Institute of Actuaries of India

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

November 2023 Examination

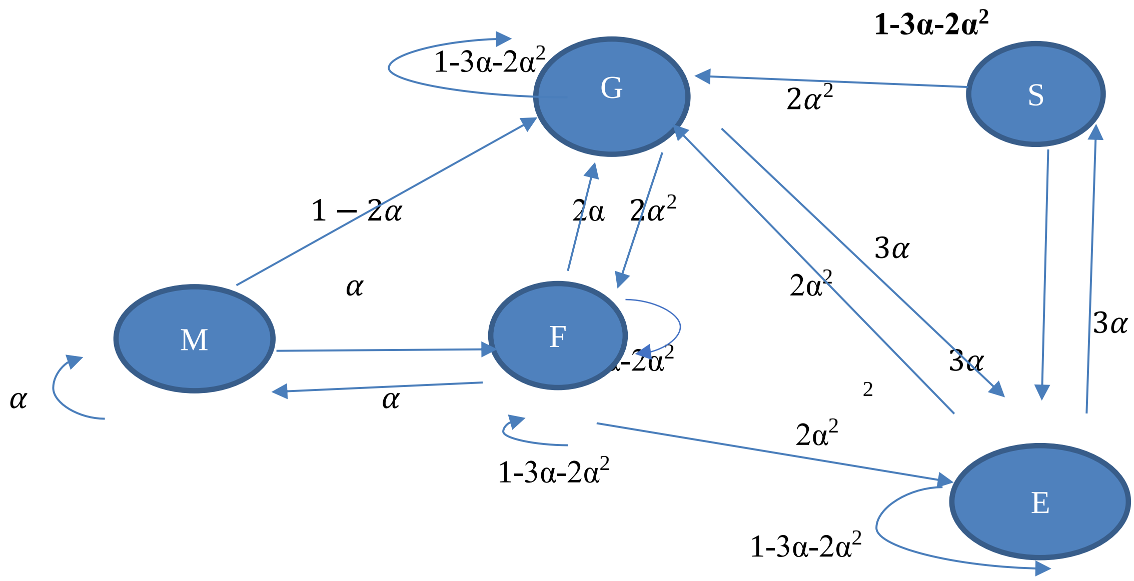
INDICATIVE SOLUTION

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)



[2]

ii) The transition matrix will be valid if the entries in each row add up to 1 (which is true) and each row entry lies in the range [0,1]

So, $0 \le \alpha \le 1$;

$0 \le 1 - 2\alpha \le 1$ => $0 \le 2\alpha \le 1$ => $0 \le \alpha \le \frac{1}{2}$

$0 \le 3\alpha \le 1$ => $0 \le \alpha \le \frac{1}{3}$

$0 \le$ 1-3α-2α^2 ≤ 1

Since $\alpha \ge 0$, it automatically implies that 1-3α-2 α^2 $\le$1

We need to find the values of α for which $0 \le 1-3\alpha-2\alpha^2$

The equation $0 = 1-3\alpha-2\alpha^2$ => $2\alpha^2+3\alpha-1=0$

- ⇨ $\alpha = \frac{(-3\pm\sqrt{17})}{4}$ = 0.280776 or -1.78078
- ⇨ $-1.78078 \le \alpha \le 0.280776$

Putting these together, we can see that all of the conditions

$0 \le \alpha \le 1$

$0 \le \alpha \le \frac{1}{2}$

$0 \le \alpha \le \frac{1}{3}$

$-1.78078 \le \alpha \le .280776$

So we must have

$0 \le \alpha \le 0.280776$ [2]

iii) The chain is irreducible since every state can be eventually reached from every other state.

If $\alpha <$ 0.280766 every state has an arrow to itself, so every state is aperiodic.

When $\alpha =$ 0.280766, none of the states has an arrow to itself. However, return to each of these states is possible in 2 or 3 or 4 steps. So, each state is aperiodic. So the chain is aperiodic. [1]

iv) When $\alpha = 0.1$ the above transition probability matrix becomes

$$\begin{bmatrix} 0.1 & 0.1 & 0.8 & 0 & 0 \\ 0.1 & 0.68 & 0.2 & 0.02 & 0 \\ 0 & 0.02 & 0.68 & 0.3 & 0 \\ 0 & 0 & 0.02 & 0.68 & 0.3 \\ 0 & 0 & 0.02 & 0.3 & 0.68 \end{bmatrix}$$

The stationary distribution is the vector of probabilities $\pi P = \pi$, where P is the transition matrix above. Writing out the equations, we have

$0.1\pi_1+0.1\pi_2 = \pi_1$------------------------(1)

$0.1\pi_1+0.68\pi_2+0.02\pi_3 = \pi_2$------------------(2)

$0.8\pi_1+0.2\pi_2+0.68\pi_3 + 0.02\pi_4 + 0.02\pi_5 = \pi_3$------------------(3)

$0.02\pi_2+0.3\pi_3 + 0.68\pi_4 + 0.3\pi_5 = \pi_4$------------------(4)

$0.3\pi_4 + 0.68\pi_5 = \pi_5$------------------(5)

We can discard equation (3) and replace with

$\pi_1+\pi_2+\pi_3 + \pi_4 + \pi_5 = 1$

From equation (1), $\pi_2 = 9\pi_1$ ----------------------(6)

Substituting in (2) ,

$0.1\pi_1 + 0.02\pi_3 = 0.32 * 9\pi_1$

$0.02\pi_3 = 2.78\pi_1$

$\pi_3 = 139\pi_1$ ------------------------------(7)

First we solve equations (4) and (5) by eliminating π_5.

