# INSTITUTE OF ACTUARIES OF INDIA

# EXAMINATIONS

## 24th November 2023

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

## Time allowed: 3 Hours 15 Minutes (14.45 – 18.00 Hours)

## Total Marks: 100

### INSTRUCTIONS TO THE CANDIDATES

1. ***Please read the instructions inside the cover page of answer booklet and instructions to examinees sent along with hall ticket carefully and follow without exception.***
2. ***Mark allocations are shown in brackets.***
3. ***Attempt all questions beginning your answer to each question on a separate sheet.***
4. ***Please check if you have received complete Question Paper and no page is missing. If so, kindly get new set of Question Paper from the Invigilator.***

**AT THE END OF THE EXAMINATION**

**Please return your answer book and this question paper to the supervisor separately. You are not allowed to carry the question paper in any form with you.**

**Q. 1)** An investor analyst assesses the rating of insurance and reinsurance companies every month. For the purpose of analysis, the ratings are grouped and classified in the order of merit as Marginal (rating symbols below B), Fair (rating symbols B and B-), Good (rating symbols B+ and B++), Excellent (rating symbols A and A-), Superior (rating symbols A+ and A++).

According to the history, the rating of the bonds evolves as a Markov chain with transition probability matrix given below for some parameter 'α' :

$$\begin{array}{c} \\ \boldsymbol{M} \\ \boldsymbol{F} \\ \boldsymbol{G} \\ \boldsymbol{E} \\ \boldsymbol{S} \end{array} \begin{array}{c} \begin{array}{ccccc} \mathbf{M} & \mathbf{F} & \mathbf{G} & \mathbf{E} & \mathbf{S} \end{array} \\ \begin{bmatrix} \alpha & \alpha & 1-2\alpha & 0 & 0 \\ \alpha & 1-3\alpha-2\alpha^2 & 2\alpha & 2\alpha^2 & 0 \\ 0 & 2\alpha^2 & 1-3\alpha-2\alpha^2 & 3\alpha & 0 \\ 0 & 0 & 2\alpha^2 & 1-3\alpha-2\alpha^2 & 3\alpha \\ 0 & 0 & 2\alpha^2 & 3\alpha & 1-3\alpha-2\alpha^2 \end{bmatrix} \end{array}$$

Note: The symbols M, F, G, E and S in the above matrix represent the rating groups Marginal, Fair, Good, Excellent and Superior respectively.

**i)** Draw the transition graph of the chain. (2)

**ii)** Determine the range of α for which the matrix is a valid transition matrix. (2)

**iii)** Explain whether the chain is irreducible and/or aperiodic. (1)

**iv)** Using the above transition probability matrix, calculate the long-run probability that an insurance company is at rating status Superior (S). Assume $\alpha = 0.1$. (4)

**[9]**

**Q. 2)** An Actuary has been asked to build a model to predict whether a particular policy will have a claim or not. Consider a sample dataset given below corresponding to past year.

| **Location/Zone** | **Policy Type** | **Business Type** | **Age Group** | **Claim** |
|---|---|---|---|---|
| ($X_1$) | ($X_2$) | ($X_3$) | ($X_4$) | (Y) |
| East | Individual | Fresh | <45 | No |
| West | Floater | Port | 45-55 | Yes |
| South | Individual | Port | >55 | Yes |
| North | Individual | Port | >55 | Yes |
| East | Floater | Fresh | <45 | No |
| North | Floater | Fresh | 45-55 | No |
| South | Floater | Fresh | 45-55 | No |
| West | Individual | Port | >55 | Yes |
| East | Individual | Port | <45 | No |
| West | Individual | Fresh | <45 | No |
| North | Floater | Port | >55 | Yes |
| South | Floater | Port | 45-55 | No |

You have been asked to use Naïve-Bayes Classifier Model to predict whether a claim will be made or not for a given set of data inputs.

**i)** What assumption will you make in order to apply Naïve-Bayes Classifier Model? (1)

**ii)** Show that the conditional probability that a policy with n characteristics gives a claim can be written as

$P\,(\text{Claim} = \text{Yes} \mid X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) \propto P(\text{Claim}=\text{Yes}) \prod_{j=1}^{n} P(X_j = x_j \mid \text{Claim} = \text{Yes})$

where $X_j$ is random variable denoting the variable j and $X_j$ is the realisation of that random variable. (3)

**iii)** Calculate the conditional probability $P(X_j \mid Y_j)$ for each $x_j$ in X and $y_j$ in Y. (7)

**iv)** Assume a new policy A of Individual Policy type from North Zone of age group 45-55 years and Port business type is written by the company. Predict whether there will be a claim or not on the given policy. Your calculation shall include proportional probabilities also. (4)

**[15]**

**Q. 3)** The time in minutes, until a squash player scoring each point (against a particular player) is exponentially distributed with parameter λ, where λ takes values 1/5, 1/2 or 1 depending upon whether the player has made no point, one point or more than one point respectively till the last score. The probability of scoring a point is not dependant on whether the player has previously scored or not. Each game is played up to 11 points.

