

# **INSTITUTE AND FACULTY OF ACTUARIES**

## **EXAMINATION**

29 September 2020 (am)

### **Subject CS2A – Risk Modelling and Survival Analysis Core Principles**

Time allowed: Three hours and fifteen minutes

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

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

**1** Consider a two-dimensional Gaussian copula function,  $C_{\text{Gauss}}(u_1, u_2)$ , with parameter  $\rho = 0$ :

- (i) Give the solution to the copula function  $C_{\text{Gauss}}(1, 1)$ . [1]
- (ii) Give the solution to the copula function  $C_{\text{Gauss}}(1, 0.2)$ . [1]
- (iii) Give the solution to the copula function  $C_{\text{Gauss}}(0.2, 0.2)$ . [1]
- (iv) Outline how your answers to parts (i), (ii) and (iii) would change if  $\rho = 1$ . [2]

An insurer uses copulas to model the dependencies between various types of claims. A statistical analysis of the insurer's claims shows that a Gumbel copula is a better representation of historic claims data than a Gaussian copula.

- (v) Discuss why the use of a Gaussian copula in a claims model could result in solvency issues for this insurer. [3]
- [Total 8]

**2** (i) Explain why Extreme Value Theory (EVT) models can be useful. [2]

A sports scientist is interested in analysing the probability that the javelin world record may be broken next year and is intending to use EVT to do this.

The sports scientist has obtained data for the distances of all javelin throws from all javelin competitions in 2019. The total number of throws recorded was 3,000. The sports scientist has carried out an EVT analysis using the Generalised Pareto Distribution by selecting only those throws that exceeded 50 metres. This resulted in the longest 150 throws being selected for the analysis.

The following parameters were obtained from the EVT analysis:

$$\begin{aligned}\beta &= 15, \\ \gamma &= 3.\end{aligned}$$

- (ii) Determine the percentage of javelin throws that would be expected to exceed 70 metres next year. [4]
  - (iii) Comment on the limitations of this analysis. [4]
- [Total 10]

3 Consider the following time series model:

$$Y_t = 1.5Y_{t-1} - 0.5Y_{t-2} + 0.7e_{t-1} + e_t$$

where  $e_t$  is a white noise process with variance  $\sigma^2$ .

(i) Express this model in terms of the backwards shift operator  $B$ , simplifying your expression as far as possible. [2]

(ii) Identify the values of  $p$ ,  $d$  and  $q$  for which the model is an ARIMA( $p$ ,  $d$ ,  $q$ ) model. [1]

(iii) Which **one** of the following options represents the correct values of  $\text{Cov}(e_t, \nabla Y_t)$  and  $\text{Cov}(e_{t-1}, \nabla Y_t)$  where  $\text{Cov}()$  is the covariance function and  $\nabla$  is the difference operator?

- A  $\text{Cov}(e_t, \nabla Y_t) = \sigma^2$ ,  $\text{Cov}(e_{t-1}, \nabla Y_t) = 1.2\sigma^2$   
 B  $\text{Cov}(e_t, \nabla Y_t) = 1.7\sigma^2$ ,  $\text{Cov}(e_{t-1}, \nabla Y_t) = 2.55\sigma^2$   
 C  $\text{Cov}(e_t, \nabla Y_t) = 1.8\sigma^2$ ,  $\text{Cov}(e_{t-1}, \nabla Y_t) = 1.6\sigma^2$   
 D  $\text{Cov}(e_t, \nabla Y_t) = 3.4\sigma^2$ ,  $\text{Cov}(e_{t-1}, \nabla Y_t) = 3.4\sigma^2$  [2]

(iv) Which **one** of the following options represents the correct values of the autocovariance function,  $\gamma$ , of  $\nabla Y_t$ ?

