Institute of Actuaries of India

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

November 2019 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)

Parametric: Easy to interpret, more efficient and accurate when the survival times follow a particular distribution. When the distribution assumption is violated, it may be inconsistent and can give sub-optimal results. [1]

Non Parametric: More efficient when no suitable theoretical distribution known. Difficult to interpret, may result in inaccurate estimates. [1]

[2]

ii)

Commonly used distributions include:

Parametric:

- Exponential distribution
- Weibull distribution
- Gompertz – Makeham formula [1]

Non Parametric:

- Kaplan-Meier model
- Nelson-Aalen model [0.5]

Semi Parametric:

- Cox PH model. [0.5]

[2]

[4 Marks]

Solution 2:

i)

Gini Index:

Gini Index is a measure of inequality of a distribution. It is calculated as the probability that, if two items are selected at random (with replacement), they will be of different types. So , if the two items are same, the probability will be zero, whereas, if they are different, the probability will be close to 1.

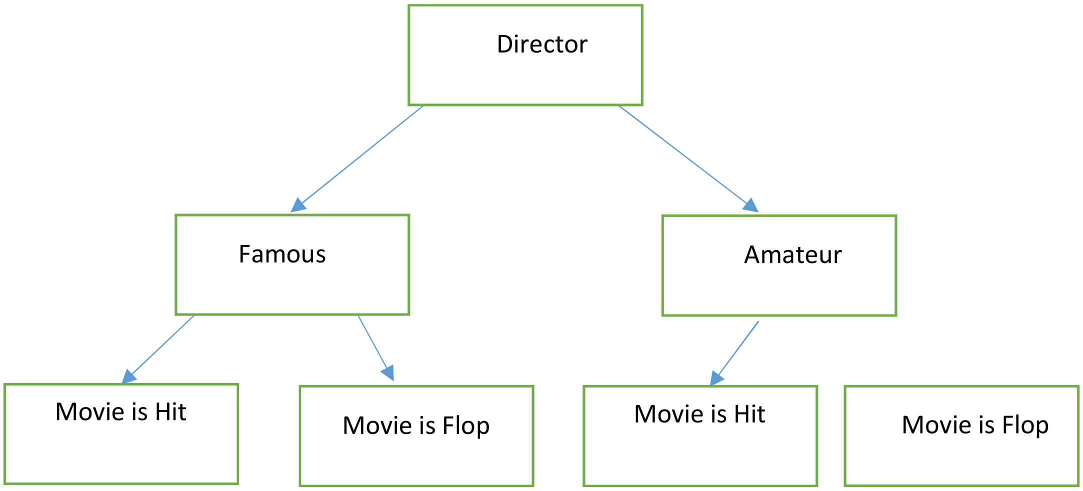
The formula for calculating the Gini Index for one of the final nodes in a Decision tree is

$$G = \sum_k \rho_k \ (1 - \rho_k)$$

Here ρ_k is the proportion of training instances with class k in the rectangle of interest. [2]

ii)

Binary Tree for Director and Result:



[1]

[1]

P(Director = Famous) = 6/10 P(Director= Amateur) = 4/10

P(Director = Famous & Result = Hit)=3/6 P(Director= Amateur & Result = Hit)=3/4

P(Director = Famous & Result = Flop)=3/6 P(Director= Amateur & Result = Flop)=1/4 [1]

Gini Index for Famous Director Node = 1- { $(\frac{1}{2})^2 + (\frac{1}{2})^2$} = 0.50

Gini Index for Amateur Director Node = 1- { $(\frac{3}{4})^2 + (\frac{1}{4})^2$} = 0.375 [1]

Gini Index for Director Node = $(\frac{6}{10}) * 0.5 + (\frac{4}{10}) * 0.375$ **= 0.45** [1]

[4]

iii)

Gini index for the Actor, Director and Budget are as follows:

Attributes	Gini Index
Actor	0.45
Director	0.45
Budget	0.4664

The Gini index for the Actor & Director node are same and having minimum value. Hence either Actor or Director can be chosen as first root for classification in decision trees. After choosing the first root, the same exercise can be repeated for the other attributes (Director & budget) to choose the subsequent roots. [2]

[8 Marks]

