INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

15th September 2021

Subject CS2B – Risk Modelling and Survival Analysis (Paper B)

Time allowed: 2 Hours (14.30 - 16.30 Hours)

Total Marks: 100

Q. 1) The Logistic function can be defined by the function:

$$f(x) = \frac{1}{(1 + \exp(-x))}$$

i) Create a function in R to calculate f(x). (3)

ii) Create a plot of this function and evaluate it at values for x ranging from -10 to 10. Add appropriate x and y-axis labels. (5)

[8]

Q. 2) A random variable follows a Weibull distribution with parameters for the shape and scale as follows:

Shape = 1.5
Scale = 100,000

Run the following code in R before beginning your answer:

```
set.seed(1234)
options(scipen=5)
```

Use the in-build functions in R to:

i) Simulate 1,000 values from a Weibull distribution with parameters given above, assign this to a vector variable called 'WeibullSim'. (3)

ii) Calculate the mean and standard deviation of the simulated values. (2)

iii) Using the mean and standard deviation estimated from (ii), estimate the μ and σ parameters of a LogNormal distribution following the method of moments approach. (5)

iv) Simulate 1,000 values from the LogNormal distribution using the parameter estimates in (iii) above and assign this to a vector variable called 'LognormalSim' and calculate the mean and standard deviation of the simulated values. (5)

v) Plot a histogram of 'WeibullSim' and 'LognormalSim' showing the frequencies and paste the plot into your answer. (3)

vi) Plot the probability density function for 'WeibullSim' and 'LognormalSim' as a line plot.

- Set the WeibullSim line to color 'red' and the LognormalSim line to color 'green'.
- Set the y-axis limits from 0 to 0.00001.

Hint: Use the density function in R to estimate densities of your simulated values. (3)

[21]

Q. 3) An Actuary in the country of Actuaria believes the actions of a football player while playing a football match, can be modelled on a ball-by-ball action basis using a Markov Chain. The Actuary has considered 3 states for a player to model the ball-by-ball action – Pass, Shoot and Dribble.
Library required – 'markovchain'

The following probabilities have been defined based on past data for each state:

1. There is a 50% chance of a Pass being made, 10% chance of Shooting and 40% chance of Dribbling in the next state of play, if the player had Passed the ball in the previous play.

2. There is a 70% chance of a Pass being made, 20% chance of Shooting and 10% chance of Dribbling in the next state of play, if the player had Shot the ball in the previous play.

3. There is a 40% chance of a Pass being made, 30% chance of Shooting and 30% chance of Dribbling in the next state of play, if the player had Dribbled the ball in the previous play.

i) Create a vector with the state space of this Markov Chain in R. (3)

ii) Create a matrix of transition probabilities in R with names of that states in the matrix. Then create a Markov Chain Object using the 'new' function and details from the matrix created. (3)

iii) Plot the transition probability matrix. (3)

iv) Calculate the steady state of this distribution. (4)

v) Calculate the probability that a player who Passes the ball initially, will Pass the ball after:

a) 2 plays have taken place (2.5)

b) 5 plays have taken place (2.5)

vi) Use the function set.seed(100). Simulate a sequence of 100 states using markovchainSequence function. Compute the frequency of the states. (4)

vii) Use a bar chart to plot the relative frequency of the states and label the chart appropriately. (2)

viii) Comment by comparing the results obtained in (vii) with the steady state distribution obtained in (iv). (3)

[27]

Q. 4) The 'sunspot' dataset contains the monthly mean relative sunspot numbers from 1749 to 1983. Collected at Swiss Federal Observatory, Zurich until 1960, then Tokyo Astronomical Observatory. Please import this dataset from the 'datasets' package in R.
Libraries required:

'tseries'
'forecast'

i) Plot the timeseries with frequency = 12, giving appropriate labels for each axis and paste the chart into your answer sheet. (3)

ii) Plot the sample Auto-correlation Function (ACF) and the sample Partial Auto-Correlation Function (PACF) of the data, giving appropriate labels for the x and y axes. (5)

iii) Estimate the ACF and PACF values at each lag. (2)

iv) Comment on the ACF based on the charts and values computed in (ii) and (iii). (2)

v) Calculate the simple seasonal difference with period 12. Print the top 12 values of the seasonal difference. (2)

vi) Decompose the trend, seasonality and randomness of the data. Plot the decomposed values. (4)

vii) Comment on the general features of your chart. (2)
[20]

Q. 5) The 'DTData.csv' file that has been provided contains details of individuals who made insurance claims in the country of Actuaria. A 'Yes' in the 'type' column indicates a claim was made. The data, additionally contains details on the age and BMI of the individuals at the time of claim. Use a decision tree approach to classify the type of claims based on BMI and age.
Library required - 'tree'

i) Load the data from the csv file into a data frame (*while importing the dataset ensure you set stringsAsFactors to TRUE in the read.csv function*). Split the dataset into 2 – one to train the model and the other to test the model. The first 70% of the data should go into the training set and the rest into the test dataset. (3)

ii) Using the 'tree' package and the 'tree' function, train a decision tree model on your training dataset. Print the results of the model. (3)

iii) Summarise the results and plot the decision tree. Use the 'text' function to add text to the decision tree. (4)

iv) Using the model trained on the 'Training' dataset, predict whether a claim will or will not occur on the test dataset. Use the 'predict' function to carry out the prediction. (3)

v) Summarise the results of your prediction. (3)

vi) Create a summary table to compare the predicted results against the actual 'type' in the test dataset and calculate an accuracy score and type 1 and type 2 errors. (4)

vii) Comment on the accuracy of your predictions and provide suggestions on how any bias in the modelling could be minimised. (4)
[24]

****