INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

25th May 2024

CS2A - Risk Modelling and Survival Analysis

Time allowed: 3 Hours 15 Minutes (14.45 – 18:00 Hours)

Total Marks: 100

Q. 1) Distribution of losses has the following density function:

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

A random sample of size five produced three losses with values 3, 6 and 14, and two losses exceeding 25. The maximum likelihood estimate of α is:

A. 0.25
B. 0.30
C. 0.34
D. 0.38
E. 0.42 **[3]**

Q. 2) For a random forest, let *p* be the total number of features and *m* be the number of features selected at each split.

Determine which of the following statements is/are true.

I. When *m* = *p*, random forest and bagging are the same procedure.
II. (p-m)/p is the probability a split will not consider the strongest predictor.
III. A typical choice of *m* is 2*p*.

A. None
B. I and II only
C. I and III only
D. II and III only
E. All are correct **[2]**

Q. 3) Following is the transition probability matrix for a Markov Chain, then

$$\begin{bmatrix} 1-2a & 2a & 0 \\ a & 1-2a & a \\ 0 & 2a & 1-2a \end{bmatrix}$$

A. Value of a is any real number
B. Value of a is any positive real number
C. Value of a is less than 0.5
D. Value of a is more than 0.5
E. Value of a is in [0, 0.5] **[2]**

Q. 4) Choose the correct relation between initial rate (q_x) of mortality and central rate of mortality (m_x).

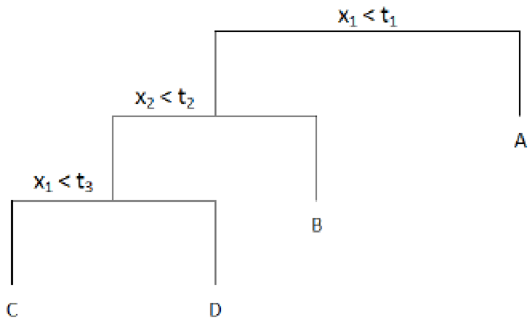
A. $m_x > q_x$
B. $m_x = q_x$
C. $m_x < q_x$
D. $m_x \geq q_x$
E. $m_x \leq q_x$ **[2]**

Q. 5) If arrivals are according to a Poisson Process then distribution of inter arrival times is

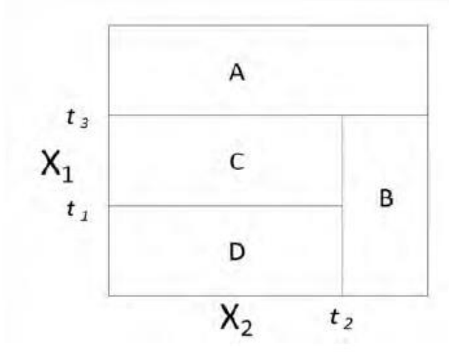
A. Gamma

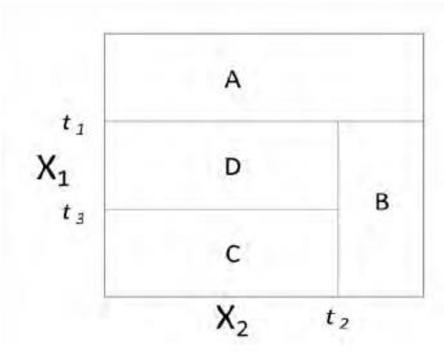
B. Chi-Square
C. Exponential
D. Normal
E. Lognormal **[1]**

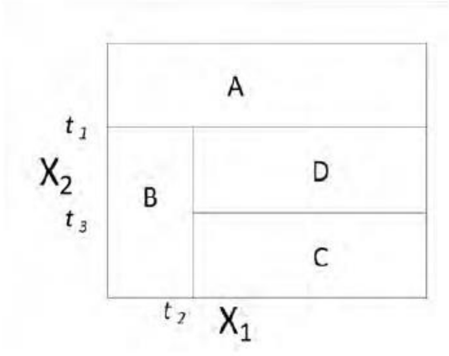
Q. 6) The following tree was constructed using recursive binary splitting with the left branch indicating that the inequality is true

