Institute of Actuaries of India

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

December 2022 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.

Solution 1:

i)

A moving average process of order q , denoted MA(q) , is a sequence {Xt} defined by the rule:

$$X_t = \mu + e_t + \beta_1 e_{t-1} + \cdots + \beta_q e_{t-q}$$

et → standard zero-mean white noise process

The moving average model explains the relationship between the Xt as an indirect effect, arising from the fact that the current value of the process results from recent past random error terms as well as the current one.

[2]

ii)

1st Auto correlation is:

$$-0.4 = \frac{ß}{1+ß\verb|^|2}$$

Solving the above equation,

∵ It is non-invertible, ß= -2 (Answer)

[2]

[4 Marks]

Solution 2:

i)

Tx is the (random) future lifetime after age x .

It is, by assumption, a continuous random variable taking values in [0, ω-x]

Its distribution function is $F_x(t) = {}_tq_x$

The curtate future lifetime of a life age x is: Kx = [Tx], where the square brackets denote the integer part. In words, Kx is equal to Tx rounded down to the integer below.

[2]

ii)

Curtate expectation of life $e_x = E(K_x) = \sum_{k=1}^{\infty} {}_kp_x$

Kx is the curtate future life time takes integer values

X is the current age of the insured

kpx indicates the probability of a life aged k surviving until he/she attains the age of x+k

[2]

[4 Marks]

Solution 3:

i)

Chain is irreducible since every state can be reached from every other state. All states will have a period of 1

[1]

ii)

Claim is irreducible and aperiodic.

Long-term bonus =∑ (i = 1 to 4) $d_i * Л_i$

Where d_i = Yearly bonus in state 'i'.

$0.2\,Л_1 + 0 + 0 + 0.4\,Л_4 = Л_1$
$0.8\,Л_1 + 0.7\,Л_2 + 0 + 0 = Л_2$
$0 + 0.3\,Л_2 + 0.4\,Л_{3} + 0 = Л_3$
$0 + 0 + 0.6\,Л_3 + 0.6\,Л_4 = Л_4$
$Л_1 + Л_2 + Л_3 + Л_4 = 1$
Solving: $Л_1 = 1/7$
$Л_2 = 8/21$
$Л_3 = 4/21$
$Л_4 = 2/7$
∴ Long-term yearly bonus = 9x1/7 + 5X8/21+0X4/21+3X2/7
= 4.0476

[4]

[5 Marks]

Solution 4:

Let S_n be the state of subscription started at time t= 0. (1)
T (X1=1 or X2 = 1 | X0 = 0)
= T (X1=1 | X0 = 0) + T (X1=0, X2 = 1 | X0 = 0)
= T (X1=1 | X0 = 0) + T (X1 = 0 | X0 = 0) X T (X1=2 | X0 = 0)
= $P_{0,0} + P_{0,00}$XP1,01
= 0.1+0.86X0.11 (1)

Expected number of cancellations by customers in next 2 years = 500* (0.1+0.86X0.11)
~~ 97 (1)

[3 Marks]

Solution 5:

i)
Conditional probability estimation:

P(Feature | Class) ➔ P(C|X) = 2/3
P(H|X) = 1/3
P(C|Y) = 1/2
P(H|Y) = ½

Prior probability estimation
P(X) = 3/5
P(Y) = 2/5

Posterior probability estimation using Naïve – Bayes:
$P(f_1, f_2, f_3, f_4, f_5,\ldots,f_n, \mid C_j) = P(C_j) \times P(f1 \mid C_j) \times P(f_2 \mid C_j) \times P(f_3 \mid C_j) \times \ldots\ldots\ldots \times P(f_n \mid C_j)$

P(C H | X) = P(X) X P(C|X) X P(H|X)
= 3/5*2/3*1/3 = 2/15
P(C H | Y) = 2/5*1/2*1/2 = 1/10
P(C H | X) > P(C H | Y) ➔ Predicted class is 'X' (Answer)

[6]

ii)
Let 'C' be the event of 1 claim made in 2021.
P(M) = 0.1
P(L) = 0.9

By Naïve-Bayes,
P(M|C) = P(C|M) X P(M) / [P(C|M) X P(M) + P(C|L) * P(L)]

$P(C|M) = e^{-0.5} X 0.5$
$P(C|L) = e^{-0.2} X 0.5$

➔ $P(M|C) = 0.6493$ (Answer)

[4]
[10 Marks]

Solution 6:

i)

