INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

19 April 2018 (am)

Subject CT6 – Statistical Methods
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. *You have 15 minutes of planning and reading time before the start of this examination. You may make separate notes or write on the exam paper but not in your answer booklet. Calculators are not to be used during the reading time. You will then have three hours to complete the paper.*
4. *Mark allocations are shown in brackets.*
5. *Attempt all 10 questions, beginning your answer to each question on a new page.*
6. *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 variable X follows a Pareto distribution with density function:

$$\frac{5}{(1+x)^6}, \quad x > 0$$

For a given estimate d of x, the loss function is defined as:

$$x^4 - 4d^2x^2 + d^4$$

(a) Show that the expected loss is given by:

$$E(L(x,d)) = 1 - \frac{2d^2}{3} + d^4$$

(b) Determine the optimal estimate for d under the Bayes rule.

[5]

2 An insurance company has a portfolio of insurance policies. Claims arise according to a Poisson process, and claim amounts have a probability distribution with parameter θ.

(i) State one assumption the insurance company is likely to make when modelling aggregate claim amounts. [1]

(ii) Explain what the Maximum Likelihood Estimate (MLE) of θ represents. [2]

(iii) State an alternative to using the MLE. [1]

(iv) Suggest two complications that may arise for the insurance company when it uses past claims data to determine the MLE of θ. [2]

[Total 6]

3 An insurance company has collected data on the number of claims arising from certain risks over the last n years. The number of claims from the i^{th} risk in the j^{th} year is denoted by X_{ij} for $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, n$.

The distribution of X_{ij} depends on an unknown parameter θ_i. The X_{ij} are independent identically distributed random variables given θ_i.

(i) Describe briefly what is meant by each of the following: $m(\theta)$, $s^2(\theta)$, $E(s^2(\theta))$, $\text{var}(m(\theta))$, and Z, when using Empirical Bayes Credibility Theory (EBCT) Model 1. [5]

(ii) Explain how the value of Z depends on the following factors: n, $E(s^2(\theta))$, $\text{var}(m(\theta))$. [5]

[Total 10]

4 The table below shows the cumulative incurred claims by year for a portfolio of general insurance policies, with all figures in £m. Claims paid to date total 13.5. The ultimate loss ratio is expected to be in line with the 2013 accident year, and claims are assumed to be fully developed by the end of Development Year 3.

	Development Year				
Accident Year	*0*	*1*	*2*	*3*	*Earned Premiums*
2013	3.01	3.38	3.85	4.00	4.32
2014	3.30	3.67	4.15		4.41
2015	3.32	3.86			4.55
2016	3.74				4.68

Calculate the total reserve required to meet the outstanding claims, using the Bornheutter-Ferguson method. [9]

5 For three different risks, an actuary is modelling the monthly claim numbers with three different Poisson distributions.

Risk	*Exposure*	*Number of claims*
Risk 1	36 months	20
Risk 2	30 months	18
Risk 3	24 months	16

(i) Derive the maximum likelihood estimates of the parameter for each of the three individual Poisson distributions fitted. [5]

(ii) Test the hypothesis that the three risks have the same monthly claim rate. [5]
[Total 10]

6 Tarik and Liam are playing a zero-sum two-person game. From a deck of three cards numbered 1, 2 and 3 Tarik selects a card, making sure Liam does not know which one it is. Liam then proceeds to guess which card Tarik has picked. If Liam is wrong, he continues to guess until he has guessed correctly, at which point the game ends.

After each guess, if Liam's guess is lower than the number on Tarik's card, Tarik says "Low", but if Liam's guess is higher than the number on Tarik's card, Tarik says "High".

At the end of each game, Liam pays Tarik \$1 for each guess he made.

You should assume that Liam will never make a guess that contradicts the information provided by Tarik – for example, if Liam guesses "2" first, and Tarik says "Low", Liam would then always guess "3", rather than "1".

Consider strategy A, where Liam will guess "1" first, and then guesses "2" if 1 is not correct.

(i) Set out the four other strategies in addition to A (labelled B to E) which Liam could adopt. [4]

(ii) Construct the payoff matrix to Tarik. [3]

(iii) Explain whether or not there is a saddle point. [2]
[Total 9]

7 Consider the following time series model:

$$(1 - \alpha B)^3 X_t = \varepsilon_t$$

where B is the backshift operator and ε_t is a standard white noise process with variance σ^2.

