INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

20 September 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 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 portfolio consists of two types of policies where the number of claims each year and the claim sizes are distributed as follows:

Policy type	*Number of claims per year*	*Individual claim amount*
A	Poisson with mean 2.5	Exponential with mean 1,500
B	Poisson with mean 3.0	Exponential with mean 1,250

You may assume all policies are independent.

Determine the mean and variance of the total annual aggregate claims on the portfolio. [2]

2 The government of a small country is interested in projecting mortality over the next 20 years. Life tables are available by single years of age for each calendar year for the past 10 years. The country has a new Chief Statistician who suggests fitting an exponential curve to the time trend in mortality at each age x and forecasting mortality separately at each age using the parameters estimated for that age.

(i) Comment on this suggestion. [3]

(ii) Suggest an alternative approach which may be more suitable for projecting mortality in this country. [1]
[Total 4]

3 A stochastic process X_t is defined as follows, over time points $t = 0, 1, 2, \ldots$

$X_0 = 0$
$\Pr[X_{t+1} - X_t = 1] = p$
$\Pr[X_{t+1} - X_t = -1] = 1 - p$.

(i) State the name, the state space and the time domain of this process. [2]

(ii) Write down the distribution of this process for $t = 1, 2$ and 3. [3]

(iii) State the distribution of the process after a large number of time periods N, including its parameters in terms of N and p. [1]
[Total 6]

4 (i) State the Markov property. [1]

The three transition matrices below describe three Markov Chains.

I

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

II

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0.75 & 0.25 \end{bmatrix}$$

III

$$\begin{bmatrix} 0.25 & 0.75 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0.75 & 0.25 \end{bmatrix}$$

(ii) State whether these Markov Chains are:

(a) irreducible

(b) periodic, stating the period where appropriate.

[3]

(iii) Calculate the stationary probability distribution (if it exists) for each of the three Markov Chains described above. [5]

[Total 9]

5 (i) Discuss the train-validation-test approach in the context of machine learning. [4]

An analyst has been tasked with using machine learning to fit a model to predict the movement of traffic in a major city in order to assist the town planners with developing a one-way traffic system. He has gathered a set of data which has been collected by a number of sensors and cameras located at key junctions throughout the city. The sensors calculate the number of vehicles which pass by each hour.

(ii) State, with reasons, whether machine learning is an appropriate approach for the analyst to take when fitting his model. [3]

The analyst is now considering how many parameters to use when fitting the model.

(iii) Discuss the advantages and disadvantages of using a large number of parameters when fitting this model. [3]

[Total 10]

6 The Eternal Life insurance company insures the lives of community leaders. Its pricing policy assumes that the mortality of its policyholders reflects the national mortality levels. Recently, however, it has been losing business to rivals, so it has commissioned an investigation into the mortality of its policyholders. Some results from the investigation are shown below.

Age x nearest birthday	*Person-years at risk, E^c_x*	*Number of deaths, d_x*	*National mortality rate*
60	750	10	0.01323
61	740	11	0.01483
62	710	12	0.01664
63	700	15	0.01870
64	680	12	0.02101
65	400	5	0.02348
66	390	5	0.02610
67	380	6	0.02893
68	360	8	0.03192
69	350	8	0.03505

(i) Carry out a general goodness-of-fit test of the hypothesis that the mortality of the company's policyholders is the same as the national mortality rate. [6]

(ii) Comment on your results in part (i). [2]

(iii) Perform one other test designed to explore further the hypothesis that the mortality of the company's policyholders is the same as the national mortality rate. [3]

(iv) Suggest explanations for the results you have obtained in parts (i) and (iii). [2]
[Total 13]

7 An insurance company has a portfolio of policies. Let X denote an individual claim amount, with probability density function $f_X(x)$ for $x > 0$.

The insurance company has an individual excess of loss reinsurance arrangement with a retention of M. Let Y be the amount paid by the insurance company net of reinsurance.

(i) (a) Express Y in terms of X.

(b) Derive an expression for the probability density function of Y in terms of X.

[3]

An actuary believes that X follows a Weibull distribution with density function $f_X(x) = 0.6cx^{-0.4}e^{-cx^{0.6}}$, where c is an unknown constant.

