Imperial College London

MATH97075 MATH70048

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)
May-June 2022

This paper is also taken for the relevant examination for the Associateship of the Royal College of Science

Survival Models

Date: 01 June 2022

Time: 09:00 – 11:30 (BST)

Time Allowed: 2:30 hours

Upload Time Allowed: 30 minutes

This paper has 5 Questions.

Candidates should start their solutions to each question on a new sheet of paper.

Each sheet of paper should have your CID, Question Number and Page Number on the top.

Only use 1 side of the paper.

Allow margins for marking.

Any required additional material(s) will be provided.

Credit will be given for all questions attempted.

Each question carries equal weight.

SUBMIT YOUR ANSWERS AS SEPARATE PDFs TO THE RELEVANT DROPBOXES ON BLACKBOARD (ONE FOR EACH QUESTION) WITH COMPLETED COVERSHEETS WITH YOUR CID NUMBER, QUESTION NUMBERS ANSWERED AND PAGE NUMBERS PER QUESTION.

1. Let T be a non-negative random variable with hazard rate

$$h(t) = a\exp(-bt) + c\exp(dt), \quad t \geq 0$$

for some $a, b, c, d > 0$.

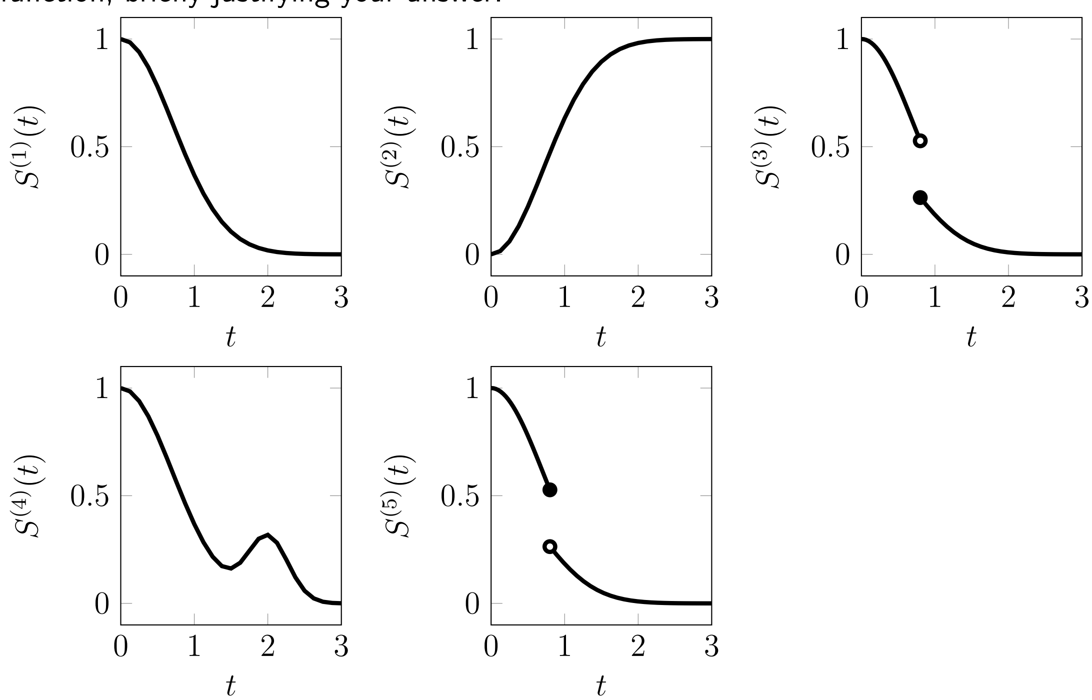
(a) Compute the integrated hazard rate $H(t)$, the survivor function $S(t)$ and the probability density function $f(t)$ of T. (6 marks)

(b) Discuss whether this distribution would be suitable for a mechanical system such as a motor, considering early or late failures. (2 marks)

(c) Give three reasons why distributions of lifetimes are often being specified via the hazard rate and not via the probability density function. Include at least one reason why it might be easier to use the hazard rate. (3 marks)

(d) Can T be written as a (non-trivial) combination of two independent random variables? If it is possible, define those random variables and how they will be combined. (4 marks)

(e) For each of the following five functions $S^{(1)}, \ldots, S^{(5)}$, state whether they could be a survivor function, briefly justifying your answer.