From (5) we get, $0.3\pi_4 = 0.32\pi_5$

$\pi_5 = 0.3/0.32\pi_4$

Substituting in (4) we get,

$0.02\pi_2+0.3\pi_3 + 0.68\pi_4 + 0.3 * 0.3/0.32\pi_4 = \pi_4$

$0.02\pi_2+0.3\pi_3 = \pi_4 * (1 - 0.68- 0.09/0.32)= 0.03875\pi_4$

Substituting for π_2 *and* π_3 from equations (6) and (7) in this equation we get

$.02*9\pi_1 + 0.3* 139\pi_1 = 0.03875\pi_4$

$\pi_4 = 1080.774\pi_1$

From (5) $0.32\pi_5 = 0.3\pi_4$

$$\pi_5 = \frac{0.3}{0.32} * 1080.774\pi_1 = 1013.226\pi_1$$

Using the relationship $\pi_1+\pi_2+\pi_3 + \pi_4 + \pi_5 = 1$ we get

$(1+9+139+1080.774+1013.226)\pi_1 = 1$

$\pi_1 = 0.000446$

$\pi_5 = 1013.226 * 0.000446 = 0.451728$ [4]

[9 Marks]

Solution 2:

i) The fundamental Naïve Bayes assumption is that each feature makes an

- **independent and**
- **equal**

contribution to the outcome. [1]

ii) Using the definition of conditional probabilities, i.e., P(X|Y) = P (X, Y)/P(Y)

P(Claim = Yes | X1 = x1, X2 = x2, ….., Xn = xn) = P(Claim = Yes, X1 = x1, X2 = x2, ….., Xn = xn) / P(X1 = x1, X2 = x2, ….., Xn = xn)

Using the same definition in reverse. i.e., P(X,Y) = P(X|Y) P(Y). we can write the numerator as :
P(Claim = Yes, X1 = x1, X2 = x2, ….., Xn = xn) = P (X1 = x1, X2 = x2, ….., Xn = xn | Claim = Yes) P (Claim = Yes)

Under the naïve bayes approach, we assume that Xj are conditionally independent given the claim outcome. This means we can write the first part of above equation as:
P(X1 = x1, X2 = x2, ….., Xn = xn | Claim = Yes) = P(X1= x1 | Claim = Yes) * P(X2= x2 | Claim = Yes) * ….. * P(Xn= xn | Claim = Yes)

$$= \prod_{j=1}^{n} P(Xj = xj \mid Claim = Yes)$$

Finally, treating the denominator as a constant of proportionality in original equation, we can write the desired probability as required:
P (Claim = Yes | $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$) ∝ P(Claim=Yes) * $\prod_{j=1}^{n} P(X_j = x_j \mid Claim = Yes)$ [3]

iii) The conditional Probability $P(X_i \mid Y_i)$ for each x_i in X and y_j in Y can be written as:

Location/Zone

	Yes	No	P(Yes)	P(No)
North	2	1	2/5	1/7
East	0	3	0/5	3/7
West	2	1	2/5	1/7
South	1	2	1/5	2/7
Total	5	7	100%	100%

Policy Type

	Yes	No	P(Yes)	P(No)
Individual	3	3	3/5	3/7
Floater	2	4	2/5	4/7
Total	5	7	100%	100%

Business Type

	Yes	No	P(Yes)	P(No)
Fresh	0	5	0/5	5/7
Port	5	2	5/5	2/7
Total	5	7	100%	100%

Age Group

	Yes	No	P(Yes)	P(No)
<45	0	4	0/5	4/7
45-55	1	3	1/5	3/7
>55	4	0	4/5	0/7
Total	5	7	100%	100%

Claim

Claim		P(Yes)/P(No)
Yes	5	5/12
No	7	7/12
Total	12	100%

[7]

iv) Let Policy A = (North Zone, Port, Individual, 45-55)

P(Claim = Yes | Policy A) = P(North Zone | Claim = Yes) P(Port | Claim = Yes) P(Individual | Claim = Yes) P(45-55 | Claim = Yes) / P(Policy A)

P(Claim = No | Policy A) = P(North Zone | Claim = No) P(Port | Claim = No) P(Individual | Claim = No) P(45-55 | Claim = No) / P(Policy A)

Since P(Policy A) is common in both probabilities, we can ignore P(Policy A) and find proportional probabilities as:

P(Claim = Yes | Policy A) $\propto$ 2/5 * 5/5 * 3/5 * 1/5 $\approx$ 0.048

P(Claim = No | Policy A) $\propto$ 1/7 * 2/7 * 3/7 * 3/7 $\approx$ 0.0075

Now, since

P(Claim = Yes | Policy A) + P(Claim = No | Policy A) = 1

These numbers can be converted into a probability by making the sum equal to 1 (normalization):

