

INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

18 April 2023 (am)

**Subject CS2A – Risk Modelling and Survival Analysis
Core Principles**

Paper A

Time allowed: Three hours and twenty minutes

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

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

- 1** A process, X_t , is created as follows. n black balls and n white balls are initially placed in two separate boxes, A and B, in such a way that each box contains n balls. An experiment is performed in which a ball is selected at random from each box at time t , ($t = 1, 2, \dots$), and the two selected balls interchanged. Let X_t be the number of white balls in box A just after time t .

(i) Explain whether X_t is irreducible. [1]

(ii) Determine the elements of the transition probability matrix of X_t . [6]

Assume that all the n balls that were initially in box A were the white balls and all the n balls that were initially in box B were the black balls.

(iii) Determine the probability that $X_n = 0$, simplifying your answer where possible. [3]

[Total 10]

- 2** Country A has recently gone through an economic crisis. As the country makes an attempt to recover, the Department of Finance is trying to estimate the rate of recovery in employment. The department has decided to use a two-state continuous-time Markov model to estimate the rate of return to employment. It has also decided to use data from one of the previous economic recoveries for the purpose. The two states are:



The employment data from a previous economic recovery was as follows:

- Waiting time to gain employment in the first year (in person-years): 30,000
- Waiting time to gain employment in the second year (in person-years): 22,000
- Number of people gaining employment in the first year: 5,000
- Number of people gaining employment in the second year: 7,000

It may be assumed that force of gaining employment in any 1 year is constant.

(i) State the likelihood function of the maximum likelihood estimator of the transition rate defining all the terms you use. [3]

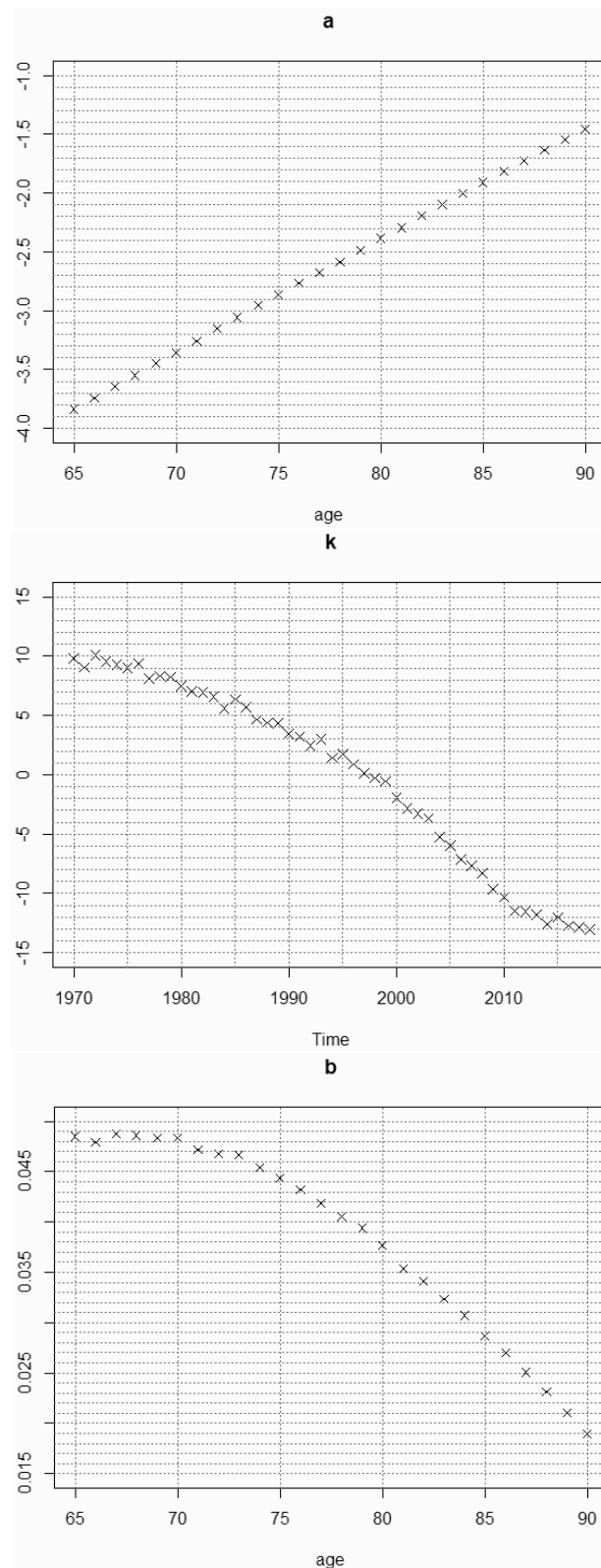
(ii) Calculate the maximum likelihood estimate of the transition rate from the state of being unemployed to gaining employment for each of the first 2 years. [2]

(iii) Estimate, by stating the expression, the variance of the second year maximum likelihood estimator. [4]

(iv) Calculate the probability of not gaining any employment in the next 2 years. [3]

[Total 12]

- 3 The figure below shows three vectors a_x , b_x and k_t , which were obtained by applying the two-factor Lee–Carter model to mortality data derived from the period 1970 to 2018 for the age range 65 to 90 years.



- (i) Discuss what each of these three vectors suggest about the characteristics of the mortality rates in the data. [4]

- (ii) Suggest a socio-economic reason for your conclusion about vector b in part (i) above. [1]
- (iii) Calculate $m_{70, 2018}$ using the output from the model (reading the values from the plots in the figure). [1]

An ARIMA(0,1,0) model (discrete random walk with drift) with Normal errors $\varepsilon \sim N(0, \sigma^2)$ was used to model projected k values. The drift term of the time series model was estimated as -0.4763 , with standard error 0.084 , and σ^2 was estimated as 0.346 .

