INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

19 September 2023 (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.

1 A leap year is a calendar year which has 366 days including 29 February. This occurs once every 4 years and the most recent leap year was 2020. A group of people who were born on 29 February in different leap years meet for dinner once every 4 years when it is a leap year to celebrate their unusual birthdays. The number of people at the dinner on 29 February 2020 was recorded by their year of birth as follows:

Year of birth	1948	1952	1956
Number attending	3	8	18

One of the group decides to estimate the survival of group members using Gompertz Law with the force of mortality at age x given by:

$$\mu_x = 0.0045\,(1.0004)^x$$

(i) Comment on the choice of formula for the force of mortality. [2]

(ii) Calculate the expected number of people that will be at the next dinner stating any assumptions you make. [5]

A new member, born on 29 February 1960, asks to join the group.

(iii) Calculate the expected cost of dinners for this member up to and including 29 February 2040 if each dinner costs £60 per person and ignoring interest and inflation. [6]

[Total 13]

2 Let X_n be a sequence of independent and identically distributed random variables.

(i) Demonstrate that the distribution function of the X_n, F, satisfies:

$$\left(F(\beta_n x + \alpha_n)\right)^n = \left(1 - \frac{1}{n}\left(1 + \frac{C(x-A)}{B}\right)^{-1/C}\right)^n$$

over the range of possible values of the X_n, in the following cases:

(a) X_n is uniformly distributed on [0,1], $\alpha_n = 1 - 1/n$, $\beta_n = 1/n$, $A = 0$, $B = 1$ and $C = -1$.

(b) X_n has the two-parameter Pareto distribution with parameters δ and λ, $\alpha_n = \lambda(n^{1/\delta} - 1)$, $\beta_n = n^{1/\delta}$, $A = 0$, $B = \lambda/\delta$ and $C = 1/\delta$.

[6]

(ii) Explain the significance of your results in part (i) by considering the limiting behaviour of the right-hand side as $n \to \infty$. [8]

[Total 14]

3 At the start of the week, Alex can choose salad, pizza or sushi for lunch. On Monday Alex starts with a uniformly random choice, but in the following days makes the choices following a discrete time Markov chain with transition matrix:

$$P = \begin{pmatrix} 0.3 & 0.2 & 0.5 \\ 0.1 & 0.3 & 0.6 \\ 0.3 & 0.2 & 0.5 \end{pmatrix}$$

where the 1st, 2nd and 3rd rows and columns correspond to salad, pizza and sushi, respectively.

(i) Derive the probability that Alex has pizza on Monday and salad on Wednesday. [4]

(ii) Calculate the probability that Alex has sushi on Wednesday and Friday. [6]

[Total 10]

4 A teacher is looking for ways to cluster her pupils into homogenous groups and decides to set up an experiment as follows.

Each day, every pupil is given a task and scores '+2' if the task is successfully completed by the end of the day and '−2' otherwise. This experiment runs for n days ($n > 1$) and at the end pupils are grouped based on their total score over n days.

Let us consider a pupil in the class who has a 50% chance of successfully completing the task each day, and we assume that their performance from day to day is independent. We denote by Z_t this pupil's score at the end of day t, and by A_t their total score from day 1 to day t.

(i) Determine the expected value and variance of A_t. [2]

(ii) Determine the probability that $A_n = k$ where k is an integer, writing your answer as a function of n and k. [4]

At the end of the experiment, a pupil will be classified as 'borderline' if their total score, A_n, is zero.

(iii) Determine $Pr(A_1 = 2 \mid A_n = 0)$ that is the conditional probability that the pupil scored +2 on the first day given they finish with a borderline classification.

[5]

[Total 11]

5 Let X_t be the process defined by:

$$X_t = \sum_{i=1}^{t} Y_i$$

where:

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

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

(i) State the values of p, d and q for which X_t is an ARIMA(p, d, q) process. [1]

(ii) Demonstrate that $\text{var}(Y_t) = (1 + b^2)\sigma^2$ and $\text{cov}(Y_t, Y_{t-1}) = b\sigma^2$. [4]

(iii) Demonstrate that

$$\text{var}(X_t) = t(1 + b^2)\sigma^2 + 2(t-1)b\sigma^2$$

and that

$$\text{cov}(X_t, X_{t-k}) = (t-k)(1 + b^2)\sigma^2 + (2(t-k) - 1)b\sigma^2$$

for $0 < k < t$. [6]

(iv) Explain what the results in part (iii) imply about the shape of the autocorrelation function of X_t. [2]

[Total 13]

6 A bank branch is responsible for cash replenishments of two cash machines (or ATMs) near its premises. It is known that the number of customers withdrawing from the first and the second cash machine follow Poisson distributions at the rate of 20 and 50 per day, respectively.

