EXAMINATION

3 October 2007 (am)

Subject CT4 — Models
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 List the factors you would consider when assessing the suitability of an actuarial model for its purpose. [4]

2 A particular baker's shop in a small town sells only one product: currant buns. These currant buns are delicious and customers travel many miles to buy them. Unfortunately, the buns do not keep fresh and cannot be stored overnight.

The baker's practice is to bake a certain number of buns, K, before the shop opens each morning, and then during the day to continue baking c buns per hour. He is concerned that:

- he does not run out of buns during the day; and
- the number of buns left over at the end of each day is as few as possible

(i) Describe a model which would allow you to estimate the probability that the baker will run out of buns. State any assumptions you make. [3]

(ii) Determine the relevant expression for the probability that the baker will run out of buns, in terms of K, c, and B_j, the number of buns bought by the day's jth customer. [1]

[Total 4]

3 A no-claims discount system has 3 levels of discount: 0%, 25% and 50%. The rules for moving between discount levels are:

- After a claim-free year, move up to the next higher level or remain at the 50% discount level.

- After a year with one or more claims, move down to the next lower level or remain at the 0% discount level.

The long-run probability that a policyholder is in the maximum discount level is 0.75.

Calculate the probability that a given policyholder has a claim-free year, assuming that this probability is constant.

[5]

4 A national mortality investigation was carried out. It was suggested that the mortality of the male population could be represented by the following graduated rates:

$$\overset{o}{\mu}_{x+\frac{1}{2}} = \mu^s_{x+2\frac{1}{2}}$$

where μ^s_x is from the standard tables, ELT15(males).

The table below shows the graduated rates for part of the age range, together with the exposed to risk, expected and actual deaths at each age. The squared standardised deviations that were calculated are also shown.

The standardised deviations were calculated as $z_x = \dfrac{\left(\theta_x - E^c_x \cdot \overset{o}{\mu}_{x+\frac{1}{2}}\right)}{\sqrt{E^c_x \cdot \overset{o}{\mu}_{x+\frac{1}{2}}}}$

Age	*Graduated rates*	*Exposed to risk*	*Expected deaths*	*Deaths*	*Squared standardised deviations*
x	$\overset{o}{\mu}_{x+\frac{1}{2}}$	E^c_x	$E^c_x \cdot \overset{o}{\mu}_{x+\frac{1}{2}}$	θ_x	z^2_x
50	0.00549	10,850	59.57	52	0.9611
51	0.00610	9,812	59.85	54	0.5724
52	0.00679	10,054	68.27	60	1.0010
53	0.00757	9,650	73.05	65	0.8872
54	0.00845	8,563	72.36	64	0.9653
55	0.00945	10,656	100.70	87	1.8637
56	0.01057	9,667	102.18	88	1.9679
57	0.01182	9,560	113.00	97	2.2653
58	0.01323	8,968	118.65	103	2.0634
59	0.01483	8,455	125.39	105	3.3150

(i) Test this graduation for overall goodness-of-fit. [5]

(ii) Comment on your findings in (i). [2]
[Total 7]

5 (i) Explain why crude mortality rates are graduated before being used for financial calculations. [3]

(ii) List two methods of graduating a set of crude mortality rates and state, for each method:

(a) under what circumstances it should be used; and
(b) how smoothness is ensured

[4]
[Total 7]

6 Below is an extract from English Life Table 15 (Males)

Age x	l_x
58	88,792
62	84,173

(i) Estimate l_{60} under each of the following assumptions:

(a) a uniform distribution of deaths between exact ages 58 and 62 years; and

(b) a constant force of mortality between exact ages 58 and 62 years

[5]

(ii) Find the actual value of l_{60} in the tables and hence comment on the relative validity of the two assumptions you used in part (i). [3]

[Total 8]

7 In order to boost sales, a national newspaper in a European country wishes to compile a "fair play league table" for the country's leading football clubs. On 1 December it undertakes a survey of all the players who play for these clubs, in which it collects the following data:

- number of games played by each player since the beginning of the season (the football season in this country begins in September); and
- for each player who had been dismissed from the field of play between the beginning of the season and 1 December (inclusive), the number of games he had played before the game in which he was first dismissed

No games were played on 1 December.

The statistic the newspaper proposes to use in order to construct its "fair play league table" is the probability that a player will not have been dismissed in any of his first 10 games. It plans to calculate this statistic for each of the 20 leading clubs.

The following table shows the data collected for the players of the club which was top of the league on 1 December.

Player	*Total number of games played*	*Number of times dismissed*	*Games played before first dismissal*
1	12	0	
2	12	0	
3	12	1	5
4	12	0	
5	12	1	7
6	12	0	
7	10	0	
8	9	1	0
9	9	1	5
10	8	0	
11	6	2	2
12	5	0	
13	5	0	
14	4	1	0
15	4	0	

(i) (a) Explain how the Kaplan-Meier estimator can be used to estimate the newspaper's statistic from these data.