<u>Solution 3</u>:

i) Mean Residual Life of two year old phone

$$= \frac{\int_x^\infty \{1 - F(y)\} dy}{1 - F(x)}$$ [1]

$$= \int_x^\infty e^{-\lambda y} dy \, / \, e^{-\lambda x}$$ [1]

$$= 1/\lambda$$

= 8 years. [1]

[3]

ii)

The expected life time of a new phone = $1/\lambda$ = 8 years. [0.5]

And the expected life time of a two year old phone is also 8 years. [0.5]

This is due to the memory loss property of exponential distribution. [0.5]

Hence the exponential distribution is not appropriate for measuring Mean Residual Life of a product. [0.5]

[2]

[5 Marks]

Solution 4:

i) The model is ARIMA(2,0,0) provided that the model is stationary. [1]

ii) The lag polynomial is $1-(\alpha+\beta)L+\alpha\beta L^2 = (1-\alpha L)(1-\beta L)$ [1]

Since the roots are $1/\alpha$ and $1/\beta$, both greater than one in absolute value the process is stationary. [1]

[2]

iii) Since the process is stationary we know that $E(Y_t)$ is equal to some constant θ independent of t. [0.5]

Taking expectations on both sides of the equation defining Y_t gives

$E(Y_t) = \mu + (\alpha+\beta) E(Y_{t-1}) - \alpha\beta E(Y_{t-2}) + E(e_t)$ [0.5]

$$E(Y_t) = \mu + (\alpha+\beta) E(Y_{t-1}) - \alpha\beta E(Y_{t-2})$$
$$\theta = \mu + (\alpha+\beta)\theta - \alpha\beta\theta$$
$$\theta = \frac{\mu}{1-(\alpha+\beta)+\alpha\beta}$$ [1]

[2]

iv)

The auto-covariance function is not affected by the constant term of μ in the equation, and this term can be ignored.

The Yule-Walker equations are

$$\gamma_0 = (\alpha+\beta)\gamma_1 - \alpha\beta\gamma_2 + \sigma^2 \quad \text{- A}$$
$$\gamma_1 = (\alpha+\beta)\gamma_0 - \alpha\beta\gamma_1 \quad \text{- B}$$
$$\gamma_2 = (\alpha+\beta)\gamma_1 - \alpha\beta\gamma_0 \quad \text{- C}$$
$$\gamma_s = (\alpha+\beta)\gamma_{s-1} - \alpha\beta\gamma_{s-2} \quad for\ s > 2$$ [1]

Dividing both sides of (B) by γ_0 and noting that $\rho_s = \frac{\gamma_s}{\gamma_0}$, we have

$$\rho_s = (\alpha+\beta) - \alpha\beta\rho_s$$
$$\rho_1 = \frac{(\alpha+\beta)}{1+\alpha\beta}$$ [1]

Dividing both sides of (C) by γ_0 and substituting ρ_1

$$\rho_2 = (\alpha+\beta)\rho_1 - \alpha\beta$$

$$\rho_2 = (\alpha+\beta)\frac{(\alpha+\beta)}{1+\alpha\beta} - \alpha\beta$$

[1]

$$\rho_3 = (\alpha+\beta)\rho_2 - \alpha\beta\,\rho_1$$

$$\rho_3 = (\alpha+\beta)\left\{(\alpha+\beta)\frac{(\alpha+\beta)}{1+\alpha\beta} - \alpha\beta\right\} - \alpha\beta\,\frac{(\alpha+\beta)}{1+\alpha\beta}$$

$$\rho_3 = \frac{(\alpha+\beta)^3}{1+\alpha\beta} - \alpha\beta\,\frac{(\alpha+\beta)}{1+\alpha\beta} - \alpha\beta(\alpha+\beta)$$ [1]

[4]

[9 Marks]

Solution 5:

i)

The likelihood is given by

$$L = K * e^{-2608(\mu+\sigma)}e^{-176(\rho+\vartheta)}\sigma^{92}\rho^{60}\mu^{12}\vartheta^{20}$$ [1]

Taking Logs

$$Log\ L = Log\ K - 2608(\mu+\sigma) - 176(\rho+\vartheta) + 92Log(\sigma) + 60Log(\rho) + 12Log(\mu) + 20Log(\vartheta)$$

[2]

