INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

18 April 2023 (am)

Subject CS2A – Risk Modelling and Survival Analysis Core Principles

Paper A

Time allowed: Three hours and twenty minutes

In addition to this paper you should have available the 2002 edition of the Formulae and Tables and your own electronic calculator.

If you encounter any issues during the examination please contact the Assessment Team on T. 0044 (0) 1865 268 873.

1 A process, X_t, is created as follows. n black balls and n white balls are initially placed in two separate boxes, A and B, in such a way that each box contains n balls. An experiment is performed in which a ball is selected at random from each box at time t, ($t = 1, 2, \ldots$), and the two selected balls interchanged. Let X_t be the number of white balls in box A just after time t.

(i) Explain whether X_t is irreducible. [1]

(ii) Determine the elements of the transition probability matrix of X_t. [6]

Assume that all the n balls that were initially in box A were the white balls and all the n balls that were initially in box B were the black balls.

(iii) Determine the probability that $X_n = 0$, simplifying your answer where possible. [3]

[Total 10]

2 Country A has recently gone through an economic crisis. As the country makes an attempt to recover, the Department of Finance is trying to estimate the rate of recovery in employment. The department has decided to use a two-state continuous-time Markov model to estimate the rate of return to employment. It has also decided to use data from one of the previous economic recoveries for the purpose. The two states are:

The employment data from a previous economic recovery was as follows:

- Waiting time to gain employment in the first year (in person-years): 30,000
- Waiting time to gain employment in the second year (in person-years): 22,000
- Number of people gaining employment in the first year: 5,000
- Number of people gaining employment in the second year: 7,000

It may be assumed that force of gaining employment in any 1 year is constant.

(i) State the likelihood function of the maximum likelihood estimator of the transition rate defining all the terms you use. [3]

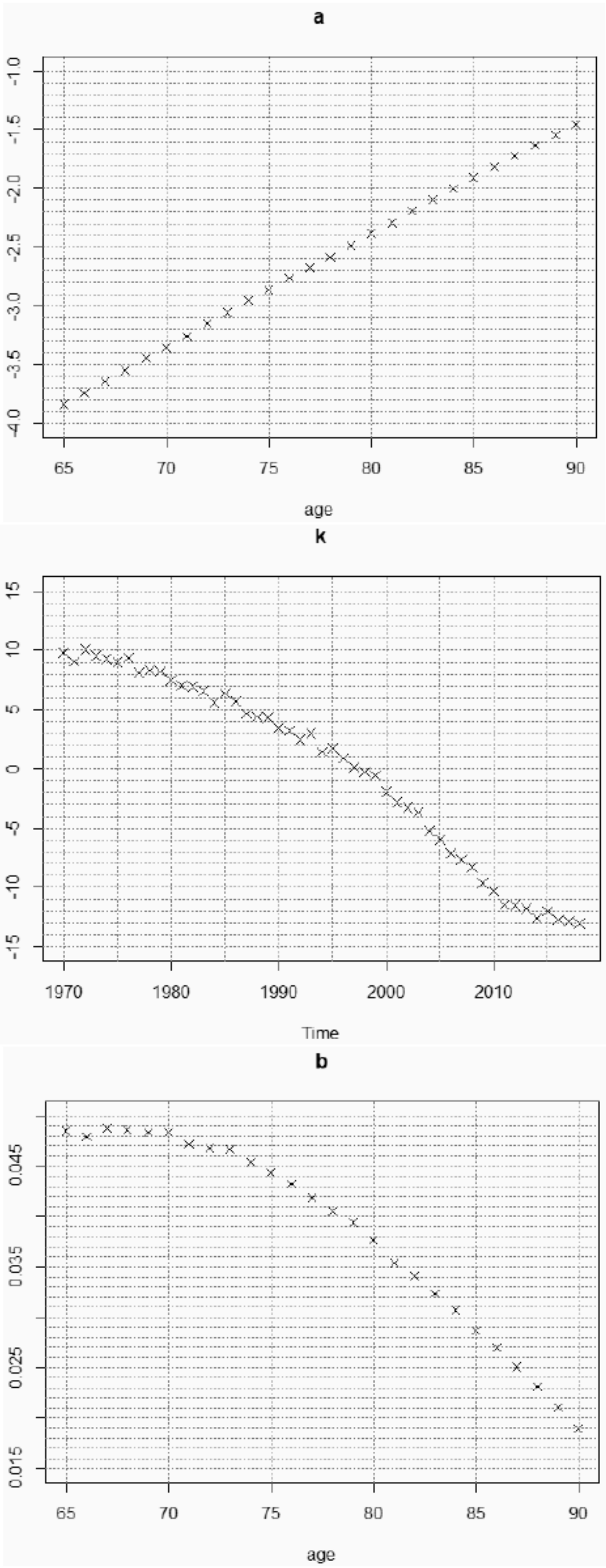
(ii) Calculate the maximum likelihood estimate of the transition rate from the state of being unemployed to gaining employment for each of the first 2 years. [2]