(5 marks)

(Total: 20 marks)

2. (a) State the Kaplan-Meier estimator, defining all variables.

In which situations is the Kaplan-Meier estimator directly related to the empirical cumulative distribution function (empirical CDF)?

Is the Kaplan-Meier estimator a parametric or a non-parametric estimator?

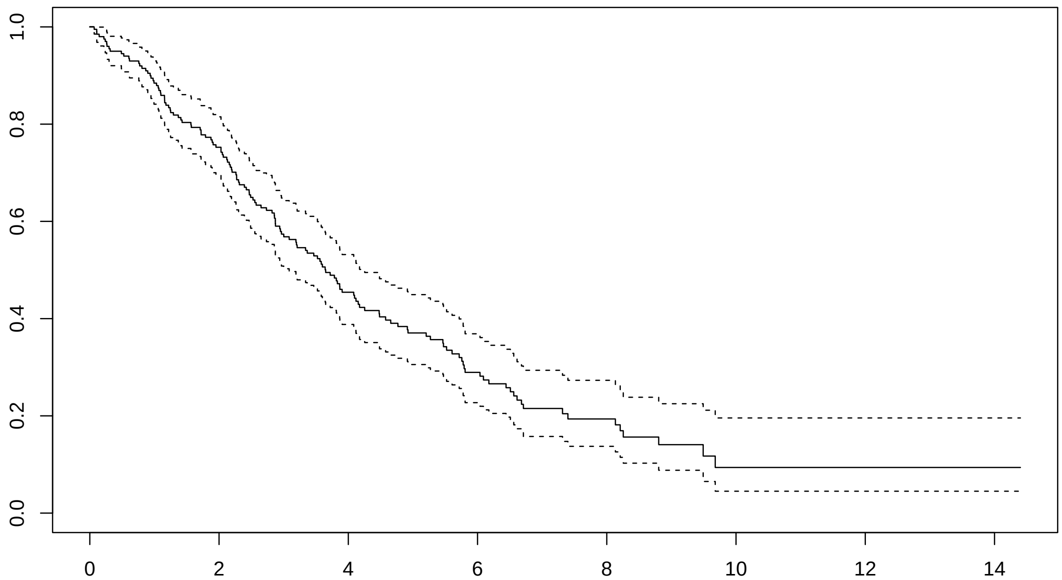
What can Greenwood's formula be used for? (5 marks)

(b) Suppose we observe the following lifetimes, where "+" indicates a right-censored observation.

$$0.1+, 0.3+, 0.7+, 1.1, 1.3+, 1.6+, 1.7, 1.7, 1.9, 2.1+, 2.5+, 3.0$$

Compute and sketch the Kaplan-Meier estimate, paying particularly care to indicate left/right continuity of the estimate. (7 marks)

(c) Below is the standard output of the function *survfit* of R for a data set with right-censoring.



In many situations, it is desirable to condense random variables into one-dimensional summaries. Often, the mean is used to get an idea of the location of random variables.

(i) State how the mean can be computed using the survivor function.

(ii) Discuss whether the mean of lifetimes can be reliably estimated based on the Kaplan-Meier estimator (mention two aspects).

(iii) Is the median a better summary for the location of survival distributions? Give an estimate of the median based on the above Kaplan-Meier estimate (no need to measure anything precisely - a rough estimate is sufficient).

(iv) Are there situations in which the median cannot be estimated? Describe how such a situation can arise.

(8 marks)

(Total: 20 marks)

3. Suppose we observe a data set of size n which contains possibly right-censored lifetimes X_i ($D_i = 1$ is uncensored, $D_i = 0$ is a censored observation), together with one-dimensional covariates z_i, $i = 1, \ldots, n$.

(a) Consider the following parametric model for the hazard rate

$$h(t|z_i) = \lambda \exp(\beta z_i), \quad i = 1, \ldots, n.$$

State which parameters are unknown and give reasonable ranges for the parameters.

Write down the log-likelihood function and simplify it.

Suppose you wish to estimate the parameters via maximum likelihood estimation, but it can be assumed that you cannot optimise the log-likelihood analytically. How would you proceed instead?

