Institute and Faculty of Actuaries

EXAMINERS' REPORT

CS2A – Risk Modelling and Survival Analysis

Core Principles

Paper A

September 2022

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.

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
December 2022

© Institute and Faculty of Actuaries

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 the same error was carried forward to later parts of the answer, candidates were given full credit for the 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 in order to maximise their performance in future CS2A examinations. The instructions applicable to this diet can be found at the beginning of the solutions contained within this document.

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

The syllabus and Core Reading for Risk Modelling and Survival Analysis Core Principles covers multiple statistical techniques and modelling approaches. Candidates' ability to evidence a strong grasp of these techniques and approaches was quite uneven across the syllabus areas with Time Series and Stochastic Process questions generally answered less well than Survival Analysis and Loss Distribution questions. Candidates are reminded that before attempting the examination they need to be thoroughly prepared across the whole syllabus and be ready to use the mathematical and modelling approaches to a range of scenarios.

The least well answered question on this paper was Question 6, on the likelihood function for a moving average process. It was particularly disappointing that a number of candidates chose not to make use of the structure for an answer given in the wording of the question itself.

To demonstrate marginal competence in this subject, candidates need to be able to evidence understanding of the statistical and modelling techniques described in the syllabus and expanded upon in the Core Reading, to apply these core techniques to a range of problems, to set out workings that demonstrate a clear structure for problem-solving and to be able to offer some commentary on the answers derived. This requires time in study to understand the statistical principles and methodologies and practice in answering questions across the syllabus.

Application skills questions were generally not answered well. With computer based examinations, these questions have become more important. Candidates should recognise

 © Institute and Faculty of Actuaries

that these are generally the questions which differentiate those candidates with a good grasp and understanding of the subject.

The comments that follow the questions in the marking schedule below concentrate on areas where candidates could have improved their performance. Candidates approaching the subject for the first time are advised to concentrate their revision in these areas.

C. Pass Mark

The Pass Mark for this exam was 55.
978 presented themselves and 195 passed.

Solutions for Subject CS2A – September 2022

Q1
Let 'X' denote Life of the battery
Under, a Weibull distribution
P (X > x) = Exp(- cx ^ gamma) , gamma>0 & c > 0 [½]
Finding the parameters of Weibull distribution
As per Golden book, Weibull distribution has 2 parameters; namely c & gamma
P (X > 400) = 0.70 [½]
Exp(- c* 400 ^ gamma) = 0.70 [½]
c*400 ^gamma = Ln(1/0.70) -- eq 1 [½]
P (X > 810) = 0.50
Exp(- c*810 ^ gamma) = 0.50 [½]
c*810 ^gamma = Ln(1/0.50) -- eq 2 [½]
Eq 2 divided by eq 1 gives
1.94336 = 810 ^gamma / 400 ^ gamma [½]
Ln(1.94336) = gamma * ln (810/400)
Gamma = 0.9417 [1]

Substituting gamma in eq 1 or eq 2 gives
c = Ln(1/0.70) / 400^0.9417 [½]
c = 0.001265 [½]

Therefore
P (X > 1000) = Exp(-0.001265 * 1000 ^ 0.9417) [1]
= 0.42944 [½]

This question was well answered. The question is a straightforward application of the Weibull distribution.

Q2
(i)
X_t is stationary if and only if the roots of the characteristic polynomial

 © Institute and Faculty of Actuaries

1 - *a* * *lambda* - ½ * *lambda*^2
are both greater than 1 in magnitude [1]
For *lambda* = 1 to be a root, *a* = ½ [½]
For *lambda* = -1 to be a root, *a* = -½ [½]
If *a* = 0, then the characteristic polynomial reduces to 1 - ½ * *lambda*^2 [½]
This has roots *lambda* = sqrt(2) and -sqrt(2), which are greater than 1 in magnitude [½]
Stationarity therefore holds for *a* = 0 [1]
Overall, stationarity holds if and only if abs(*a*) < ½ [1]

(ii)
X_t is invertible if and only if the value of *lambda* satisfying
1 + *b* * *lambda* = 0
is greater than 1 in magnitude, [1]
i.e. if and only if -1 / *b* is greater than 1 in magnitude [½]
Hence invertibility holds if and only if abs(*b*) < 1 [½]

(iii)
The condition for *X_t* to be *I*(1) is that the characteristic polynomial should have a root equal to 1
[1]
This is the case if and only if *a* = ½ [1]
[Total 9]

Part (i) was well answered by the majority of candidates. Rather than arguing the values of alpha from lambda = +/- 1, it is also possible to solve the inequality algebraically.