- A  $\gamma_0 = 0.5\gamma_1 + 1.84\sigma^2$ ,  $\gamma_1 = 0.5\gamma_0 + 0.7\sigma^2$ ,  $\gamma_k = 0.5\gamma_{k-1}$  for  $k > 1$   
 B  $\gamma_0 = 0.5\gamma_1 + 2.92\sigma^2$ ,  $\gamma_1 = 0.5\gamma_0 + 2.86\sigma^2$ ,  $\gamma_k = 0.5\gamma_{k-1}$  for  $k > 1$   
 C  $\gamma_0 = 0.5\gamma_1 + 3.49\sigma^2$ ,  $\gamma_1 = 0.5\gamma_0 + 1.19\sigma^2$ ,  $\gamma_k = 0.5\gamma_{k-1}$  for  $k > 1$   
 D  $\gamma_0 = 0.5\gamma_1 + 5.78\sigma^2$ ,  $\gamma_1 = 0.5\gamma_0 + 5.78\sigma^2$ ,  $\gamma_k = 0.5\gamma_{k-1}$  for  $k > 1$  [3]

(v) Which **one** of the following options represents the correct values of  $\gamma_0$ , and  $\gamma_k$  for  $k \geq 1$ ?

- A  $\gamma_0 = 2.92\sigma^2$ ,  $\gamma_k = 0.5^{k-1} (2.16\sigma^2)$  for  $k \geq 1$   
 B  $\gamma_0 = 5.44\sigma^2$ ,  $\gamma_k = 0.5^{k-1} (3.91\sigma^2)$  for  $k \geq 1$   
 C  $\gamma_0 = 5.8\sigma^2$ ,  $\gamma_k = 0.5^{k-1} (5.76\sigma^2)$  for  $k \geq 1$   
 D  $\gamma_0 = 11.56\sigma^2$ ,  $\gamma_k = 0.5^{k-1} (11.56\sigma^2)$  for  $k \geq 1$  [2]

[Total 10]

- 4 Zebras in a large African wildlife park have recently become susceptible to a particular disease called Zebra rabies. Zebras often die as a direct result of contracting this disease. An investigation to monitor deaths due to this disease was carried out between 1 January 2018 and 1 January 2019.

A researcher was interested in the rate at which zebras die once contracting this disease and decided to monitor the health of each zebra on the first day of each month. All zebras in the wildlife park were tagged to ensure that they were identifiable.

14 zebras were diagnosed with rabies during 2018. The data recorded on these 14 zebras are set out below:

<i>Reference tag</i>	<i>Date of diagnosis</i>	<i>Date of death</i>	<i>Reason for death</i>
1	1 Jan 2018	1 Jun 2018	Rabies
3	1 Jan 2018	1 Dec 2018	Rabies
4	1 Apr 2018	1 Jul 2018	Killed by lion
7	1 Apr 2018	1 Jun 2018	Rabies
8	1 Apr 2018	1 Dec 2018	Rabies
10	1 Jul 2018	1 Sep 2018	Rabies
11	1 Aug 2018	1 Oct 2018	Rabies
12	1 Aug 2018	1 Nov 2018	Rabies
19	1 Sep 2018	1 Oct 2018	Rabies
20	1 Oct 2018	1 Nov 2018	Rabies

Two zebras (reference tags 9 and 21) escaped from the wildlife park on 1 December 2018 having been diagnosed with rabies on 1 July 2018 and 1 November 2018, respectively.

In addition, the following two zebras that contracted the disease were still alive at the end of the investigation:

<i>Reference tag</i>	<i>Date of diagnosis</i>
13	1 Aug 2018
25	1 Dec 2018

- (i) Discuss whether or not the different types of censoring present in the above investigation are likely to be informative. [3]

- (ii) Determine the Kaplan–Meier estimate of the survival function, where the decrement of interest is death due to rabies. [8]

[Total 11]

- 5 An investigation was undertaken into the mortality of policyholders for a large life insurance company. The crude mortality rates were graduated using a formula of the form:

$$\mu_x = \exp(ax + bx^2).$$

An extract of the results is set out below. All data have been collated between 1 January 2018 and 1 January 2019 on an ‘age last birthday’ basis:

<i>Age</i>	<i>Exposed-to-risk (years)</i>	<i>Observed deaths</i>	<i>Graduated rates</i>
50	23,308	70	0.00368
51	19,316	58	0.00379
52	16,914	54	0.00391
53	21,082	90	0.00402
54	14,820	70	0.00415
55	24,084	96	0.00428
56	28,076	114	0.00441
57	22,958	86	0.00455
58	24,960	102	0.00469
59	21,134	86	0.00485
60	18,374	94	0.00500