- (iv) Calculate a 90% confidence interval for $m_{70, 2019}$ based on the value of $m_{70, 2018}$ calculated in part (iii). [5]

On analysing the past data over a larger age range (from 20 to 90 years) it was determined that there was an increase in rates for 20–30 year olds in the 1970s due to an illness that almost exclusively affected the younger population. The higher mortality rates only lasted until 1980 when a cure was found and introduced to the population.

- (v) Explain how this cohort effect may manifest itself in the output from the two-factor Lee–Carter model, fitted to ages 20–90 and calendar years 1970–2018, in terms of a , b and k . [5]
- [Total 16]

4 A Markov jump process model is used to describe the recovery of people bitten by a certain type of poisonous snake. There are three states:

- Sick and receiving medical care following the snake bite
- Fully recovered
- Recovered but with long-term health effects from the bite.

- (i) Explain why a time in-homogeneous Markov jump process model is more suitable than a simpler time homogeneous multi-state model in this scenario. [1]

The transition rates from the sick state in this model t days after being bitten by the snake are:

$$e^{-2.5t} \quad \text{for the transition to fully recovered and}$$

$$0.05 - e^{-2.5t} \quad \text{for the transition to recovered but with long-term health effects.}$$

- (ii) Comment on the key features of this model including the transition rates. [3]
- (iii) Determine the probability that a person just bitten by a snake will eventually make a full recovery without any long-term health effects. [6]
- [Total 10]

- 5** A sample of size n is taken from a process, X_t , which is believed to be an ARMA(1,1) process of the form

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

where $|a|, |b| < 1$. The sample autocorrelations at lag 1 and lag 2 are 0.65 and 0.325, respectively.

- (i) Estimate the parameters a and b by equating the sample autocorrelations to the theoretical values. [6]

Fisher's transformation states that the sample correlation coefficient, r , between two random variables, Y and Z , is such that $\frac{1}{2}\log\left(\frac{1+r}{1-r}\right)$ is approximately Normally distributed with mean $\frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right)$ and variance $\frac{1}{n-3}$, where ρ is the theoretical correlation coefficient between Y and Z and n is the sample size.

- (ii) Determine the minimum value of n necessary to reject the null hypothesis that $b = 0$ in favour of the alternative $b > 0$ at the 95% significance level. You should assume that a is equal to the value determined in part (i) and use Fisher's transformation on the autocorrelation at lag 1. [6]
[Total 12]

- 6** A hydroelectric company is managing a water reservoir created from a dam in a river valley. The dam was originally chosen so that the water level would exceed a threshold of 50 metres in about 2 days in every 300 days. In these extreme events, the excess water is left to escape the reservoir so that the water level is kept below the safety 50-metre limit.

It is believed that the daily water level in the reservoir follows an exponential distribution with mean μ .

- (i) Estimate the value of μ . [3]
(ii) Determine the expected threshold exceedance of the water level over the 50-metre threshold. [2]

In order to better manage the excess water, it is now assumed that the excess water level follows a Generalised Pareto distribution with scale parameter $\beta = 1$.

- (iii) Explain the circumstances in which the Generalised Pareto distribution would be preferred to the exponential distribution. [2]
(iv) Estimate the value of the parameter γ if the expected threshold exceedance is the same as that in part (ii). [3]
[Total 10]

- 7 A mountain rescue service is looking to introduce a new training programme for volunteers who wish to join the service. Each Saturday for 10 weeks trainee rescuers are asked to join a mountain climb. Only those who successfully complete the climb are invited back the following week.

The rescue service will recruit those trainees who successfully complete a certain number of Saturday climbs. To decide on how many successful weeks should be required for a new recruit, the rescue service conducts a trial with 20 volunteers. The table below shows how many of these volunteers fail to complete the climb each week and the number who are eligible but do not arrive for the beginning of each climb.

<i>Week</i>	<i>Eligible but do not arrive</i>	<i>Arrive but fail to complete the climb</i>
1	0	1
2	0	2
3	1	2
4	0	0
5	0	1
6	4	1
7	0	2
8	0	1
9	0	2
10	0	1

- (i) Explain why the Kaplan–Meier estimate is a suitable way to evaluate this training programme. [2]

The rescue service would like to recruit 30% of the volunteers who start the programme.

- (ii) Calculate the number of successful weeks the service should require trainees to complete using the Kaplan–Meier estimate. [8]
- (iii) Discuss what concerns the rescue service should have about using this study to set the recruitment criteria for all future volunteers. [3]

[Total 13]

8 Consider the time-series model:

$$y_t = a y_{t-2} + e_t + b e_{t-1} \quad (\text{A})$$

where e_t is a white noise process with mean 0 and variance σ^2 .

(i) Derive the possible values of a and b for which the process y_t is stationary and invertible. [4]

(ii) State the values of p and q for which y_t is an ARMA(p, q) process. [1]

If $b = 0$ the original model (A) reduces to

$$y_t = a y_{t-2} + e_t \quad (\text{B})$$

(iii) Derive the autocorrelation function for this model while stationarity is assumed to hold. [8]

An actuary attempts to fit the model (A) to some time series data but concludes that the simpler model (B) is more appropriate.

(iv) Discuss how this conclusion could have been reached. [4]

[Total 17]

END OF PAPER