Part (ii) was also well answered.

Part (iii) was not well answered with many candidates omitting the condition required.

Q3
(i)
X_t is stationary since it is an AR(1) process. In particular
E(X_t)=E\sum_0^infinity 0.5^i E(e_{t-i})=0 [½]
Cov(X_t,X_{t-s})= E(\sum_0^infinity 0.5^i E(e_{t-i}),
\sum_0^infinity 0.5^i E(e_{t-s-i})=0+ 0.5^s E(\sum_0^infinity 0.5^i 0.5^i e^2_{t-s-i})
=0.5^s \sigma^2 /(1-0.5^2) [1½]
As Xt=0.5 X_{t-1}+e_t, the distribution of Xt depends on X_{t-1} and only the information at time t-1 and not any other r.v. before that point [1]
Hence we have the Markov property [1]

(ii)
The stationarity of Yt is implied from that of Xt as
E(Y_t)= E(X_t)-0.3 E(X_{t-1})=0 [1]

and

 © Institute and Faculty of Actuaries

Cov(Y_t,Y_{t-s})= Cov(X_t-0.3 X_{t-1},X_{t-s} -0.3 X_{t-s-1})=

Cov(Xt,X_{t-s})-0.3 Cov(Xt,X_{t-s-1}) -0.3 Cov(X{t-1},X_{t-s})+0.09

Cov(X{t-1},X_{t-s-1})

All these four components do not depend on t due to the stationarity of Xt [1]

For the Markov property Yt however this is not the case:
As Xt=0.5 X_{t-1}+e_t, [1]
Y_t=0.5 X_t-1+et-0.3 X_{t-1}=et+0.2 X_{t-1}= et+0.2 *0.5 X_{t-2}+0.2 e_{t-1} [1]

Substituting X_{t-2}=1/0.2*(Y_{t-1}-e_{t-1})
Y_t=0.5 Y_{t-1}+e_t-0.3 e_{t-1} [1]

In this form one can see that the prediction for Y_t depends not only on Y_{t-1} but also on the information contained in e_{t-1} [1]
Hence the Markov property is NOT satisfied [1]

[Marks available 7, maximum 6]
[Total 10]

Overall, this question was not well answered.

For both parts this was often because candidates rushed to a conclusion about whether the Markov property was satisfied rather than setting out structured reasoning. Similarly, in both parts a number of candidates stated that Stationarity held without any reasoning.

In part (i) this reasoning could have either been derived from the Covariance term or by observing that the 0.5 coefficient means that the variance is infinite.

Q4

(i)
q_x is the probability someone age x dies before age x+1 [½]
m_x is the probability of death in that same year per person year lived [½]
m_x is higher than q_x [½]
unless $q_x = 0$ [½]

(ii)
m_x is estimated as Deaths / Central exposed to risk [1]

x	m_x
60	0.007968127
61	0.008917513
62	0.00983837
63	0.010932775
64	0.011707814

[1]

(iii)

 © Institute and Faculty of Actuaries

if the force of mortality is assumed constant in each individual year of age μ_x [½]
then $m_x = \mu_x$ for each x [½]
with constant force of mortality we can use the Exponential model within a year of age [1]
and then chain together five exponential survival probabilities [½]
so ${}_5p_{60}$ = exp(-0.007968127).exp(-0.008917513).exp(-0.00983837).exp(-0.010932775) .exp(-0.011707814) [1]
= 0.951834 [½]

[Total 8]

Part (i) was generally well answered and is straightforward.

Part (ii) was well answered.

Part (iii) was generally well answered. Candidates that did less well in this part often tried to calculate the survival probability direct from the central rates of mortality rather than recognise that a constant force of mortality assumption allows use of the exponential model.

Q5

(i)

Null hypothesis: the graduated rates are the true rates underlying the observed data [½]
Alternative hypothesis: The graduated rates are NOT the true rates underlying the observed data [½]

i.z_x = (Expected deaths – Observed deaths) / sqrt(Expected deaths) [½]

ii.