P(Claim = Yes | Policy A) = $\frac{0.048}{0.048+0.0075}$ = 0.865

P(Claim = No | Policy A) = $\frac{0.0075}{0.048+0.0075}$ = 0.135

Since, P(Claim = Yes | Policy A) > P(Claim = No | Policy A)

So, prediction that Claim will be there on the given policy is **Yes.** [4]

[15 Marks]

<u>Solution 3</u>

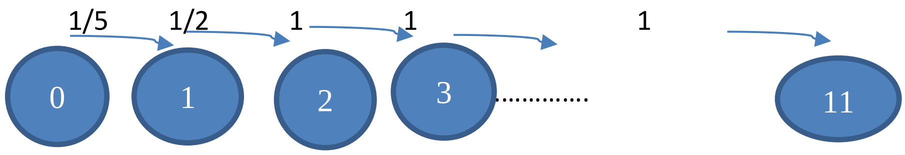
i) This is a 12-state Markov jump process. The states are (1) 0 point, (2) 1 points, (3) 2 points,........(12) 11 points.

The Markov assumption is that the probability of jumping to any particular state depends only on knowing the current state that is occupied and the transition rates between states are constant over time. [1]

ii) Let N(t) be the number of points scored up to time t. Then the state space N(t) is the set {0,1,2,3......................11}

Since the game starts at point 0 we have N(0)=0.

The transition diagram is



The required probability is
P(N(6)> 2) = 1- P(N(6) =0)- P(N(6) =1)- P(N(6) =2)

First, we have:

P(N(6) =0) = $e^{-6*1/5}$ = 0.301194,
<u>Calculating P(N(6) =1)</u>

Forward equation for P(N(t) =1) = $P_{01}(t)$ is

$\frac{d}{dt}P_{01}(t) = \frac{1}{5}P_{00}(t) - \frac{1}{2}P_{01}(t)$

Solving by integrating factor method, by first writing in the form:

$$\frac{d}{dt}P_{01}(t) + \frac{1}{2}P_{01}(t) = \frac{1}{5}P_{00}(t)$$

We have $P_{00}(t) =$ P(N(t) =0) $e^{-t/5}$

Substituting for $P_{00}(t)$, the differential equation becomes

$\frac{d}{dt}P_{01}(t) + \frac{1}{2}P_{01}(t) = \frac{1}{5}e^{-t/5}$

Multiplying by the integrating factor $e^{t/2}$ gives:

$e^{t/2} * \frac{d}{dt}P_{01}(t) + e^{t/2} * \frac{1}{2}P_{01}(t) = \frac{1}{5}e^{3t/10}$

LHS can be written as:

$\frac{d}{dt}(e^{t/2} * P_{01}(t))$

Integrating, we get,

$e^{t/2} * P_{01}(t) = 1/5 * 10/3 * e^{3t/10} + C$

$e^{t/2} * P_{01}(t) = 2/3 * e^{3t/10} + C$

When t= 0: $P_{01}(0) = 0$

Substituting t=0,

0 = 2/3 + C

C= -2/3

$e^{t/2} * P_{01}(t) = 2/3 * e^{3t/10} - 2/3$

$P_{01}(t) = 2/3 * e^{-t/5} - 2/3\, e^{-t/2}$

P(N(6) =1) = $P_{01}(6) = 2/3 * (e^{-6/5} - e^{-6/2}) = 0.167605$

Calculating P(N(6) =2)

Forward equation for P(N(t) =2) = $P_{02}(t)$ is

$\frac{d}{dt}P_{02}(t) = P_{00}(t)\mu_{02} + P_{01}(t)\mu_{12} + P_{02}(t)\mu_{22}$

Using the transition rates $\mu_{02} = 0$, $\mu_{12} = \frac{1}{2}, \mu_{22} = -1$

$\frac{d}{dt}P_{02}(t) = 1/2 * P_{01}(t) - P_{02}(t)$

$$\frac{d}{dt}P_{02}(t) + P_{02}(t) = \frac{1}{2}P_{01}(t)$$

Substituting for $P_{01}(t)$,

$\frac{d}{dt}P_{02}(t) + P_{02}(t) = \frac{1}{2}$ * {2/3 * $e^{-t/5}$ -2/3 $e^{-t/2}$ }

= 1/3* { $e^{-t/5}$ - $e^{-t/2}$ }

Multiplying by the integrating factor e^{t}, we get
$\frac{d}{dt}\{P_{02}(t)e^{t}\} =$ 1/3* { $e^{4t/5}$ - $e^{t/2}$ }

Integrating,

$P_{02}(t)e^{t} =$ 1/3* {5/4* $e^{4t/5}$ - 2*$e^{t/2}$} + C

Since , $P_{02}(0) = 0$

$0 = \frac{1}{3} * \left\{\frac{5}{4} - 2\right\}$ + C
0= -1/4 + C
C= ¼

$P_{02}(t)e^{t} =$ 1/3* {5/4* $e^{4t/5}$ - 2*$e^{t/2}$} + ¼

$P_{02}(t) =$ 1/3* {5/4* $e^{-t/5}$ - 2*$e^{-t/2}$} + ¼ * e^{-t}

P(N(6) =2) = $P_{02}(6) =$ 1/3* {5/4* $e^{-6/5}$ - 2*e^{-3}} + ¼ * e^{-6}

= 0.092926

Probability of scoring more than two points in next 6 minutes

= 1- P(N(6) =0)- P(N(6) =1)- P(N(6) =2)

= 1-0.301194 -0.167605-0.092926

=0.438275 [6]

iii) For i= 0,1,2,……10 the expected holding time in state i is $1/\lambda_i$ where λ_i is the total force out of State i.