- (i) Perform the signs test to explore the hypothesis that the graduated mortality rates are the true rates underlying the observed data from the life insurance company. [5]
- (ii) State the main limitation of the signs test in assessing the suitability of the graduated rates. [1]
- (iii) Perform a chi-square goodness-of-fit test to explore further the hypothesis that the graduated mortality rates are the true rates underlying the observed data from the life insurance company. [4]
- (iv) Comment on your answers to parts (i) and (iii). [2]
- [Total 12]

- 6 An online professional examination paper is sat by a large number of candidates. After the examination, each script is marked electronically by two examiners in sequence.

All scripts are initially sent to a First Marking folder (F) for marking by a first examiner.

Once the first examiner has marked a script, that script is immediately sent to a Second Marking folder (S) for marking by a second examiner.

Once both examiners have marked a script the marks are assessed to determine whether the script is in the ‘borderline’ category. A script is ‘borderline’ if one of the marks is below the pass mark and the other one is above.

If the script is **not** placed in the ‘borderline’ category, it is automatically moved to the Completed folder (C). If the script is placed in the ‘borderline’ category, it is sent to the Borderline folder (B). A third examiner will then mark the ‘borderline’ script and forward it to the Completed folder once finished.

The first and second examiners mark scripts at a rate  $\mu$  per day. The third examiner marks ‘borderline’ scripts at a rate  $\beta$  per day. A proportion  $b$  of the scripts are deemed to be ‘borderline’.

The Kolmogorov forward equation for  $p^{SS}(t)$  is  $\frac{d}{dt}p^{SS}(t) = -\mu p^{SS}(t)$ .

- (i) Which **one** of the following options represents the Kolmogorov forward equation for  $p^{SB}(t)$ ?

- A  $\frac{d}{dt}p^{SB}(t) = b\mu p^{SS}(t) - \mu p^{SB}(t)$   
B  $\frac{d}{dt}p^{SB}(t) = b\mu p^{SS}(t) - \beta p^{SB}(t)$   
C  $\frac{d}{dt}p^{SB}(t) = b\mu p^{BB}(t) - \mu p^{SB}(t)$   
D  $\frac{d}{dt}p^{SB}(t) = b\mu p^{BB}(t) - \beta p^{SB}(t)$

[2]

In a certain examination marking process,  $\mu = 0.2$ ,  $\beta = 0.3$  and  $b = 0.25$ .

- (ii) Which **one** of the following options represents the correct probability that a script now awaiting second marking will be found in the Borderline folder in  $t$  days’ time?

- A  $0.5\exp(-0.2t) - 0.5\exp(-0.3t)$   
B  $0.05t\exp(-0.3t)$   
C  $0.5\exp(-0.2t)$   
D  $0.05t\exp(-0.2t)$

[4]

- (iii) Which **one** of the following options represents the correct probability that a script now awaiting second marking will be found in the Completed folder in 10 days' time?

- A 0.7970
- B 0.8219
- C 0.8398
- D 0.8647

[2]

The Examination Board, who are responsible for overseeing the examination process, wants to investigate ways of speeding up the script marking process. One of the members of the Board has suggested that more examiners could be added to the pool of first and second examiners.

- (iv) Comment on the Board member's suggestion.

[4]

[Total 12]

- 7 The transition matrix,  $P$ , describes a Markov chain for the state space  $S = \{1, 2, 3, 4\}$ :

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

- (i) Explain whether this Markov chain is irreducible. [3]
- (ii) Determine all of the stationary probability distributions,  $\pi$ , for the Markov chain defined above, if any such probability distributions exist. [3]

The hitting probability,  $h_{ij}$ , is defined as the probability of **ever** reaching state  $j$ , starting from the initial state  $i$ .

- (iii) Determine the hitting probabilities for state 4 ( $h_{i4}$ ) for all states  $i$  in  $S$ . [6]
- [Total 12]

- 8 A computer manufacturing company is monitoring the quality of the electronic circuit boards it produces. It has two factories, which are located in Oxford and London. The Oxford factory was opened only 12 months ago whereas the London factory has been producing circuit boards for 10 years.