Age, x	Expected deaths	z_x_	z_x^2
60	78.4	0.1810	0.0327
61	85.4	-0.0392	0.0015
62	91.9	0.8410	0.7073
63	100.0	0.4964	0.2464
64	108.2	-0.0154	0.0002
65	117.7	-0.0626	0.0039
66	125.6	-1.3938	1.9427
67	134.0	-0.2566	0.0659
68	145.0	-0.0004	0.0000
69	159.9	-1.5703	2.4659

[1]

 © Institute and Faculty of Actuaries

The test statistics is $X = \sum_x z_x^2$

Under the null hypothesis, X has a Chi2 distribution with m degrees of freedom, where m is the number of age groups less one for each parameter fitted, so in this case $m = 10 - 3$ [½]

The number of degrees of freedom is therefore 7 [½]

The critical value of the Chi2 distribution with 7 degrees of freedom at the 5% level is 14.067 [½]

The observed value of X is 5.47 [½]

This is lower than the critical value; there is not enough evidence to reject the null hypothesis at 5% level [½]

(ii)

Under the null hypothesis, the test statistic

$$Y = \frac{\sum Observed\ deaths\ -\ Expected\ deaths}{\sqrt{\sum Expected\ deaths}} \sim \mathcal{N}(0,1)$$

[1]

$$Y = \frac{-25.023 + 150 - 10000\hat{\mu}_{70}}{\sqrt{1146.023 + 10000\hat{\mu}_{70}}} = \frac{124.977 - 10000\hat{\mu}_{70}}{\sqrt{1146.023 + 10000\hat{\mu}_{70}}}$$

[1]

We are looking for Y such that $-1.96 < Y < 1.96$ [1]

That is

$$\frac{(124.977 - 10000\hat{\mu}_{70})^2}{1146.023 + 10000\hat{\mu}_{70}} < 1.96^2$$

[½]

$$(124.977 - 10000\hat{\mu}_{70})^2 < 1.96^2 \times (1146.023 + 10000\hat{\mu}_{70})$$

$$10000^2\, \mu_{70}^2 - 2537956\, \mu_{70} + 11216.69 < 0$$

[1]

The roots are 0.0057 and 0.0197 [1]

[Total 11]

Part (i) was very well answered. It is a straightforward application of the Chi-squared test. Candidates are reminded when answering this type of question to set out clearly their hypotheses, the calculation of the test statistic, the reasoning for the number of degrees of freedom, and the conclusion as it relates to the null hypothesis.

© Institute and Faculty of Actuaries

Part (ii) was also well answered. This is more challenging than part (i) but candidates who structured their answer carefully gained the majority of marks available. There are a number of different ways to calculate z_70 here, all of which were credited. However some candidates then failed to relate this z_70 calculation back to the Cumulative Deviations Test statistic.

Q6

(i)

We have $e_1 = x_1 - b * e_0$ [½]

Since $e_0 = 0$ by assumption [½]

$e_1 = x_1$ [½]

Suppose we have shown that $e_t = x_t + \text{Sum}(1,t\text{-}1)[(-b)\text{^}i\ x_t\text{-}i]$ [½]

Then $e_t+1 = x_t+1 - b * e_t$, [½]

i.e. $e_t+1 = x_t+1 - b * (x_t + \text{Sum}(1,t\text{-}1)[(-b)\text{^}i\ x_t\text{-}i])$, [½]

i.e. $e_t+1 = x_t+1 + \text{Sum}(1,t)[(-b)\text{^}i\ x_t\text{-}i]$ [1]

Hence $e_t = x_t + \text{Sum}(1,t\text{-}1)[(-b)\text{^}i\ x_t\text{-}i]$ for $t = 1,2,\ldots,T$

The likelihood function is given by

constant * Product(1,*T*)[1 / sqrt(*sigma*^2) * exp(-½*(*e_t*/*sigma*)^2)] [1]

The log-likelihood function is given by

constant + Sum(1,*T*)[- ½*log(*sigma*^2) - ½ * (*e_t*/*sigma*)^2], [½]

which reduces to the required expression on substituting for *e_t* [½]

(ii)

If $r_1 = 0$, then $bhat = 0$ [½]

Otherwise, *bhat* satisfies the quadratic equation $r_1 * bhat\text{^}2 - bhat + r_1 = 0$ [½]

The roots of this quadratic are (1 - sqrt(1 - 4 * *r_1*^2)) / (2 * *r_1*) [1]

We require the root less than or equal to 1 in magnitude [1]

Thus for $r_1 > 0$, *bhat* = (1 - sqrt(1 - 4 * *r_1*^2)) / (2 * *r_1*) [½]

For $r_1 < 0$, *bhat* = (1 + sqrt(1 - 4 * *r_1*^2)) / (2 * *r_1*) [½]

(iii)

When abs(*r_1*) > ½, there is no real value of *beta* equating the theoretical autocorrelation at lag 1, *beta* / (1 + beta^2), to *r_1* [1]

Consideration should be given to collecting more data, [½]

as it is possible that the observed value of *r_1* has been distorted by sampling error [½]

Otherwise, it should be concluded that the data come from a different type of model [1]

[Total 13]

This question as a whole was very poorly answered. Candidates are reminded of the importance of being able to apply the Core Reading material in this part of the syllabus to unfamiliar situations.

