INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

16th March 2022

Subject CS2A – Risk Modelling and Survival Analysis (Paper A)

Time allowed: 3 Hours 30 Minutes (9.30 - 13.00 Hours)

Total Marks: 100

Q. 1)

i) For a discrete time stochastic process, define the terms;

a) Stationary
b) Weakly stationary
c) Increments
d) Markov property
e) Martingale (5)

ii) In a simple discrete time model for the price of a share, the change in price at time t, X_t is assumed to be independent of anything that happened before time t and to have

$$X_t = \begin{cases} 1 & with\ a\ probability \quad p \\ -1 & with\ a\ probability \quad 1-p \end{cases}$$

Let $S_n = S_0 + X_1 + \cdots + X_n$ be the price after n time units and $S_0 = m$ be the original price of the share. Explain with reasons which of properties stated in 1(i) is satisfied by S_n. (6)

[11]

Q. 2) An insurance company employs 300 agents grouped into four grades labelled as A, B, C and D. Agents move between the grades according to whether they meet their weekly business target. Agents employed at the start of any week remain employed throughout that week. At the end of each week, agents are considered for promotion to the next grades, or they leave employment. If an agent leaves the company he is instantly replaced by a new one in grade A.

- Agents in grade A at the beginning of a week get promoted to grade B with probability 0.04, leave the company with probability 0.03 or continue in the same grade at the beginning of the next week.
- Agents in grade B at the beginning of a week get promoted to grade C with probability 0.03, leave the company with probability 0.06 or continue in the same grade at the beginning of the next week.
- Agents in grade C at the beginning of a week get promoted to grade D with probability 0.005, leave the company with probability 0.01 or continue in the same grade at the beginning of the next week.
- Agents in the grade D at the beginning of a week leave the company with the probability 0.02 or continue in the same grade at the beginning of the next week.

i) Define Stationary probability distribution and its role in different Markov chains. (2)

ii) Which of the following is Agents' movements transition probability matrix P (in terms of the grade they are in, and when they leave the company) of a discrete time Markov chain.

A. $$\begin{array}{c} \begin{matrix} A & \quad B & \quad C & \quad D \end{matrix} \\ \begin{bmatrix} 0.96 & 0.04 & 0 & 0 \\ 0.06 & 0.97 & 0.03 & 0 \\ 0.01 & 0 & 0.995 & 0.005 \\ 0.02 & 0 & 0 & 0.98 \end{bmatrix} \end{array}$$

B. $$\begin{array}{c} \begin{matrix} L & \quad A & \quad B & \quad C & \quad D \end{matrix} \\ \begin{bmatrix} 0.03 & 0.96 & 0.04 & 0 & 0 \\ 0.06 & 0.97 & 0.03 & 0 & 0 \\ 0.01 & 0 & 0.985 & 0.005 & 0 \\ 0.02 & 0 & 0 & 0 & 0.98 \end{bmatrix} \end{array}$$

C. $$\begin{array}{c} \begin{matrix} L & \quad A & \quad B & \quad C & \quad D \end{matrix} \\ \begin{bmatrix} 0.03 & 0.93 & 0.04 & 0 & 0 \\ 0.06 & 0.03 & 0.97 & 0 & 0 \\ 0.01 & 0 & 0 & 0.995 & 0.005 \\ 0.02 & 0 & 0 & 0 & 0.98 \end{bmatrix} \end{array}$$

D. $$\begin{array}{c} \begin{matrix} A & \quad B & \quad C & \quad D \end{matrix} \\ \begin{bmatrix} 0.96 & 0.04 & 0 & 0 \\ 0.06 & 0.91 & 0.03 & 0 \\ 0.01 & 0 & 0.985 & 0.005 \\ 0.02 & 0 & 0 & 0.98 \end{bmatrix} \end{array}$$

(2)

iii) Suppose the company has 300 agents at the beginning of week 1 distributed as 150 in grade A, 75 in grade B, 45 in grade C and 30 in grade D. Calculate the expected number of agents in each grade at the beginning of week 3. Explain the key steps in the solution. (5)

iv) Calculate the expected number of agents in each grade in steady state. (5)

[14]

Q. 3) We are investigating the survival times of patients undergoing kidney transplant by three separate surgical procedures A, B and C. The following data has been recorded for each patient:

Z1 = Age in years less 30
Z2 = 0 for females
= 1 for males
Z3 = 1 if procedure A adopted
= 0 if any other procedure
Z4 = 1 if procedure B adopted
= 0 if any other procedure

A Cox proportional hazard model (Model 1) is used. $\lambda(t) = \lambda_0(t)e^{\beta Z^T}$ where "t" is the time in weeks and $\lambda_0(t)$ is the baseline hazard. The parameters have been estimated as –
$\hat{\beta}_1 = 0.003, \hat{\beta}_2 = 0.021, \hat{\beta}_3 = -0.025, \hat{\beta}_4 = 0.015$

i) Which class of patient does the baseline hazard represent? (2)

ii) Compare the survival period of a male aged 40 with surgical procedure A with that of a female aged 45 with surgical procedure B. (3)

iii) A second model (Model 2) to further the above study is developed which also studies the two way interaction between age and surgical procedure adopted.

$$Z5 = Z1 \times Z3$$
$$Z6 = Z1 \times Z4$$

The new parameters are estimated as $\hat{\beta}_5 = 0.0021 \quad \hat{\beta}_6 = 0.00033$