State i must be followed by state i+1 . So the expected time to complete the game is

$$\sum_{i=0}^{10} 1/\lambda_i = 5+ 2+ (1*9) = 16 \text{ minutes}$$

[1]

[8 Marks]

Solution 4:

i) Claims of type Silver occur according to a Poisson process with a mean of 50%* 4 = 2 per day.

So the waiting time until the first claim of type Silver has an exp(2) distribution and the expected waiting time is 1/2 days. [1]

ii) Let N(t) denote the number of claims during the interval [0,t]. Then:

P[N(2)>=9| N(1)=7]
= P[N(2)- N(1)>=2| N(1)- N(0)=7]
= P[N(2)- N(1)>=2]

Since N(0)=0 and assuming number of claims in non-overlapping time intervals are independent.

N(2)-N(1) ~ *Poisson*(4), so:

P(N(2)>=9| N(1)=7) =P(*Poisson*(4)>=2)
=1- P(*Poisson*(4)<=1)
= 1- exp(-4) (1+4/1!)
=1- exp(-4)*5
= 1- 0.018316*5
= 0.908422 [2]

iii) Let $N_G(t)$ denote the number of claims of Gold Type in the interval [0,t].

P(N_G(1) >=5, N_G(2))>=7)

If we have 7 or more claims during the first day, then the second condition is automatically satisfied.

If we have exactly 6 claims on the first day, then we need at least 1 claim on the second day.

If we have exactly 5 claims on the first day, then we need at least 2 claims on the second day.

So the required probability is:

P(N_G(1) >=7) + P(N_G(1) =6, N_G(2)- N_G(1)>=1)
+ P(N_G(1) =5, N_G(2)- N_G(1)>=2) ---------------------------------(1)

Now N_G(1) and N_G(2)- N_G(1) are both independent Poisson with mean 0.3*4 = 1.2 .

P(N_G(1) >=7) = 1- (N_G(1) <=6) = 1- exp(-1.2)* (1+ 1.2/1!+1.2^2/2! + 1.2^3/3!+ 1.2^4/4! + 1.2^5/5!+ 1.2^6/6!)
=1- .999749
=0.000251

P(N_G(1) =6, N_G(2)- N_G(1)>=1) = P(N_G(1) =6)* P(N_G(2)- N_G(1))>=1)

= P(N_G(1) =6)* { 1- P(N_G(2)- N_G(1))= 0)
= {exp(-1.2)* 1.2^6/6! } * {1- exp(-1.2)}
= 0.000873

P(N_G(1) =5, N_G(2)- N_G(1)>=2) = P(N_G(1) =5)* { 1- P(N_G(2)- N_G(1)) <=1}
= {exp(-1.2)* 1.2^5/5! } *
{1- exp(-1.2)(1+1.2)}
=0.006246*0.337373
=0.002107

Substituting in (1),

The required probability = 0.000251+ 0.000873+0.002107
= 0.003231 [4]

[7 Marks]

Solution 5:

i) A rate interval is a period of one year during which a life's recorded age remains the same.

Definition of x	Rate interval	$\hat{\mu}$ estimates	$\hat{q}$ estimates
Age last birthday	[x,x+1]	$\mu_{x+1/2}$	q_x
Age nearest birthday	[x-1/2,x+1/2]	μ_x	qx-1/2
Age next birthday	[x-1,x]	$\mu_{x-1/2}$	q_{x-1}

[3]

ii)
a) Graphical as no possibility of standard table
b) Parametric formula as large experience is available
c) With reference to standard table as the standard table includes experience for females
d) Spline function to be used as it special case of parametric formula and there are different knots for this age group with infant mortality, accident hump, etc. [2]

iii)

a)
μ = d/v

we observe 2 deaths i.e. d=2

'v' is for all 10 lives and computation of v on annual basis
Exposure period in age 50 to 51 under investigate to be determined

Life (i)	Date of birth	Date of entry into observation	Date of exit from observation	v_i(years)
1	1st May 1972	1st April 2022	1st April 2023	11/12
2	1st July 1972	1st April 2022	1st April 2023	9/12
3	1st October 1972	1st July 2022	1st April 2023	6/12
4	1st January 1972	1st April 2022	1st July 2022	3/12
5	1st March 1973	1st April 2022	1st April 2023	1/12
6	1st December 1971	1st July 2022	1st October 2022	3/12
7	1st August 1972	1st April 2022	1st April 2023	8/12
8	1st August 1972	1st April 2022	1st September 2022	1/12
9	1st November 1972	1st October 2022	1st April 2023	5/12
10	1st June 1972	1st October 2022	1st April 2023	6/12