 © Institute and Faculty of Actuaries

In part (i) the most common omission was a structured approach that begins with the relationship between e_t+1 and e_t in terms of x, and then moves to a likelihood function and then its log.

In part (ii) which was now well answered, many candidates failed to take heed of the instruction in the question to consider three different conditions for r separately. As well as the methodology shown above it is also possible to derive the roots algebraically.
In part (iii) many candidates simply stated that the equation from part (ii) has no real solution rather than demonstrating why and then engaging with the possible reasons for this.

Q7

(i)

Using $(u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}$

F_X(3) = 1 - exp(-0.08 * 3) [½]

= 0:2134, [½]

F_Y(8) = P(Z < -1) where Z ~ N(0, 1) [½]

= 0.1587 [½]

$C(0.2134; 0.1587) = (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}$ [½]

$= (0.2134^{-2} + 0.1587^{-2} - 1)^{-½} = 0.1284$ [½]

(ii)

Using $(u^{-\alpha} + v^{-\alpha} + w^{-\alpha} - 1)^{-1/\alpha}$

F_Z(20) = P(Z < 0) where Z ~ N(0, 1) [½]

= 0.5 [½]

$C(0.2134; 0.1587; 0.5) = (u^{-\alpha} + v^{-\alpha} + w^{-\alpha} - 1)^{-1/\alpha}$ [1]

$= (0.2134^{-2} + 0.1587^{-2} + 0.5^{-2} - 2)^{-½} = 0.1253$ [1]

(iii)

F_X(10) = 1 - exp(-0.08 * 10) = 0.5507 [½]

F_Y(12) = P(Z < 1) where Z ~ N(0, 1)

= 0.8413 [½]

$C(0.5507; 0.8413) = (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}$

$= (0.5507^{-2} + 0.8413^{-2} - 1)^{-½} = 0.5191$ [½]

Then using the survival copula, the required probability is

1 - F_X(10) - F_Y(12) + C(0.5507; 0.8413) [1]

= 1 – 0.5507 – 0.8413 + 0.5192 = 0.1271 [½]

(iv)

It is true that part of the benefit of copulas is that they allow us to model different degrees of dependency [1]

However, copulas also exhibit different patterns of dependency [1]

As well as different degrees of dependency in the body of the distributions, copulas exhibit different degrees of dependency in the tails [1]

Copulas allow the dependency structure between random variables to be modelled separately from the marginal distributions [1]

[Marks available 4, maximum 3]

[Total 12]

 © Institute and Faculty of Actuaries

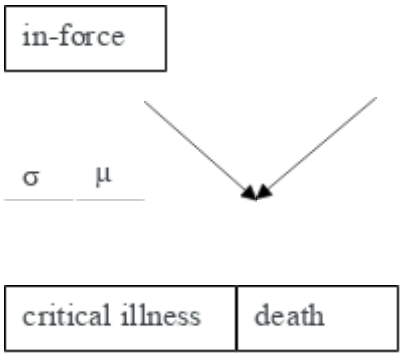
This question was well answered. In all four parts credit was given either for stating the copula algebraically or for being able to evidence the correct copula from the numbers used in calculating the probabilities required.

In part (ii) it is possible to take the answer from part (i) and insert that into a two-part copula rather than derive the three-part copula directly.

