# EXAMINATION

20 April 2010 (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. *Mark allocations are shown in brackets.*

4. *Attempt all 11 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** A coin is biased so that the probability of throwing a head is an unknown constant $p$. It is known that $p$ must be either 0.4 or 0.75. Prior beliefs about $p$ are given by the distribution:

$$P(p = 0.4) = 0.6 \qquad P(p = 0.75) = 0.4$$

The coin is tossed 6 times and 4 heads are observed.

Find the posterior distribution of $p$. [5]

**2** An insurance company is modelling claim numbers on its portfolio of motor insurance policies using a Poisson distribution, whose mean depends on the age and gender of the policyholder.

(i) Suggest a link function for fitting a generalised linear model for the mean of the Poisson distribution. [1]

(ii) Specify the corresponding linear predictor used for modelling the age and gender dependence as:

(a) age + gender
(b) age + gender + age × gender

[4]
[Total 5]

**3** The following two models have been suggested for representing some quarterly data with underlying seasonality.

Model 1 $Y_t = \alpha Y_{t-4} + e_t$
Model 2 $Y_t = \beta e_{t-4} + e_t$

Where $e_t$ is a white noise process in each case.

(i) Determine the auto-correlation function for each model. [4]

The observed quarterly data is used to calculate the sample auto-correlation.

(ii) State the features of the sample auto-correlation that would lead you to prefer Model 1. [1]
[Total 5]

**4** The number of claims $N$ on a portfolio of insurance policies follows a binomial distribution with parameters $n$ and $p$. Individual claim amounts follow an exponential distribution with mean $1/\lambda$. The insurer has in place an individual excess of loss reinsurance arrangement with retention $M$.

(i) Derive an expression, involving $M$ and $\lambda$, for the probability that an individual claim involves the reinsurer. [2]

Let $I_i$ be an indicator variable taking the value 1 if the $i$th claim involves the reinsurer and 0 otherwise.

(ii) Evaluate the moment generating function $M_{I_i}(t)$. [1]

Let $K$ be the number of claims involving the re-insurer so that $K = I_1 + \cdots + I_N$.

(iii) (a) Find the moment generating function of $K$.

(b) Deduce that $K$ follows a binomial distribution with parameters that you should specify.

[4]
[Total 7]

**5** An insurance company has issued life insurance policies to 1,000 individuals. Each life has a probability $q$ of dying in the coming year. In a warm year, $q = 0.001$ and in a cold year $q = 0.005$. The probability of a warm year is 50% and the probability of a cold year is 50%. Let $N$ be the aggregate number of claims across the portfolio in the coming year.

(i) Calculate the mean and variance of $N$. [4]

(ii) Calculate the alternative values for the mean and variance of $N$ assuming that $q$ is a constant 0.003. [2]

(iii) Comment on the results of (i) and (ii). [2]
[Total 8]

**6** Observations $y_1, y_2, \ldots, y_n$ are made from a random walk process given by:

$$Y_0 = 0 \text{ and } Y_t = a + Y_{t-1} + e_t \text{ for } t > 0$$

where $e_t$ is a white noise process with variance $\sigma^2$.

(i) Derive expressions for $E(Y_t)$ and $Var(Y_t)$ and explain why the process is not stationary. [3]

(ii) Show that $\gamma_{t,s} = Cov(Y_t, Y_{t-s})$ for $s < t$ is linear in $s$. [2]

(iii) Explain how you would use the observed data to estimate the parameters $a$ and $\sigma$. [3]

(iv) Derive expressions for the one-step and two-step forecasts for $Y_{n+1}$ and $Y_{n+2}$. [2]

[Total 10]

**7** The truncated exponential distribution on the interval $(0, c)$ is defined by the probability density function:

$$f(x) = ae^{-\lambda x} \quad \text{for } 0 < x < c$$

where $\lambda$ is a parameter and $a$ is a constant.

(i) Derive an expression for $a$ in terms of $\lambda$ and $c$. [1]

(ii) Construct an algorithm for generating samples from this distribution using the inverse transform method. [3]

Suppose that $0 < \lambda < c$ and consider the truncated Normal distribution defined by the probability density function:

$$g(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}[\Phi(c) - 0.5]} \quad \text{for } 0 < x < c$$

where $\Phi(z)$ is the cumulative density function of the standard Normal distribution.