Note: The player who first scores 11 points wins. Assume that the player is going to win without giving a chance for the opponent to score.

**i)** Explain how you could model this as a Markov process, commenting on any assumptions made. (1)

**ii)** Calculate the probability that the player will score more than two points in next 6 minutes (Using differential equation). (6)

**iii)** Calculate the expected time in minutes to complete the game. (1)

**[8]**

**Q. 4)** A health insurance policy offers three types of hospital cash benefits options - Silver, Gold and Platinum, having different lumpsum hospital cash benefits payable.

Claims on these portfolios of policies occur according to a Poisson process with a mean rate of 4 claims per day, irrespective of the type of benefit. 50% of claims are of type Silver, 30% are of type Gold and remaining of type Platinum.

**i)** Calculate the expected waiting time until the first claim of type Silver. (1)

**ii)** Calculate the probability that there are at least 9 claims during the first 2 days, given that there were exactly 7 claims during the first day. State any assumptions made. (2)

**iii)** Calculate the probability that there are at least 5 claims of Gold type during the first day and at least 7 claims of amount Gold type during the first 2 days. (4)

**[7]**

**Q. 5)**

**i)** Explain rate interval and three possible definition of age, rate interval, μ rate of mortality and q probability of death. (3)

**ii)** How would an Actuary graduate the crude rates of mortality for following classes of lives:

a) patients dealing with Chronic myeloid leukemia (CML).

b) male population of all civil services of entire country.

c) females with life insurance for large developed country.

d) a life insurance product to be sold to lives under age 30. (2)

**iii)** An investigation took place into the mortality of diabetes patients between exact age 50 and 51, the force of mortality $\mu_{50}$ is assumed to be constant.

The investigation began on 1st April 2022 and ended on 1st April 2023. Table below gives the data collected in this investigation for 10 lives.

| Life | Date of birth | Date of entry into observation | Date of exit from observation | Exit due to death (Yes-1, No-0) |
|---|---|---|---|---|
| 1 | 1st May 1972 | 1st April 2022 | 1st April 2023 | 0 |
| 2 | 1st July 1972 | 1st April 2022 | 1st April 2023 | 0 |
| 3 | 1st October 1972 | 1st July 2022 | 1st April 2023 | 0 |
| 4 | 1st January 1972 | 1st April 2022 | 1st July 2022 | 1 |
| 5 | 1st March 1973 | 1st April 2022 | 1st April 2023 | 0 |
| 6 | 1st December 1971 | 1st July 2022 | 1st October 2022 | 0 |
| 7 | 1st August 1972 | 1st April 2022 | 1st April 2023 | 0 |
| 8 | 1st August 1972 | 1st April 2022 | 1st September 2022 | 1 |
| 9 | 1st November 1972 | 1st October 2022 | 1st April 2023 | 0 |
| 10 | 1st June 1972 | 1st October 2022 | 1st April 2023 | 0 |

**a)** Estimate the constant force of mortality and probability of death using a two-state model and the data of 10 lives in the table above. (4)

**b)** Assuming the maximum likelihood estimate of the constant force using a Poisson model is exactly same as estimate using the two-state model, outline the differences between the two-state model and the Poisson model when used to estimate the force of mortality or otherwise. (2)

**iv)** A graduation of mortality experience has been carried out for the given population.

| **Age last birthday** | **$\hat{q}_x$ using census data** |
|---|---|
| 37 | 0.0072 |
| 38 | 0.0078 |
| 39 | 0.0081 |
| 40 | 0.0104 |
| 41 | 0.0112 |
| 42 | 0.0114 |
| 43 | 0.0120 |

Perform test of goodness of fit for the graduated $\hat{q}_x$ where Exposed Actual death data to be captured using following information:

| **Age nearest birthday** | **Year 1** | **Year 2** | **Beginning of Year 1** | **Beginning of Year 2** | **Beginning of Year 3** |
|---|---|---|---|---|---|
| 37 | 6 | 8 | 1050 | 1043 | 1053 |
| 38 | 6 | 8 | 1130 | 1123 | 1133 |
| 39 | 8 | 8 | 1060 | 1052 | 1062 |
| 40 | 11 | 9 | 1235 | 1225 | 1235 |
| 41 | 14 | 12 | 1160 | 1149 | 1159 |
| 42 | 13 | 13 | 1370 | 1357 | 1367 |
| 43 | 14 | 16 | 1240 | 1225 | 1235 |

Explicitly state any assumption of rate interval and degrees of freedom. (4)

**[15]**

**Q. 6)** Claims on a mediclaim policy follow the lognormal distribution with parameters $\mu = 10$ and $\sigma = 0.9$. The insurer effects an individual excess of loss reinsurance treaty with a retention limit of Rs.15,000.