Q8

(i)

Three-state model with transition intensities



[2]

where μ is the transition intensity from in-force to death [½]
σ is the transition intensity from in-force to critical illness [½]

(ii)

There are 75000/15000 = 5 critical illness transitions and (115000 – 75000)/10000
= 4 death claims [½]
The waiting time in the in-force state can be estimated by census as
½ (887 + 849) = 868 years [1]

We assume that transition intensities are constant throughout a single year of age [½]
and are not policy duration dependent [½]
The likelihood of the model is $L(\mu,\sigma) = \exp(-(\mu + \sigma)868).\mu 4.\sigma 5$ [1]
and the log-likelihood is $\log L = 4 \log \mu + 5 \log \sigma - 868\mu - 868\sigma$ [½]

The maximum likelihood estimates are found by differentiating with respect to μ and σ respectively, setting to zero and solving
gives
μ_hat = 4/868 = 0.004608 [½]
σ_hat = 5/868 = 0.005760 [½]

(iii)

We need the probability of transition to critical illness
Pr[remain in in-force state] = exp(-0.004608-0.00576) = 0.989685 [½]
Pr[making a transition] = 1 – 0.989685 = 0.010315 [½]
Pr[transition to critical illness] found by ratio of transition intensities [½]
= 0.010315 x 0.005760 / (0.004608+0.005760)
= 0.005731 [½]

 © Institute and Faculty of Actuaries

so expected cost of claims = 15000 x 0.005731 = £85.96 [1]

(iv)
Should be cautious because:
number of transitions is low and a small change in observations would lead to a large change in transition intensities [1]
the census approximation might not be valid since the number of policies might not have varied linearly over the year [1]
the large fall in policy numbers suggests censoring is important [1]
would need to investigate whether this censoring is informative [1]
the assumption that transition rates are constant over the year of age should be validated [1]
the assumption that transition rates do not depend on duration should be validated [1]
[Marks available 6, maximum 4]
[Total 15]

This question was well answered.

In part (i) candidates are reminded of the need to label and define all transition intensities when displaying a two-state or multi-state model.

In part (ii) a number of candidates failed to apply the Census method to the calculation of the waiting time.

The expected cost in part (iii) requires both a survival probability and a transition intensity whereas some candidates only included the latter.

Q9
(i)
There are more data points in the central part of a distribution than in the tails [1]
Therefore, maximum likelihood estimation gives more weight to the central part of the distribution than to the tails [1]
and hence tends to result in estimated distributions that fit the tails more poorly than the central part of the actual distribution [1]
This will result in inaccurate estimates of the probabilities of extreme currency movements [1]
The best-fitting model under maximum likelihood estimation may be inappropriate for other reasons [1]
[Marks available 5, maximum 4]

(ii)
The formula for the GPD CDF is
$G(x) = 1 – (1 + x / (gamma * beta))$ ^ $(-gamma)$. [½]
Hence
$G(0.02) = 1 – (1 + 0.02 / (5.417 * 0.009785))$ ^ (-5.417) [1]
= 0.823 [½]
This is the probability that a loss is less than 4.2% given that it is greater than 2.2% [1]

 © Institute and Faculty of Actuaries

(iii)
The probability that a loss is less than 5% given that it is greater than 2.2% is $G(0.028)$ [½]
= 0.8995. [½]
The probability that a loss is greater than 5% given that it is greater than 2.2% is therefore 1 – 0.8995 = 0.1005
[½]
The estimated probability that a loss is greater than 5% is therefore
0.1005 * 150 / 3,000 [1]
= 0.0050 [½]

(iv)
The number of standard deviations is 0.05 / 0.014 = 3.57 [½]
From the Tables, the required probability is therefore 1 – 0.99982 = 0.00018 [½]
Hence the probability of a loss exceeding 5% is significantly lower under the Normal distribution than under the GPD distribution [½]
Given the expectation of leptokurtic behaviour, there is reason to believe that the GPD distribution is more appropriate than the Normal distribution [1]
Using the Normal distribution would therefore be expected to significantly underestimate the risk of extreme currency movements [½]

(v)
The effect of using a different threshold should be investigated [½]
The goodness of fit of the GPD distribution should be investigated [½]
e.g. by using a chi-squared test [½]
The joint behaviour of movements in different currencies should be investigated [½]
The assumption that the distribution of currency movements remains constant over time should be validated [½]
[Marks available 2½, maximum 2]
[Total 15]

This question was not well answered.

In part (i) a number of candidates (presumably having read on through the whole question) rushed to discuss the Generalised Pareto Distribution and missed the points about the nature of the data set and the limitations of maximum likelihood in this type of scenario.

In parts (ii) and (iii) a number of candidates completed a GPD calculation but then failed to state the probability that was asked for.

Part (iv) was generally well answered.

In part (v) it is important to relate answers back to the scenario given in the question, in this case currency movement data and also to the assumptions made.

[Paper Total 100]

END OF EXAMINERS' REPORT

© Institute and Faculty of Actuaries

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

© 2021 Institute and Faculty of Actuaries