INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

7 October 2015 (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 11 questions, beginning your answer to each question on a new page.*
5. *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 A random sample of 20 claim amounts (x, in £) made on a certain type of travel insurance policy (type A) was selected and gave the following data summaries:

$\sum x = 11{,}860$
$\sum x^2 = 8{,}438{,}200$
sample mean = 593
sample standard deviation = 271.95.

It was later discovered that two of these 20 claims were made on a different incorrect type of policy, and the corresponding amounts were £770 and £510.

These claims are going to be replaced by two claims made on the correct type of policy, with corresponding amounts £1,000 and £280.

(i) Determine the sample mean and standard deviation of the revised sample. [4]

(ii) Comment on how your answers compare with the original sample mean and standard deviation. [3]
[Total 7]

2 Consider the following measure of skewness for a unimodal distribution:

$$\zeta = \frac{\text{mean} - \text{mode}}{\text{standard deviation}}.$$

(i) Determine the value of ζ for a gamma distribution with parameters $\alpha = 1.6$ and $\lambda = 0.2$. [5]

(ii) Comment on why ζ is a suitable measure of skewness for distributions with one mode. [3]
[Total 8]

3 Random samples are drawn from two different normally distributed populations with variances σ_1^2 and σ_2^2. From the first population a sample of 25 measurements is taken with a sample variance of $s_1^2 = 2.4$. For the second population the sample size is 13 with a sample variance of $s_2^2 = 1.5$.

Determine a 95% confidence interval for the ratio of the true variances σ_1^2 / σ_2^2. [3]

4 During a particular year, it was found that 83 claims were made in a sample of 500 insurance policies. Policies are assumed to be independent from each other and the number of claims per policy is identically distributed for all policies according to a Poisson distribution, with parameter λ denoting the claim rate per policy per year.

(i) Calculate an approximate 95% confidence interval for λ. [3]

(ii) Comment on the validity of this confidence interval. [1]
[Total 4]

5 An insurance company is accused of delaying payments for large claims. To investigate this accusation a sample of 25 claims is considered. In each case the claim size x_i (in £) and the time y_i (in days) taking to pay the claim are recorded. Assume that the claim size and the time taken to pay the claim are normally distributed. In the sample the following statistics have been observed:

$$\sum_{i=1}^{25}(x_i - \bar{x})^2 = 5{,}116{,}701, \qquad \sum_{i=1}^{25}(y_i - \bar{y})^2 = 61.44\,.$$

$$\sum_{i=1}^{25}(x_i - \bar{x})(y_i - \bar{y}) = 2{,}606.96\,.$$

(i) Calculate the correlation coefficient between the claim sizes, x_i, and the times taken to pay the claim, y_i. [1]

(ii) Perform a statistical test of the hypothesis that the correlation between claim size and time until payment is zero against the alternative that the correlation is different from zero. [3]
[Total 4]

6 Consider a survey of alcohol consumption in three different locations in the UK. In each of the three locations 50 men are asked about the units of alcohol they consumed during the week preceding the survey. The results are summarised in the following table:

Location code	A	B	C
Average number of units	26	22	27
Sample standard deviation	7	6	9

Perform a one-way analysis of variance test to test the hypothesis that the location has no impact on alcohol consumption. [6]

7 X and Y are discrete random variables with joint distribution given below.

	$Y = -1$	$Y = 0$	$Y = 1$
$X = 1$	0	1/4	0
$X = 0$	1/4	1/4	1/4

(i) Determine the conditional expectation $E[Y|X = 1]$. [1]

(ii) Determine the conditional expectation $E[X|Y = y]$ for each value of y. [3]

(iii) Determine the expected value of X based on your conditional expectation results from part (ii). [2]

[Total 6]

8 Consider three groups of policyholders: A, B and C. Denote by X_A the random variable for the number of claims that a randomly chosen policyholder in group A submits during any particular calendar year. X_B and X_C denote the corresponding random variables for policyholders in groups B and C.

Assume that X_A, X_B and X_C have Poisson distributions with parameters $\lambda_A = 0.2$, $\lambda_B = 0.1$ and $\lambda_C = 0.05$, depending on the group.

Each policyholder belongs to exactly one group and group membership does not change during the lifetime of a policyholder.

Assume that:

- any individual policyholder submits a claim during any year independently of claims submitted by other policyholders.
- the number of claims a policyholder submits during a year depends on the group the policyholder belongs to, but given which group the policyholder is a member of, the number of claims submitted during a year is independent of the number of claims the policyholder submitted in previous years.