Describe one method to test $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$ at the level 5%.

(8 marks)

(b) As an alternative to the model in (a), we could use the semi-parametric Cox model.

Define this model. State how it differs from the model in part (a).

State the partial likelihood function, defining all variables.

(6 marks)

(c) Suppose now that we work with a specific data set of size 500. Using the model from part (a), we get an estimate of β of 0.593 and a standard error of the estimator of 0.067.

Below is the output from the coxph function in R for the same data set.

```
Call:
coxph(formula = Surv(X, Delta) ~ Z)

     coef exp(coef) se(coef)     z      p
Z 1.00424   2.72983  0.08168 12.29 <2e-16

Likelihood ratio test=158  on 1 df, p=< 2.2e-16
n= 500, number of events= 230
```

Compare the results from these two models and give one possible explanation for your observation.

(6 marks)

(Total: 20 marks)

4. (a) Consider the following 5 state Markov model, through which we assume we move via a continuous Markov jump process $X(t)$. All connections not present in the graph have transition rate 0.

$$\begin{tikzcd} 1 \arrow[r] & 2 \arrow[r] \arrow[d] & 4 \arrow[d] \\ & 3 & 5 \end{tikzcd}$$

(i) Which states are absorbing?

(ii) Write down the generator matrix G.

(iii) State the Kolmogorov forward equations for the transition probability $p^{ij}(t)$ (which is defined as in the lecture notes).

(iv) Using the the Kolmogorov forward equations, derive $p^{22}(t)$ and $p^{23}(t)$.

(13 marks)

(b) Consider again the model of part (a) and assume $\mu^{12} = \mu^{45} = 1$, $\mu^{23} = \mu^{24} = 1/2$. Suppose we start in state 1 at time 0. What is the distribution of

$$S := \inf\{t \geq 0 : X(t) \in \{3, 4\}\}?$$

(3 marks)

(c) For $t \geq 0$, let

$$\mathbf{P}_t = \begin{pmatrix} \exp(-\lambda t) & 1 - \exp(-\lambda t) \\ 1 - \exp(-\rho t) & \exp(-\rho t) \end{pmatrix}$$

for some $\lambda, \rho > 0$.

Can $\mathbf{P}_t$ be the transition matrix of a two-state continuous time Markov jump process?

(4 marks)

(Total: 20 marks)

5. This question is based on the following paper:

Balan, T.A. and Putter, H. (2020) 'A tutorial on frailty models', Statistical methods in medical research, 29(11), pp. 3424–3454. doi:10.1177/0962280220921889.

The notation used in this question follows the conventions used in that paper.

(a) Consider a Cox proportional hazard model $\lambda(t|x) = \lambda_0(t)\exp(\beta x)$ with a single covariate x following a standard normal distribution $x \sim N(0,1)$, and with an effect size of $\beta < 0$. Describe how the set of individuals at risk changes over time. (2 marks)

(b) Consider the frailty model in equation (3) of the paper, i.e. $\lambda(t|Z) = Z\lambda(t)$ with a gamma distributed frailty Z with mean 1.

What is the marginal hazard rate $\bar{\lambda}(t)$? How does it compare to $\lambda(t)$? (2 marks)

(c) Suppose we want to model the survival of 100 twin pairs (i.e. 200 individuals), having as only covariate the age of the individuals, where a_i is the age of the ith pair. Write down an appropriate model for this setting, by giving the hazard rate, the range of indices and stating which parameters are to be estimated in the model. (3 marks)

(d) Suppose we want to use the survival package in R to fit a recurrent event model. Consider an individual, who is at risk from time 0 to time 5 and who has events at times 1 and 3.

Name two different encodings for recurrent events data and represent the data about the individual in these encodings. (3 marks)

(e) Consider the frailty model in equation (3) of the paper, i.e.

$$\lambda(t|Z) = Z\lambda(t).$$

Suppose that the frailty Z is discrete-valued with $\mathsf{P}(Z = 2) = \mathsf{P}(Z = 0) = 0.5$.

Compute the marginal survival function $\bar{S}(t)$ and the marginal hazard rate $\bar{\lambda}(t)$, simplifying your answer.

