INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

3 October 2013 (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 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 The stem and leaf plot below shows 40 observations of an exchange rate.

1.21	9
1.22	4569
1.23	2679
1.24	3467889
1.25	011222345677778
1.26	00346688
1.27	
1.28	1

For these data, $\sum x = 50.000$.

(i) Find the mean, median and mode. [3]

(ii) State, with reasons, which measure of those considered in part (i) you would prefer to use to estimate the central point of the observations. [1]
[Total 4]

2 An insurance company experiences claims at a constant rate of 150 per year.

Find the approximate probability that the company receives more than 90 claims in a period of six months. [4]

3 The random variable X has a distribution with probability density function given by

$$f(x) = \begin{cases} \dfrac{2x}{\theta^2} & ; \quad 0 \le x \le \theta \\ 0 & ; \quad x < 0 \text{ or } x > \theta \end{cases}$$

where θ is the parameter of the distribution.

(i) Derive expressions in terms of θ for the expected value and the variance of X. [3]

Suppose that $X_1, X_2, ..., X_n$ is a random sample, with mean $\bar{X}$, from the distribution of X.

(ii) Show that the estimator $\hat{\theta} = \dfrac{3\bar{X}}{2}$ is an unbiased estimator of θ. [2]
[Total 5]

4 An actuary is considering statistical models for the observed number of claims, X, which occur in a year on a certain class of non-life policies. The actuary only considers policies on which claims do actually arise. Among the considered models is a model for which

$$P(X=x) = -\frac{1}{\log(1-\theta)}\frac{\theta^x}{x}, \quad x=1, 2, 3, \ldots$$

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

Suppose that the actuary has available a random sample $X_1, X_2, \ldots, X_n$ with sample mean $\bar{X}$.

(i) Show that the method of moments estimator (MME), $\tilde{\theta}$, satisfies the equation

$$\bar{X}\left(1-\tilde{\theta}\right)\log\left(1-\tilde{\theta}\right)+\tilde{\theta}=0.$$ [3]

(ii) (a) Show that the log likelihood of the data is given by

$$l(\theta) \propto -n\log\left\{-\log\left(1-\theta\right)\right\}+\sum_{i=1}^{n} x_i \log(\theta).$$

(b) Hence verify that the maximum likelihood estimator (MLE) of θ is the same as the MME. [4]

(iii) Suggest two ways in which the MLE of θ can be computed when a particular data set is given. [1]

[Total 8]

5 Consider a random sample consisting of the random variables $X_1, X_2, ..., X_n$ with mean μ and variance σ^2. The variables are independent of each other.

(i) Show that the sample variance, S^2, is an unbiased estimator of the true variance σ^2. [3]

Now consider in addition that the random sample comes from a normal distribution, in which case it is known that $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.

(ii) (a) Derive the variance of S^2 in terms of σ and n.

(b) Comment on the quality of the estimator S^2 with respect to the sample size n. [4]

[Total 7]

6 A researcher obtains samples of 25 items from normally distributed measurements from each of two factories. The sample variances are 2.86 and 9.21 respectively.

(i) Perform a test to determine if the true variances are the same. [3]

(ii) For each factory calculate central 95% confidence intervals for the true variances of the measurements. [4]

(iii) Comment on how your answers in parts (i) and (ii) relate to each other. [1]
[Total 8]

7 A motor insurance company has a portfolio of 100,000 policies. It distinguishes between three groups of policyholders depending on the geographical region in which they live. The probability p of a policyholder submitting at least one claim during a year is given in the following table together with the number, n, of policyholders belonging to each group.. Each policyholder belongs to exactly one group and it is assumed that they do not move from one group to another over time.

Group	*A*	*B*	*C*
p	0.15	0.1	0.05
n (in 1000s)	20	20	60

It is assumed that any individual policyholder submits a claim during any year independently of claims submitted by other policyholders. It is also assumed that whether a policyholder submits any claims in a year is independent of claims in previous years conditional on belonging to a particular group.