	O	I
O	-0.25	0.25
I	0.75	-0.75

[2]

ii)

Transition probabilities do not depend on history. Prior to coming in & Out form. (1)

Therefore, it is Markov. Continuous time operation & state space is discrete. (1)

[2]

iii)

$dP_{OO}(t) /dt = 0.75 X P_{OI}(t) – 0.25 X dP_{OO}(t)$ (1)

$P_{OI}(t) + P_{OO}(t) = 1$
Substituting:
$dP_{OO}(t) /dt + P_{OO}(t) = 0.75$ (1)
$d e^t X P_{OO} (t) /dt = 0.75 X e^t$ (1)
➔ $e^t X P_{OO}(t) = 0.75 X e^t + k$ (1)
$P_{OO}(0) = 1$ ➔ $k = 0.25$ (1)
$P_{OO}(t) = 0.75 + 0.25 X e^{-t}$ (1)

[Max 6]
[10 Marks]

Solution 7:

i)

a)

Total claim 'C' = $C_1 + C_2 + C_3 + C_4 + C_5 + ... + C_n$
C_i : claim for i^{th} policyholder
Assumptions:

- Fixed no. of risks
- Independent claims
- Nol of claims either 0 or 1

I: indicator random variable whether claim paid or not
P(I=0) = 1-q, P(I=1) = q; b (constant) = 10000 or 8500
I ~ Bin(1,q)
E(I) = q (1)
E(X) = qb (1)

[2]

b)

Var (I) = q X (1-q) (1)
Var (X) = var (bI) = b^2 var(I) = $b^2q(1-q)$ (1)

[2]

c)

C is a compound binomial random variable and consider its CGF.
No. of claims N ~ Bin(1,q)
$M_c(t) = M_N(\ln M_Y(t))$, Y = Individual claim amount random variable

$M_N(t) = (1-q) + qe^t$

$M_Y(t) = E(e^{tY}) = e^{tb}$
$M_C(t) = (1\text{-}q) + qe^{\ln} M_Y(t)$
$= (1\text{-}q) + qM_Y(t)$
$= (1\text{-}q) + qe^{tb}$ (1)
CGF ➔ $C_c(t) = \ln M_C(t) = \ln = (1\text{-}q) + qe^{tb}$
Skewness is 3rd derivative of CGF at t = 0
$C'c(t) = qbe^{bt} / [(1\text{-}q) + qe^{bt}]$
$C''c(t) = [(1\text{-}q) + qe^{bt}]\, qb^2e^{bt} - qbe^{bt}\, qbe^{bt} / [(1\text{-}q) + qe^{bt}]^2$
$= q(1\text{-}q)b^2e^{bt} / [(1\text{-}q) + qe^{bt}]^2$
$C'''c(t) = q(1\text{-}q)b^2e^{bt}\, q(1\text{-}q)b^3e^{bt\,-}\, q(1\text{-}q)\, qb^2e^{bt}\, 2((1\text{-}q) + qe^{bt})\, qbe^{bt} / [(1\text{-}q) + qe^{bt}]^4$
$= q(1\text{-}q)(1\text{-}2q)b^3qe^{bt}$ (2)
Skewness = $q(10q)(1\text{-}2q)b^3$ (1)
[4]

ii)
Mean = 10000*880*0.0012 + 8500*765*0.008 (1)
= 62580

Variance = 10000^2 *880*0.0012*0.9988 + 8500^2*765*0.008*0.992 (1)
= 544105920

Skewness = 10000^3 *880*0.0012*0.9988*0.9976 + 8500^3*765*0.008*0.992*0.984
= 1052201949080 (1)

Coefficient of Skewness = Skewness/Variance^3/2 = 0.0829037 (1)
[4]

iii)
S ~ N(62580, 544105920)
P(100000<S<200000) = P[N(0,1) < (200000-62580)/sqrt(544105920)] - P[N(0,1) < (100000-62580)/sqrt(544105920)]
= 1-0.94567
=0.05433 [3]
[15 Marks]