(b) Comment on the way in which censoring arises and on the type of censoring produced. [4]

(ii) Calculate the newspaper's statistic using the data above. [4]

[Total 8]

8 (i) Describe the difference between the central exposed to risk and the initial exposed to risk. [2]

The following data come from an investigation of the mortality of participants in a dangerous sport during the calendar year 2005.

Age x	*Number of lives aged x last birthday on:* 1 January 2005	1 January 2006	*Number of deaths during 2005 to persons aged x last birthday at death*
22	150	160	20
23	160	155	25

(ii) (a) Estimate the initial exposed to risk at ages 22 and 23.
(b) Hence estimate q_{22} and q_{23}.
[4]

Suppose that in this investigation, instead of aggregate data we had individual-level data on each person's date of birth, date of death, and date of exit from observation (if exit was for reasons other than death).

(iii) Explain how you would calculate the initial exposed-to-risk for lives aged 22 years last birthday. [4]
[Total 10]

9 In a game of tennis, when the score is at "Deuce" the player winning the next point holds "Advantage". If a player holding "Advantage" wins the following point that player wins the game, but if that point is won by the other player the score returns to "Deuce".

When Andrew plays tennis against Ben, the probability of Andrew winning any point is 0.6. Consider a particular game when the score is at "Deuce".

(i) Show that the subsequent score in the game can be modelled as a Markov Chain, specifying both:

(a) the state space; and
(b) the transition matrix [3]

(ii) State, with reasons, whether the chain is:

(a) irreducible; and
(b) aperiodic [2]

(iii) Calculate the number of points which must be played before there is more than a 90% chance of the game having been completed. [3]

(iv) (a) Calculate the probability that Andrew wins the game.
(b) Comment on your answer. [4]
[Total 12]

10 (i) Compare the advantages and disadvantages of fully parametric models and the Cox regression model for assessing the impact of covariates on survival. [3]

You have been asked to investigate the impact of a set of covariates, including age, sex, smoking, region of residence, educational attainment and amount of exercise undertaken, on the risk of heart attack. Data are available from a prospective study which followed a set of several thousand persons from an initial interview until their first heart attack, or until their death from a cause other than a heart attack, or until 10 years had elapsed since the initial interview (whichever of these occurred first).

(ii) State the types of censoring present in this study, and explain how each arises. [2]

(iii) Describe a criterion which would allow you to select those covariates which have a statistically significant effect on the risk of heart attack, when controlling the other covariates of the model. [4]

Suppose your final model is a Cox model which has three covariates: age (measured in age last birthday minus 50 at the initial interview), sex (male = 0, female = 1) and smoking (non-smoker = 0, smoker = 1), and that the estimated parameters are:

Age	0.01
Sex	-0.4
Smoking	0.5
Sex *x* smoking	-0.25

where "sex *x* smoking" is an additional covariate formed by multiplying the two covariates "sex" and "smoking".

(iv) Describe the final model's estimate of the effect of sex and of smoking behaviour on the risk of heart attack. [3]

(v) Use the results of the model to determine how old a female smoker must be at the initial interview to have the same risk of heart attack as a male non-smoker aged 50 years at the initial interview. [3]

[Total 15]

11 The following data have been collected from observation of a three-state process in continuous time:

State occupied	*Total time spent in state (hours)*	*Total transitions to:* *State A*	*State B*	*State C*
A	50	Not applicable	110	90
B	25	80	Not applicable	45
C	90	120	15	Not applicable

It is proposed to fit a Markov jump model to this data set.

(i) (a) List all the parameters of the model.
(b) Describe the assumptions underlying the model. [4]

(ii) (a) Estimate the parameters of the model.
(b) Give the estimated generator matrix. [4]

The following additional data in respect of secondary transitions were collected from observation of the same process.

Triplet of successive transitions	*Observed number of triplets* n_{ijk}	*Triplet of successive transitions*	*Observed number of triplets* n_{ijk}
ABC	42	BCA	38
ABA	68	BCB	7
ACA	85	CAB	64
ACB	4	CAC	56
BAB	50	CBA	8
BAC	30	CBC	7

(iii) State the distribution of the number of transitions from state i to state j, given the number of transitions out of state i. [1]

(iv) Test the goodness-of-fit of the model by considering whether triplets of successive transitions adhere to the distribution given in (iii). [5]

[Hint: Use the test statistic $\chi^2 = \sum_i \sum_j \sum_k \frac{(n_{ijk} - E)^2}{E}$ where E is the expected number of triplets under the distribution in (iii)]

(v) Identify two other aspects of the appropriateness of the fitted model that could be tested, stating suitable tests in each case. [2]

(vi) Outline two methods for simulating the Markov jump process, without performing any calculations. [4]
[Total 20]

END OF PAPER