INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

19 April 2022 (am)

Subject CS2 – Risk Modelling and Survival Analysis
Core Principles

Paper A

Time allowed: Three hours and twenty minutes

In addition to this paper you should have available the 2002 edition of the Formulae and Tables and your own electronic calculator.

If you encounter any issues during the examination please contact the Assessment Team on T. 0044 (0) 1865 268 873.

© Institute and Faculty of Actuaries

1 (i) Write down the two-factor Lee–Carter model, clearly defining each of the terms you use. [3]

A statistician is using the two-factor Lee–Carter model to project future mortality rates and has fitted the model to a set of mortality data. The statistician has observed that the fitted forces of mortality do not vary regularly with each calendar year but vary more regularly with age.

Therefore, the statistician suggests that, before projecting future mortality rates, the k parameters in the Lee–Carter model should be smoothed using penalised regression splines but the a and b parameters should be kept as separate, unsmoothed parameters for each age.

(ii) Discuss the statistician's suggestion by considering each of the three parameters of the Lee–Carter model, k, a and b, in turn. [3]

[Total 6]

2 An international pensions provider is interested in quantifying the force of mortality at certain ages for a particular country, for the period from 1 January 2017 through to 1 January 2020.

In this country, deaths are recorded by calendar year and classified by age at last birthday on the date of death, and the annual population censuses are completed on 30 June each year, classifying the population by age at last birthday.

Derive an expression for the crude force of mortality, $\hat{\mu}_{x+\frac{1}{2}}$, for a given age interval $[x, x+1]$, for the period from 1 January 2017 to 1 January 2020 using the population and death data available, clearly stating all assumptions and defining each of the terms you use. [7]

3 During a football match, the referee can caution players if they commit an offence by showing them a yellow card. If a player commits a second offence that the referee deems worthy of a caution, they are shown a red card, and are sent off the pitch and take no further part in the match. If the referee considers a particularly serious offence to have been committed, a red card will be shown to a player who has not previously been cautioned, and the player will be sent off immediately.

The football team manager can also decide to substitute one player for another at any point in the match so that the substituted player takes no further part in the match. Due to the risk of a player being sent off, the manager is more likely to substitute a player who has been shown a yellow card. Experience shows that players who have been shown a yellow card play more carefully to try to avoid a second offence.

The rate at which uncautioned players are shown a yellow card is 1/8 per hour.

The rate at which players who have already been shown a yellow card are shown a red card is 1/12 per hour.

The rate at which uncautioned players are shown a red card is 1/20 per hour.

The rate at which players are substituted is 1/10 per hour if they have not been shown a yellow card, and 1/5 if they have been shown a yellow card.

An actuary decides to model a player's state during a match using a Markov jump process and the following four states:

- State U: uncautioned
- State Y: yellow card shown
- State R: red card shown
- State S: substituted.

(i) Write down the generator matrix of the Markov jump process that is used in the compact form of Kolmogorov's forward equations. [2]

A football match lasts 1.5 hours.

(ii) Solve Kolmogorov's forward equation for the probability that a player who starts the match remains in the game for the whole match without being substituted and without being shown a yellow card or a red card. [3]

(iii) Determine the probability that a player who starts the match is sent off during the match without having previously been cautioned. [4]

[Total 9]

4 An investigation was carried out into the effects of a newly developed medication to treat a potentially fatal illness, including an analysis of the effects of delaying the treatment after symptoms are first reported. A placebo that was designed to look like the real medication but that contained no active treatment was also used among the patients in the study to act as a control to test the effectiveness of the newly developed medication.

A Cox proportional hazards model was used to model the rate at which patients recovered from the illness. The following two covariates were used in the model:

X_1 = Medication indicator (0 = Placebo administered, 1 = Medication administered)
X_2 = Treatment delay indicator (0 = Treated on first day after symptoms developed
1 = Treated on second day after symptoms developed
2 = Treated after second day after symptoms developed)

The investigation followed a sample of 600 patients for 10 days after treatment. The number of patients in each of the six covariate groupings was as follows:

	$X_1 = 1$	$X_1 = 0$
$X_2 = 0$	100	100
$X_2 = 1$	100	100
$X_2 = 2$	100	100

The first two patients to recover were both treated with medication on the second day after symptoms developed. The first individual recovered 2 days after treatment and the second individual recovered 3 days after treatment. There were no censoring events prior to the second individual recovering.

(i) State the term in the partial likelihood expression that relates to the second individual recovering, clearly defining each of the terms you use. [4]

Following completion of the investigation, the coefficients for the two covariates were estimated as follows:

Covariate	*Coefficient*	*Standard error*
X_1	+0.15	0.02
X_2	–0.02	0.02

(ii) Comment on the impact of the two covariates implied by these results. [5]
[Total 9]

5 A banking regulator in a developing country is assessing the probability of insolvencies in the country's banking sector. Of the country's ten banks, two banks, A and B, are considered to be 'very large', and two banks, C and D, are considered to be 'large'.