(i) Determine for which values of α the process X_t is stationary. [2]

Now assume that X_t is stationary.

(ii) Calculate the autocorrelation function for the first two lags: ρ_1 and ρ_2, using the Yule-Walker equations. [7]

(iii) State the formulae, in terms of ρ_1 and ρ_2, for the first two values of the partial auto correlation function ϕ_1 and ϕ_2. [1]

Now assume that $\alpha = 1$.

(iv) Explain how to fit the parameter of this model, given the time series observations $X_1, X_2, ..., X_T$. [2]
[Total 12]

8 Claim events on a portfolio of insurance policies follow a Poisson process with parameter λ. Individual claim amounts, X, follow a Normal distribution with parameters $\mu = 500$ and $\sigma^2 = 200$.

The insurance company calculates premiums using a premium loading factor of 20%.

(i) Show that the adjustment coefficient, $r = 0.000708$ to three significant figures. [4]

The insurance company's initial surplus is 5,000.

(ii) Calculate an upper bound on the probability of ruin, using Lundberg's inequality. [1]

The insurance company actuary believes that claim amounts are better modelled using an exponential distribution. You may assume that the mean μ is unchanged, and is now equal to the standard deviation.

(iii) Calculate the new upper bound of the probability of ruin. [5]

(iv) Give a reason why claim amounts on insurance policies are not usually modelled using a Normal distribution, and suggest an alternative distribution, other than the exponential. [2]

[Total 12]

9 A health actuary is modelling the impact of a new infection which occurs in hospitals. He is studying 100 long term patients in different hospitals across a country. Infections occur according to a Poisson process, and the additional cost incurred due to an infection is a random variable, X, with mean 250 and variance 200.

Based on previous infections, the health actuary believes that in all hospitals across the country, the following applies:

Patient Type	*Proportion of patients*	*Poisson parameter*
High resistance	1 in 2	0.1
Moderate resistance	1 in 3	0.3
Low resistance	1 in 6	0.9

There is no way of knowing in advance which particular patients are more resistant. For a given patient, let λ_i represent the Poisson parameter, and let S_i represent the additional cost.

(i) Explain whether this is a heterogeneous or homogeneous group of risks. [2]

(ii) (a) Calculate the mean and variance of λ_i.

(b) State the formulae for the mean and variance of S_i, conditional on λ_i, in terms of the moments of X.

(c) Show that the unconditional mean of S_i is 75, and determine the variance of S_i.

(d) Calculate the mean and variance of the aggregate additional costs for all 100 patients.

[5]

In another country, a health actuary is modelling the same infection, again for 100 patients, but in a single hospital where it is believed each patient has the same level of resistance, and hence the Poisson parameter for each patient is the same. It is not known whether the Poisson parameter is 0.2 or 0.4, each being equally likely. The additional costs due to infection have the same distribution as above.

(iii) Calculate the mean and variance of the aggregate additional costs. [4]

(iv) Comment on your answers to parts (ii) (d) and (iii). [2]

[Total 13]

10 Consider the following probability density function:

$$h(x) = \lambda e^{-\lambda x} \quad x > 0$$

(i) Set out an algorithm for sampling from $h(x)$ using the inverse transform method. [3]

Now also consider the following density function, the "half-normal" distribution:

$$f(x) = \frac{\sqrt{2}}{\sigma\sqrt{\pi}} e^{\frac{-x^2}{2\sigma^2}}, \quad x > 0$$

Let M be the maximum of $\dfrac{f(x)}{h(x)}$

(ii) Show that $M = \dfrac{\sqrt{2}}{\lambda\sigma\sqrt{\pi}} e^{\frac{\lambda^2\sigma^2}{2}}$ [2]

(iii) Set out an acceptance-rejection algorithm, using $h(x)$, which generates samples from the half normal distribution $f(x)$. [3]

(iv) Determine the value of λ for which the algorithm is the most efficient (i.e. on average require fewest samples from $h(x)$ to generate samples from $f(x)$). [3]

(v) Show that the samples from part (iii) can be used to generate samples from the normal distribution with mean μ and variance σ^2. [3]

[Total 14]

END OF PAPER