Reassess (ii) using model 2. (2)

iv) If the maximised log likelihood for model 1 and model 2 are -64.34 and -64.21 respectively, state and explain which one of the models is preferable. (2)

[9]

Q. 4) **i)** Why mortality rates are expected to be smooth? How would you test the smoothness of the given rates? (3)

ii) What are the advantages or disadvantages of using following tests while testing the adherence of data to the mortality rates?

a) Chi-square test

b) Standardized deviation test

c) Sign test

d) Cumulative deviation test

e) Grouping of sign test

f) Serial correlation test (6)

iii) Do you feel that rates adapted on the basis of above tests completely reflect the inherent feature of the data? Give Reasons. (1)

iv) How would your approach be different in adopting these rates for pricing assurance product or annuity products? (2)

[12]

Q. 5) **i)** Choose the odd one out from the options below

a. Random variability

b. Uncertainty in parameter estimate

c. Model mis-specification

d. Accurate data (1)

ii) Choose the odd one out from the options below

a. policy number

b. gender

c. smoker/non-smoker

d. age (1)

iii) Choose the odd one out from the options below

a. force of mortality
b. central rate of mortality
c. frequency of event
d. probabilities of survival or death (1)

iv) Choose the odd one out from the options below

a. To analyse crude rates
b. To produce smooth set of rates
c. To improve reliability of estimates based on adjacent ages
d. To remove random sampling errors (1)

[4]

Q. 6) Three bonds in the same industry with similar credit rating and a similar maturity period of 10 years have been purchased by an investor. The individual default probability for each of the bonds during the first year is 20%.

i) Using an independence copula, calculate the probability that all three bonds default within the first year. (3)

ii) Using a Gumbel copula with parameter α = 2, calculate the probability that all three bonds default within the first year. (3)

iii) Using a Clayton copula with parameter α = 2, calculate the probability that all three bonds default within the first year. (3)

iv) Among the (i), (ii) and (iii) above, state with reasons, which copula model is more appropriate to model the probability of default of bonds in the first year? (3)

[12]

Q. 7) X is following a log normal distribution with parameters $X \sim logN(10, 0.9^2)$, calculate:

i) $\int_{15000}^{50000} f(x)dx$ (2)

ii) $\int_{0}^{15000} x * f(x)dx$ (2)

iii) $\int_{50000}^{\infty} x^2 * f(x)dx$ (2)

[6]

Q. 8) **i)** An extreme event is

(a) High frequency, low severity event

(b) High frequency, high severity event

(c) Low frequency, high severity event

(d) Low frequency, low severity event

(2)

ii) As the sample size increases, the maximum values of a distribution (when appropriately standardised) converge to:

(a) Elliptical Distributions

(b) Generalized Pareto Distribution

(c) Generalized Extreme Value Distributions

(d) Exponential Distributions (2)

iii) In your organization, you have observed that catastrophic risk has been modelled using Normal distribution. State with reasons, the improvements that you would suggest to this model. (3)

[7]

Q. 9) **i)** Which of the following cannot be an ARIMA (1, 1, 1) model? In the series, "e" denotes the white noise process

(a) $Y_t = 0.7Y_{t-1} + 0.3Y_{t-2} + e_t + 0.7e_{t-1}$

(b) $Y_t = 0.6Y_{t-1} + 0.4Y_{t-2} + e_t + 0.6e_{t-1}$

(c) $Y_t = 0.35Y_{t-1} + 0.3Y_{t-2} + e_t + 0.35e_{t-1}$

(d) $Y_t = 0.1Y_{t-1} + 0.9Y_{t-2} + e_t + 0.9e_{t-1}$ (3)

ii) Which of the following models is more appropriate for the given time series, where "e" denotes the white noise process

$$X_t = 2.5X_{t-1} - 2X_{t-2} + 0.5X_{t-3} + e_t + 0.4e_{t-1}$$

(a) ARIMA (3,0,1)

(b) ARIMA (2,1,1)

(c) ARIMA (1,2,1)

(d) ARIMA (0,3,1) (4)

iii) Which of the following type of series do the following expression belong to?

$$X_t = 0.2t + 0.4X_{t-1} + e_t$$

(a) Series with a Trend

(b) Stationary Series

(c) Alternating Series

(d) White Noise process (2)

iv) Cricket players in the Gully Premier League are to be classified into 5 groups based on their batting and bowling averages. The cluster centres (Batting Average, Bowling Average) obtained are C1 (6,48), C2 (15,33), C3 (20,26), C4 (36,11) and C5 (44,9). Into which cluster would you classify a player like Ramesh Sachin whose Batting Average is 30 and who's Bowling Average is 41 by using Euclidean distance?

(a) C1

(b) C2

(c) C3

(d) C4 (2)

v) If instead of Euclidean distance in (iv) above, Manhattan distance is used, which cluster will be the most appropriate for Ramesh Sachin?

(a) C2

(b) C3

(c) C4

(d) C5 (2)

[13]

Q. 10) Consider the time series model defined by: $Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \alpha_3 Y_{t-3} + \epsilon_t$

i) Derive the formula for autocorrelation coefficient with lag 1 and lag 2 in terms of α_1, α_2 *and* α_3. (5)

ii) By considering $\alpha_1 = \alpha_2 = \alpha_3$= 0.3, compute the value of autocorrelation coefficient with lag 1 and lag 2. (2)

iii) From (ii) or otherwise, compute the PACF with lag 1 and lag 2. (2)

iv) Comment on the ACF and PACF values computed in (ii) and (iii) above. (3)

[12]

****