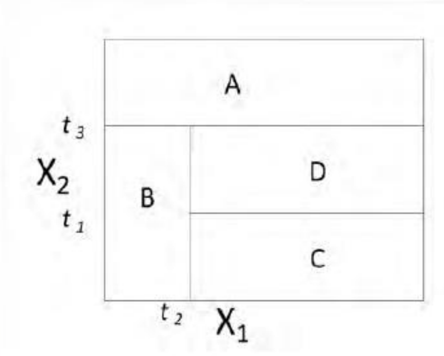


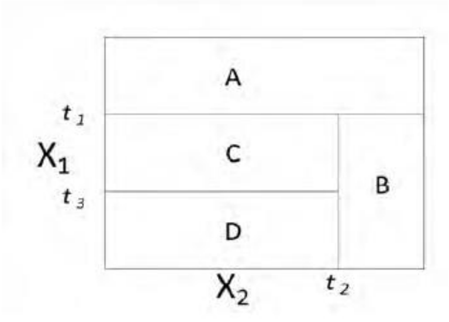
Determine which of the following plots represents this tree.

(A) 

(B) 

(C) 

(D) 

(E) 

[3]

Q. 7) If $\{N_1(t)\}$ and $\{N_2(t)\}$ are two independent Poisson process with parameters λ_1 and λ_2 respectively then $N_1(t) - N_2(t)$ is a

A. Poisson process with parameter $\lambda_1 + \lambda_2$

B. Poisson process with parameter $\lambda_1 - \lambda_2$
C. Poisson process with parameter λ_1 / λ_2
D. Poisson process with parameter $\lambda_1 * \lambda_2$
E. It is not a Poisson process **[1]**

Q. 8) The relationship between complete expectation of life & curtate expectation of life is

A. $\dot{e} = e_x + ½$
B. $e_x = \dot{e}_x + ½$
C. $\dot{e} \approx e_x + ½$
D. $e_x \approx \dot{e}_x + ½$
E. $\dot{e} = E[K_x]$ **[1]**

Q. 9) A time-inhomogeneous Markov Jump process has state space {A,B} and the transition rate for switching between states equals 4t. regardless of the state currently occupied, where t is time. The process starts in State A at t = 0.

The probability that the process remains in state A until at least time s.

A. e^{-2s^2}
B. e^{-s^2}
C. e^{-4s^2}
D. e^{-s}
E. e^{-2s} **[2]**

Q. 10) A mortality table, which follows Gompertz Law for middle and older ages has :

$\mu_{50} = 0.0125$
$\mu_{60} = 0.0180$

Calculate the probability that a life aged 70 will survive for 30 years.

A. 0.1901
B. 0.2112
C. 0.2257
D. 0.2436
E. 0.2679 **[3]**

Q. 11) Consider the Healthy (H), Sick (S) & Dead (D) Model with below transition probability matrix

$$\begin{bmatrix} & H & S & D \\ H & 0.9 & 0.06 & 0.04 \\ S & 0.06 & 0.8 & 0.14 \\ D & 0 & 0 & 1 \end{bmatrix}$$

The probability that a healthy life will never become sick in the future is:

A. 0.9
B. 0.04
C. 0.94
D. 0.375
E. 0.4 **[2]**

Q. 12) Consider a partial likelihood function L of Cox hazard model.

The second differentiation of partial log-likelihood is given as

$$\frac{d^2 \log L}{d\beta^2} = -\frac{2e^{\beta}}{(e^{\beta}+5)} - e^{\beta}$$

We have calculated the value of $\hat{\beta}$ (the maximum likelihood estimate of β) as 0.5ln(8).

Use this information to construct an approximate 95% confidence interval for β.

A. (-0.00039, 2.07983)
B. (-0.00039, 1.95321)
C. (0.00039, 2.07983)
D. (0.00039, 1.95321)
E. (0.00093, 2.07983) **[4]**

Q. 13) An investigation has revealed that 70 percent of newborn babies survived to exact age 40 years and 40 percent of newborn baby survived to exact age 60 years. Assuming that force of mortality μ is constant at ages over 20 years. Calculate μ.

