INSTITUTE OF ACTUARIES OF INDIA

CS2 - Risk Modelling and Survival Analysis (Paper A)

May 2024 Examination

INDICATIVE SOLUTION

Introduction

The indicative solution has been written by the Examiners with the aim of helping candidates. The solutions given are only indicative. It is realized that there could be other points as valid answers and examiner have given credit for any alternative approach or interpretation which they consider to be reasonable.

Q. 1) Answer: A **[3]**

$F(x) = \int_1^x \alpha t^{-\alpha-1}$
$= 1-x^{-\alpha}$

L = f(3)f(6)f(14) [1-F(25)]^2
$= \alpha 3^{-\alpha-1}\ \alpha 6^{-\alpha-1}\ \alpha 14^{-\alpha-1}(25^{-\alpha})$^2
$\propto \alpha^3[\,3*6*14*25\text{^}2]^{-\alpha} * [\,3*6*14*25\text{^}2]^{-1}$

Taking Logs

Log L = 3 log α – α log 157500 +C
$\frac{d\ log\ L}{d\alpha} = 3\ \alpha^{-1}$ – log 157500 = 0

α = 0.25

Q. 2) Answer: B **[2]**

I is true. Random forest differs from bagging by setting m < p.
II is true, p – m represents the split not chosen.
III is false, typical choices are the square root of p or p/3.

Q. 3) Answer: E **[2]**

1 – 2a ≥ 0
1 ≥ 2a
a ≤ ½

1 -2a ≤ 1
2a ≥ 0
a ≥ 0

therefore, a is in the range of 0 to 0.5 inclusive of 0 and 0.5

Q. 4) Answer : D **[2]**

$m_x = \frac{q_x}{\int_0^1 {}_t p_x \ dt}$
$\int_0^1 {}_t p_x \ dt \leq 1$

Therefore $m_x \geq q_x$

Q. 5) Answer: C **[1]**

Q. 6) Answer: B **[3]**

The first split is X1 < t1. This requires a horizontal line at t1 on the vertical axis. Graphs B and E have such a line.

The second split is the case where the first split is true. That means all further action is below the line just described. Both graphs do that. The second split is x2 < t2. This requires a vertical line at t2 on the horizontal axis with the line only going up to t1. Again, both graphs have this.

The third split is when the second split is true. That means all further action is to the left of the line just described. The only difference between graphs B and E is which part relates to node C and which to node D. The third split indicates that node C is the case when x1 < t3. Only graph B has this region marked as C.

Q. 7) **Answer: Either B or E**

If t1 > t2 , then Option B would be correct.

If t2 > t1, then it is not a poisson process and Option E would be correct

[1]

Q. 8) **Answer : C** **[1]**

Q. 9) **Answer: A** **[2]**

$= \exp(-\int_0^s \mu_{AB}(t)\, dt)$

$= \exp(-\int_0^s 4t\, dt)$

= exp ([-2t^2] from 0 to s

$= e^{-2s^2}$

Q.10) **Answer : D** **[3]**

$\frac{\mu_{60}}{\mu_{50}} = \frac{Bc^{60}}{Bc^{50}} = \frac{0.018}{0.0125}$

c^{10}= 1.44

c = 1.03714

B = 0.00201882

g = $\exp(\frac{-B}{In\ C})$

g= 0.94614

${}_{30}p_{70} = g^{c^{70}(c^{30}-1)}$

= 0.2436

Q. 11) **Answer: E** **[2]**

It will stay in Healthy state or move to Dead state

Therefore,

0.04 + 0.9*0.04 + 0.9^2 *0.04 + 0.9^3 * 0.04………

0.04 / (1-0.9) = 0.4

Q.12) **Answer : A** **[4]**

$\text{Var}(\hat{\beta}) = [-\frac{d^2 \log L}{d\beta^2}]^{-1}$

$= \left[\frac{2e^{\beta}}{(e^{\beta}+5)} + e^{\beta}\right]^{-1}$

$= \left[\frac{2\sqrt{8}}{(\sqrt{8}+5)} + \sqrt{8}\right]^{-1}$

$= 0.28161$

So,

CI is $\hat{\beta} \pm 1.96\sqrt{var(\hat{\beta})}$

(-0.00039, 2.07983)