Withdrawals of cash per customer from the two cash machines follow normal distributions with mean \$1,500 and \$1,000 per customer and standard deviation \$300 and \$200 per customer, respectively.

Assuming that the total cash withdrawals from each of the cash machines independently follow compound Poisson distributions:

(i) Determine the probability that combined cash withdrawal per customer from both the cash machines is less than \$1,400 per customer. [6]

(ii) Calculate the mean and standard deviation of the combined cash withdrawal per day from both the cash machines. [8]

[Total 14]

7 A survival study followed twelve patients, for a maximum of 10 days each, following a major surgical operation. From previous similar studies, around one-third of patients survived 10 days. The condition of all patients (i.e. whether a patient was alive or dead) was monitored daily. The results are set out below where $S(t)$ is the Kaplan–Meier estimate of the survival function:

Time since operation (days)	$S(t)$
$0 \le t < 2$	1
$2 \le t < 4$	0.9
$4 \le t < 5$	0.7
$5 \le t < 7$	0.56
$7 \le t < 10$	0.373

(i) Calculate the number of deaths and the number of patients who were censored, stating the times of all deaths and censoring events. [8]

An expert analyst has voiced concerns about the accuracy of the data, thinking there is likely to have been an error.

(ii) Briefly explain the likely source of the expert's concerns. [1]

(iii) State ways in which this study could be improved. [2]

[Total 11]

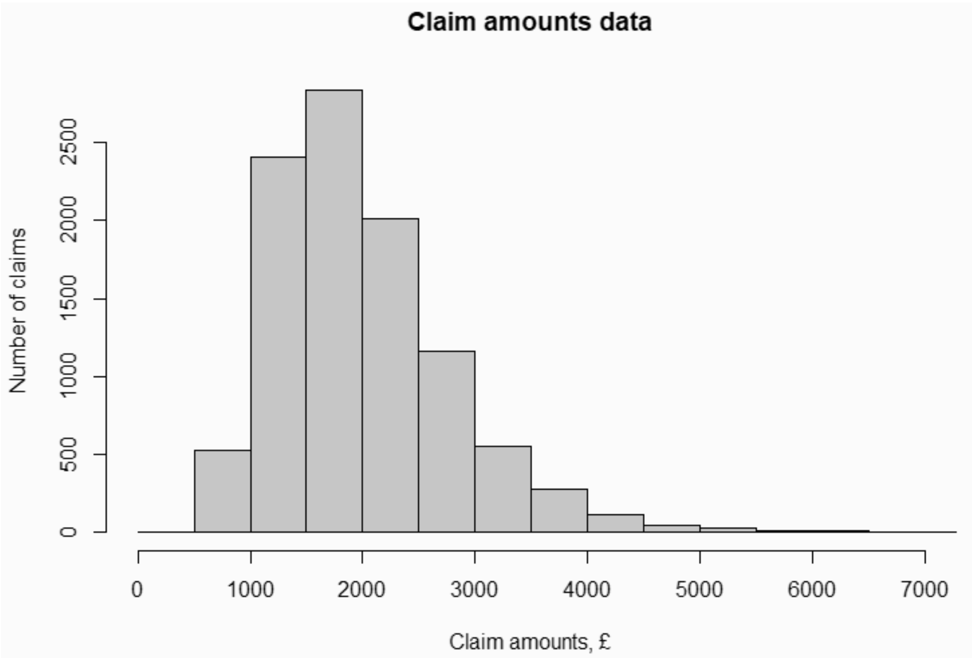
8 An insurer has incurred 10,000 claims under a portfolio of home insurance policies. These claims have a mean size of £2,000 and a standard deviation of £800. One hundred of these claims have exceeded the excess of loss limit on a reinsurance policy that the insurer has in place.

(i) Using a Lognormal distribution, estimate the excess of loss limit on this reinsurance policy. State your assumptions and show all your working clearly. [5]

(ii) Calculate the number of claims that would be expected to be less than £1,000. [1]

It has been proposed by the insurance regulator that statutory solvency calculations should be based on modelling claims using a suitable Normal distribution.

(iii) Comment on the appropriateness of using a Normal distribution under various conditions. (You may use the information about the portfolio of policies above to illustrate your answer.) [3]



(iv) Comment briefly on which of the following alternative distributions should be considered, in addition to the Lognormal distribution, when fitting a suitable model to these claims, given the histogram showing the claims data in the figure above, and how you may decide which distribution to include in your final model.

- gamma
- exponential
- Weibull.

[5]
[Total 14]

END OF PAPER