INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

8 April 2019 (am)

Subject CS2B – Risk Modelling and Survival Analysis Core Principles

Time allowed: One hour and forty-five minutes

INSTRUCTIONS TO THE CANDIDATE

1. *You are given this question paper.*
2. *Mark allocations are shown in brackets.*
3. *Attempt all questions. Each question is to be answered on a new page.*

If you encounter any issues during the examination, please contact the Examinations Team at T. +44 (0) 1865 268 255

1 Data is available for the movement of taxis in London. The city is divided into three zones “North”, “South” and “West”. The movement of a taxi from one zone to another will depend only on its current position. The following probabilities have been determined for taxi movements:

- Of all taxis in the North zone, 30% will remain in North and 30% will move to South, with the remaining 40% moving to West.
- In the South zone, taxis have a 40% chance of moving to North, 40% chance of staying in South and 20% chance of moving to West.
- Of all the drivers in the West zone, 50% will move to North and 30% to South with the remaining 20% staying in West.

The movement of taxis in London will be modelled in R using a Markov Chain.

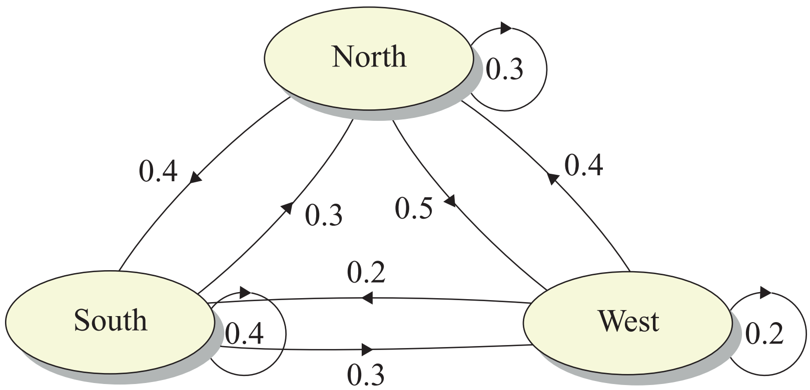
(i) Create a vector with the state space of the Markov Chain, using R code. You should print this to the screen and paste into your answer. [3]

(ii) Construct a transition matrix of the zone movement probabilities. You should print this to the screen and paste into your answer. [3]

(iii) Load the R package for Markov Chains and paste your coding into your answer. [3]

(iv) Create a Markov Chain object with state space equal to your vector in part (i) and transition matrix from part (ii). You should print this to the screen and paste into your answer. [3]

The transition diagram for the Markov Chain is shown below:



(v) Calculate the probability that a driver currently in the North zone will be in the North zone after:

(a) two trips
(b) three trips.

[4]

(vi) Determine the stationary state of the Markov Chain. [4]

[Total 20]

2 An exponential distribution has a parameter of $\lambda = 0.4$.

(i) Use the in-built functions in R to:

(a) Simulate 1,000 values from this distribution, assigning this to a variable called `Exp_Vector` and calculate the mean and variance of the simulated values. Paste the results of your calculation into your answer.

(b) Plot a histogram of `Exp_Vector` showing the frequencies and paste the plot into your answer.

(c) Plot the probability density function for this distribution as:

1 a scatter plot

2 a line graph.

Paste your plots into your answer.

[10]

A lognormal distribution has parameters $\mu = 0$ and $\sigma^2 = 1$.

(ii) Use the in-built functions in R to:

(a) Simulate 1,000 values from this distribution, assigning this to a vector called `LNorm_Vector` and calculate the mean and variance of the simulated values. Paste the results of your calculation into your answer.

(b) Plot a histogram of `LNorm_Vector` showing the frequencies and paste the plot into your answer.

(c) Plot a second histogram in a new graph of `LNorm_Vector` showing the probability densities, setting the *y*-axis range from 0 to 0.7 for this graph and paste the plot into your answer.

(d) Add the cumulative density function of `LNorm_Vector` to the chart in part (ii)(c) and paste the plot into your answer.

(e) Add the theoretical lognormal (0,1) distribution to the chart in part (ii)(d) to highlight the difference to the sample, including appropriate labels and legend and paste the plot into your answer.

[12]

The Pareto distribution is one for which there is no built-in R code. The probability density function can be written as:

$$f(x) = \frac{\alpha\lambda^{\alpha}}{(\lambda + x)^{\alpha+1}}, x > 0$$

(iii) (a) Write down the R code for a function to simulate variables from the Pareto distribution, denoting it `rpareto` and paste your coding into your answer.

(b) Simulate, using `rpareto`, 1,000 values from a Pareto distribution with parameters $\alpha = 3$, $\lambda = 1$, assigning the simulation to a vector called `LNorm_Vector` and calculate the mean and variance of the simulated values. Paste the results of your calculation into your answer.

[8]
[Total 30]

3 You have been asked to analyse the percentage changes in quarterly personal consumption expenditure and personal disposable income in the US from 1970–2010. This information is contained in a time series called "**consumption**" in R's fpp library.

The data can be loaded into R using the command:

```
install.packages("fpp")
library(fpp)
consumption <- usconsumption[,1]
```

(i) Plot this time series giving appropriate labels for each axis and paste the R code and the chart into your answer. [4]

(ii) (a) Plot the Autocorrelation function (ACF) and Partial Autocorrelation function (PACF), giving appropriate labels for each axis and paste the R code and the charts into your answer. [8]

(b) Comment on your plots in part (ii)(a), making reference to the stationarity. [3]

(iii) Fit the most appropriate ARIMA model based on the results in part (ii)(a), stating the equation of the model and justifying your choice. [5]

The dataset usconsumption also includes the quarterly personal disposable income from 1970 to 2010. The data can be loaded into R using the command:

```
income <- usconsumption[,2]
```

(iv) (a) Compare the performance of the ARIMA model you have chosen in part (iii) with a linear regression model of consumption on income, by computing the root mean square error (RMSE) for the fitted values of each model. Paste your coding and R output into your answer.

(b) Explain why the RMSE is not the ideal measure to compare these models, including a recommendation of a more appropriate measure.

[15]

An analyst has suggested that neither model in part (iv) is a good fit to the data and has asked you to propose an alternative model.

(v) (a) Suggest a suitable alternative model to fit to the data.

(b) Fit the model you have suggested in part (v)(a) to the data, stating the equation used.

(c) Compare the results of this model to the models fitted in part (iv).

[15]

[Total 50]

END OF PAPER