A. 0.05596
B. 0.01399
C. 0.00932
D. 0.55962
E. 0.02798 **[2]**

Q. 14) Time Series are an example of which type of models

A. Discrete time & Discrete state space
B. Discrete time & continuous state space
C. Continuous time & Discrete state space
D. Continuous time & continuous state space
E. Mixed process i.e. a mixture/combination of two process **[1]**

Q. 15) Calculate the curtate expectation of life for a newborn baby subject to constant force of mortality of 0.01 per annum

A. 100 years
B. 99.5 years
C. 100.5 years
D. 99 years
E. 99.9 years **[2]**

Q. 16) Consider the confusion matrix given below :

		Predicted	
		Yes	*No*
Actual	Yes	70	10
	No	1	19

Choose the correct F1 Score & Accuracy

A. 95% and 70%
B. 89% and 92.7%
C. 90.2% and 89%
D. 95% and 89%
E. 92.7% and 89% **[3]**

Q. 17) The force of mortality is given by

	μ_x
$50 < x \leq 60$	0.010
$60 < x \leq 70$	0.015
$70 < x \leq 80$	0.025
$80 < x \leq 85$	0.050
$x > 85$	0.080

Calculate the probability that a life aged exactly 65 will die between exact ages 85 and 88.

A. 0.4426
B. 0.0592
C. 0.1201
D. 0.0987
E. 0.0435 **[2]**

Q. 18) A study of mortality of 10 laboratory rats were undertaken. These rats were observed from birth until either they died or period of study ended, at which point those rats still alive were treated as censored. Any rats that escaped were also treated as censored. No additional rat joins the study after time 0.

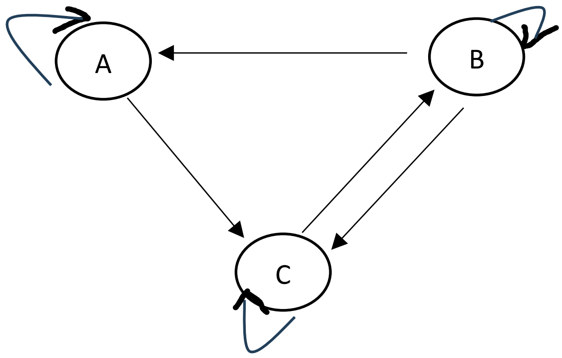
The following table shows the Kaplan – Meier estimate of the survival function, based on the data from 10 rats.

t (weeks)	S(t)
$0 \leq t < 1$	1.0000
$1 \leq t < 3$	0.9000
$3 \leq t < 5$	0.6750
$5 \leq t$	0.5400

Calculate the total number of rats that died in the study.

A. 3
B. 4
C. 5
D. 6
E. 7 **[4]**

Q. 19) A three state process with state space {A, B, C} is believed to follow a Markov chain with the following possible transitions:



An instrument was used to monitor this process, but it was set up incorrectly and only recorded the state occupied after every two time periods. From these observations the following two-step transition probabilities have been estimated:

$P^2[AA] = 0.5625$
$P^2[AB] = 0.125$
$P^2[BA] = 0.475$
$P^2[CC] = 0.4$

Calculate the one-step transition matrix consistent with these estimates. **[5]**

Q. 20) A large insurance company conducts an investigation into the mortality of its older male population. The crude mortality rates are graduated using a standard table by subtracting a constant from the rates given in the table.

A new trainee who has recently joined the company has been asked to test the goodness-of-fit of the proposed graduation using a chi-squared test.

The trainee's workings are reproduced below (in *italics*):

"*Test H0: good fit against H1: bad fit.*

Age	*Actual Deaths (A)*	*Expected Deaths (E)*	*(A-E)^2/A*
70	*9*	*9.53*	*0.024544*
71	*8*	*10.71*	*0.918013*
72	*11*	*11.52*	*0.199127*
73	*11*	*14.70*	*1.244545*
74	*13*	*14.11*	*0.094777*
75	*15*	*17.97*	*0.58806*
Test Statistic			*3.069067*

Age range is 75-70= 5 years so 5 degrees of freedom.

*Two-tailed test so take 2 * 2.66413 = 5.32826 and compare against tabulated value of chi-square distribution with 5 degrees of freedom at 2.5% level, which is 12.833.*

So we accept the null hypothesis."