(iii) Estimate, by stating the expression, the variance of the second year maximum likelihood estimator. [4]

(iv) Calculate the probability of not gaining any employment in the next 2 years. [3]

[Total 12]

3 The figure below shows three vectors a_x, b_x and k_t, which were obtained by applying the two-factor Lee–Carter model to mortality data derived from the period 1970 to 2018 for the age range 65 to 90 years.



(i) Discuss what each of these three vectors suggest about the characteristics of the mortality rates in the data. [4]

(ii) Suggest a socio-economic reason for your conclusion about vector b in part (i) above. [1]

(iii) Calculate $m_{70,\,2018}$ using the output from the model (reading the values from the plots in the figure). [1]

An ARIMA(0,1,0) model (discrete random walk with drift) with Normal errors $\varepsilon \sim N(0, \sigma^2)$ was used to model projected k values. The drift term of the time series model was estimated as −0.4763, with standard error 0.084, and σ^2 was estimated as 0.346.

(iv) Calculate a 90% confidence interval for $m_{70,\,2019}$ based on the value of $m_{70,\,2018}$ calculated in part (iii). [5]

On analysing the past data over a larger age range (from 20 to 90 years) it was determined that there was an increase in rates for 20–30 year olds in the 1970s due to an illness that almost exclusively affected the younger population. The higher mortality rates only lasted until 1980 when a cure was found and introduced to the population.

(v) Explain how this cohort effect may manifest itself in the output from the two-factor Lee–Carter model, fitted to ages 20–90 and calendar years 1970–2018, in terms of a, b and k. [5]

[Total 16]

4 A Markov jump process model is used to describe the recovery of people bitten by a certain type of poisonous snake. There are three states:

- Sick and receiving medical care following the snake bite
- Fully recovered
- Recovered but with long-term health effects from the bite.

(i) Explain why a time in-homogeneous Markov jump process model is more suitable than a simpler time homogeneous multi-state model in this scenario. [1]

The transition rates from the sick state in this model t days after being bitten by the snake are:

$e^{-2.5t}$ for the transition to fully recovered and

$0.05 - e^{-2.5t}$ for the transition to recovered but with long-term health effects.

(ii) Comment on the key features of this model including the transition rates. [3]

(iii) Determine the probability that a person just bitten by a snake will eventually make a full recovery without any long-term health effects. [6]

[Total 10]

5 A sample of size n is taken from a process, X_t, which is believed to be an ARMA(1,1) process of the form

$$X_t = aX_{t-1} + e_t + be_{t-1}$$

where $|a|, |b| < 1$. The sample autocorrelations at lag 1 and lag 2 are 0.65 and 0.325, respectively.

(i) Estimate the parameters a and b by equating the sample autocorrelations to the theoretical values. [6]

Fisher's transformation states that the sample correlation coefficient, r, between two random variables, Y and Z, is such that $\frac{1}{2}\log\left(\frac{1+r}{1-r}\right)$ is approximately Normally distributed with mean $\frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right)$ and variance $\frac{1}{n-3}$, where ρ is the theoretical correlation coefficient between Y and Z and n is the sample size.

(ii) Determine the minimum value of n necessary to reject the null hypothesis that $b = 0$ in favour of the alternative $b > 0$ at the 95% significance level. You should assume that a is equal to the value determined in part (i) and use Fisher's transformation on the autocorrelation at lag 1. [6]

[Total 12]

6 A hydroelectric company is managing a water reservoir created from a dam in a river valley. The dam was originally chosen so that the water level would exceed a threshold of 50 metres in about 2 days in every 300 days. In these extreme events, the excess water is left to escape the reservoir so that the water level is kept below the safety 50-metre limit.