V = $\sum v_i$ = 4.4167
μ = 2/ 4.4167 = 0.45283
q=1-exp^(-μ)

q = 1-0.635826
q= 0.364174 [4]

b)

- Poisson model is not an exact model, it allows for non-zero probability of more than n deaths in an sample size of n.
- The variance of MLE of two-state model is only available asymptotically whereas for Poisson model is available in term of true μ
- Two-state model extents to processes with increments, whereas Poisson model does not.
- The Poisson model is a less satisfactory approximation to the multiple state model when transition rates are high. [2]

iv) Null Hypothesis: The graduated rates are the true underlying mortality rates for the population
Standardised deviation at each age is computed below using formula

$Z_x = \frac{\theta_x - E_x\widehat{q_x}}{\sqrt{E_x\widehat{q_x}}}$:

To determine initial exposure and observed deaths adjustment is required at each age for last birthday.

Two approaches:

1st approach:

Exposure = (0.5*Year1 beginning + Year2 beginning + 0.5*Year3 beginning)/2
No of death = avg of 2 years

OR

2nd approach:

Exposure = (0.5*Year1 beginning + Year2 beginning + 0.5*Year3 beginning)
No of death = sum of 2 years

1st approach:

The corresponding goodness of fit to be performed.

Age	Initial Exposure to Risk E_x	Observed Deaths θ_x	$\hat{q}_x$	Expected Deaths $E_x*\hat{q}_x$	Z_x	Z_x^2
37	1047.25	7	0.0072	7.54	-0.19673	0.03870
38	1127.25	7	0.0078	8.79	-0.60452	0.36545
39	1056.50	8	0.0081	8.56	-0.19063	0.03634
40	1230.00	10	0.0104	12.79	-0.78063	0.60939
41	1154.25	13	0.0112	12.93	0.02014	0.00041
42	1362.75	13	0.0114	15.54	-0.64325	0.41377
43	1231.25	15	0.012	14.78	0.05854	0.00343

OR

2nd approach:

[4]

Age	Initial Exposure to Risk E_x	Observed Deaths θ_x	$\hat{q}_x$	Expected Deaths $E_x*\hat{q}_x$	Z_x	Z_x^2
37	2094.50	14	0.0072	15.08	-0.27821	0.07740
38	2254.50	14	0.0078	17.59	-0.85493	0.73090
39	2113.00	16	0.0081	17.12	-0.26959	0.07268
40	2460.00	20	0.0104	25.58	-1.10398	1.21877
41	2308.50	26	0.0112	25.86	0.02848	0.00081
42	2725.50	26	0.0114	31.07	-0.90969	0.82753
43	2462.50	30	0.012	29.55	0.08278	0.00685

1st approach:
Test statistics = $\sum Z_x^2$ = 1.47

OR

2nd approach:
$\sum Z_x^2$ = *2.94*

Degrees of freedom = 7-1 = 6

At 5% level, the critical value = 12.59
1.47/2.94<12.59 and no evidence to reject null hypothesis

[15 Marks]

Solution 6:

i) Let X denote the individual claim amount random variable. Then X ~ log N(10, 0.81).
The claim amount paid by the reinsurer exceeds Rs.5000 if X > 20,000

Using the CDF of the lognormal distribution we have:

P(X> 20,000) = $\int_{20000}^{\infty} f(x)dx$, where f(x) is p.d.f of log N(10, 0.81)

= $\Phi(U_0) - \Phi(L_0)$

$U_0 = \infty$, $L_0 =$ (ln(20000)-10)/0.9 = -0.10724

So, the required probability is equal to

= $\Phi(\infty) - \Phi(-0.10724)$ = 1- 0.457299 =0.542701 [2]

ii) In the previous arrangement, i.e. with the old retention limit (Rs.15,000), the probability the a claim involves the reinsurer

P(X> 15000) = $\int_{15000}^{\infty} f(x)dx$ = $\Phi(\infty) - \Phi(L'_0)$

$L'_0 =$ (ln(15000)-10)/.9 = -0.42688

Probability that a claim involves the reinsurer (with the old retention limit)

= 1- $\Phi(-0.42688)$
= 1-0.334732 = 0.665268

Next year claims are expected to increase by 10%.

Let Z be the new retention limit. The claim amount random variable for next year is 1.1X
So, the probability that a claim involves the reinsurer is

P(1.1X > Z) = P(X> Z/1.1) = $\int_{\frac{Z}{1.1}}^{\infty} f(x)dx$
= $\Phi(\infty) - \Phi(L^*_0)$ =1- $\Phi(L^*_0)$,

where $L^*_0 =$ (ln(Z/1.1) -10)/ 0.9

Any one of the two possible interpretations of new probability stated below:

1st: 0.665268+.05 = 0.715268
2nd: 0.665268*1.05 = 0.698531

1st approach:
The choice of Z is such that probability that a claim involves the reinsurer (with the new retention limit) is increased by 5%.