Identify the errors in the trainee's workings, without performing any detailed calculations. **[4]**

Q. 21)

i) Derive the auto covariance and auto correlation function of the AR(1) process given by:

$$X_t = \alpha X_{t-1} + e_t$$

where $|\alpha| < 1$ and the e_t form a white noise process. (4)

ii) A researcher is using Box Jenkins Approach to model a time series Z_t that is believed to follow a ARIMA (1,d,0) process for some value of d. The time series $Z_t^{(k)}$ is obtained by differencing k times and the sample auto correlations, $\{r_i : 1, \ldots\ldots, 10\}$, are shown in the table below for various values of k.

	k=0	k=1	k=2	k=3	k=4	k=5
r_1	100%	100%	83%	-3%	-45%	-64%
r_2	100%	100%	66%	-12%	-5%	13%
r_3	100%	100%	54%	-11%	-4%	-5%
r_4	100%	99%	45%	-1%	6%	4%
r_5	100%	99%	37%	-3%	4%	5%
r_6	100%	99%	30%	-12%	-12%	-12%
r_7	99%	98%	27%	3%	7%	9%
r_8	99%	98%	24%	3%	0%	-4%
r_9	99%	97%	19%	3%	5%	6%
r_{10}	99%	97%	13%	7%	-5%	-4%

Suggest with reasons, appropriate value for d and the average value of parameter $'\alpha'$ in the underlying AR(1) process (which is of the form as specified in part (i) above). (3)

iii) Having selected an appropriate value of d and carried out some further calculations, the researcher has decided that a zero-mean ARMA (1,1) model provides a good description of the series Delta d_x

The 100 residents for this model are found to contain 74 turning points. The sample auto correlations at lags 1,2,…..,5 are calculated to be:

+0.16 – 0.05 +0.10 +0.12 – 0.02

Carry out each of the following tests, explaining what each test is designed to check for:

a) Ljung – Box (portmanteau) test

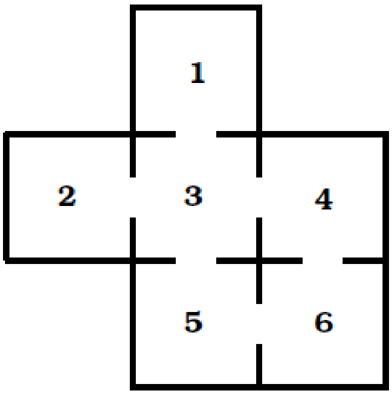
b) turning point test

c) Inspection of the sample auto correlation function

State your conclusions clearly. (9)

[16]

Q. 22) A rat "Jerry" runs through the maze shown below. At each step it leaves the room it is in by choosing at random one of the doors out of the room.



i) Is the above process a Markov Chain? (1)

ii) Give the transition matrix P for the above process. (2)

iii) Is the chain irreducible? (1)

iv) Is the chain aperiodic? (1)

v) Find the stationary distribution. (5)

vi) What is the expected time to return to room 1 if Jerry is currently in Room 1. (2)

vii) Suppose that a piece of mozzarella cheese is placed in a deadly trap in Room 5, Jerry starts in Room 1. Find the expected number of steps before reaching Room 5 for the first time, starting in Room 1. (6)

[18]

Q. 23) An insurance company sells a 1-year insurance policy covering automobile repair cost. In FY 2023-24, the policy has no deductible.

The number of claims is assumed to follow a Poisson Distribution with an expected value of 1.9. The cost of each claim is assumed to follow a Pareto distribution with parameters of $\alpha = 4$ and $\lambda = 900$. The number of claims and cost of each repair are assumed to be mutually independent.

There are loss adjudication and investigation expenses ('claim expenses') involved in setting a claim. It is a random variable uniformly distributed between 50 and 100. The amount of these expenses is independent of the amount of associated claim.

Let the random variable 'S' represent the total aggregate of claims and expenses in one year from this portfolio. Suppose this portfolio comprises of 'n' independent policies.

i) Calculate the mean of S. (3)

ii) Show the variance of S is 609583.33. (6)

In FY 2024-25, the cost of claims is expected to increase uniformly by 20% due to inflation. However, there is no increase expected in claim expenses for FY 2024-25. The number of claims is assumed to have the same distribution in FY 2024-25 as in FY 2023-24.

iii) Calculate the revised mean of S. (3)

iv) The insurers is planning to impose deductible amount of 'd' per claim that would result in insurer's expected aggregate payments being the same in 2024 as on 2023. Calculate d.

Hint:

Given $X \sim Pa(\alpha, \lambda)$

If $W = (X-d) \mid (X > d)$, *then* $f_W(w) = \frac{f_X(w+d)}{P(X>d)}$ (5)

[17]

****