It is believed that the daily water level in the reservoir follows an exponential distribution with mean μ.

(i) Estimate the value of μ. [3]

(ii) Determine the expected threshold exceedance of the water level over the 50-metre threshold. [2]

In order to better manage the excess water, it is now assumed that the excess water level follows a Generalised Pareto distribution with scale parameter $\beta = 1$.

(iii) Explain the circumstances in which the Generalised Pareto distribution would be preferred to the exponential distribution. [2]

(iv) Estimate the value of the parameter γ if the expected threshold exceedance is the same as that in part (ii). [3]

[Total 10]

7 A mountain rescue service is looking to introduce a new training programme for volunteers who wish to join the service. Each Saturday for 10 weeks trainee rescuers are asked to join a mountain climb. Only those who successfully complete the climb are invited back the following week.

The rescue service will recruit those trainees who successfully complete a certain number of Saturday climbs. To decide on how many successful weeks should be required for a new recruit, the rescue service conducts a trial with 20 volunteers. The table below shows how many of these volunteers fail to complete the climb each week and the number who are eligible but do not arrive for the beginning of each climb.

Week	*Eligible but do not arrive*	*Arrive but fail to complete the climb*
1	0	1
2	0	2
3	1	2
4	0	0
5	0	1
6	4	1
7	0	2
8	0	1
9	0	2
10	0	1

(i) Explain why the Kaplan–Meier estimate is a suitable way to evaluate this training programme. [2]

The rescue service would like to recruit 30% of the volunteers who start the programme.

(ii) Calculate the number of successful weeks the service should require trainees to complete using the Kaplan–Meier estimate. [8]

(iii) Discuss what concerns the rescue service should have about using this study to set the recruitment criteria for all future volunteers. [3]

[Total 13]

8 Consider the time-series model:

$$y_t = a\, y_{t-2} + e_t + b\, e_{t-1} \qquad \text{(A)}$$

where e_t is a white noise process with mean 0 and variance σ^2.

(i) Derive the possible values of a and b for which the process y_t is stationary and invertible. [4]

(ii) State the values of p and q for which y_t is an ARMA(p, q) process. [1]

If $b = 0$ the original model (A) reduces to

$$y_t = a\, y_{t-2} + e_t \qquad \text{(B)}$$

(iii) Derive the autocorrelation function for this model while stationarity is assumed to hold. [8]

An actuary attempts to fit the model (A) to some time series data but concludes that the simpler model (B) is more appropriate.

(iv) Discuss how this conclusion could have been reached. [4]

[Total 17]

END OF PAPER

Institute and Faculty of Actuaries

EXAMINERS' REPORT

CS2 - Risk Modelling and Survival Analysis

Core Principles

Paper A

April 2023

Introduction

The Examiners' Report is written by the Chief Examiner with the aim of helping candidates, both those who are sitting the examination for the first time and using past papers as a revision aid and also those who have previously failed the subject.

The Examiners are charged by Council with examining the published syllabus. The Examiners have access to the Core Reading, which is designed to interpret the syllabus, and will generally base questions around it but are not required to examine the content of Core Reading specifically or exclusively.

For numerical questions the Examiners' preferred approach to the solution is reproduced in this report; other valid approaches are given appropriate credit. For essay-style questions, particularly the open-ended questions in the later subjects, the report may contain more points than the Examiners will expect from a solution that scores full marks.

For some candidates, this may be their first attempt at answering an examination using open books and online. The Examiners expect all candidates to have a good level of knowledge and understanding of the topics and therefore candidates should not be overly dependent on open book materials. In our experience, candidates that spend too long researching answers in their materials will not be successful either because of time management issues or because they do not properly answer the questions.

Many candidates rely on past exam papers and examiner reports. Great caution must be exercised in doing so because each exam question is unique. As with all professional examinations, it is insufficient to repeat points of principle, formula or other text book works. The examinations are designed to test "higher order" thinking including candidates' ability to apply their knowledge to the facts presented in detail, synthesise and analyse their findings, and present conclusions or advice. Successful candidates concentrate on answering the questions asked rather than repeating their knowledge without application.