So the new probability becomes 0.665268+.05 = 0.715268

So, 1- $\Phi(L^*_0)$ = 0.715268

$\Phi(L^*_0)$ = 1- 0.715268 = 0.284732

L^*_0 = -0.56884

L^*_0 = (ln(Z/1.1) -10)/ 0.9

ln(Z/1.1) = 10+0.9* (-0.56884) = 9.48804

Z/1.1 = exp(9.48804) = 13200.95

So, the new retention limit is equal to Z =13200.95*1.1 =14521.04

Or

2nd approach

The choice of Z is such that probability that a claim involves the reinsurer (with the new retention limit) is increased by 5%.

So the new probability becomes 0.665268*1.05 = 0.698531

So, 1- Φ(L*0) = 0.698531

Φ(L*0) = 1- 0.698531= 0.301469

L*0 = -0.52018

L*0 = (ln(Z/1.1) -10)/ 0.9

ln(Z/1.1) = 10+0.9* (-0.52018) = 9.531837

Z/1.1 = exp(9.48804) = 13791.9

So, the new retention limit is equal to Z =13200.95*1.1 =15171.09

[4]

iii) Let Z' denote the reinsurer's claim payment random variable next year. Then:

Z' = 0 if 1.1X<= 14521.04 (ie X<= 14521.04/1.1)
= 1.1X- 14521.04 if 1.1X > 14521.04 (ie. X> 14521.04/1.1)

As 14521.04/1.1= 13200.95,

$$E(Z') = \int_{13200.95}^{\infty} (1.1X - 14521.04) f(x)dx$$

$$= 1.1 \int_{13200.95}^{\infty} Xf(x)dx - 14521.04 \int_{13200.95}^{\infty} f(x)dx \text{----------(1)}$$

$$\int_{13200.95}^{\infty} Xf(x)dx = \exp(10 + 0.5 * 0.81) * (\Phi(\infty) - \Phi(L1))$$

L1 = (ln(13200.95) -10)/0.9- 0.9
= -1.46884

$$\int_{13200.95}^{\infty} Xf(x)dx = \exp(10 + 0.5 * 0.81) * (1 - \Phi(-1.46884))$$

= 33024.64*(1- 0.070938)
= 30681.65

$$\int_{13200.95}^{\infty} f(x)dx = \Phi(\infty) - \Phi(L0)$$

L0 = (ln(13200.95) -10)/.9= -0.56884

$\int_{13200.95}^{\infty} f(x)dx = 1 - \Phi(-0.56884)$ = 1-0.284732 = 0.715268

Substituting in (1),

Reinsurer's expected claim payment next year is equal to

= 1.1* 30681.65 – 14521.04* 0.715268
= Rs. 23363.38 [4]

[10 Marks]

Solution 7:

i) µ = 5954.39/100= 59.54

Using the known expression of the auto covariance function for AR(1) process
$\rho_k = \alpha^k$
$\widehat{\alpha^k} = \rho_1$ = 3628.34/3832.26 = 0.9468

To estimate variance σ
Taking variance on both sides of X_t - µ= α (X_{t-1} - µ) + e_t
The fact is Var(X_t - µ)= Var(X_{t-1} - µ)= γ_0
Thus, $\gamma_0=\alpha^2\gamma_0+\sigma^2$

$\hat{\sigma}^2 = \widehat{\gamma_0}\,(1 - \widehat{\alpha^2})$
=(3832.26/100)*(1-0.9468^2)
=3.9699
$\hat{\sigma}$=1.99 [3]

ii) Null Hypothesis: Estimate of errors e_t follows white noise process
We will reject the null hypothesis if the 48 falls outside the Confidence interval of 95%.

Mean = 2(N-2)/3 = 2(100-2)/3 = 65.33
Variance = (16N-29)/90=17.45 and thus S.D. =4.177985

The CI is (65.33± 1.96*4.177985) and thus CI is (57.1, 73.5)
Since 48 is outside confidence interval of 95%, e_t does not follow white noise process [3]

[6 Marks]

Solution 8:

i) Since γ>0, this is a Fréchet-type GEV distribution.

The key characteristic is they are 'heavy tail' whose higher moments can be infinite. [1]

ii) a = 2 * 50$^{1/5}$ -2=2.37345 → location parameter
b = 2 * 50$^{1/5}$/5 = 0.87469 →scale parameter
c=1/5=0.2 →shape parameter

P(X_{M}>4) ≈ 1 – F(X_M=4) = 1- exp{-[1+c*($\frac{4-a}{b}$)]^$(-1/c)$}

P(X_{M}>4) ≈ 1 – exp{-[1+0.2*(4-2.37345)/0.87469]^(1/0.2)} = 1-0.814027 = 0.18597

P(X_{M}>4) = 1- F_x(4) = 1-[1-(λ/(λ+4))$^\alpha$]^n = 1-[1-(2/(2+4))^5]^{50}
=1-[1-(1/3)^5]^{50} = 0.18632 [2]

The probabilities are very close, suggesting the GEV is reasonably approximate for this block size.