Assuming that $\Lambda(t) \to \infty$ as $t \to \infty$, what is the limiting behaviour of $\bar{S}(t)$ and $\bar{\lambda}(t)$ as $t \to \infty$? Intuitively explain this result. (10 marks)

(Total: 20 marks)

BSc and MSci EXAMINATIONS (MATHEMATICS)

May 2022

This paper is also taken for the relevant examination for the Associateship.

MATH60048/70048/97075

Survival Models (Solutions)

Setter's signature	Checker's signature	Editor's signature
....	....	....

1. (a) $\boxed{\text{meth seen} \Downarrow}$

$$\begin{aligned} H(t) &= \int_0^t h(s)ds = -\frac{a}{b}\exp(-bs) + \frac{c}{d}\exp(ds)\,|_{s=0}^t \\ &= -\frac{a}{b}(\exp(-bt)-1) + \frac{c}{d}(\exp(dt)-1) \end{aligned}$$

$$S(t) = \exp(-H(t)) = \exp\left(\frac{a}{b}(\exp(-bt)-1) - \frac{c}{d}(\exp(dt)-1)\right)$$

$$f(t) = h(t)S(t) = (a\exp(-bt) + c\exp(dt))\exp\left(\frac{a}{b}(\exp(-bt)-1) - \frac{c}{d}(\exp(dt)-1)\right)$$

$\boxed{6, \text{B}}$

(b) The hazard rate can accommodate a "bathtub" shaped distribution of the hazard rate, allowing for early failures (e.g. through production defects) through the first term and then for an increase in the hazard later on to allow for fatigue/degradation over time. $\boxed{\text{unseen} \Downarrow}$

$\boxed{2, \text{C}}$

(c) Possible points mentioned in the lecture notes - three of these should be mentioned, at least one of points (iii) and (v) should be mentioned. $\boxed{\text{seen} \Downarrow}$

(i) It may be physically enlightening to consider the immediate risk.

(ii) Comparisons of groups of individuals are sometimes most incisively made via the hazard.

(iii) Hazard-based models are often convenient when there is censoring.

(iv) When fitting parametric models the form of the hazard function can be enlightening about the assumptions made by the model: e.g. Exponential $\implies$ constant hazard.

(v) The hazard rate does not need to satisfy as many conditions as pdf/CDF.

$\boxed{3, \text{A}}$

(d) The hazard rate of T is the sum of two terms, $a\exp(-bt)$ and $c\exp(dt)$. In one of the problem sheets we have shown that such a random variable can be written as the minimum of two independent random variables S_1 and S_2 with hazard rates $a\exp(-bt)$ and $c\exp(dt)$. $\boxed{\text{meth seen} \Downarrow}$

$\boxed{4, \text{B}}$

(e) $S^{(1)}$ and $S^{(3)}$ are survivor functions - they are decreasing and right-continuous, with values in [0,1].

$S^{(2)}$ and $S^{(4)}$ cannot be survivor functions - $S^{(2)}$ as it is increasing everywhere, $S^{(4)}$ as it is increasing just before $t = 2$.

$S^{(5)}$ cannot be a survivor function as it is not right-continuous. $\boxed{5, \text{A}}$

2. (a)

$$\hat{S}(t) = \prod_{j:t_j \le t} \left(1 - \frac{d_j}{n_j}\right),$$

where $t_1 < t_2 < \ldots < t_k$ are the ordered death times, d_j denotes the number of deaths at time t_j, and n_j is the total number of individuals still 'in view' (or at risk) as we reach time t_j. 2, A

If there is no censoring then $\hat{S}(t) = 1 - \hat{F}(t)$, where $\hat{F}$ is the empirical cdf. 1, A

It is a non-parametric estimator. 1, A

Greenwood's formula can be used to estimate the standard error of the KM-estimate. 1, A