The report is written based on the legislative and regulatory context pertaining to the date that the examination was set. Candidates should take into account the possibility that circumstances may have changed if using these reports for revision.

Sarah Hutchinson
Chair of the Board of Examiners
July 2023

A. General comments on the *aims of this subject and how it is marked*

The aim of the Risk Modelling and Survival Analysis Core Principles subject is to provide a grounding in mathematical and statistical modelling techniques that are of particular relevance to actuarial work, including stochastic processes and survival models.

Some of the questions in this paper admit alternative solutions from those presented in this report, or different ways in which the provided answer can be determined. All mathematically correct and valid alternative solutions received credit as appropriate.

In cases where an error was carried forward to later parts of the answer, candidates were given full credit for those later parts.

In higher order skills questions, where comments were required, well-reasoned comments that differed from those provided in the solutions also received credit as appropriate.

Candidates are advised to take careful note of all instructions that are provided with the exam to maximise their performance in future CS2A examinations.

B. Comments on *candidate performance in this diet of the examination.*

The number of questions on the paper was reduced from nine to eight in response to previous candidate feedback related to time pressure when answering the paper using MS Word. However, candidates should note that the syllabus and core reading for this subject is extensive, and the examiners are concerned that a considerable number of candidates present themselves without sufficient preparation and understanding across the whole syllabus. Although this exam is set twice a year, the examiners feel that many candidates are likely to need more than 4-6 months preparation time given the length of the syllabus.

Although this examination is "open-book," the examiners would recommend that candidates prepare for this paper as if they were sitting a traditional "closed-book" exam. There are two reasons for this. Firstly, access to resources in a time limited exam is no substitute for understanding and revision. Secondly, this examination tests understanding of statistics and risk models asking candidates to apply these techniques to certain scenarios or data sets given in the questions. Successful answers to these questions require candidates to enter the examination with the necessary understanding of the statistical methods and then apply a careful problem-solving approach to the data or scenario given. It is highly unlikely that consulting resources in the examination time available will help here.

The examination team note that the higher order questions were found by candidates to be challenging in this sitting. Again, the key here is application of knowledge (for example about censoring in the Kaplan-Meier estimate or the parameters in the Lee-Carter model in two of the questions on this paper) to the scenario presented or the results obtained in earlier parts of the question. It is hoped that the solutions presented below will help candidates in future sessions appreciate the type of answers that can reach the highest marks.

C. Pass Mark

The Pass Mark for this exam was 51.
1226 candidates presented themselves and 376 passed.

Solutions for Subject CS2A - April 2023

Q1
(i)
The chain is irreducible because every state can be reached from any other state [1]

(ii)
We are interested in:
$P(X_{t+1} = j \mid X_t = i)$, for $i,j = 0,1,\ldots,n$

If $i = 0$, $P(X_{t+1} = j \mid X_t = i) = \begin{cases} 1 & \text{if } j = 1 \\ 0 & \text{if } j \neq 1 \end{cases}$ [1]

If $i = n$, $P(X_{t+1} = j \mid X_t = i) = \begin{cases} 1 & \text{if } j = n-1 \\ 0 & \text{if } j \neq n-1 \end{cases}$ [1]

For $i = 1,2,\ldots,n-1$

$P(X_{t+1} = i-1 \mid X_t = i) = \frac{i}{n} \times \frac{i}{n} = (\frac{i}{n})^2$ [1]

$P(X_{t+1} = i \mid X_t = i) = (\frac{i}{n} \times \frac{n-i}{n}) + (\frac{n-i}{n} \times \frac{i}{n}) = \frac{2i(n-i)}{n^2}$ [1]

$P(X_{t+1} = i+1 \mid X_t = i) = \frac{n-i}{n} \times \frac{n-i}{n} = (\frac{n-i}{n})^2$ [1]

$P(X_{t+1} = j \mid X_t = i) = 0$ if $j \notin \{i-1, i, i+1\}$ [1]

(Candidates do not need to use the i, j notation - other formats are acceptable including an explanation of where there are zero entries without listing them all)