iii) The key advantage is the generalised Pareto distribution makes use of all the data in the tail whereas generalised extreme value distribution might exclude some data values. [1]

[4 Marks]

Solution 9:

i) The characteristic polynomial is given as below

$$1-\frac{39}{28}t+\frac{3}{7}t^2-\frac{1}{28}t^3$$

i.e. $28-39t+12t^2-t^3 = 0$
i.e. $t^3-12t^2+39t-28=0$
one root is 1
$(t-1)*(a+bt+ct^2)=0$

i.e.
$a+bt+ct^2-at-bt^2-ct^3 = 0$
$ct^3 + bt^2+ at-a-bt-ct^2=0$
$ct^3 + (b-c)t^2+ (a-b)t-a=0$
c=1
b-c=-12
b=-11
a-b=39
a=28
$(t-1)*(28-11t+t^2)==0$
$(t-1)*(t-4)*(t-7)=0$
Other 2 roots are 4 and 7

All the roots are greater than 1 one of the root is 1 and the process is not stationary [3]

ii) $Cov(X_t, Z_t) = 1$

and,

$Cov(X_t, Z_{t-1}) = 9/15 – 2/7 = 33/105$

We need to generate three distinct equations linking $\gamma_0, \gamma_1, \gamma_2$.

$\gamma_0 = Cov(X_t, X_t) = Cov(1 + 9/15\ X_{t-1} – 2/15\ X_{t-2} + Z_t - 2/7\ Z_{t-1}, X_t)$
$= 9/15\ \gamma_1 – 2/15\ \gamma_2 + 1 – 2/7 \times 33/105$
$= 9/15\ \gamma_1 – 2/15\ \gamma_2 + 669/735$

$\gamma_1 = Cov(X_t, X_{t-1}) = Cov(1 + 9/15\ X_{t-1} – 2/15\ X_{t-2} + Z_t - 2/7\ Z_{t-1}, X_{t-1})$
$= 9/15\ \gamma_0 – 2/15\ \gamma_1 – 2/7$

$\gamma_2 = Cov(X_t, X_{t-2}) = Cov(1 + 9/15\ X_{t-1} – 2/15\ X_{t-2} + Z_t - 2/7\ Z_{t-1}, X_{t-2})$
$= 9/15\ \gamma_1 – 2/15\ \gamma_0$

Solving these equations by substitution,

$\gamma_0 = 9/15\ \gamma_1 – 2/15(9/15\ \gamma_1 – 2/15\ \gamma_0) + 669/735$
$= 9/15\ \gamma_1 – 18/225\ \gamma_1 + 4/225\ \gamma_0 + 669/735$

$221/225\ \gamma_0 = 117/225\ \gamma_1 + 669/735$

$\gamma_0 = 117/221\ \gamma_1 + 10035/10829$

Now,

$\boldsymbol{\gamma}_1$ = 9/15 (117/221 $\boldsymbol{\gamma}_1$ + 10035/10829)– 2/15 $\boldsymbol{\gamma}_1$ – 2/7
= 351/1105 $\boldsymbol{\gamma}_1$ + 6021/10829 – 2/15 $\boldsymbol{\gamma}_1$ – 2/7
= 611/3315 $\boldsymbol{\gamma}_1$ + 2927/10829

2704/3315 $\boldsymbol{\gamma}_1$ = 2927/10829

$\boldsymbol{\gamma}_1$ = 0.3314
$\boldsymbol{\gamma}_0$ = 1.1021
$\boldsymbol{\gamma}_2$ = 0.0519

Finally,

$\boldsymbol{\rho}_0$ = 1

$\boldsymbol{\rho}_1 = \boldsymbol{\gamma}_1 / \boldsymbol{\gamma}_0$

= 0.3007

$\boldsymbol{\rho}_2 = \boldsymbol{\gamma}_2 / \boldsymbol{\gamma}_0$

= 0.0471 [5]

iii)

a) Since e_t are independent from $X_t, X_{t-1},$ .... and E(e_t) = 0 we have

E(X_t) = E($\mu + e_t\sqrt{\alpha_0 + \alpha_1(X_{t-1} - \mu)^2}$)
= μ + E($e_t\sqrt{\alpha_0 + \alpha_1(X_{t-1} - \mu)^2}$)
= μ + E(e_t) E($\sqrt{\alpha_0 + \alpha_1(X_{t-1} - \mu)^2}$) … Since e_t and X_{t-1} are independent
= μ + 0 X E($\sqrt{\alpha_0 + \alpha_1(X_{t-1} - \mu)^2}$)
= μ

Cov(X_t, X_{t-s}) = E($X_t X_{t-s}$) – E(X_t)E(X_{t-s})
= E(($\mu + e_t\sqrt{\alpha_0 + \alpha_1(X_{t-1} - \mu)^2}$)($\mu + e_{t-s}\sqrt{\alpha_0 + \alpha_1(X_{t-s-1} - \mu)^2}$)) - μ^2
= E(μ^2 + $\mu e_t\sqrt{\alpha_0 + \alpha_1(X_{t-1} - \mu)^2}$) + $\mu e_{t-s}\sqrt{\alpha_0 + \alpha_1(X_{t-s-1} - \mu)^2}$) + $e_t e_{t-s}\sqrt{\alpha_0 + \alpha_1(X_{t-s-1} - \mu)^2}\sqrt{\alpha_0 + \alpha_1(X_{t-1} - \mu)^2}$) - μ^2