<u>Solution 8:</u>
Using truncated moments of lognormal distribution:
$E[Z] = \int_{900}^{1700}(C-900)f(c)dc + \int_{1700}^{\infty}(800)f(c)dc$ (1)
$= \int_{900}^{1700} cf(c)dc - \int_{900}^{1700}(900)f(c)dc + \int_{1700}^{\infty}(800)f(c)dc$
$= \int_{900}^{1700} cf(c)dc$ – 900P(900<C<1700) + 800 [P(C) > 1700] (1)
Using truncated moments of lognormal distribution:
$E[Z] = e^{\mu + \;\; 2}$ [Φ{(ln 1700 – μ) / σ)– σ - Φ {(ln 900 – μ) / σ)– σ]
- 900 [Φ{(ln 1700 – μ) / σ)– σ - Φ {(ln 900 – μ) / σ)– σ]
+ 900 [1- Φ{(ln 1700 – μ) / σ] (2)

= e^8 [Φ (-1.28) – Φ (-1.599)] – 900[Φ (0.72) – Φ (0.4)] + 800[1- Φ (0.72)] (1)
= e^8 (0.10027-0.05491) – 900(0.76424-0.65542) +800(1-0.76424)
= 135.216-97.938+188.608
= 225.886 (1)
[6 Marks]

Solution 9:

i)

a)

$x = -\ln\{(e^{-\alpha t} - 1) / (e^{-\alpha} - 1)\}$

$x = \Psi(t) = -\ln\{(e^{-\alpha t} - 1) / (e^{-\alpha} - 1)\}$ (1)

➔ $t = \Psi^{-1}(t)$

➔ $x = (-1/\alpha) \ln\{1+(e^{-\alpha} - 1)e^{-x}\}$ (1)

[2]

b)

Copula Function:

$C[u,v] = \Psi^{-1}[-\ln\{(e^{-\alpha u} - 1) / (e^{-\alpha} - 1)\} -\ln\{(e^{-\alpha v} - 1) / (e^{-\alpha} - 1)\}]$ (1)

$= \Psi^{-1}[-\ln\{(e^{-\alpha u} - 1) / (e^{-\alpha} - 1) X (e^{-\alpha v} - 1) / (e^{-\alpha} - 1)\}]$

$= (-1/\alpha) \ln [1+(e^{-\alpha u} - 1) (e^{-\alpha v} - 1) / (e^{-\alpha} - 1)\}$ (1)

[2]

ii)

$\alpha = 2.9$

$P(X<=100, Y<=100) = C[u,v]$

$= (-1/2.9) \ln [1+(e^{-2.9x0.1247} - 1) (e^{-2.90.2748} - 1) / (e^{-2.9} - 1)\}$

$= 0.066917$

[2]

[6 Marks]

Solution 10:

i)

$(1-3\alpha\beta + 3\alpha^2\beta^2 - \alpha^3\beta^3)X_{t} = e_t$

$X_t - 3\alpha X_{t-1} + 3\alpha^2 X_{t-2} - \alpha^3 X_{t-3} = e_t$ (1)

AR (3) process: (1)

$\alpha_1 = 3\alpha$

$\alpha_2 = -3\alpha^2$

$\alpha_3 = \alpha^3$

Yule-Walker for ρ_1: (1)

$\rho_1 = \alpha_1 + \alpha_2 \rho_1 + \alpha_3 \rho_2$

$\rho_1 (1- \alpha_2) = \alpha_1 + \alpha_3 \rho_2$

Yule-Walker for ρ_2: (1)

$\rho_2 = \alpha_1 \rho_1 + \alpha_2 + \alpha_3 \rho_1$

$\rho_2 = (\alpha_1 + \alpha_3)\rho_1 + \alpha_2$

$\rho_1 = (\alpha_1 + \alpha_3 \alpha_2) / (1- \alpha_2 - \alpha_1 \alpha_3 - \alpha_3^2) = (3\alpha - 3\alpha^5) / (1+3\alpha^2-3\alpha^4 - \alpha^6)$ (1)

$\rho_2 = (\alpha_1 + \alpha_3) \rho_1 + \alpha_2 = (3\alpha + \alpha^3) (3\alpha - 3\alpha^5) / (1+3\alpha^2-3\alpha^4 - \alpha^6)$ (1)

[Max 6]

ii)

Null hypothesis (H0): Residuals follow white noise with zero mean (1 Mark)

(a)

Ljung-Box Statistic:

$= 278 X 280 [0.03^2/277 + 0.09^2/276 +(-0.11)^2/275+(-0.02)^2/274+0.02^2/273]$

= 6.19 (1 Mark)

ARMA (1,1)

$\therefore X^2$ @ 5% upper = 7.815 > 6.19

∴ H_0 can't be rejected. (1)

(b) Turning points test
E(T) = 2/3 * 278 = 185
Var(T) = 16*278-29/90 = 491/10 = 49.1 (1)