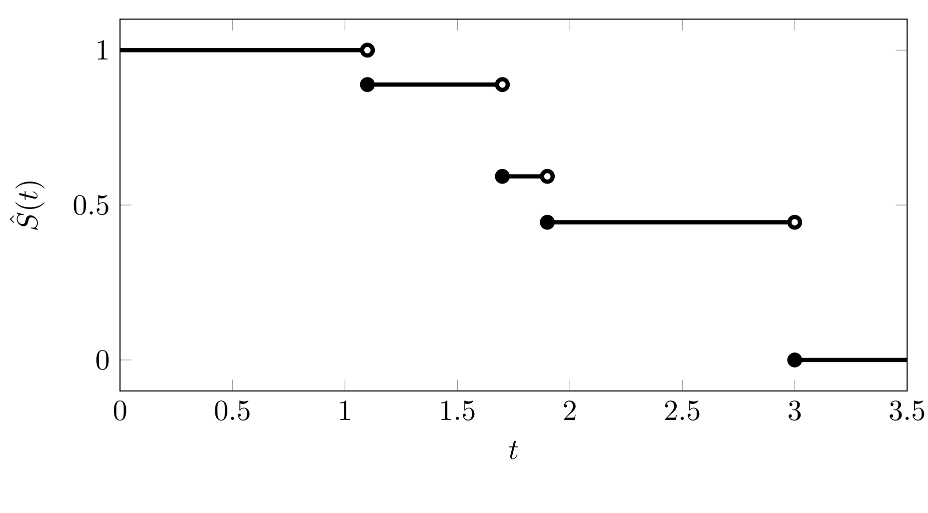
(b) Candidates might use a table of the following type: sim. seen ⇓

j	Death time	Censoring time	n_j	d_j	$\hat{S}(t)$
		0.1			1
		0.3			1
		0.7			1
1	1.1		9	1	8/9
		1.3			8/9
		1.6			8/9
2	1.7		6	2	16/27=$0.\overline{592}$
3	1.9		4	1	4/9
		2.1			4/9
		2.5			4/9
4	3.0		1	1	0

To get full credit, the estimate for all t should be given, e.g. in the following way:

$$\hat{S}(t) = \begin{cases} 1 & \text{for } t < 1.1 \\ 8/9 & \text{for } t \in [1.1, 1.7) \\ 16/27 & \text{for } t \in [1.7, 1.9) \\ 4/9 & \text{for } t \in [1.9, 3.0) \\ 0 & \text{for } t \ge 3.0 \end{cases}$$

4, A



3, B

(c) (i) If $S(t)$ is a survivor function then the mean of a corresponding random variable X is

$$\mathbb{E}[X] = \int_0^\infty S(s)ds.$$

seen ⇓

2, A

(ii) Using the Kaplan-Meier estimate for estimating the mean is problematic: For larger lifetimes it can become unstable (high variance) as few individuals are at risk. If the last observation is censored then the expected mean lifetime would be infinite *(candidates can also say "undefined")*.

unseen ⇓

2, D

(iii) The median does usually not suffer from the above two problems, as one merely needs to evaluate when the survivor function crosses 0.5. In this particular example it crosses roughly between 3 and 4 (*any number in this range is acceptable*).

2, D

(iv) The median can not be estimated if the estimated survivor function does not cross 0.5. Such a situation can arise if there is a large amount of censoring.

2, D

3. (a) λ and β are unknown. A reasonable parameter range for λ is $(0, \infty)$ or $[0, \infty)$, and for β a reasonable range is $\mathbb{R}$. $\boxed{\text{meth seen} \Downarrow}$ $\boxed{2, \text{A}}$

The likelihood is as follows:

$$\begin{aligned} l(\lambda, \beta) &= \sum_{i=1}^{n} D_i \log(h(X_i|z_i)) - \sum_{i=1}^{n} \int_0^{X_i} h(s|z_i)ds \\ &= \sum_{i=1}^{n} D_i(\log \lambda + \beta z_i) - \sum_{i=1}^{n} X_i \lambda \exp(\beta z_i) \\ &= \log \lambda \sum_{i=1}^{n} D_i + \beta \sum_{i=1}^{n} D_i z_i - \lambda \sum_{i=1}^{n} X_i \exp(\beta z_i) \end{aligned}$$

$\boxed{2, \text{B}}$

We first would maximise it numerically to find the maximum likelihood estimator $(\hat{\lambda}, \hat{\beta})$, e.g. using the R function optim. $\boxed{1, \text{A}}$

Both Wald tests and likelihood ratio tests were covered in the course - candidates should describe one of these.