An insurance company has a portfolio of policies with 20% of policyholders belonging to group A, 20% belonging to group B and the remaining policyholders belonging to group C.

The insurance company randomly chooses one of its policyholders.

(i) Calculate the probability that this policyholder will submit at least two claims in a particular year given that he belongs to group A. [2]

Now assume that the insurance company does not know to which group the randomly selected policyholder belongs.

(ii) Show that the probability that the randomly selected policyholder submits exactly one claim in any particular year is approximately 0.0794. [3]

(iii) Determine the probability that the randomly selected policyholder belongs to group A given that the policyholder submitted exactly one claim in the previous year. [2]

(iv) Determine the probability that the randomly chosen policyholder will submit one claim during the current year given that he submitted one claim in the previous year. [5]
[Total 12]

9 A survey team is using satellite technology to measure the height of a mountain. This is an established technology and the variability of measurements is known. On each satellite pass over a mountain they get a measurement that they know lies within ±5m of the true height with a 95% probability. The survey height is given by the mean of the measurements. They assume that all measurements are independent and follow a normal distribution with mean equal to the true height.

(i) (a) Show that the standard deviation of a single measurement is 2.551m.

(b) Determine how many satellite passes over a mountain are required to have a 95% confidence interval for the true height with width less than 1m.
[5]

In a mountain range there are two summits which appear to have a similar height. The team manages to get 20 measurements for each summit and finds there is a difference of 1.6m between the mean survey height of the two summits.

(ii) Perform a statistical test of the null hypothesis that the summits' true heights are the same, against the alternative that they are different. [4]

At the same time the team is testing a new system on these two summits. They again get 20 measurements on each summit with an estimated standard deviation on the first summit of 2.5m and on the second of 2.6m. This system also measures the difference in survey heights between the two peaks to be 1.6m.

(iii) Perform a statistical test of the same hypotheses as in part (ii) when heights are measured by the new system, justifying any assumptions you make. [5]

(iv) Comment on your answers to parts (ii) and (iii). [3]
[Total 17]

10 The random variables $X_1, X_2, ..., X_n$ are independent from each other and all follow a Poisson distribution with parameter λ.

(i) Derive the maximum likelihood estimator of λ based on $X_1, X_2, ..., X_n$. You are not required to verify that your answer corresponds to a maximum. [4]

(ii) Derive an expression for an approximate 95% confidence interval for λ under the situation in part (i), using the Cramer-Rao lower bound. [4]

Suppose that instead of observing the values of $X_{1,} X_{2,} \ldots, X_n$ precisely, we only observe that for K of these variables we have $X_i = 0$, while for the remaining variables we have $X_i > 0$.

(iii) (a) Derive the maximum likelihood estimator of λ when only this information is available. You are not required to verify that your answer corresponds to a maximum. [6]

(b) Explain why we need to observe at least one variable to be equal to zero for the estimator in part (iii) (a) to provide a sensible answer. [1]

(iv) State, with reasons, whether you would prefer to use the estimator of λ in part (i) or that in part (iii). [2]

[Total 17]

11 A property agent carries out a study on the relationship between the age of a building and the maintenance costs, X, per square metre per annum based on a sample of 86 buildings. In the sample denote by x_i the annual maintenance costs per square metre for building i. In a first step the sample is divided into new and old buildings. The maintenance costs are summarised in the following table:

	sample size n	$\sum x_i$	$\sum x_i^2$
new buildings	25	100	800
old buildings	61	300	2200

(i) Perform a test for the null hypothesis that the variance of the maintenance costs of new buildings is equal to the variance of the maintenance costs for old buildings, against the alternative that the variance of the maintenance costs of new buildings is larger. Use a significance level of 5%. [6]

(ii) Perform a test of the null hypothesis that the mean of the maintenance costs of new buildings is equal to the mean of the maintenance costs for old buildings, against the alternative of different means. Use a significance level of 5%. [4]

To obtain further insight into the relationship between age and maintenance costs for old buildings the agent wishes to carry out a linear regression analysis. Let A denote the age of a building and X denote the annual maintenance costs per square metre. The agent uses the model $E[X] = \gamma A + \beta$. The agent has the following summary data for the age a_i and costs x_i of the 61 old buildings in the sample.

$$\sum_{i=1}^{61} a_i = 4{,}500\,,\ \sum_{i=1}^{61} a_i x_i = 30{,}000 \text{ and } \sum_{i=1}^{61} a_i^2 = 506{,}400\,.$$

(iii) Estimate the correlation coefficient $\rho(A, X)$ between age A and maintenance costs X. [4]

(iv) Estimate the parameters γ and β. [2]

[Total 16]

END OF PAPER