The regulator wants to understand the likelihood that both of the two largest banks, A and B, will become insolvent in the next 12 months. Bank A has calculated its probability of insolvency in the next 12 months to be 10% while Bank B has calculated this to be 5%. A banking analyst at the regulator has commented that the probability of both becoming insolvent in the next 12 months is therefore 1 in 200.

(i) Discuss the analyst's comment. [4]

The regulator is now analysing the probability of insolvency of all four of the largest banks, A–D, in the next 12 months. Banks C and D have calculated their individual probabilities of insolvency over the next 12 months to be 3% and 8%, respectively. The regulator decides to model the dependency between the insolvency of the banks with the lower tail of a Clayton copula with a parameter value, $\lambda = 6$.

The generator function for the Clayton copula is:

$$\Psi(t) = \frac{1}{\lambda}\left(t^{-\lambda} - 1\right) \quad \text{where } -1 \leq \lambda < \infty, \lambda \neq 0$$

(ii) Determine, by first deriving a formula for the four-dimensional Clayton copula $C(u_1, u_2, u_3, u_4)$ in terms of the generator function and its pseudo-inverse function, the probability that all four banks will become insolvent in the next 12 months. [7]

[Total 11]

6 An actuary is working on a machine learning modelling project involving a very large number of covariates. A colleague suggests that the best approach is to include all the covariates in the model and use the method of maximum likelihood to estimate the regression parameters.

(i) Explain why this approach may not work well. [3]

The actuary consults a data scientist who suggests using penalised regression. This involves maximising a penalised log-likelihood of the form:

$$\log L\left(\beta_0, \beta_1, \ldots, \beta_p \mid x_1, \ldots, x_p\right) - \lambda P\left(\beta_1, \ldots, \beta_p\right)$$

where L is the standard likelihood function, P is the penalty function and λ is the regularisation parameter.

(ii) Explain the consequences of using too large or too small a value of lambda in the penalised log-likelihood function when estimating the regression parameters. [3]

The actuary decides to use the following penalty function P:

$$P\left(\beta_1, \ldots, \beta_p\right) = a\sum_{i=1}^{p}\beta_i^2 + (1-a)\sum_{i=1}^{p}\left|\beta_i\right|$$

where a is a hyperparameter with $0 \le a \le 1$.

(iii) State the type of model obtained by the actuary if $a = 0$ and if $a = 1$. [2]

The actuary fits the model to two data sets:

- Data set A, in which the actuary believes that only a small number of covariates have a non-negligible impact on the response
- Data set B, in which the actuary believes that most of the covariates have a non-negligible impact on the response.

(iv) Explain which data set the actuary should use a higher value of a for. [3]

[Total 11]

7 The annual aggregate claim amount, S, arising on a short-term insurance portfolio follows a compound Poisson distribution with parameter 5. Individual claim amounts follow a two-parameter Pareto distribution with parameters α and λ. A sample of individual claim amounts was taken and the sample mean and standard deviation were 10,000 and 15,000, respectively.

(i) Estimate the parameters of the Pareto distribution of the individual claim amounts using the method of moments. [5]

(ii) Determine the variance and the third central moment of S, using the estimated Pareto parameters from part (i). [6]

[Total 11]

8 The number of customers, N_t, in a queue at each integer time t is modelled using a Markov chain model.

At the start of each time interval, a number of customers following a Poisson distribution with parameter p join the queue, subject to the constraint that the number of customers in the queue can be no more than N_{max}.

At the end of each time interval, a number of customers following a Poisson distribution with parameter q, where $q > p$, are served and leave the queue, subject to the constraint that the number of customers in the queue cannot be negative.

(i) Comment on the limitations of this model. [2]

An analyst sets the model's parameters to $N_{max} = 2$, $p = 0.5$ and $q = 1$.

(ii) Verify, by separately considering the two transition matrices of customers joining the queue and customers leaving the queue, respectively, that the transition matrix of N_t is:

$$\begin{array}{c} 0 \\ 1 \\ 2 \end{array} \begin{pmatrix} 0.82207 & 0.14475 & 0.03318 \\ 0.48737 & 0.36788 & 0.14475 \\ 0.26424 & 0.36788 & 0.36788 \end{pmatrix}$$ [8]

(iii) Determine the stationary distribution of N_t. [8]

[Total 18]

9 A zero-mean, first-order moving average process is defined by the following equation:

$$X_t = e_t + be_{t-1}$$

where e_t is a sequence of independent and identically distributed $N(0, \sigma^2)$ random variables.

(i) Derive, in terms of b, the value of p that minimises:

$$E[(X_t - pX_{t-1})^2]$$ [7]

(ii) Comment on your answer to part (i) in the case where $b = 1$. [3]

(iii) Determine, in the case where $b = 1$, the values of q and r that minimise:

$$E[(X_t - qX_{t-1} - rX_{t-2})^2].$$ [8]

[Total 18]

END OF PAPER