Wald test: The optimiser also returns an approximation of the Hessian matrix $\hat{H}$ at the optimum. An estimate of the Fisher Information matrix is $-\hat{H}$ and an estimate $\hat{\Sigma}$ of the covariance matrix Σ of $(\hat{\lambda}, \hat{\beta})$ is the inverse of the negative of the Hessian matrix, i.e.

$$\hat{\Sigma} = (-\hat{H})^{-1}.$$

A Wald test would then reject the two-sided null hypothesis if $\hat{\beta}/\sqrt{\hat{\Sigma}_{2,2}}$ is not in $(q_{0.025}, q_{0.975})$, where q_x is such that $\mathrm{P}(Y < q_x) = x$ for $Y \sim N(0, 1)$.

Likelihood ratio test: A second optimisation of l under the restriction $\beta = 0$ is needed. This could be done analytically (as this is just estimating the parameter of an exponential distribution) or numerically. Say this yields the estimate λ^*.

If $T = 2(l(\hat{\lambda}, \hat{\beta}) - l(\lambda^*, 0)) > c$, where c is the 95% quantile oft a χ^2_1 distribution, then H_0 is rejected. $\boxed{3, \text{C}}$

(b) *Seen, apart from the comparison to the model from a).*

The semi-parametric Cox-model assumes the hazard rate

$$h(t; z) = h_0(t) \exp(\beta z),$$

where $\beta \in \mathbb{R}$ is an unknown parameter and the baseline h_0 is a completely unspecified function. $\boxed{2, \text{A}}$

The difference to the model in (a) is that model in (a) assumes that the baseline is constant. $\boxed{1, \text{C}}$

The partial likelihood function is

$$L(\beta) = \prod_{i \in U} \frac{\exp(\beta z_i)}{\sum_{j \in R_{t_i}} \exp(\beta z_j)},$$

where U is the set of uncensored observations (death times) and R_{t_i} is the risk set; that is, the set of individuals "at risk" (or "in view", so not dead or censored) at time t_i. $\boxed{3, \text{A}}$

(c) The two methods give very different estimates for β (0.593 for model in part (a) and 1.00424 for the Cox model). Crudely constructed 95% confidence intervals of β in these two models (estimate $\pm$ 2 standard errors) do not overlap, reinforcing this statement. $\boxed{\text{unseen} \Downarrow}$ $\boxed{3, \text{B}}$

One key difference between the models is that the model in part (a) assumes a constant hazard rate, whereas the semi-parametric Cox model makes no such assumption. Hence, it could be that the baseline hazard is indeed time-dependent and that thus the data has not been sampled from a model of the type describe in part (a). $\boxed{3, \text{D}}$

4. (a) (i) States 3 and 5 are absorbing states. $\boxed{\text{seen/sim.seen} \Downarrow}$

(ii) $\boxed{1, \text{A}}$

$$G = \begin{pmatrix} -\mu^{12} & \mu^{12} & 0 & 0 & 0 \\ 0 & -\mu^{23}-\mu^{24} & \mu^{23} & \mu^{24} & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\mu^{45} & \mu^{45} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$\boxed{3, \text{A}}$

(iii) The forward equations state $\boxed{\text{seen} \Downarrow}$

$$\frac{d}{dt}p^{ij}(t) = \sum_k p^{ik}(t)\mu^{kj}.$$

$\boxed{1, \text{A}}$

(iv) (*the calculations below are identical to the two-decrement model, which was a question on one of the problem sheets during the course.)*
So for $p^{22}(t)$,

$$\begin{aligned}\frac{d}{dt}p^{22}(t) &= \sum_{k=1}^{5} p^{2k}(t)\mu^{k2} = \mu^{12}p^{21}(t) + \mu^{22}p^{22}(t) \qquad (\text{since } \mu^{32} = \mu^{42} = 0) \\ &= \mu^{22}p^{22}(t) \qquad (p^{21}(t) = 0 \text{ since (1) cannot be reached from (2)}) \\ &= -(\mu^{23}+\mu^{24})p^{22}(t)\end{aligned}$$