= E(μ^2) + μE(e_t) E($\sqrt{\alpha_0 + \alpha_1(X_{t-1} - \mu)^2}$)+ μE(e_{t-s}) E($\sqrt{\alpha_0 + \alpha_1(X_{t-s-1} - \mu)^2}$) + E($e_t e_{t-s}\sqrt{\alpha_0 + \alpha_1(X_{t-s-1} - \mu)^2}\sqrt{\alpha_0 + \alpha_1(X_{t-1} - \mu)^2}$) - μ^2

= μ^2 + μ X 0 X E($\sqrt{\alpha_0 + \alpha_1(X_{t-1} - \mu)^2}$)+ μ X 0 X E($\sqrt{\alpha_0 + \alpha_1(X_{t-s-1} - \mu)^2}$) + E($e_t$) E($e_{t-s}\sqrt{\alpha_0 + \alpha_1(X_{t-s-1} - \mu)^2}\sqrt{\alpha_0 + \alpha_1(X_{t-1} - \mu)^2}$) - μ^2

= μ^2 + 0 + 0 + 0 X E($e_{t-s}\sqrt{\alpha_0 + \alpha_1(X_{t-s-1} - \mu)^2}\sqrt{\alpha_0 + \alpha_1(X_{t-1} - \mu)^2}$) - μ^2
= μ^2 + 0 + 0 + 0 - μ^2
= 0

Thus, X_t, X_{t-s} are uncorrelated.

[4]

b) The conditional variant of X_t/X_{t-1} is

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

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

So, the values of X_{t-1} are affecting the variance of X_t

If we apply this recursively, it can be seen that the variance of X_t will be affected by the value of X_{t-s}.

So, X_t and X_{t-s} are not independent.

[3]

iv) It can be used to model financial time series.
For e.g. - If Z_t is the price of an asset at the end of the t th trading day, it is found that the ARCH model can be used to model $X_t = \ln(Z_t/Z_{t-1})$, interpreted as the daily return on day t. [1]

[16 Marks]

Solution 10:

i) Random censoring as the time of test termination is not known in advance
Right censoring as the test is cut short after discharge of 8 batteries
Type II censoring as the predetermined count of failures lead to termination of test [2]

ii)

j	t_j	d_j	c_j	n_j	$\lambda_j = d_j/n_j$	$1-\lambda_j = (n_j-d_j)/n_j$	$\prod_{t_j \le t}^{n} 1-\lambda_j$
1	22	2	0	12	2/12	10/12	10/12
2	29	3	0	10	3/10	7/10	7/12
3	32	2	1	6	2/6	4/6	7/18
4	34	1	0	4	¼	¾	7/24

The Kaplan-Meier estimate is then
$\hat{S}(t) = \prod_{t_j \le t}^{n} 1-\lambda_j =$

t	$\hat{S}(t)$
0 ≤ t < 22	1
22 ≤ t < 29	10/12
29 ≤ t < 32	7/12
32 ≤ t <34	7/18
34 ≤ t	7/24

[3]

iii) Assuming the MLE asymptotically follows normal distribution, 95% confidence interval for $\tilde{S}(33)$
$\tilde{S}(33) \pm 1.96\sqrt{\text{var}[\tilde{S}(33)]}$

Greenwood's formula:
$\text{var}[\tilde{F}(t)] \approx (1-\widehat{F(t)})^2 \sum_{t_j \le t} \frac{d_j}{n_j(n_j-d_j)}$

$\text{var}[\tilde{S}(33)] = \text{var}[1-\tilde{F}(33)] = \text{var}[\tilde{F}(33)]$

Using Greenwood's formula,
$\text{var}[\tilde{S}(33)] \approx (1-\widehat{F(33)})^2 \sum_{t_j \le 33} \frac{d_j}{n_j(n_j-d_j)}$
$\text{var}[\tilde{S}(33)] \approx (7/18)^2 \sum_{t_j \le 33} \frac{d_j}{n_j(n_j-d_j)}$

j	t_j	d_j	n_j	$\frac{d_j}{n_j(n_j-d_j)}$
1	22	2	12	2/(12*10)
2	29	3	10	3/(10*7)
3	32	2	6	2/(6*4)
$\sum_{t_j \le 33} \frac{d_j}{n_j(n_j-d_j)}$				0.142857

var$[\tilde{S}(33)] \approx 0.021605$

Thus, $\tilde{S}(33) \pm 1.96\sqrt{\text{var}[\tilde{S}(33)]}$
$(7/18) \pm 1.96\sqrt{0.021605}$
CI is (0.10,0.68)

[4]

iv)
1. Battery life is independent and assume to follow same model for failure.
2. We do not have information of remaining batteries and the censoring is non-informative.
3. The battery exploded after 30th month assumed to be censored before 32nd month observations are captured.

[1]

[10 Marks]

****