EXAMINATION

19 April 2007 (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.

1 (i) State two conditions for a risk to be insurable. [2]

(ii) Describe briefly three distinct examples of financial loss insurance policies. [3]
[Total 5]

2 (i) Explain the concept of cointegrated time series. [3]

(ii) Give two examples of circumstances when it is reasonable to expect that two processes may be cointegrated. [2]
[Total 5]

3 The random variable X has an exponential distribution with mean 1000. Individual claim amounts on a certain type of insurance policy, Y, are such that

$$Y = X \quad \text{for } 0 < X < 2000$$

and $$P(Y = 2000) = P(X \geq 2000).$$

The insurer applies a deductible of 100 on claims from this type of insurance.

Calculate the mean of the distribution of individual claim amounts paid by the insurer. [5]

4 A casino operator moving into a country for the first time must apply to the casino regulator for a licence. There are three types of licence to choose from — slots, dice and cards — each with different running costs. The casino operator has to pay a fixed amount annually (£1,300,000) to the regulator, plus a variable annual licence cost.

The variable licence cost and expected revenue per customer for each type of game are as follows:

	Variable Licence cost £	*Expected Revenue per customer* £
Slots	250,000	60
Dice	550,000	120
Cards	1,150,000	160

The casino operator is uncertain about the number of customers and decides to prepare a profit forecast based on cautious, best estimate and optimistic numbers of customers. The figures are 14,000; 20,000 and 23,000 respectively.

(i) Determine the annual profits under each possible combination. [2]

(ii) Determine the minimax solution for optimising the profits. [2]

(iii) Determine the Bayes criterion solution based on the annual profit given the probability distribution P(cautious) = 0.2, P(best estimate) = 0.7 and P(optimistic) = 0.1. [2]

[Total 6]

5 The delay triangles given below relate to a portfolio of motor insurance policies.

The cost of claims settled during each year is given in the table below:

(Figures in £000s)

Accident Year	*Development year* 0	*1*	*2*
2004	4,144	694	183
2005	4,767	832	
2006	5,903		

The corresponding number of settled claims is as follows:

Accident Year	*Development year* 0	*1*	*2*
2004	581	75	28
2005	626	71	
2006	674		

Calculate the outstanding claims reserve for this portfolio using the average cost per claim method with grossing-up factors, and state the assumptions underlying your result.

[7]

6 (i) Explain the main advantage of the Polar method compared with the Box-Muller method for generating pairs of uncorrelated pseudo-random values from a standard normal distribution. [2]

(ii) Pseudo-random numbers are generated using the Box-Muller method in order to simulate values of $Y = \sqrt{X}$, where X has a lognormal distribution with parameters $\mu = 5$ and $\sigma = 2$. The quantity of interest is $\theta = E[Y]$.

(a) Calculate the value of Y when the number generated by the Box-Muller method is 0.9095.

(b) The variance of Y has been estimated as 26.3. Calculate how many simulations should be performed in order to ensure that the discrepancy between Y and $\hat{\theta}$, measured by the absolute error, is less than 1 with probability at least 0.9. [5]

[Total 7]

7 The total claims arising from a certain portfolio of insurance policies over a given month is represented by

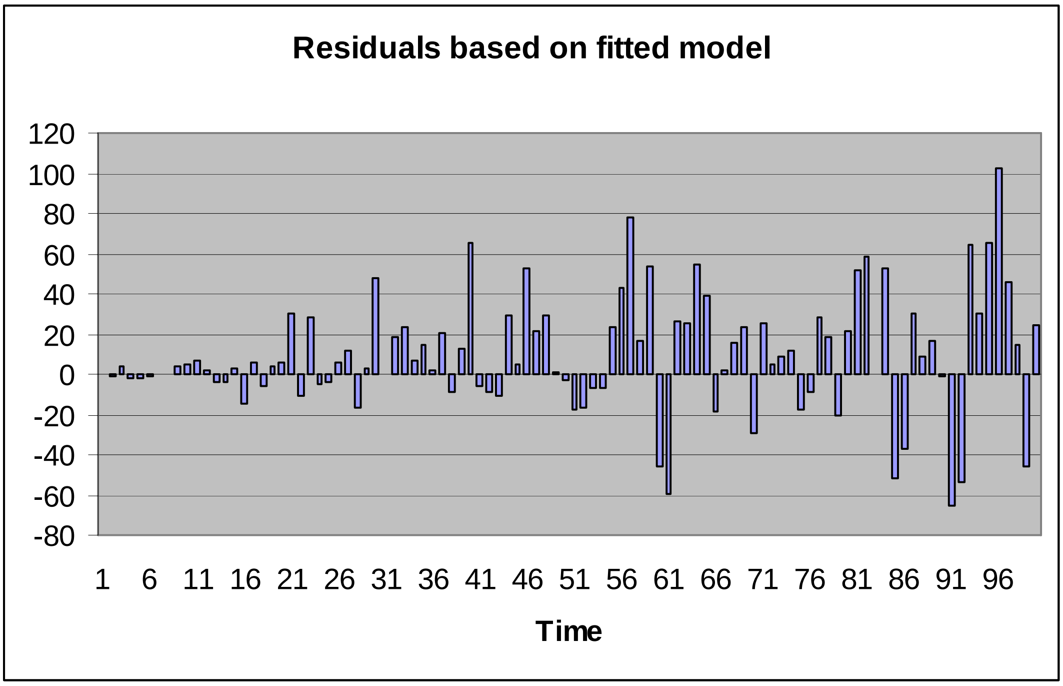
$$S = \begin{cases} \sum_{i=1}^{N} X_i & \text{if} \quad N > 0 \\ 0 & \text{if} \quad N = 0 \end{cases}$$

where N has a Poisson distribution with mean 2 and $X_1, X_2, \ldots, X_N$ is a sequence of independent and identically distributed random variables that are also independent of N. Their distribution is such that $P(X_i = 1) = 1/3$ and $P(X_i = 2) = 2/3$. An aggregate reinsurance contract has been arranged such that the amount paid by the reinsurer is $S - 3$ (if $S > 3$) and zero otherwise.