(iii)
Starting with all the white balls in A, getting $X_n = 0$ will require that a white ball is drawn from A and a black ball from B for each $t = 1,2,\ldots,n$ [1]
Thus, the probability

$$\left(\frac{n}{n}\frac{n}{n}\right) \times \left(\frac{n-1}{n}\frac{n-1}{n}\right) \times \cdots \times \left(\frac{1}{n}\frac{1}{n}\right) = \frac{n!^2}{n^{2n}}$$ [2]

[Total 10]

Parts (ii) and (iii) of this question were not well answered by many candidates. In fact, as a proportion of the marks available, the average mark was lowest for the whole paper. This is a surprise as the question does not require specialist technical knowledge, but rather basic knowledge of transition probabilities and a careful approach to problem-solving. This question is an excellent example of the comment at the beginning of this report about preparation for open-book examinations.

Q2
(i)
The likelihood function for the ith year [½]
L (*mu*_i;di,vi) = exp(-*mu*_i *vi)* *mu*_i ^di [1]

Where:
mu_i is the transition rate from unemployed to employed in the ith year [½]

di is the number of transitions from state "unemployed" to state "employed" in i^{th} year [½]
vi is the total observed waiting time in State "unemployed" in i^{th} year [½]

(ii)
Which results in maximum likelihood estimate of:
*mu*_i^hat = di / vi
Therefore,
*mu*_1^hat = 5000 /30000 = 0.16667 [1]
*mu*_2 ^hat = 7000/22000 = 0.31818 [1]

(iii)
The maximum likelihood estimator *mu*_2^hat has a variance equal to:
*Mu*2/E[V] [2]
Where:
*Mu*2 is the true transition rate in the second year
E[V] is the expected waiting time of being unemployed
*Mu*2 ~= *mu*_2^hat = 0.31818
E[V] ~=v2 = 22000 [1]
Variance = 0.000014 [1]

(iv)
Estimating
P (not getting employed in the next 2 years) = 2p0 = exp(- int(o to 2) mu x+s ds) [1]
= Exp(-mu1)*exp(-mu2) [1]
= 0.8465* 0.7275 [½]
= 0.61579 [½]

[Total 12]

This question was well answered, and the average mark was the highest across the whole paper.

In parts (i) and (ii) credit was given to candidates who only started to differentiate between years 1 and 2 in the second part, and to candidates who expressed all their answers numerically without full notation. The most common mistake was to calculate a blended transition rate across the two years for which partial credit was given if subsequent calculations proceeded correctly with that rate.

Q3
(i)
From the 3 plots in Figure 1
a
The increasing values of *a* imply that Mortality rates increase with age [½]
Visually the plot appears linear [½]
which implies that the increase is exponential [1]

k
Given that the *b*-values are positive [½]

the decreasing values of *k* imply that mortality has been improving over the period [½]
either: Visually the plot appears somewhat linear [½]
which implies that improvement rates have been constant [½]
or: the rate of decrease appears less pronounced at the earliest and latest years suggesting different rates of improvement over the period [1]

b
The increasing values of *b* imply that mortality improvements have been greatest for younger ages [½]
The values appear constant in ages 65 - 70 and decrease thereafter [½]
(½ mark for any other reasonable observation about the nature of the graphs)
[Marks available 6, maximum 4]

(ii)
Medical improvements have been greatest for younger ages
Education has had the greatest effect on younger population e.g. related to smoking advice, general health
Or any other reasonable comment [1]
(Award ½ mark if reasonable comment about older ages instead)

(iii)
exp(-3.35 + 0.0484 * (-13)) = 0.0187 accept any 0.0185 - 0.0189 [1]
(Award ½ mark if parameters read correctly or calculation performed correctly but not both)

(iv)
projected k, in 1 years: $-0.4763 \pm 1.64 * (0.084^2 + 0.346)^{1/2}$ [1]
$k = -0.4763 \pm 0.9745$ [1]
k = (-1.4508, 0.4982) [1]
Therefore, the lower limit of the required confidence interval is
m = 0.0187 exp(0.0484 * -1.4508) = 0.0174 [1]
The upper limit is m = 0.0187 exp(0.0484 * 0.4982) = 0.0192 [1]

(v)
a_x is a measure of the average rate at each age over the investigation period. Values would therefore be relatively high for the 20-30 year group. A plot of a_x may show an "illness bump" [2]