Test Statistic = (197.5-184) / sqrt(49.1) (1)

= 1.9266 < 1.96

Considering the upper and lower tail – 2-tailed test of +/- 1.96
∴ H_0 can't be rejected. (1)

(c)
95% of confidence interval for ρ_k, k >=1 is +/- 1.96 / sqrt(278) = +/- .11755 (1 Mark)

All of the SACF are within +/- 0.11755 (1)

∴ H_0 can't be rejected, residuals follow a while noise and ARMA (1,1) is fitted well. (1)

[Max 9]
[15 Marks]

Solution 11:

i)

- produce a smooth set of rates that are suitable for a particular purpose
- remove random sampling errors
- use the information available from adjacent ages.
- make them fit for the purpose for which they are intended

[2]

ii)

Age x	Exposed to Risk	Observed deaths	Graduated mortality rates	Expected deaths	z_x	z_x^2
35	5444	80	0.01658	90	-1.0801	1.1666
36	5355	102	0.01787	96	0.6446	0.4156
37	5268	88	0.01894	100	-1.1789	1.3898
38	5197	110	0.01988	103	0.6575	0.4324
39	4978	91	0.02022	101	-0.9624	0.9262
40	4831	106	0.02154	104	0.1902	0.0362
41	4654	123	0.02365	110	1.2327	1.5196
42	4521	107	0.02811	127	-1.7817	3.1744
43	4487	122	0.02957	133	-0.9272	0.8598
44	4321	125	0.03069	133	-0.6610	0.4369
45	4101	140	0.03081	126	1.2142	1.4742
46	4021	145	0.03166	127	1.5683	2.4596
47	3951	140	p			Sum: 14.2912

$X^2_{(13-4)}$ test-Statistic = 21.67 (Upper-tail @ 1%)
Z_x = (O-E)/sqrt(E)

$\sum Z_x^2 = 14.2912 + (140\text{-}3951p)^2 / 3951p$

X^2 test-statistic degree of freedom: 13 – 4
Critical-value is 21.67
∴ $14.2912 + (140\text{-}3951p)^2 / 3951p < 21.67$
➔ $(140\text{-}3951p)^2 / 3951p < 29.153.64p$
➔ $15610401p^2 - 1135433.64p + 19600$
Range of p € (0.028, 0.044)

[8]

iii)

- Cumulative Deviations test
- Signs test
- Grouping of Signs test
- Serial Correlations test

[2]

iv)

- One factor in the age-period-cohort is linearly dependent on the other two
- Intense data demands, for example, for all age groups, more than 100 years of data may be required

[2]
[14 Marks]

Solution 12:

i)
Model for force of elimination is

$$\mu(t,Z) = \mu_0(t)\exp(-0.15Z_1 - 0.08Z_2 - 0.2Z_3 - 0.13Z_4$$

Where t is the time since movie underwent the screening procedure
$\mu_0(t)$ is the baseline hazard at time t

Z1, Z2, Z3 and Z4 are covariates of the model

Z1 = 1 if the movie is a non-English movie, else a zero
Z2 = 1 if the movie is an action movie, else a zero
Z3 = 1 if the movie is a comedy movie, else a ero
Z4 = 1 if the runtime of the movie is high, else a zero

[3]

ii)
The model is a proportional hazards model since the hazards of different movies with covariate vectors are in the same proportion at all times and does not depend on t.

[1]

iii)
The baseline hazard refers to the movies whose Z values are all 0, ie to English drama movies with a low run time (1)

. Group with lowest force of elimination

Here we must make the power in the exponential as negative as possible.
The movies with the lowest force of elimination according to this model are those for which Z1 = 1 , Z2 = 0 , Z3 = 1 and Z4 = 1 , ie Non English comedy movies with a higher run time. (1)

[2]

iv)

According to the model, the force of elimination at time t since the evaluation for an English Action movie with a low run time is $\mu(0,1,0,0) = \mu_0(t)\exp(-0.08)$ (0.5)

Similarly the force of elimination at time t since the evaluation for a non-English Comedy movie with a high run time is $\mu(1,0,1,1) = \mu_0(t)\exp(-0.48)$ (0.5)

Dividing the first of these expressions by the second, we obtain: exp(0.4) = 1.4918
So the force of elimination for English Action movie with a low runtime exceeds or fall short of a non-English comedy movie with a high run time by 49.18%. (1)

[2]

[8 Marks]

****