The insurance company has an excess of loss reinsurance arrangement with retention \$1,000. The following claims are observed:

Claims below retention: \$156, \$208, \$232, \$270, \$378, \$486, \$540, \$700, \$822, \$982
Number of claims above retention: 5
Total number of claims: 15

(ii) Estimate c using the method of maximum likelihood. [7]

(iii) Estimate c applying the method of percentiles to the median. [4]

[Total 14]

8 An actuary is modelling a set of observations, $y_1, y_2, \dots, y_n$. Her model has the following form:

$$y_t = ay_{t-1} + \varepsilon_t$$

where ε_t is a standard Gaussian white-noise process, with variance σ^2.

(i) State the distribution of y_t conditional on y_{t-1}. [1]

(ii) Show that the likelihood function of this model for the given observations is:

$$L\left(a, \sigma^2\right) = p\left(y_1\right)\frac{1}{\left(\sqrt{2\pi\sigma}\right)^{n-1}}e^{\frac{-\sum_{t=2}^{n}\left(y_t - \alpha y_{t-1}\right)^2}{2\sigma^2}}$$

where $p(y_1)$ is the pdf of the marginal distribution of y_1. [2]

(iii) State the log-likelihood function, assuming $p(y_1)$ may be ignored. [1]

(iv) Derive the maximum likelihood estimates for α and σ^2. [5]

(v) Derive expressions for the estimates of α and σ^2 using the Yule-Walker equations (you may leave the expression for σ^2 in terms of α). [3]

(vi) Comment on the difference between the estimates of α in parts (iv) and (v). [2]

[Total 14]

9 An investigation was undertaken into the effect of marriage on the risk of death for males aged 30–39 years. A model with three states was used: 1 Single, 2 Married and 3 Dead. It is assumed that transition rates between states are constant.

(i) Sketch a diagram showing the possible transitions between states. [2]

(ii) Write down an expression for the likelihood of the data in terms of transition rates and waiting times, defining all the terms you use. [3]

(iii) Derive the maximum likelihood estimator of the transition rate from Single to Dead. [3]

The following data were collected from information on males aged 30–39 years.

Years spent in Single state	11,343
Years spent in Married state	39,098
Number of transitions from Married to Single	1,512
Number of transitions from Single to Dead	13
Number of transitions from Married to Dead	30

(iv) Estimate:

(a) the constant transition rate from Single to Dead; and
(b) its variance.

[2]

(v) Test the hypothesis that the death rate for Single and Married men is the same. [4]

[Total 14]

10 A firm owns two garages that fitted new exhaust systems on cars. It was suggested that the exhausts fitted by Garage A lasted longer than those fitted by Garage B. In September 2018 the firm commissioned a study to determine this.

The study collected data from a sample of cars which had exhausts fitted by each garage. The data consisted of the month the exhaust system was fitted and the month the exhaust system wore out (where this was known). For some cars, the exhaust systems were still working in September 2018.

Below are given data for eight cars from each garage. You can assume that all dates refer to the mid-point of the relevant month.

Garage	*Date exhaust fitted*	*Date exhaust wore out (mm/yyyy)*
A	01/2012	01/2014
A	03/2013	05/2017
A	07/2014	06/2017
A	10/2014	Still working in 09/2018
A	12/2014	10/2015
A	01/2015	Still working in 09/2018
A	06/2015	Still working in 09/2018
A	10/2015	06/2018
B	01/2012	01/2015
B	03/2013	05/2018
B	04/2013	06/2017
B	10/2014	10/2015
B	12/2014	Still working in 09/2018
B	01/2015	09/2017
B	06/2015	Still working in 09/2018
B	10/2015	Still working in 09/2018

(i) Calculate the Kaplan-Meier estimates of the survival function of wearing out of exhausts for cars from each garage. [8]

(ii) Plot the Kaplan-Meier estimates of the survival functions for the two garages on the same graph. [3]

It is suggested that the difference between the lifetimes of exhaust systems from the two garages could be measured using a proportional hazards model.

(iii) Comment on the appropriateness of this suggestion. [3]
[Total 14]

END OF PAPER