(i) Show that the probability that a randomly selected policyholder will submit a claim in a particular year is 0.08. [2]

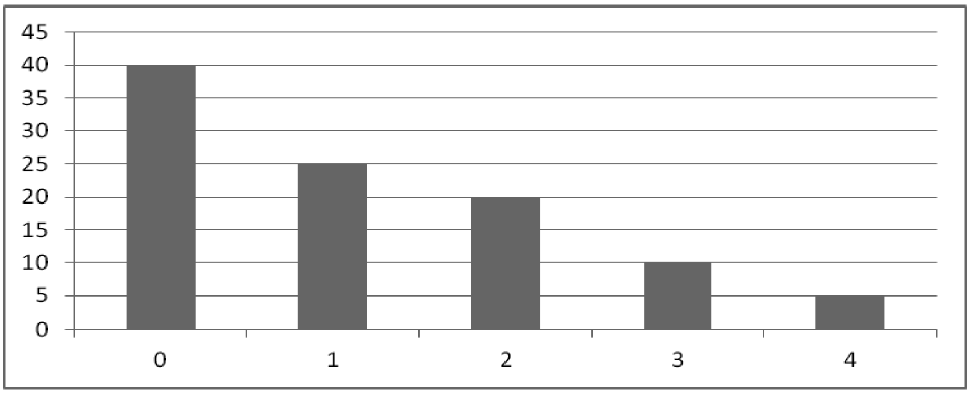
(ii) Calculate the probability that a randomly selected policyholder will submit a claim in a particular year given that the policyholder is not in group C. [2]

(iii) Calculate the probability for a randomly selected policyholder to belong to group A given that the policyholder submitted a claim last year. [2]

(iv) Calculate the probability that a randomly selected policyholder will submit a claim in a particular year given that the policyholder submitted a claim in the previous year. It is assumed that the insurance company does not know to which group the policyholder belongs. [3]

(v) Calculate the probability that a randomly selected policyholder will submit a claim in two consecutive years. [2]
[Total 11]

8 The following graph shows the number of policyholders who made 0, 1, 2, 3 or 4 claims during the last year in a group of 100 policyholders.



(i) Calculate the sample mean, median, mode and standard deviation of the number of claims per policyholder. [5]

Assume that the number of claims X per policyholder per annum from this group of policyholders has a Poisson distribution with unknown parameter λ.

(ii) Calculate an approximate 95% confidence interval for the unknown parameter λ using the data in the above graph, justifying the validity of your approach. [4]

The following table shows the average claim size for each group of number of claims that a policyholder made during the last year.

Number of claims per policyholder	0	1	2	3	4
Average claim size (£)	--	1000	1100	930	980

Assume that the claim size is independent of the number of claims, and that policyholders make claims independently. Also assume that the size of each claim is normally distributed with estimated standard deviation $s = £120$.

(iii) Estimate the expected size of a single claim. [2]

(iv) State the type of the distribution of the total amount claimed in the group of the 100 policyholders. [1]

Now assume that the number of claims per policyholder has a Poisson distribution with true parameter $\lambda = 1.15$ and that the true expected value of the size of a single claim is £1,010 and its true standard deviation is £120.

(v) Calculate the expected value of the total amount claimed in the group of the 100 policyholders and its standard deviation. [4]

[Total 16]

9 The random variables Y_A and Y_B describe the number of hours per month that a randomly selected household in Cities A and B, respectively, uses its car. Both cities recently decided to introduce measures to reduce road congestion. To investigate the effect of these measures ten households in each city were randomly selected and asked about the hours per month that they use their car before and after the measures were introduced. The random variables Z_A and Z_B describe the hours of car usage after the measures have been introduced, and $X_A = Y_A - Z_A$ and $X_B = Y_B - Z_B$ denote the reduction in car usage. The following table shows the summary statistics for the ten households in the two cities.

	Sample size n	$\bar{y}$	s_Y	$\bar{z}$	s_Z	s_X
City A	10	33	7.5	28.5	7	2
City B	10	29	8	28	7	2.5

Here, $\bar{y}$ and $\bar{z}$ denote the sample means of Y and Z in the two cities, and s_Y, s_Z and s_X denote the sample standard deviations for Y, Z and X respectively.

You can assume that the random variables Y_A and Y_B are independent and approximately normally distributed

(i) Perform a statistical test at a 5% significance level to test the null hypothesis that expected car usage in City A was the same as expected car usage in City B before the measures were introduced. State all other assumptions that you make and justify them. [6]

An actuary wishes to investigate whether the measures to reduce road congestion have been effective.