**i)** Calculate the probability that the claim paid by the reinsurer is more than Rs. 5,000. (2)

**ii)** Next year the claim amounts on these policies are expected to increase by 10%. The reinsurer wants to change the retention limit so that the probability that a claim involves the reinsurer is 5% higher than the previous arrangement. Calculate the new retention limit. (4)

**iii)** With the new retention limit in (ii), calculate the reinsurer's expected claim payment per claim next year . (4)

**[10]**

**Q. 7)**

**i)** The following data is observed from n = 100 realisations from a time series:

$$\sum_{i=1}^{100} x_i = 5954.39\ ,\ \sum_{i=1}^{100}(x_i - \bar{x})^2 = 3832.26\ ,\ \sum_{i=1}^{99}(x_i - \bar{x})((x_{i+1} - \bar{x}) = 3628.34$$

Estimate using the data above, the parameters μ, α and σ from the model:

$$X_t - \mu = \alpha\,(X_{t-1} - \mu) + e_t$$

where $e_t$ is a white noise process with variance $\sigma^2$. (3)

**ii)** After fitting the model with the parameters found in (i), it was calculated that the number of turning points of the residual's series $e_t$ is 48. Explain and perform significance test at 95% confidence interval on turning points whether there is evidence that $e_t$ is not generated from a white noise process. (3)

**[6]**

**Q. 8)** A generalised extreme value distribution used to model flood damage in Actuaria town, in particular the shape parameter '$\gamma$' used is 1.

**i)** Identify the type of extreme value distribution and state it's key characteristic. (1)

**ii)** Individual losses X of flood in the Actuaria town, follow a Pa(5,2) distribution. The maximum claim in the sample of 50 claims is defined as $X_M$ which follows approximate distribution of GEV(a,b,c) where $a=\lambda n^{1/\alpha}-\lambda$, $b=\lambda n^{1/\alpha}/\alpha$ and $c=1/\alpha$.
Calculate the approximate and exact probability that the maximum claim is more than Rs. 4 crores and comment on GEV distribution. (2)

**iii)** State the key advantage that the generalised Pareto distribution has over the generalised extreme value distribution for the extreme event of flood damage. (1)

**[4]**

**Q. 9)**

**i)** The time series $X_t$ of change in temperature of country Originia is assumed to follow ARIMA(p,d,q) process defined by

$$X_t = 1 + \frac{39}{28}X_{t-1} - \frac{3}{7}X_{t-2} + \frac{1}{28}X_{t-3} + et$$

where $e_t$ follows N(0,1) random variables.

Derive roots of the characteristic equation given that one of the root is 1. Also, comment on stationarity of the process. (3)

**ii)** A time series $X_t$ is assumed to be stationary and to follow an ARMA (2,1) process defined by:

$$X_t = 1 + 9/15\, X_{t-1} - 2/15\, X_{t-2} + Z_t - 2/7\, Z_{t-1}$$

where $Z_t$ are independent N (0,1) random variables.
Find the autocorrelation function for lags 0,1,2. (5)

**iii)** Consider the ARCH(1) process:

$$X_t = \mu + e_t\sqrt{\alpha_0 + \alpha_1(X_{t-1} - \mu)^2}$$

where $e_t$ are independent normal random variables with variance 1 & mean 0.

Show that for s = 1,2,…..,t-1. $X_t$ , $X_{t-s}$ are :

**a)** Uncorrelated (4)

**b)** Not independent (3)

**iv)** Give one real life example where ARCH model can be used. (1)

**[16]**

**Q. 10)** In the Country of Electronia, only 1 type of vehicle is used. The Government wanted to introduce electronic vehicle and thus a small scale study was conducted to analyse the performance of batteries. The initial study was conducted using 12 batteries, all the batteries were charged uniformly for 6 hours and then run them till discharge everyday. The batteries were tested for failing. The test was terminated by the Government after failure of the 8th Battery after 34th month. The details of battery failures were reported as below:

- 2 batteries had failed after 22nd month
- 3 further batteries failed after 29th month
- 2 further batteries failed after 32nd month
- 1 further battery failed after 34th month

Additionally, 1 battery exploded after 30th month.

**i)** State with reason the type of censoring present in this study. (2)

**ii)** Calculate the Kaplan-Meier estimate of the survival function based on the data provided above. (3)

**iii)** Additionally, assuming Government increases the sample size also calculate approximate 95% confidence interval for $\tilde{S}(33)$. To estimate var[$\tilde{S}(33)$], state and use Greenwood's formula. (4)

**iv)** State the assumptions made for above. (1)

**[10]**

**********************