INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

9 April 2019 (pm)

Subject CS2A – Risk Modelling and Survival Analysis Core Principles

Time allowed: Three hours and fifteen minutes

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 questions, begin your answer to each question on a new page.*

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 State why data are divided into homogeneous groups when conducting mortality investigations. [2]

2 (i) Describe two features of short-term insurance contracts. [2]

(ii) Suggest two examples of short-term insurance contracts. [2]
[Total 4]

3 (i) Describe the difference between supervised and unsupervised learning. [2]

(ii) Give an example of each type of learning mentioned in part (i). [2]
[Total 4]

4 (i) Define a general random walk. [2]

(ii) Show that a general random walk is not stationary. [3]
[Total 5]

5 (i) State the principle of correspondence as it applies to death rates. [1]

A country uses an administrative data system which shows the estimated population aged x last birthday on 1 January each year. Deaths are registered on an age nearest birthday basis.

(ii) Derive a formula which may be used to estimate an appropriate exposed to risk for calculating the average death rate at age x exact in the two-year period 1 January 2016 – 1 January 2018, defining all the terms you use. Assume that birthdays are evenly distributed across the calendar year. [4]

Someone remarks that, in this country, only one third of births take place in the first half of the calendar year.

(iii) Discuss the implications of this for the formula you have derived in part (ii). [3]
[Total 8]

6 (i) State the properties of a valid generator matrix for a Markov jump process. [2]

An insurance company has a very old product which is no longer sold and for which there are only two policies still in force.

(ii) Give the state space for the number of this type of policy still in force at a future time. [1]

The transition rate at which an individual policy terminates is 0.2 per annum, and the policies are independent of each other.

Let $P_i(t)$ be the probability that i policies are in force at time t.

(iii) Draw a transition graph for the process for the number of policies in force. [2]

(iv) Give the Kolmogorov forward equations for $\frac{d}{dt}P_{22}(t)$, $\frac{d}{dt}P_{21}(t)$ and $\frac{d}{dt}P_{20}(t)$. [2]

The two policies in force are the last two policies on an administration system which costs £25,000 per year to manage. The insurance company wishes to discontinue this system. It has received a quotation from a third party to take over the administration of these policies. Once the third party has taken over these policies, it will continue to administer them for a negligible cost.

The insurance company will accept the third party's quotation if it is likely to lead to a decrease in overall costs by so doing. While the insurance company is considering whether to take up the third party's quotation, one of the policies lapses.

(v) Estimate the maximum amount the insurance company should be willing to pay the third party to take over the one remaining policy. [4]

[Total 11]

7 (i) Write down Sklar's theorem. [2]

(ii) Explain, in words, the meaning of the following copula expression:
$C(u_1, u_2, u_3)$ [2]

The Gumbel copula has a generating function:

$$_{\text{Gumbel}}\psi_{\alpha}\,[F(x)] = (-\ln F(x))^{\alpha}$$

(iii) Derive an expression for the Gumbel (Hougaard) copula for the case where there are three variables. [3]

A student has fitted a Gumbel copula to investment returns from three developing markets, and has calculated a value for the dependency parameter, α, of 4.0.

She has separately determined that the probability of making a loss over the next calendar year (i.e. the probability that the return is less than 0%) in each of the three markets is 5%, 7.5% and 10% respectively.

(iv) Calculate the probability that all three markets have returns of less than 0% over the next calendar year. [3]

(v) State what type of copula is equivalent to a Gumbel copula if $\alpha = 1.0$. [1]

(vi) Calculate the probability that all three markets have returns of less than 0% over the next calendar year, assuming that each of the markets were independent. [1]

[Total 12]

8 (i) State two advantages of the Cox regression model for assessing the impact of risk factors on a hazard. [2]

An exercise company called FlexPexApps is developing a computer program to investigate the effect of certain factors on the incidence of a common medical condition which affects millions of people in early middle age. It has identified three factors which appear to have a large impact on the onset of the disease and has set up a Cox regression model for the hazard as follows:

$$h(t) = h_0(t) \exp(\beta_A A + \beta_E E + \beta_D D)$$

where:

- A is the age of the individual minus 40 years
- E is an exercise indicator and takes the value of 1 if the person exercises, which in this case means they follow a set regime for 30 minutes each day, and 0 otherwise
- D is a diet indicator and takes the value of 1 if the person diets, which in this case means they consume fewer than 2,000 calories per day, and 0 otherwise
- β_A, β_E and β_D are the parameters to be estimated.