Diffrentiating w.r.t σ

$$\frac{d}{d\sigma}(Log\ L) = -2608 + \frac{92}{\sigma}$$

Equating to zero we get

$$\hat{\sigma} = \frac{92}{2608} \quad = 0.0353\ p.a.$$ [1]

Diffrentiating again

$$\frac{d^2}{d\sigma^2}(Log\ L) = -\frac{92}{\sigma^2} < 0\ for\ above\ \hat{\sigma}$$ [1]

Therefore $\hat{\sigma}$ is the maximum likelihood estimate

[5]

ii)

The variance of $\hat{\sigma}$ *is given by*

$$\frac{-1}{\frac{d^2}{d\sigma^2}(Log\ L)} = \frac{\sigma^2}{92}$$ [1]

We can estimate this using $\hat{\sigma}$

Therefore, the estimated standard deviation of $\hat{\sigma}$ is $\frac{\hat{\sigma}}{2\sqrt{23}}$ = 0.00368 [1]

[2]

[7 Marks]

Solution 6:

i)

Life A contribution

$E^c_{29,10}$ = 31+28+31+30+31+30+31 [1]

$E^c_{30,11}$ = 31+30+31+30+24 [1]

Life B contribution

$E^c_{28,5}$ = 31 [1]

$E^c_{29,5}$ = 28+30 [1]

[4]

ii)
Reduction factor formula:
$R_{x,t} = \alpha_x + (1 - \alpha_x)(1 - f_{n,x})^{\frac{t}{n}}$ [2]

We can first calculate α_{62}
$$\alpha_{62} = \frac{0.001}{0.005916} = 0.16903$$

We are also given that $f_{8,62} = 0.55$, so we need:
$$R_{62,15} = \alpha_{62} + (1 - \alpha_{62})(1 - f_{8,62})^{\frac{15}{8}}$$

$R_{62,15} = 0.16903 + (1 - 0.16903)(1 - 0.55)^{\frac{15}{8}} = 0.35497$

The projected mortality rate for lives aged 62 in 15 years time is 0.005916 x 0.35497 = 0.0021 [3]
[5]
[9 Marks]

Solution 7:

i)
We wish to find m_{24}

Now, $m_{i4} = \begin{cases} 0, \ if \ i = 4, \\ 1 + \sum_{j \neq 4} p_{ij} \, m_{j4}, \ if \ i \neq 4 \end{cases}$ [1]

Thus, $m_{24} = 1 + (1 \times m_{34})$

$m_{44} = 0$

$m_{34} = 1 + (\frac{1}{3} \times m_{44}) + (\frac{2}{3} \times m_{24})$

$m_{34} = 1 + 0 + \frac{2}{3}(1 + m_{34})$

$m_{34} = 5$

Hence, $m_{24} = 1 + 5 = 6 \ steps$ [2]
[3]

ii)
$P(1,2,3,2,3,4) = P(X_0 = 1) \times (p_{12}) \times (p_{23}) \times (p_{32}) \times (p_{23}) \times (p_{34})$

$= \frac{3}{4} \times \frac{3}{5} \times 1 \times \frac{2}{3} \times 1 \times \frac{1}{3}$

$= \frac{1}{10}$ [2]
[5 Marks]

Solution 8:

i)
Concordance: If two random variables are concordant if small vales of one are likely to be associated with small values of the other and vice versa. [1]

Spearman's rho and Kendall's tau are the two measures of concordance of random variables. [1]

[2]

ii)

Spearman's rho:

€	Rank	£	Rank	di	di^2
70	6	75	7	-1	1
73	4	73	9	-5	25
78	1	71	10	-9	81
74	3	74	8	-5	25
72	5	78	5	0	0
76	2	77	6	-4	16
68	8	86	2	6	36
60	10	85	3	7	49
65	9	87	1	8	64
69	7	83	4	3	9
				Total	**306**

[2]

$_s\rho = 1 - \frac{6}{10(10^2-1)} * 306 = -0.86$ [2]

[4]

[6 Marks]

<u>Solution 9</u>:

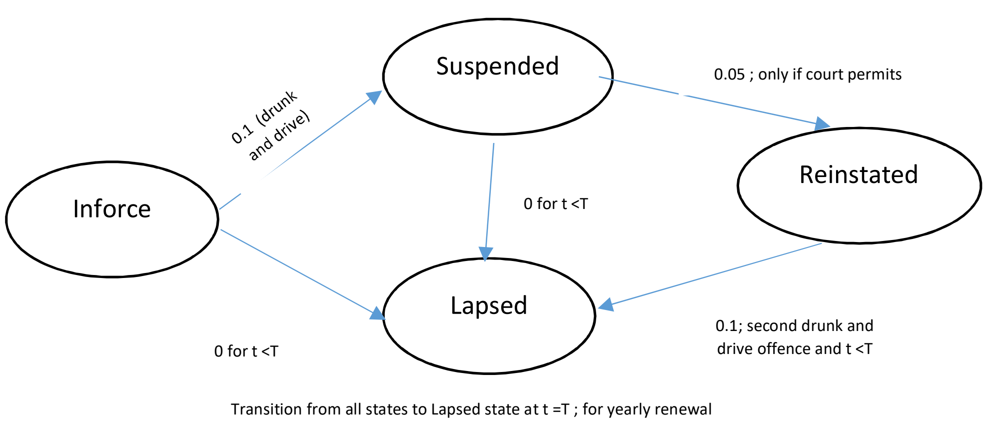
i)

A time inhomogeneous model should be used. Transition probabilities out of the "Suspended" state between times s and t may depend not only on the time difference t – s but on the duration s the policy has been in that state. [2]

ii)

A model with given state space would not satisfy the Markov property because a license can only be reinstated once, so if in state inforce we would need to know if the license has previously been Suspended. A Markov model could be obtained by expanding the state space to {Inforce, Suspended, Reinstated, Lapsed}. In this case the future transitions will depend only on the state currently occupied and duration, irrespective of previous states. [3]

iii)



[2]

iv)
Labelling states as I, S, R and L
$P_{II}(0,t) = P_{\overline{II}}(0,t)$ as no return to this state is possible

$\frac{d}{dt} P_{\overline{II}}(0,t) = -0.1 * P_{\overline{II}}(0,t)$, By Markov theorem [0.5]

$$\frac{\frac{d}{dt} P_{\overline{II}}(0,t)}{P_{\overline{II}}(0,t)} = -0.1$$

$\frac{d}{dt} Log\left(P_{\overline{II}}(0,t)\right) = -0.1$ [0.5]

$Log\left(P_{\overline{II}}(0,t)\right) = -0.1t + C\ ;\ when\ t = 0\ then\ P_{\overline{II}}(0,t) = 0$ => C =0

$P_{\overline{II}}(0,t) = e^{-0.1t}$ [1]

[2]

v)
To be in S at time t, must have remained in state I until some time w, then transitioned to S at time w, then remained in state S until t time. Probabilities are $P_{\overline{II}}(0,w)$, 0.1dw, and $P_{\overline{SS}}(w,t)$ respectively.
[0.5]

Integrating over the possible values of w and t <T;

$P_{IS}(0,t) = \int_0^t P_{II}(0,w) * 0.1 * P_{SS}(w,t)\mathrm{dw}$ [1]

$P_{SS}(w,t) = e^{-0.05\,(t-w)}$ (by extension of part (d)) [0.5]

$P_{IS}(0,t) = \int_0^t e^{-0.1w} * 0.1 * e^{-0.05\,(t-w)}\mathrm{dw}$ [1]
$P_{IS}(0,t) = 0.1 * e^{-0.05t} \int_0^t e^{-0.05w}\mathrm{dw}$
$P_{IS}(0,t) = -2e^{-0.05t}(e^{-0.05t} - 1)$
$P_{IS}(0,t) = 2e^{-0.05}\ (1 - e^{-0.05}\ )$ [1]

[4]

[13 Marks]

<u>Solution 10</u>:

i)
Probability of claim involves reinsurer:
Let X denote the individual claim amount random variable.
Then X ~ Pa(5,10000)
We need to find P(X>4000) = $(\frac{10000}{10000+400})^5$ = 0.186 [2]

ii)
Insurer's expected claim payment per claim:

Let Y denote the amount of a claim paid by insurer. Then

$$Y = \begin{cases} X, & X \leq 4000 \\ 4000, & X > 4000 \end{cases}$$ [0.5]