The owner of the company wants to know whether the quality of production differs materially between the two factories. In addition, the owner wants to test whether a new manufacturing process improves the production quality. Over the last 6 months, half of the circuit boards in both factories have been made using the new process.

An analyst has been employed to test whether the location and the manufacturing process affect the quality of the circuit boards. The quality is measured by recording how long each circuit board lasts before developing a fault. The analyst decides to use the Cox proportional hazards model in the analysis.

The two covariates used are:

Location: London = 0, Oxford = 1,                      Process: Old = 0, New = 1.

- (i) State the type of circuit board that represents the baseline hazard. [1]

During the investigation, 6,000 boards were tested, equally split between the four possible permutations of location and process. The first board to fail was one made in Oxford using the new process. No previous censoring events had occurred.

- (ii) State the term in the partial likelihood expression that relates to this first failure, clearly defining all notation used. [3]

The following coefficients were obtained from the investigation using the Cox proportional hazards model:

Location = 0.01,                      Process = -0.30.

- (iii) Comment on the coefficient values obtained. [2]
- (iv) Identify the additional information that would be required before being able to reach a conclusion on the effects of location and process on the quality of the circuit boards produced. [1]

Prior to the introduction of the new process, the probability that a circuit board made in London failed in the first year was estimated to be 20%. The company plans to make a batch of 10,000 circuit boards next year in London.

- (v) Which **one** of the following options represents the best estimate of the probability that a circuit board made in London using the new process does **not** fail in the first year?

- A      0.5927  
B      0.6965  
C      0.7982  
D      0.8476

[3]

(vi) Which **one** of the following options represents the best estimate of the expected absolute difference in the number of circuit board failures in the first year if the new process is used, compared to if the old process is used?

- A 18
- B 476
- C 1,035
- D 2,073

[2]

[Total 12]

- 9 An insurance company offers annual home insurance policies in partnership with a bank. The distribution deal involves taking part in the bank's loyalty scheme called '1234'. Under '1234', a customer gets a discount when buying or renewing the policy according to how many bank accounts they hold, as follows:

<i>Number of bank accounts</i>	<i>Discount</i>
One	5%
Two	10%
Three	15%
Four or more	25%

An analysis of the data suggests that the transition matrix for the number of bank accounts held at annual intervals is as follows:

$$\begin{array}{l}
 \text{One} \\
 \text{Two} \\
 \text{Three} \\
 \text{Four+}
 \end{array}
 \begin{pmatrix}
 0.5 & 0.2 & 0.2 & 0.1 \\
 0.2 & 0.4 & 0.3 & 0.1 \\
 0.2 & 0.2 & 0.4 & 0.2 \\
 0 & 0.2 & 0.2 & 0.6
 \end{pmatrix}$$

A customer takes out a policy in January 2017 at the 10% discount level.

- (i) Calculate the probability that the customer remains at the 10% discount level for all of their renewals up to and including 2020. [2]

- (ii) Calculate the probability that the customer receives a discount of at least 15% in 2019. [2]

- (iii) Which **one** of the following options represents the correct stationary distribution,  $\pi$ , of the transition matrix above?

- A  $\pi_1 = \frac{34}{45}\pi_3, \quad \pi_2 = \frac{8}{9}\pi_3, \quad \pi_3, \quad \pi_4 = \frac{41}{45}\pi_3$   
 B  $\pi_1 = \frac{34}{45}\pi_3, \quad \pi_2 = \frac{8}{9}\pi_3, \quad \pi_3, \quad \pi_4 = \pi_3$   
 C  $\pi_1 = \pi_3, \quad \pi_2 = \frac{8}{9}\pi_3, \quad \pi_3, \quad \pi_4 = \frac{41}{45}\pi_3$   
 D  $\pi_1 = \pi_3, \quad \pi_2 = \pi_3, \quad \pi_3, \quad \pi_4 = \pi_3$

[3]

- (iv) Which **one** of the following options represents the correct average long-term level of discount?

- A 13.60%  
 B 13.75%  
 C 14.19%  
 D 14.45%

[3]

- (v) Comment on the commercial implications for the insurer of the bank's loyalty scheme. [3]

[Total 13]

**END OF PAPER**