From the data FlexPexApps has managed to acquire, it has established that:

- a 53 year old who exercises but does not diet has a hazard of contracting the condition half that of a 48 year old who does not exercise but diets
- a 55 year old who does not exercise but diets has a hazard of contracting the condition 1.5 times that of a 55 year old who neither exercises nor diets
- a 58 year old who diets but does not exercise has a hazard of contracting the condition double that of a 43 year old who neither diets nor exercises.

(ii) Calculate β_A, β_E and β_D. [5]

(iii) Explain what the values you have calculated in part (ii) say about the relative impact of age, diet and exercise on contracting this affliction. [2]

FlexPexApps has created an advertisement based on the above findings, but the Advertising Regulator has contacted them on the grounds that their model was not sufficiently complex to take into account all the relevant factors. They have suggested four additional factors which might materially impact the hazard of contracting the condition.

(iv) Explain how FlexPexApps could extend their model to see if any one of the suggested additional factors materially impacts the hazard of contracting the condition. [3]

[Total 12]

9 Consider the time series model:

$$Y_t = 0.5 + 0.9Y_{t-1} - 0.14Y_{t-2} + e_t$$

where e_t is a white noise process with variance σ^2.

(i) Determine whether Y_t is a stationary process. [3]

(ii) Identify the model as an ARIMA(p,d,q) process. [1]

(iii) Calculate $E(Y_t)$. [3]

(iv) Calculate the variance, $Var(Y_t)$, of the process, and auto-correlation values ρ_1 and ρ_2. [6]

[Total 13]

10 A new screening method for a rare but serious disease is being tested. The new method is based on a questionnaire and is designed as an alternative to the existing method, which involves an invasive clinical procedure.

Under the new method, patients are asked a series of questions, and an algorithm is applied to their answers to predict whether they have the disease. In the test, both the questionnaire and the existing clinical procedure are given to 100 patients whose disease status is known. Some results from the test are shown in the table below:

True status	***Prediction from clinical procedure***		***Prediction from questionnaire test***	
	Has disease	*Does not have disease*	*Has disease*	*Does not have disease*
Has disease	45	5	40	10
Does not have disease	15	35	10	40

(i) Calculate, for both the clinical procedure and the questionnaire test:

(a) the precision
(b) the recall
(c) the F1 score.

[5]

(ii) Show that the F1 score expresses the true positives as a proportion of the true positives plus the average of those incorrectly classified. [3]

(iii) Comment on your results from part (i), and on the usefulness of the test. [4]

[Total 12]

11 A life insurance company is investigating the mortality of its policyholders over the past year. It wishes to compare the current mortality rates with those obtained from a similar investigation ten years ago. The following is an extract of the data:

Current investigation			***Previous investigation***
Age x last birthday	*Exposed to risk*	*Observed deaths*	*Mortality rate*
50	5,368	25	0.00479
51	4,986	26	0.00538
52	4,832	30	0.00603
53	5,298	37	0.00675
54	5,741	45	0.00756
55	4,866	46	0.00844
56	4,901	52	0.00942
57	5,003	63	0.01050
58	3,952	45	0.01169
59	2,786	45	0.01299

(i) Sketch a graph, showing clearly both the current and the previous mortality rates. [3]

(ii) Carry out a goodness-of-fit test on the data. [5]

(iii) Carry out the following additional statistical tests:

- Signs test.
- Grouping of Signs test.

[6]

(iv) Comment on your answers to part (iii) in the light of your sketch in part (i). [3]

[Total 17]

END OF PAPER