So,

E(Y) = $\int_0^{4000} xf(x)dx + \int_{4000}^{\infty} 4000f(x)dx$ [0.5]

= $\int_0^{4000} x \frac{5*10000^5}{(10000+x)^6}$ dx + 4000 P(X>4000) [1]

= $\{-x \frac{10000^5}{(10000+x)^5}\}_0^{4000} + \int_0^{4000} \frac{5*10000^5}{(10000+x)^5}$dx + 4000*0.186 (using integration by parts)

[1]

= -4000*0.186 + = $\{- \frac{10000^5}{4*(10000+x)^4}\}_0^{4000}$ + 4000*0.186

= 1849.23 [1]

[4]

iii)

The claim amounts are expected to increase by 20% and X ~ Pa(5,10000)

⇨ X' = Next year claim amount = 1.2 X

⇨ P(X'>4000) = P(X>4000/1.2)

⇨ $(\frac{10000}{10000+4000/1.2})^5$ = 0.2373 [3]

[9 Marks]

Solution 11:

Clearly, (X_0, X_1, X_2, . . .) is a Markov chain with state space

S = {0, 1, 2, 3, 4, 5}

i)

If, at some point of time, X_n = x (i.e. the number of white balls in the left urn is x) then there are 5 − x black balls in the left urn, while the right urn contains x black and 5 − x white balls. Clearly,

$p_{x,x+1} = P(X_{n+1} = x + 1 | X_n = x)$

= P(pick a white ball from the right urn and a black ball from the left urn) [1]

$$= \frac{5-x}{5} \times \frac{5-x}{5},$$

as long as $x < 5$. On the other hand,

$p_{x,x-1} = P(X_{n+1} = x - 1|X_n = x)$

$= P$(pick a white ball from the left urn and a black ball from the right urn)

$$= \frac{x}{5} \times \frac{x}{5},$$

as long as $x > 0$. When $0 < x < 5$, we have

$$p_{x,x} = 1 - p_{x,x+1} - p_{x,x-1},$$

because there is no chance that the number of balls change by more than 1 ball. Summarising, the answer is:

$$p_{x,y} = \begin{cases} \left(\frac{5-x}{5}\right)^2, & \text{if } 0 \le x \le 4, \quad y = x+1 \\ \left(\frac{x}{5}\right)^2, & \text{if } 1 \le x \le 5, \quad y = x-1 \\ 1 - \left(\frac{5-x}{5}\right)^2 - \left(\frac{x}{5}\right)^2, & \text{if } 1 \le x \le 4, \quad y = x, \\ 0, & \text{in all other cases.} \end{cases}$$

[2]

[3]

ii)

To compute the stationary distribution, cut between states x and x − 1 and equate the two flows, as usual:

$$\pi(x)p_{x,x-1} = \pi(x-1)p_{x-1,x},$$

i.e.

$$\pi(x)\left(\frac{x}{5}\right)^2 = \pi(x-1)\left(\frac{5-(x-1)}{5}\right)^2$$

which gives

$$\pi(x) = \left(\frac{6-x}{x}\right)^2 \pi(x-1)$$

[1]

We thus have

$$\pi(1) = \left(\frac{5}{1}\right)^2 \pi(0) = 25\pi(0)$$
$$\pi(2) = \left(\frac{4}{2}\right)^2 \pi(1) = \left(\frac{4}{2}\right)^2\left(\frac{5}{1}\right)^2 \pi(0) = 100\pi(0)$$
$$\pi(3) = \left(\frac{3}{3}\right)^2 \pi(2) = \left(\frac{3}{3}\right)^2\left(\frac{4}{2}\right)^2\left(\frac{5}{1}\right)^2 \pi(0) = 100\pi(0)$$
$$\pi(4) = \left(\frac{2}{4}\right)^2 \pi(3) = \left(\frac{2}{4}\right)^2\left(\frac{3}{3}\right)^2\left(\frac{4}{2}\right)^2\left(\frac{5}{1}\right)^2 \pi(0) = 25\pi(0)$$
$$\pi(5) = \left(\frac{1}{5}\right)^2 \pi(4) = \left(\frac{1}{5}\right)^2\left(\frac{2}{4}\right)^2\left(\frac{3}{3}\right)^2\left(\frac{4}{2}\right)^2\left(\frac{5}{1}\right)^2 \pi(0) = \pi(0).$$