(iii) Extend the algorithm in (ii) to use samples generated from $f(x)$ to produce samples from $g(x)$ using the acceptance / rejection method. [6]

[Total 10]

**8** The table below shows the incremental claims paid on a portfolio of insurance policies together with an extract from an index of prices. Claims are fully paid by the end of development year 3.

| | *Development Year* | | | | | *Price index* |
|---|---|---|---|---|---|---|
| *Accident Year* | *0* | *1* | *2* | *3* | *Year* | *(mid year)* |
| 2006 | 103 | 32 | 29 | 13 | 2006 | 100 |
| 2007 | 88 | 21 | 16 | | 2007 | 104 |
| 2008 | 110 | 35 | | | 2008 | 109 |
| 2009 | 132 | | | | 2009 | 111 |

Calculate the reserve for unpaid claims using the inflation-adjusted chain ladder approach, assuming that future claims inflation will be 3% p.a. [11]

**9** On the tapas menu in a local Spanish restaurant customers can order a dish of 20 roasted chillies for £5. There is always a mixture of hot chillies and mild chillies on the dish and these cannot be distinguished except by taste. The restaurant produces two types of dish: one containing 4 hot and 16 mild chillies and one containing 8 hot and 12 mild chillies. When the dish is served, the waiter allows the customer to taste one chilli and then offers a 50% discount to customers who correctly guess whether the dish contains 4 hot chillies or 8 hot chillies.

A hungry actuary who regularly visits the restaurant is trying to work out the best strategy for guessing the number of hot chillies.

(i) List the four possible decision functions the actuary could use. [3]

(ii) Calculate the values of the risk function for the two different chilli dishes and each decision function. [6]

(iii) Determine the optimum strategy for the actuary using the Bayes criterion and work out the average price he will pay for a dish of chillies if the restaurant produces equal numbers of the two types of dish. [3]
[Total 12]

**10** Claims on a portfolio of insurance policies arrive as a Poisson process with annual rate $\lambda$. Individual claims are for a fixed amount of 100 and the insurer uses a premium loading of 15%. The insurer is considering entering a proportional reinsurance agreement with a reinsurer who uses a premium loading of 20%. The insurer will retain a proportion $\alpha$ of each risk.

(i) Write down and simplify the equation defining the adjustment coefficient $R$ for the insurer. [3]

(ii) By considering $R$ as a function of $\alpha$ and differentiating show that

$$(120\alpha - 5)\frac{dR}{d\alpha} + 120R = \left(100R + 100\alpha\frac{dR}{d\alpha}\right)e^{100\alpha R}$$ [3]

(iii) Explain why setting $\frac{dR}{d\alpha} = 0$ and solving for $\alpha$ may give an optimal value for $\alpha$. [3]

(iv) Use the method suggested in part (iii) to find an optimal choice for $\alpha$. [4]

[Total 13]

**11** An actuary has, for three years, recorded the volume of unsolicited advertising that he receives. He believes that the number of items that he receives follows a Poisson distribution with a mean which varies according to which quarter of the year it is. He has recorded $Y_{ij}$ the number of items received in the $i$th quarter of the $j$th year ($i = 1, 2, 3, 4$ and $j = 1, 2, 3$). The actuary wishes to estimate the number of items that he will receive in the 1st quarter of year 4. He has recorded the following data:

| | $Y_{i1}$ | $Y_{i2}$ | $Y_{i3}$ | $\bar{Y}_i = \frac{1}{3}\sum_j Y_{ij}$ | $\sum_j \left(Y_{ij} - \bar{Y}_i\right)^2$ |
|---|---|---|---|---|---|
| $i = 1$ | 98 | 117 | 124 | 113 | 362 |
| $i = 2$ | 82 | 102 | 95 | 93 | 206 |
| $i = 3$ | 75 | 83 | 88 | 82 | 86 |
| $i = 4$ | 132 | 152 | 148 | 144 | 224 |

(i) Estimate $Y_{1,4}$ the number of items that the actuary expects to receive in the first quarter of year 4 using the assumptions of EBCT model 1. [5]

The actuary believes that, in fact, the volume of items has been increasing at the rate of 10% per annum.

(ii) Suggest how the approach in (i) can be adjusted to produce a revised estimate taking this growth into account. [2]

(iii) Calculate the maximum likelihood estimate of $Y_{1,4}$ (based on the quarter 1 data already observed and the 10% p.a. increase described above). [5]

(iv) Compare the assumptions underlying the approach in (i) and (ii) with those underlying the approach in (iii). [2]

[Total 14]

## END OF PAPER