For k_t there would be a rapid decline in values around 1980, due to the rapid fall off of deaths in the 20-30 age range following the cure being introduced to the population. It may lead to a general underestimation of projected mortality rates if a simple linear model is adopted for projecting k. The effect will depend on the relative weightings of deaths in that age group [2]

bx will show large numbers in the 20-30 year band; these characteristics will be incorrectly projected into the future (assuming the illness has been eradicated), with improvements at these ages being greatly exaggerated [2]

In summary projected mortality rates are likely to be too low, with the greatest effect on the 20-30 year age band [1]

Other reasonable comments on ages 30+ [1]
[Marks available 8, maximum 5]
[Total 16]

This question was generally poorly answered, particularly parts (i), (iv) and (v).

In part (i) a large range of sensible points about a, b and k were given credit. Many candidates simply recited definitions of these parameters rather than applying those definitions to the evidence of the graphs in the question. The best answers combined a description of the plots with an understanding of the model parameters.

In part (iv) partial credit was given to a wide range of approaches to calculating a confidence interval. In particular a number of candidates derived a value for m70;2019 and then built a confidence interval for that rather than building the interval around k.

In part (v), candidates were not given credit for discussion of the advantages and disadvantages of cohort models given one is assumed in the question.

Q4

(i)
Allows path to recovery to vary with time since snake bite [½]
Constant transition intensities would seem inappropriate here [½]

(ii)
Transition rate to full recovery falls with duration [½]
Given -2.5 parameter probability of full recovery quickly becomes negligible [½]
Transition rate to recovery with long term effects increases with duration [½]
It seems reasonable that as the duration of sickness increases, the probability of recovery without long-term health effects decreases and the probability of recovery with long-term health effects increases [½]
As t increases this transition rates trends to 0.05 [½]
There is no upper limit to the time taken to recover in this model [½]
There is no death state [½]
The transition rate can go negative at some durations which is unrealistic [½]
Other reasonable observations [½]
[Marks available 5½, maximum 3]

(iii)
Pr (person bitten eventually fully recovered)
= integral(0,) Pr(remains sick from 0 to t)
* (transition rate to fully recovered at t) dt [2]
Pr(remains sick from 0 to t) = exp(-integral(0,t)(exp(-2.5u) + 0.05 - exp(-2.5u))du) = exp(-0.05t) [1]
so integral becomes:
= integral(0,∞) exp(-0.05t) exp(-2.5t) dt = integral(0,∞) exp-(2.55t) dt [1]
= [-exp-(2.55t) / 2.55]:(0,∞) [1]
= 1 / 2.55 = 0.392 [1]

(Full credit should be awarded to candidates who give the correct numeric answer and show some working but not necessarily all of the steps above)

[**Total 10]**

This is a straightforward Markov jump process question that was reasonably well answered.

In part (ii) a wide range of suitable comments attracted credit. This (somewhat akin to question 3 above) is an example of the need to show understanding of mathematical concepts by applying them to the scenario given in the question. Being successful in this is one of the key differences between candidates who passed and those who did not. Once again, taking a closed-book rather than open-book approach would pay dividends here as the necessary step of applying knowledge to the scenario is unlikely to be found in resources consulted during an examination.

Q5

(i)

We have (1 + *a* * *b*) * (*a* + *b*) / (1 + *b*^2 + 2 * *a* * *b*) = 0.65 (1) [1]
and (1 + *a* * *b*) * (*a* + *b*) / (1 + *b*^2 + 2 * *a* * *b*) * *a* = 0.325 (2) [1]
Dividing (2) by (1) gives *a* = 0.5 [½]

Substituting in (1) gives:
(1 + 0.5 * *b*) * (0.5 + *b*) = 0.65 * (1 + *b*^2 + 2 * 0.5 * *b*) [½]
i.e. 0.5 * *b*^2 + 1.25 * *b* + 0.5 = 0.65 * (*b*^2 + *b* + 1) [½]
i.e. 0.15 * *b*^2 - 0.6 * *b* + 0.15 = 0. [½]
The roots of this quadratic are 0.268 and 3.732 [1]
We require the root less than 1 in magnitude [½]
which is 0.268 [½]

(ii)