[2]

We find π(0) by normalisation:

$\pi(0) + \pi(1) + \pi(2) + \pi(3) + \pi(4) + \pi(5) = 1$

$\Rightarrow \pi(0) = 1/(1 + 25 + 100 + 100 + 25 + 1) = 1/252.$ [1]

Putting everything together, we have

$$\pi(0) = \frac{1}{252},\quad \pi(1) = \frac{25}{252},\quad \pi(2) = \frac{100}{252},\quad \pi(3) = \frac{100}{252},\quad \pi(4) = \frac{25}{252},\quad \pi(5) = \frac{1}{252}.$$

[1]

[5]

iii)

π(x) is a hyper geometric distribution. [1]

[9 Marks]

<u>Solution 12</u>:

i)

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

$$E(X_t) = \mu + E(e_t\sqrt{\alpha + \beta(X_{t-1} - \mu)^2})$$
$$E(X_t) = \mu + E(e_t)E(\sqrt{\alpha + \beta(X_{t-1} - \mu)^2})$$
$$E(X_t) = \mu + 0 * E(\sqrt{\alpha + \beta(X_{t-1} - \mu)^2})$$
$$E(X_t) = \mu$$ [1]

$$Cov(X_t, X_{t-s}) = E(X_t X_{t-s}) - E(X_t)E(X_{t-s})$$

$$Cov(X_t, X_{t-s}) = E(\left(\mu + e_t\sqrt{\alpha + \beta(X_{t-1} - \mu)^2}\right)\left(\mu + e_{t-s}\sqrt{\alpha + \beta(X_{t-s-1} - \mu)^2}\right)) - \mu^2$$

Let $A_t = \sqrt{\alpha + \beta(X_{t-1} - \mu)^2}$
$Cov(X_{t,}X_{t-s}) = E((\mu + e_t A_t)(\mu + e_{t-s}A_{t-s})) - \mu^2$

$Cov(X_{t,}X_{t-s}) = E(\mu^2 + \mu e^t A_t + \mu e^{t-s}A_{t-s} + e_t e_{t-s}A_t A_{t-s}) - \mu^2$

$Cov(X_{t,}X_{t-s}) = E(\mu^2) + \mu E(e^t)E(A_{t)} + \mu E(e^{t-s})(A_{t-s)} + E(e_t e_{t-s}A_t A_{t-s}) - \mu^2$
$Cov(X_{t,}X_{t-s}) = \mu^2 + \mu * 0 * E(A_{t)} + \mu * 0 * (A_{t-s)} + E(e_t e_{t-s}A_t A_{t-s}) - \mu^2$
$Cov(X_{t,}X_{t-s}) = E(e_t e_{t-s}A_t A_{t-s})$ [2.5]

Now e_t is independent of X_{t-1} as above, and that e_t is independent of $e_{t-s}A_t A_{t-s}$
$Cov(X_{t,}X_{t-s}) = E(e_t)E(e_{t-s}A_t A_{t-s})$
$Cov(X_{t,}X_{t-s}) = 0 * E(e_{t-s}A_t A_{t-s})$
$Cov(X_{t,}X_{t-s}) = 0$ [1]

Thus $X_{t,} and\ X_{t-s}$ are uncorrelated. [0.5]

[5]

ii)

The conditional variance of $X_t\ |X_{t-1}$ is

$Var(X_t|X_{t-1}) = Cov(X_{t,}X_t) = E(e_t e_t A_t A_t)$ from previous expression [0.5]

$Var(X_t|X_{t-1}) = E(e_t^2 A_t^2)$ [0.5]
$Var(X_t|X_{t-1}) = E(e_t^2)E(A_t^2)$
$Var(X_t|X_{t-1}) = Var(e^t)\mathrm{E}(\alpha + \beta(X_{t-1} - \mu)^2)$
$Var(X_t|X_{t-1}) = 1 * \mathrm{E}(\alpha) + \beta E(X_{t-1} - \mu)^2)$
$Var(X_t|X_{t-1}) = \alpha + \beta(X_{t-1} - \mu)^2$ [1]

Therefore, variance of X_t depends on X_{t-1}. 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. [1]

[3]

[8 Marks]

****