Q.13) Answer : E **[2]**

${}_{40}p_0 = {}_{20}p_0\, e^{-20\mu} = 0.7$

${}_{60}p_0 = {}_{20}p_0\, e^{-40\mu} = 0.4$

Dividing the first equation by the second

$e^{20\mu} = 1.75$

$\mu = 0.02798$

Q. 14) Answer: B **[1]**

Q.15) Answer : B **[2]**

$e_0 = \sum_{k=1}^{\infty} {}_k p_0$

$= \sum_{k=1}^{\infty} e^{-0.01k}$

$= \frac{e^{-0.01}}{1-e^{-001}}$

$=$ 99.5 years

Q. 16) Answer: E **[3]**

Precision = TP / (TP + FP) = 70/71 = 98.6%

Recall = TP/ (TP + FN) = 70/80 = 87.5%

F1 Score = (2 * P * R) / (P+R) = (2*98.6 %*87.5%)/(98.6%+87.5%) = 92.7%

Accuracy = (TP +TN)/ (TP +TN + FP +FN) = 89%

Q.17) Answer : C **[2]**

${}_{20|3}q_{65}$ = 5p65 * 10p70 * 5p80 *(1- 3p85)

$= e^{-5(0.015)} * e^{-10(0.025)} * e^{-5(0.05)} * (1- e^{-3(0.08)})$

$= e^{-0.575} * (1 - e^{-0.24})$

$= 0.1201$

Q.18) Answer : B **[4]**

The estimate of the discrete hazard at time 1 is $\hat{\lambda}_1$, is

1 – 0.9 = 0.1 = 1/10

There are 10 rats exposed to risk of death at time 0. The number of deaths at time 1 can only be a positive integer so feasible value of n1 is 10 and d1 is 1.

The estimate of the discrete hazard at time 3 is $\hat{\lambda}_2$, is

$(1 – 1/10)\ (1- \hat{\lambda}_2) = 0.675$

$\hat{\lambda}_2 = 0.25 = 2/8$

There are 9 rats exposed to risk of death at time 1. The number of rats exposed to risk of death immediately before time 3 must be n2 ≤ 9. The feasible value of n2 is 8 and d2 is 2.

The estimate of the discrete hazard at time 5 is $\hat{\lambda}_3$, is

$(1 – 1/10)\ (1 – 2/8)\ (1- \hat{\lambda}_3) = 0.54$

$\hat{\lambda}_3 = 0.2 = 1/5$

There are 6 rats exposed to risk of death at time 5. The number of rats exposed to risk of death immediately before time 5 must be n3 ≤ 6. The feasible value of n3 is 5 and d3 is 1.

Therefore, number of rats died = 1+2+1 = 4

Q. 19) $P^2AA = 0.5625 \Rightarrow a^2 = 0.5625 \Rightarrow a = 0.75$

0.5 mark

Rows of transition matrix must sum to 1.
So, a + c = 1
and c = 0.25

0.5 mark

$P^2AB = 0.125 \Rightarrow ch = 0.125 \Rightarrow h = 0.5$

0.5 mark

h + i = 1
so i=0.5

0.5 mark

$P^2CC = 0.4 \Rightarrow f \times 0.5 + 0.5^2 = 0.4 \Rightarrow f = 0.3$

0.5 mark

$P^2BA = 0.475 \Rightarrow d\ (0.75 + e) = 0.475$
Rows sum to 1 so, d + e =0.7
Substitute for e:
$d(1.45 - d) = 0.475 \Rightarrow d^2 - 1.45d + 0.475 = 0$
Solving using standard quadratic formula:
d = 0.95 or 0.5
0.95 is not possible because e would need to be negative

1.5 mark

So d = 0.5 and e = 0.2

0.5 mark

Transition matrix is:

0.75	0	0.25
0.5	0.2.	0.3
0	0.5	0.5

(Max 5)

Q. 20) The null hypothesis is poorly expressed – should be "underlying rates are the graduated rates" or similar.

0.5 mark

The test statistic is incorrect – the denominator should be expected deaths.