For autocorrelation at lag 1 the sample size to be used in the formula for Fisher's transformation is *n* - 1 [1]
The test statistic ½ * log ((1 + *r*_1) / (1 - *r*_1)) is approximately Normally distributed with mean ½ * log ((1 + *rho*_1) / (1 - *rho*_1)) and variance 1 / (*n* - 4), where *r*_1 is the sample autocorrelation at lag 1 and *rho*_1 is the theoretical autocorrelation at lag 1 [1]
For *a* = 0.5 and *b* = 0, *rho*_1 = 0.5 [½]
The 95th percentile of the standard Normal distribution is 1.645 [½]
We therefore require the least positive integer *n* such that
½ * log ((1 + 0.65) / (1 - 0.65)) - ½ * log ((1 + 0.5) / (1 - 0.5)) > 1.645 / sqrt(*n* - 4), [1]
i.e. such that
n > 4 + (1.645 / (½ * log ((1 + 0.65) / (1 - 0.65)) - ½ * log ((1 + 0.5) / (1 - 0.5))))^2. [1]
The least positive integer *n* satisfying this inequality is 57 [1]
(Full credit was given to candidates who use n-3 instead of n-4. In this case the final numeric answer will be 56)

[Total 12]

This question was not well answered and for the third consecutive session candidates have not answered Time Series questions as well as expected. A lot of candidates spent valuable exam time deriving the autocorrelation formulae rather than applying them, which the examiners suspect was due to the derivations being consulted during the open-book exam rather than the application of the formulae having been revised beforehand.

In part (i) the first marks available are for restating the autocorrelation formulae in terms of the ARMA(1,1) model and then proceeding from there.

Q6

(i)
Since the exponential distribution with parameter \lambda and with expectation \mu=1/\lambda has tail probability
Exp(-x/ \mu) then [1]
Exp(-50/\mu)=2/300 so
-50/\mu =log(2/300)=-5.010635 [1]
So
\mu=-50/5.010635=9.978775 [1]

(ii)
Since the threshold exceedance distribution for the exponential distribution is the same as the original distribution then [1]
[or since the exponential distribution is memoryless, then …]
the random variable U=X-50|X>50 has the same expectation as above, i.e. 9.978775 [1]

(iii)
GPD is preferred if extreme weather events are becoming more likely [1]
and therefore the exceedance distributions are expected to have fatter tails than those of the exponential [1]
modelling of the tails is seen as more important in a scenario such as this [1]
other sensible comments contrasting the GPD and the exponential [1]
[Total marks 4, maximum 2]

(iv)
If beta =1 the Pareto distribution will have expectation the same as the expected exceedance amount
\gamma /(gamma -1)= 9.978775 [1]
or
\gamma = (gamma -1)*9.978775 [1]
\gamma= 9.978775/(9.978775-1)= 1.111374 [1]
[Total 10]

The first three parts of this question on loss distributions is relatively straightforward, so again it is disappointing that these were not well answered.

Part (iv) is more demanding.

In part (i) as with other calculation questions in this paper, full credit was awarded for solutions with the correct final numerical answer (9.9878) and some but not each step of working given.

Full marks were also awarded in (ii) for reference to the memoryless property of the exponential distribution.

Q7

(i)

We have discrete data [½]
The hazard depends on duration / time [½]
There is [right] censoring [½]
There is non-informative censoring [½]
The data is suited to a non-parametric approach [½]
Other sensible comment on data suited to K-M approach [½]

[Marks available, 3, maximum 2]

(ii)

At duration t weeks, let d_t be the number who fail the task that week
c_t be the number censored that week
(see below for application of censoring to this problem)
n_t be the "risk set" - the number of volunteers still on the program
then h_t is the hazard of failing the task in week t where $h_t = d_t / n_t$

[1]

and the Kaplan Meier survival function is S(t) where

$$S(t) = \prod_{t_j \le t} (1 - h_j)$$ [½]

