(b). [2]

[Total 16]

10 A life insurance company runs a statistical analysis of mortality rates. The company considers a population of 100,000 individuals. It assumes that the number of deaths X during one year has a Poisson distribution with expectation $E[X] = \mu$. Over four years the company has observed the following realisations of X (number of deaths).

Year	1	2	3	4
Number of deaths (*per 100,000 lives*)	1,140	1,200	1,170	1,190

The maximum likelihood estimator for the parameter μ of the Poisson distribution is given by $\bar{X}$.

(i) Obtain the maximum likelihood estimate of the parameter μ using these data. [1]

To obtain a more realistic model, it is proposed that the number of deaths should depend on the age of the population. To this end the total population is divided into four age groups of equal size and the number of deaths in each group during the following year is counted. The observed values are given in the following table. The total number of deaths is again per 100,000 lives.

Middle age (t) in group	25	35	45	55
Number of deaths (x) *in age group*	84	113	255	727

For these data we obtain: $\Sigma t = 160$, $\Sigma t^2 = 6,900$, $\Sigma x = 1,179$, $\Sigma x^2 = 613,379$ and $\Sigma xt = 57,515$

(ii) (a) Calculate the correlation coefficient between the middle age t in a group and the number of deaths x in that group, and comment briefly on its value.

(b) Perform a linear regression of the number of deaths x as a function of the middle age t of the group. [10]

A statistician suggests using a Poisson distribution for the number of deaths per year in each group, where the parameter μ depends on the middle age in that group. Under the suggested model the number of deaths in the group with middle age t_i is given by $X_i \sim \text{Poisson}(\mu_i)$ with $\mu_i = wt_i$, where t_i is the middle age of the group that the individual belongs to at the time of death.

(iii) Derive a maximum likelihood estimator for the parameter w and estimate the value of w from the data in the above table. [9]

[Total 20]

END OF PAPER