0.5 mark

Number of ages is 6 not 5.

0.5 mark

However fewer than 6 degrees of freedom is appropriate because 1 should be deducted for estimated parameter and some for choice of standard table.

1 mark

This is a one-tailed test not two-tailed.

0.5 mark

Even if it were two-tailed, multiplying test statistic by 2 is inappropriate.

0.5 mark

The trainee has not stated the level of significance e.g. 5%, 1% etc.

0.5 mark

Does not explain that the reason for conclusion is 12.833 > 5.32826.

0.5 mark

The null hypothesis should never be "accepted" rather it is "not rejected".

0.5 mark

The trainee has not stated his or her conclusion in terms of the null hypothesis.

0.5 mark

All the graduated rates are above the crude rates so although the graduation has been accepted it is suspect.

0.5 mark

Cannot comment on figures in table as no access to workings.

0.5 mark

(Max 4)

Q. 21) **i)**

$\gamma_k = \text{Cov}(X_t, X_{t-k})$

$= \text{Cov}(\alpha X_{t-1} + e_t, X_{t-k})$

$= \alpha\gamma_{k-1}$

1 mark

and

$\gamma_0 = \text{Cov}(X_t, X_t)$

$= \text{Cov}(\alpha X_{t-1} + e_t, \alpha X_{t-1} + e_t)$

$= \alpha^2 \text{Cov}(X_{t-1}, X_{t-1}) + \text{Cov}(e_t, e_t)$

$= \alpha^2\gamma_0 + \sigma^2$

0.5 mark

and hence

$\gamma_0(1-\alpha^2) = \sigma^2$

i.e. $\gamma_0 = \frac{\sigma^2}{1-\alpha^2}$

0.5 mark

therefore,

$$\gamma_k = \frac{\alpha^k\sigma^2}{1-\alpha^2}$$

1 mark

The autocorrelation function is given by

$\rho_k = \frac{\gamma_k}{\gamma_0}$

$= \alpha^k$

1 mark

(4)

ii) The autocorrelation should decay exponentially as *i* increases. Looking at the table this behaviour occurs after differencing 2 times, suggesting the value of $d = 2$.

1 mark

We know that the ratio of successive *r's* should be α. We can form these ratios as follows:

r_2/r_1	80%
r_3/r_2	82%
r_4/r_3	83%
r_5/r_4	82%
r_6/r_5	81%
r_7/r_6	90%
r_8/r_7	89%
r_9/r_8	79%
r_{10}/r_9	68%

1.5 marks

Average **81.6%**

0.5 mark

Therefore, $\alpha = 0.816$

(3)

iii) For each of these tests, the null hypothesis is

H_0 : the residuals form a white noise process with zero mean.

0.5 mark

a) <u>Ljung Box Test</u>

This test checks for correlation between the residuals. If residuals form a white noise process, then the sample autocorrelations will be small.

1 mark

$$n(n+2)\sum_{k=1}^{m}\frac{r_k^2}{n-k} = 100 * 102 * \left[\frac{0.16^2}{99}+\frac{-0.05^2}{98}+\frac{0.1^2}{97}+\frac{0.12^2}{96}+\frac{-0.02^2}{95}\right]$$

$$= 5.522$$

1 mark

Since the fitted model is ARMA (1,1) i.e. p=1 and q =1, we compare the test statistic with the χ_3^2 distribution.

One tailed test, upper χ_3^2 distribution is 7.815

0.5 mark

Since 5.522 < 7.815, we have insufficient evidence at 5% level to reject the null hypothesis. So we conclude that the residuals are uncorrelated.

0.5 mark

b) Turning Point test

This test checks whether the residuals are patternless.

0.5 mark

E(T) = 2/3 * 98 = $65\frac{1}{3}$

0.5 mark

Var (T) = (16 *100 - 29)/90 = 1571/90 = 17.456

0.5 mark

Using Continuity Correction, and using normal approximation

$$\frac{73\frac{1}{2}-65\frac{1}{3}}{\sqrt{\frac{1571}{90}}} = 1.955$$ 1 mark

Two tailed test (as too many or too few turning points suggests non-random pattern)
Upper and lower 2.5% points of standard normal distribution are ± 1.96

0.5 mark

Since 1.955 < 1.96, there is insufficient evidence to reject the null hypothesis at 5% significance level.