$\boxed{2, \text{B}}$

This differential equation can be solved as follows:

$$p^{22}(t) = Ce^{-(\mu^{23}+\mu^{24})t} = e^{-(\mu^{23}+\mu^{24})t}$$

where $C = 1$ since we need $p^{22}(0) = 1$. $\boxed{2, \text{C}}$
For $p^{23}(t)$,

$$\begin{aligned}\frac{d}{dt}p^{23}(t) &= \sum_{k=1}^{5} p^{2k}(t)\mu^{k3} \\ &= p^{22}(t)\mu^{23} \qquad (\text{since } \mu^{13} = \mu^{33} = \mu^{43} = \mu^{53} = 0) \\ &= \mu^{23}e^{-(\mu^{23}+\mu^{24})t}\end{aligned}$$
$$\implies p^{23}(t) = -\frac{\mu^{23}}{\mu^{23}+\mu^{24}}e^{-(\mu^{23}+\mu^{24})t} + C.$$

Substituting $t = 0$, we get $0 = C - \frac{\mu^{23}}{\mu^{23}+\mu^{24}}$ and hence

$$p^{23}(t) = \frac{\mu^{23}}{\mu^{23}+\mu^{24}}\left[1 - e^{-(\mu^{23}+\mu^{24})t}\right].$$

$\boxed{4, \text{C}}$

(b) From the course, we know that the holding time in a state i is exponentially distributed with rate $-\mu^{ii}$. S can be written as the sum of two such holding times. Both of which are exponentially distributed with rate 1. unseen ⇓

2, D

The sum of two Exponential distributed random variables with rate 1 is a Gamma distribution with rate parameter 1 and shape 2 (or an Erlang distribution with rate 1 and shape 2). 1, D

(A precise statement would involve conditioning on the transition time to state 2 and then arguing conditionally using the Markov property. This is not required here.) unseen ⇓

(c) No. The Chapman-Kolmogorov equations $\mathbf{P}_{s+t} = \mathbf{P}_s \mathbf{P}_t$ need to be satisfied. The following example (with $s = t = 1$) shows that they are not satisfied:

$$(\mathbf{P}_1 \mathbf{P}_1)_{11} = \exp(-2\lambda) + (1 - \exp(-\lambda))(1 - \exp(-\rho)) > (\mathbf{P}_2)_{11}$$

4, D

5. (a) *(The article discusses this on p. 3425 for $\beta > 0$ - candidates need to adapt this to the case $\beta < 0$ - no marks will be awarded without this.)*

Individuals with smaller values of x have a *higher* hazard. At time $t = 0$, the mean and variance of x are 0 and 1, respectively. Because individuals with higher hazards tend to be the first to die, as time passes, the risk set progressively comprises individuals with *higher* values of x. For this reason, the average value of x *increases* and the sample variance of x decreases among the individuals at risk over time. $\boxed{2}$

(b) *(Candidates need to put equation (6) on page 3429 and the displayed equation in the middle of page 3431 together.)*

$$\bar{\lambda}(t) = \lambda(t)\mathbb{E}[Z|T \geq t] = \lambda(t)\frac{\theta}{\theta + \Lambda(t)}$$

As $\Lambda(t) \geq 0$ and $\theta > 0$, we have $\bar{\lambda}(t) \leq \lambda(t)$. $\boxed{2}$

(c) Example answer using the notation of Section 3.2.1 of the article:

This can be modelled as a frailty model with $N = 100$ clusters with $n = 2$ individuals in each cluster (the twins). A Cox-model can be used to account for the covariate, so overall the hazard rate could be

$$\lambda_{ij}(t|Z_i) = Z_i \exp(\beta a_i)\lambda_0(t), \quad i = 1, \ldots, 100, \ j = 1, 2.$$

where β, $\lambda_0(t)$ and any parameters of the frailty distribution are the unknown parameters. One can use a gamma distribution with mean 1 for the frailty variables Z_i, which would give one parameter for the frailty distribution. $\boxed{3}$

(d) As described in Section 4.2, this event history is represented by a collection of observations via tuples (tstart, tstop, status).