(ii) Perform a statistical test at the 5% significance level, where the alternative hypothesis is that car usage in City A has been reduced as a result of the measures. [4]

(iii) Calculate a 95% confidence interval for the expected reduction in car usage for City B. [3]

To investigate further the impact of measures to reduce road congestion, a third city, City C, is included in the study. The following table contains the data for 10 randomly selected households in City C:

	Sample size n	$\bar{y}$	s_Y	$\bar{z}$	s_Z	s_X
City C	10	37	9	33	8	3

Let x_{ij} denote the observed reduction in car usage in city i for household j.

(iv) Confirm that $\sum_{j=1}^{10} x_{Aj} = 45$ and $\sum_{j=1}^{10} x_{Aj}^2 = 238.5$. [2]

You are also given $\sum_{j=1}^{10} x_{Bj} = 10$, $\sum_{j=1}^{10} x_{Cj} = 40$, $\sum_{j=1}^{10} x_{Bj}^2 = 66.25$ and $\sum_{j=1}^{10} x_{Cj}^2 = 241$.

(v) Perform an analysis of variance to test at a 5% significance level the null hypothesis that there is no difference in the mean reduction in car usage between the three cities. [6]

[Total 21]

 PLEASE TURN OVER

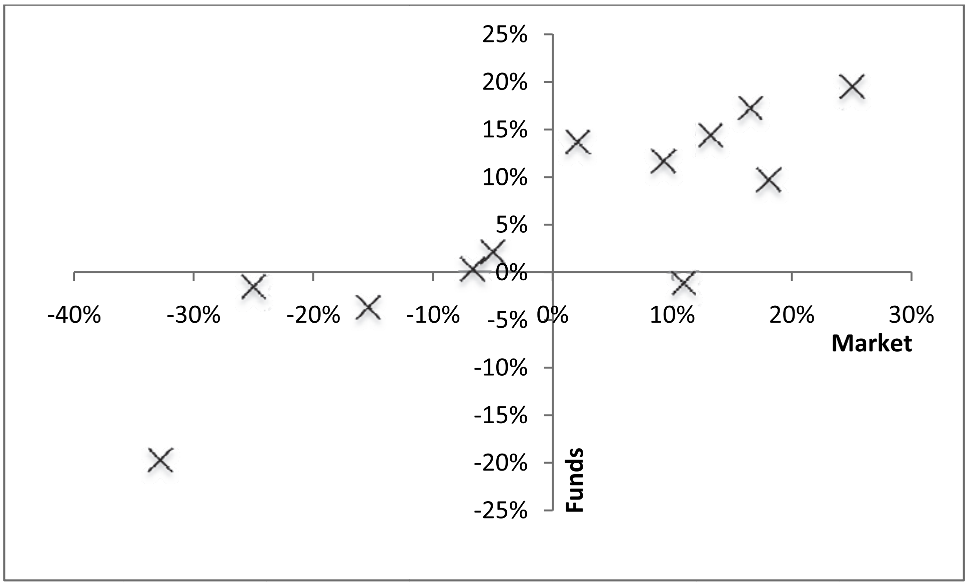
10 An analyst wishes to compare the results from investing in a certain category of hedge funds, f, with those from the stock market, x. She uses an appropriate index for each, which over 12 years each produced the following returns (in percentages to one decimal place).

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Market (x)	–5.0	–15.4	–25.0	16.6	9.2	18.1	13.2	2.0	–32.8	25.0	10.9	–6.7
Funds (f)	2.1	–3.7	–1.6	17.3	11.6	9.7	14.4	13.7	–19.8	19.5	–1.2	0.3

$$\sum x = 0.101, \sum x^2 = 0.3612, \sum f = 0.622, \sum f^2 = 0.1710, \sum xf = 0.1989$$

It is assumed that observations from different years are independent of each other.

Below is a scatter plot of market returns against fund returns for each year.



(i) Comment on the relationship between the two series. [1]

The hedge fund industry often claims that hedge funds have low correlation with the stock market.

(ii) (a) Calculate the correlation coefficient between the two series.

(b) Test whether the correlation coefficient is significantly different from 0. [7]

(iii) Calculate the parameters for a linear regression of the fund index on the market index. [2]

(iv) Calculate a 95% confidence interval for the underlying slope coefficient for the linear model in part (iii). [4]

(v) Comment on your answers to parts (ii)(b) and (iv). [2]
[Total 16]

END OF PAPER