So we conclude that residuals are patternless. 0.5 mark

Alternative Solution:

This test checks whether the residuals are patternless.

0.5 mark

*E(T) = 2/3 * 98 =* $65\frac{1}{3}$

0.5 mark

*Var (T) = (16 *100 - 29)/90 = 1571/90 = 17.456*

0.5 mark

Calculating p value:

$$2\,P\,(T>74) = 2\,P\left(N(0,1) > \frac{73\frac{1}{2} - 65\frac{1}{3}}{\sqrt{\frac{1571}{90}}}\right)$$

$$= 2\,[\,1 - \Phi(1.9547)]$$

$$= 2\,[1-0.9747]$$

$$= 0.0506$$

1.5 mark

There seems insufficient evidence to reject H_0 as the above is more than 5%.

0.5 mark

c) Inspection of the sample auto correlation function

This test checks whether the residuals are uncorrelated.

0.5 mark

ACF of residuals should be 0 for all lags except 0. An approximate 95% Confidence interval for ρ_k, $k \geq 1$ of a white noise process is $\pm\, 1.96/\sqrt{n}$
$= \pm 0.196$

0.5 mark

Since all values lies within the confidence interval, residuals appear to be consistent with white noise process

0.5 mark

All the above tests suggest residuals form a white noise process and hence fitted ARMA (1,1) model is satisfactory.

0.5 mark

(Max 9)

[16]

Q. 22) **i)** The Markov Property means given the present state of the process $X_m = I$, the additional knowledge of the past is irrelevant for the calculation of the probability distribution of the future values of the process. Markov Chain is a process with Markov Property in discrete time with a discrete state space.

The process in the question has discrete state space and is in discrete time and future development of the process can be predicted from the present state alone and thus is a Markov Chain. (1)

ii)

$$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1/4 & 1/4 & 0 & 1/4 & 1/4 & 0 \\ 0 & 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix}$$

(2)

iii) The chain is irreducible, because it is possible to go from any state to any other state.
(1)

iv) It is not aperiodic, because the highest common factor of the return times for the state is greater than 1. For a chain to aperiodic d = 1 which is not the case here.
(1)

v) Writing the equations for π_j:

$\pi_1 = \frac{1}{4}\pi_3$
$\pi_2 = \frac{1}{4}\pi_3$
$\pi_3 = \pi_1 + \pi_2 + \frac{1}{2}\pi_4 + \frac{1}{2}\pi_5$
$\pi_4 = \frac{1}{4}\pi_3 + \frac{1}{2}\pi_6$
$\pi_5 = \frac{1}{4}\pi_3 + \frac{1}{2}\pi_6$
$\pi_6 = \frac{1}{2}\pi_4 + \frac{1}{2}\pi_5$

(2)

From above.
$\pi_1 = \pi_2$
$\pi_4 = \pi_5$

From last equation and above equalities
$\pi_6 = \frac{1}{2}\pi_4 + \frac{1}{2}\pi_4$
$\pi_4 = \pi_5 = \pi_6$
(1)

From 4th equation above
$\pi_4 = \frac{1}{4}\pi_3 + \frac{1}{2}\pi_6$
$\pi_4 = \frac{1}{4}\pi_3 + \frac{1}{2}\pi_4$
$\pi_4 = \frac{1}{2}\pi_3$

Writing the equations in terms of π_3

$\pi = \pi_3\,(\frac{1}{4}, \frac{1}{4}, 1, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$

(1)

And $\sum_j \pi_j = 1$

$\pi = \frac{1}{4}\pi_3\,(1 + 1 + 4 + 2 + 2 + 2)$

$\pi_3 = 1/3$

Therefore stationary distribution is (1/12 , 1/12 , 4/12, 2/12 , 2/12 , 2/12)

(1)
(5)

vi) From the stationary distribution, the probability of being in Room 1 is 1/12.
1 mark

Therefore, expected time to return to Room 1 is 1/ (1/12) i.e. 12.
1 mark
(2)

vii) Let m_k be the expected number of steps until Jerry reached state 5, given current state is is in state k.