The Kaplan Meier estimate assumes that censoring occurs after failure therefore volunteers who do not arrive for week j need to be included in c_{j-1} not c_j [1]

t	n_t	d_t	c_t	h_t	$1-h_t$	S(t)
1	20	1	0	0.05	0.95	0.95
2	19	2	1	0.105263158	0.894736842	0.85
3	16	2	0	0.125	0.875	0.74375
4	14	0	0	0	1	0.74375
5	14	1	4	0.071428571	0.928571429	0.690625
6	9	1	0	0.111111111	0.888888889	0.613888889
7	8	2	0	0.25	0.75	0.460416667
8	6	1	0	0.166666667	0.833333333	0.383680556
9	5	2	0	0.4	0.6	0.230208333
10	3	1	0	0.333333333	0.666666667	0.153472222
	[½]	[½]	[1]	[½]	[½]	[1]

We seek the largest t at which S(t) >= 0.3 [½]

The number of weeks required is 8 [1]

(iii)
hazard unlikely to be zero at week 4 [½]
large amount of censoring between weeks 5 & 6 - would want to investigate why [1]
relatively small sample size [½]
would knowledge of the required number of weeks change behaviour? [½]
the right censoring may well be informative [½]
other types of censoring may be present [½]
different training programs may not be uniformly difficult [½]
other sensible comments [½]
(To obtain full marks on this part some discussion of the censoring here is required)
[Marks 4 available, maximum 3]
[Total 13]

This Kaplan Meier Estimate question was reasonably well answered.

The key to a full correct solution in part (ii) and then to a strong answer to part (iii) is to understand the role of censoring in this scenario. The ordering of failure and censoring is one of the K-M assumptions and well prepared candidates recognised this in their answers.

A wide variety of layouts for the calculations and answers in (ii) were given full credit. Candidates are reminded of the importance of defining terms when completing survival model calculations.

Part (iii) was less well answered and again asks candidates to apply knowledge of the model to the scenario in the question.

Q8
(i)
Using the backshift operator one can show that the corresponding polynomials are
1-a B^2 [1]
and
1+bB [1]
The roots need to be in absolute value less than 1
abs(a)<1 and abs(b)<1 [2]

(ii)
ARMA(2,1) [1]

(iii)
The Yule-Walker equations are
gamma_0=a gamma_2+sigma^2 [1]
and
gamma_k=a gamma_{k-2} for k >= 1 [1]
So

gamma_1=a gamma_1 [1]
gamma_2=a gamma_0 [1]
These imply that
gamma_1=0, gamma_2=a gamma_0 and in general [1]
gamma_k =0 for k odd [1]
gamma_k = a^{k/2} gamma_0 for k even [1]
therefore
rho_k=0 for k odd [½]
rho_k=a^(k/2) for k even [½]
(There are no marks available for deriving the Yule Walker equations from first principles)

(iv)
Sample acf of the data could have indicated insignificant spikes for odd lags as for b=0 case those values are zero [2]
AIC/BIC could have also been used to confirm the statistical preference between the two models [1]
In the parameter estimation process for model (1), some low t-values could have been produced, particularly for the parameter *b*, indicating over-parametrisation. [1]
other sensible comments contrasting the fit of the two models [1]

[Marks available 5, maximum 4]

[Total 17]

This question was much better answered than the earlier Time Series question.

Parts (i) and (ii) were well answered.

In part (iii) some candidates used valuable exam time to show a derivation of the Yule Walker equations which was not required (and gained no marks). Again, in an open-book exam, application not knowledge based ais required for marks.

In part (iv) a wider range of sensible comments attracted marks including alternative tests that could have been performed.

[Paper Total 100]

END OF EXAMINERS' REPORT

Institute and Faculty of Actuaries

Beijing

14F China World Office 1 · 1 Jianwai Avenue · Beijing · China 100004
Tel: +86 (10) 6535 0248

Edinburgh

Level 2 · Exchange Crescent · 7 Conference Square · Edinburgh · EH3 8RA
Tel: +44 (0) 131 240 1300

Hong Kong

1803 Tower One · Lippo Centre · 89 Queensway · Hong Kong
Tel: +852 2147 9418

London (registered office)

7th Floor · Holborn Gate · 326-330 High Holborn · London · WC1V 7PP
Tel: +44 (0) 20 7632 2100

Oxford

1st Floor · Belsyre Court · 57 Woodstock Road · Oxford · OX2 6HJ
Tel: +44 (0) 1865 268 200

Singapore

5 Shenton Way · UIC Building · #10-01 · Singapore 068808
Tel: +65 8778 1784

www.actuaries.org.uk