INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

20 April 2022 (am)

Subject CS2 – Risk Modelling and Survival Analysis Core Principles

Paper B

Time allowed: One hour and fifty 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 The annual aggregate claim amount of an insurer follows a compound Poisson distribution with parameter 1,000. Individual claim amounts follow a Gamma distribution with shape parameter $\alpha = 750$ and rate parameter $\lambda = 0.25$.

(i) Generate 20,000 simulated aggregate claim values for this insurer, using a random number generator seed of 914. Use the R function, `head()`, to display the first seven simulated claim values in your answer script. [7]

(ii) Determine the mean and the standard deviation of the simulated aggregate claim values from part (i). [3]

(iii) Plot the empirical density function of the simulated aggregate claim values from part (i), setting the x-axis range from 2,600,000 to 3,300,000 and the y-axis range from 0 to 0.0000045. [5]

(iv) Suggest a suitable distribution, including its parameters, that approximates the simulated aggregate claim values from part (i). [2]

(v) Generate 20,000 values from your suggested distribution in part (iv) using a random number generator seed of 914. Use the R function, `head()`, to display the first seven generated values in your answer script. [3]

(vi) Plot the empirical density function of the simulated values in part (v) as a different coloured line in the chart that was produced in part (iii). [4]

[Total 24]

2 Before answering this question, the data set x, representing $n = 100$ observations from an ARMA(2,0) model, must be generated in R using the following code:

```
set.seed(12456)
x = arima.sim(n = 100, model = list(ar = c(0.7,0.2)))
```

(i) Plot the sample autocorrelation function (sample ACF) and sample partial autocorrelation function (sample PACF) for the data set x. [3]

(ii) Comment on the general features of the two plots in part (i) with emphasis on how they compare to the theoretical behaviour of the corresponding functions for an ARMA(2,0) process. [2]

(iii) Fit the following three models to the generated data set x, displaying their Akaike Information Criterion (AIC) values in the R output in your answer script:

ARMA(1,0)
ARMA(1,1)
ARMA(2,0)

[4]

(iv) Explain, using the results from part (iii), which of the fitted models is the most appropriate for modelling the data set x. [2]

(v) Comment on your answer to part (iv). [3]

(vi) Explain, with reasons, how the value of n in the R code for generating data set x may be changed to ensure that the ARMA(2,0) model becomes the best fitting model. [2]

(vii) Generate a new data set, y, using your suggested change to the value of n in part (vi). Set the same random number generator seed as above before generating y. [2]

(viii) Fit the three models to the new data set y, displaying their AIC values in the R output in your answer script. [3]

(ix) Explain, using the results from part (viii), why the ARMA(2,0) model is the most appropriate model for modelling data set y. [1]

(x) Comment on the differences between the suggested models fitted for data sets x and y in parts (iv) and (ix), respectively. [3]

[Total 25]

3 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. The 'CS2B_A22_Qu_3_Data.csv' file contains mortality data from a recent mortality investigation. The deaths and exposure data are all in respect of the period 1 January 2020 to 31 December 2020 inclusive. The file contains the following six variables:

Age	Age last birthday, x, in single ages from 55 to 95 inclusive
Deaths	Number of observed deaths at age x, D_x
Exposure	Central exposed to risk at age x, E_x^c (measured in years)
Graduation1	Central mortality rate at age x, m_x, derived from graduation method 1
Graduation2	Central mortality rate at age x, m_x, derived from graduation method 2
Graduation3	Central mortality rate at age x, m_x, derived from graduation method 3

The graduation methods are set out below:

Graduation 1	Graduation by parametric formula
Graduation 2	Graduation obtained by removing one parameter from the formula underlying Graduation 1
Graduation 3	Graduation obtained by adding one parameter to the formula underlying Graduation 1

Before answering this question, the 'CS2B_A22_Qu_3_Data.csv' file should be loaded into R and assigned to a data frame called *graduation*.

(i) Construct three new columns in *graduation*, called zx1, zx2 and zx3, containing the standardised deviations:

$$z_x = \frac{D_x - m_x E_x^c}{\sqrt{m_x E_x^c}}$$

for each of the graduations.

Use the R function, `head()`, to display the first seven values of columns zx1, zx2 and zx3 ONLY in your answer script. [6]

(ii) Determine, using R, the p-value of the chi-square goodness-of-fit test for each of the three graduations given that the numbers of degrees of freedom are as follows:

Graduation	1	2	3
Degrees of freedom	36	37	35

[8]

(iii) Comment on your results in part (ii), with reference to the suitability of the three graduations for the observed mortality rates. [6]

(iv) Determine, using R, the numbers of positive and negative deviations for each of the three graduations. [8]

(v) Determine, using R, the numbers of groups of positive deviations for each of the three graduations. [12]

(vi) Determine, using R, the *p*-value of the grouping of signs test for each of the three graduations.

[**Hint:** If you are using the binomial coefficients $\binom{n}{r}$ in your calculations, these may be obtained in R as `choose(n, r)`.] [9]

(vii) Comment on whether your conclusions in part (iii) have changed in light of the results in part (vi). [2]

[Total 51]

END OF PAPER