INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

20 April 2021 (am)

Subject CS2 – Risk Modelling and Survival Analysis Core Principles

Paper B

Time allowed: One hour and forty-five 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 Before answering this question, the R package for Markov chains should be loaded into R using the following code:

```
install.packages("markovchain")
library(markovchain)
```

X_t is a Markov chain on the state space {1,2,3} with the following transition matrix:

$$\begin{pmatrix} p & 1-p & 0 \\ q & 0 & 1-q \\ 0 & r & 1-r \end{pmatrix}$$

where $0 < p, q, r < 1$.

(i) Construct an R function, with arguments p, q and r, that generates a Markov chain object for X_t. [5]

Assume that $p = 0.75$ and $r = 0.25$.

(ii) Construct R code that calculates the stationary distribution of X_t for values of q from 0.1 to 0.9 inclusive, at intervals of 0.1, pasting your results into your answer script. [7]

(iii) Plot a graph showing the stationary distributions of X_t calculated in part (ii) as a function of q. You should include three separately coloured lines on your graph, each line representing the stationary probabilities of each state. [8]

[Total 20]

2 Before answering this question, construct the function, *rpareto*, in R using the following code:

```
rpareto = function(n, alpha, lambda){
rp = lambda * ((1 - runif(n))^(-1/alpha) - 1)
rp}
```

The *rpareto* function generates a random sample of size n from a two-parameter Pareto distribution with parameters α and λ.

(i) Generate, using *rpareto*, a random sample of size 25,000 from the two-parameter Pareto distribution with parameters $\alpha = 3$ and $\lambda = 1$, assigning the simulated values to a vector called *A_vec*. You should set a random number generator seed of 123 before generating *A_vec* and use the R function, `head()`, to display the first eight values of *A_vec* in your answer script. [3]

(ii) Construct an R function, called *A_exceed_u*, with two arguments, *A* and *u*, that returns only the non-zero entries of a vector, *E*, where vector *E* is of length 25,000, with entries defined as:

$$E_i = \max\{A_i - u, 0\}.$$

Use the R function, `head()`, to display in your answer script the first eight values of *A_exceed_u* when *A* = *A_vec* and *u* = 1. [8]

(iii) Construct another R function, called *F_u*, with one argument, *A_greater_than_u,* that returns a vector of length 101, containing the probabilities:

$$P(A_i - u \leq x \mid A_i > u), \text{ where } x \geq 0,$$

for values of *x* from 0 to 10 inclusive, at intervals of 0.1.

Use the R function, `head()`, to display in your answer script the first eight values of *F_u* when *A_greater_than_u* = *A_exceed_u* with arguments *A* = *A_vec* and *u* = 1. [9]

(iv) Plot, on a single graph, four line graphs of the values of *F_u* against *x* for the values of *x* specified in part (iii) when *A_greater_than_u* = *A_exceed_u,* *A* = *A_vec* and *u* = 1, 2, 3 and 4. You should use separate colours to identify each line graph and you should clearly specify which value of *u* is represented by each of the four line graphs. [10]

(v) Comment on the graph produced in part (iv). [6]

[Total 36]

3 Before answering this question, the 'survival' package should be loaded into R using the following code:

```
install.packages("survival")
library(survival)
```

A recently developed drug, MediCo, has been used over the past 12 months to treat a potentially fatal disease. MediCo was approved by a medical regulator last year following initial trials and mortality data has been collected over the past 12 months, since approval, to continue reviewing the drug's effectiveness.

This mortality data has been compared with mortality data collected from infected patients, prior to MediCo's approval, who were NOT administered the drug.

It is suspected that gender may be a significant covariate on the mortality rate of infected patients.

The 'CS2B_Apr_21_Qu_3_Data.csv' file contains the combined mortality data from this investigation for 4,400 infected patients. The file contains the following five variables:

Life	Unique patient identifier (integers 1, 2, …, 4,400)
Drug	Drug indicator (1 = received drug, 0 = did not receive drug)
Gender	Gender indicator (1 = female, 0 = male)
Status	Status indicator (1 = death due to disease, 0 = censoring event occurred)
Time	Duration in days at which death/censoring occurred (integers with a range of 1–365, with 1 = first day of the investigation and 365 = last day of the investigation).

Before answering this question, the 'CS2B_Apr_21_Qu_3_Data.csv' file should be loaded into R and assigned to a data frame called *mortalitydata*.

(i) Plot the Kaplan–Meier survival function estimate for all patients, together with its two-sided 99.5% confidence interval. [8]

(ii) Determine, using the estimated survival function from part (i), the probability that a patient survived from the beginning of the investigation to the end of the investigation. [3]

(iii) Evaluate the appropriateness of the probability value, calculated in part (ii), for assessing MediCo's effectiveness. [3]

(iv) Plot, on a single graph, four Kaplan–Meier survival function estimates without any confidence intervals, where each estimate represents one of the four possible patient group combinations of drug and gender. You should use separate colours to identify each survival function. [9]

(v) Comment on the graph produced in part (iv). [3]

(vi) Estimate a Cox proportional hazards model with death as the event of interest using the two covariates, drug and gender, with no interaction term, pasting your results into your answer script. You should use the Breslow method for tie handling. [5]

(vii) Comment on the results produced in part (vi) with reference to the effects of the two covariates, drug and gender, on the mortality rate. [5]

(viii) Update the Cox proportional hazards model in part (vi) to include an interaction term between drug and gender, pasting your results into your answer script. [3]

(ix) Analyse the effectiveness of MediCo, commenting on any differences between males and females. [5]

[Total 44]

END OF PAPER