Then for k = 1,2,3,4,5,6
$m_1 = 1 + m_3$
$m_2 = 1 + m_3$
$m_3 = 1 + ¼\, m_1 + ¼\, m_2 + ¼\, m_4 + ¼\, m_5$
$m_4 = 1 + ½\, m_6 + ½\, m_3$
$m_5 = 0$
$m_6 = 1 + ½\, m_5 + ½\, m_4$

2 marks

We need to find m_1 that is the expected number of steps until jerry reaches 5, given currently he is in State 1.

Therefore,
$m_6 = 1 + 0 + ½\, m_4$

1 mark

$m_4 = 1 + ½\,(1 + ½\, m_4) + ½\, m_3$

$m_4 = 2 + 2/3\, m_3$

1 mark

$m_3 = 1 + ¼\,(1 + m_3) + ¼\,(1 + m_3) + ¼\,(2 + 2/3\, m_3) + 0$

$m_3 = 6$

1 mark

therefore $m_1 = 7$

1 mark
(6)
[18]

Q. 23) **i)** Let X_i be the amount of ith claim and Y_i be the amount of associated expense. Let N be the total number of claims from the portfolio.

N ~ Poi (1.9)

S = $\sum_{i=1}^{n}(X_i + Y_i)$

Here $(X_i + Y_i)$ are a sequence of IID random variables independent of N

Thus, S has a compound Poisson Distribution

1 mark

Therefore,

E[S] = 1.9 * E$[X_i + Y_i]$

E$[X_i]$ = $\frac{\lambda}{\alpha - 1}$ = 300

0.5 mark

E$[Y_i]$ = 75

0.5 mark

E[S] = 712.5

1 mark

(3)

ii)

Var(S) = E[N] Var[$X_i + Y_i$] + Var[N] (E[$X_i + Y_i$])^2 1 mark

= 1.9 [{E[$X_i + Y_i$]^2}- {E[$X_i + Y_i$]}^2] + 1.9 (E[$X_i + Y_i$])^2

= 1.9 [E[$X_i + Y_i$]^2] 1 mark

(E[$X_i + Y_i$])^2 = E[X_i^2]+ E[Y_i^2]+ 2 E[X_i] E[Y_i]

1 mark

E[X_i^2] = 270000 1 mark

E[Y_i^2] = 17500/3 1 mark

(E[$X_i + Y_i$])^2 = 962500/3 1 mark

Var(S) = 609583.33

(6)

iii) If cost of repairs increases by 20%, the revised mean would be

E[X'$_i$] = 1.2 E[X_i]

$=1.2 \frac{\lambda}{\alpha-1}$

$== \frac{1.2\lambda}{\alpha-1}$

Therefore, new distribution would be Pa (4, 1.2 * 900) i.e. Pa (4, 1080) 1.5 marks

E[X'$_i$] = 1.2 E[X_i]

= 1.2 * 300 = 360 0.5 mark

New E[S] = 1.9* E[$X'_i + Y_i$]

=1.9 * [360+75]

= 826.5 1 mark

(3)

iv) Since we want $E[S]_{2023} = E[S]_{2024}$

712.5 = E[N] (Pr[X>d] E[X-d/X>d]+E[Y_i]) 1 mark

712.5 = 1.9 ((Pr[X>d] E[X-d/X>d]+75)

300 = Pr[X>d] E[X-d/X>d] 1 mark

Now,

Using the Hint given in the question

For Pa (α, λ)

If W = X-d / X> d

Then it is Pa$(\alpha, \lambda + d)$ 1 mark

Therefore,

E[X-d/X>d] = $\frac{\lambda+d}{\alpha-1}$ 0.5 mark

Substituting in the equation above

300 = $\left(\frac{\lambda}{\lambda+d}\right)^{\alpha} * \left(\frac{\lambda+d}{\alpha-1}\right)$ 1 mark

Substituting $\lambda = 1080$ and α = 4

d = 67.7 0.5 mark

(5)

[17]

****