The aggregate claims paid by the direct insurer and the reinsurer are denoted by S_I and S_R, respectively.

Calculate $E(S_I)$ and $E(S_R)$. [8]

8 A modeller has attempted to fit an ARMA(p,q) model to a set of data using the Box-Jenkins methodology. The plot of residuals based on this proposed fit is shown below.



(i) Under the assumptions of the model, the residuals should form a white noise process.

(a) By inspection of the chart, suggest two reasons to suspect that the residuals do not form a white noise process.

(b) Define what is meant by a turning point.

(c) Perform a significance test on the number of turning points in the data above. (There are 100 points in the data and 59 turning points.)

[6]

(ii) On your suggestion, the original fitted model is discarded, and re-parameterised to:

$$X_{n+2} = 5 + 0.9(X_{n+1} - 5) + e_{n+2} + 0.5e_n .$$

Given the following observations:

$X_{99} = 2,\qquad X_{100} = 7$
$\hat{e}_{99} = -0.7,\qquad \hat{e}_{100} = 1.4$

Use the Box-Jenkins methodology to calculate the forward estimates $X_{100}(1), X_{100}(2)$ and $X_{100}(3)$.

[4]
[Total 10]

9 An insurer's NCD scale for motor policies has 3 levels of discount: 0%, 25% and 40%. The rules for moving between these levels are as follows:

- following a claim-free year, a policyholder moves to the next higher level of discount, or remains at 40% discount
- following a year of one or more claims, a policyholder at 40% discount moves to 25% discount while a policyholder at 25% or 0% moves to or stays at 0% discount

The full premium for each policyholder is £1,000. Following an accident, policyholders decide whether or not to claim by considering total outgoing over the next two years, assuming no further claims in this period and ignoring interest.

(i) Find the claim threshold for each level of discount. [3]

(ii) The probability of no accidents in any year for each policyholder is 0.88 and individual losses are assumed to have a lognormal distribution with $\mu = 6.0$ and $\sigma = 3.33$. Ignoring the possibility of more than 1 accident occurring in a year, calculate the transition matrix. [3]

(iii) Calculate the stationary distribution. [3]

(iv) Derive the stationary distribution under the alternative assumption that a policyholder always claims after a loss (regardless of the size of the claim). [3]

(v) Comment on the difference between the results of (iii) and (iv). [1]

[Total 13]

10 (i) The Gamma distribution with mean μ and variance μ^2/α has density function

$$f(y) = \frac{\alpha^\alpha}{\mu^\alpha \Gamma(\alpha)} y^{\alpha-1} e^{-\frac{y\alpha}{\mu}} \qquad (y > 0)$$

(a) Show that this may be written in the form of an exponential family.

(b) Use the properties of exponential families to confirm that the mean and variance of the distribution are μ and μ^2/α. [9]

(ii) Explain the difference between a continuous covariate and a factor. [3]

(iii) A company is analysing its claims data on a portfolio of motor policies, and uses a gamma distribution to model the claim severities. The company uses three rating factors:

policyholder age (as a continuous variable);
policyholder gender;
vehicle rating group (as a factor).

(a) Write down the form of the linear predictor when all rating factors are included as main effects.

(b) State how the linear predictor changes if an interaction between policyholder age and gender is included.

[4]
[Total 16]

11 The number, X, of claims on a given insurance policy over one year has probability distribution given by

$$P(X = k) = \theta^k (1-\theta) \quad k = 0, 1, 2, \ldots$$

where θ is an unknown parameter with $0 < \theta < 1$.

Independent observations $x_1, \ldots, x_n$ are available for the number of claims in the previous n years. Prior beliefs about θ are described by a distribution with density

$$f(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\alpha-1}$$

for some constant $\alpha > 0$.

(i) (a) Derive the maximum likelihood estimate, $\hat{\theta}$, of θ given the data $x_1, \ldots, x_n$.

(b) Derive the posterior distribution of θ given the data $x_1, \ldots, x_n$.

(c) Derive the Bayesian estimate of θ under quadratic loss and show that it takes the form of a credibility estimate

$$Z\hat{\theta} + (1 - Z)\mu$$

where μ is a quantity you should specify from the prior distribution of θ.

(d) Explain what happens to Z as the number of years of observed data increases.

[11]

(ii) (a) Determine the variance of the prior distribution of θ.

(b) Explain the implication for the quality of prior information of increasing the value of α. Give an interpretation of the prior distribution in the special case $\alpha = 1$.

[3]

(iii) Calculate the Bayesian estimate of θ under quadratic loss if $n = 3$, $x_1 = 3, x_2 = 3, x_3 = 5$ and

(a) $\alpha = 5$
(b) $\alpha = 2$

Comment on your results in the light of (ii) above. [4]

[Total 18]

END OF PAPER