If the calendar time convention is used, then this would results in

tstart	tstop	status
0	1	1
1	3	1
3	5	0

If the gap time convention is used then this would be

tstart	tstop	status
0	1	1
0	2	1
0	2	0

$\boxed{3}$

(e) The Laplace transform of Z is

unseen $\Downarrow$

$$\mathcal{L}(c) = \mathbb{E}[\exp(-cZ)] = \mathsf{P}(Z=0)\exp(-0) + \mathsf{P}(Z=2)\exp(-2c) = \frac{1}{2}(1+\exp(-2c))$$

3

Thus, using the second displayed equation on p.3430 of the article, we get

$$\bar{S}(t) = \frac{1}{2}(1+\exp(-2\Lambda(t))).$$

1

By the third displayed equation on the same page, we have

$$\bar{\lambda}(t) = -\frac{\mathcal{L}'(\Lambda(t))}{\mathcal{L}(\Lambda(t))}\lambda(t).$$

As $\mathcal{L}'(c) = -\exp(-2c)$ we get

$$-\frac{\mathcal{L}'(c)}{\mathcal{L}(c)} = 2\frac{\exp(-2c)}{1+\exp(-2c)} = \frac{2}{\exp(2c)+1}$$

and thus

$$\bar{\lambda}(t) = \frac{2}{\exp(2\Lambda(t))+1}\lambda(t).$$

2

As $t \to \infty$ (assuming $\Lambda(t) \to \infty$), we get

$$\bar{\lambda}(t) \to 0 \text{ and } \bar{S}(t) \to \frac{1}{2}.$$

2

The population of individuals consist of two equally sized groups - those with $Z = 2$, which have an increased hazard rate by a factor of 2, and those with $Z = 0$, which have a hazard rate of 0. As t increases, more and more of the individuals with $Z = 2$ experience the event, thus leaving mostly those with $Z = 0$. These never experience the event - and thus the marginal hazard rate should go to 0. All individuals with $Z = 2$ experience the event and those with $Z = 0$ never experience the event - as the groups are of the same size the marginal survivor function should converge to 1/2. 2

Review of mark distribution:
Total A marks: 32 of 32 marks
Total B marks: 20 of 20 marks
Total C marks: 12 of 12 marks
Total D marks: 16 of 16 marks
Total marks: 80 of 80 marks

If your module is taught across multiple year levels, you might have received this form for each level of the module. You are only required to fill this out once for each question.

Please record below, some brief but non-trivial comments for students about how well (or otherwise) the questions were answered. For example, you may wish to comment on common errors and misconceptions, or areas where students have done well. These comments should note any errors in and corrections to the paper. These comments will be made available to students via the MathsCentral Blackboard site and should not contain any information which identifies individual candidates. Any comments which should be kept confidential should be included as confidential comments for the Exam Board and Externals. If you would like to add formulas, please include a separate pdf file with your email.

ExamModuleCode	QuestionNumber	Comments for Students
Survival Models_MATH60048 MATH97075 MATH70048	1	Generally well answered, particulary parts (a), (c), (d). In part (b) the ability of the distribution to accommodate a decreasing initial rate for "early" failures due to eg production faults was often not discussed. In part (d), candidates often did not discuss how the two random variables need to be combined.
Survival Models_MATH60048 MATH97075 MATH70048	2	Parts (a) and (b) were generally fairly well done, though only a handful of students were awarded full marks. For (a), marks were most often lost for giving an incorrect answer to when the KM estimator is directly related to the empirical CDF. For (b), marks were most often lost for failing to specify the estimate for all t (i.e. not just at death/censoring times). Across the cohort, part (c) was reasonably well-done, though there were common mistakes that led to marks being dropped. For part (i), a significant number of students gave the mean as an integral with the -t S'(t) in the integrand; for full marks, this should have been developed to give the integrand in terms of S(t). Another common mistake in part (i) was to give an expression for the sample mean, whereas the question discussed the "location of random variables", i.e. the population mean. Most students managed to get a reasonable estimate of the median of the given data, although some students mistakenly used the data from part (b) to calculate a median. For those that reached part (iv), a common mistake was to suggest that the median could not be calculated when a discontinuity in the survival function crossed over 0.5, i.e. where S(t) is right-continuous, but not left-continuous at its median; this is incorrect, and missed the key point, that the survival function could stay above 0.5 for the whole observation period.
Survival Models_MATH60048 MATH97075 MATH70048	3	For Q3, part (a) was generally done quite well. For the final component of part (a), marks were awarded for a correct description of a suitable test statistic, for its sampling distribution and for a correct description of when the corresponding test could be rejected; quite a few students dropped at least one of these marks. For part (b), a surprising number of students approached this by copying a significant chunk of their notes, which they deemed to have the relevant information; often, however, this gave insufficient direct comparison between the models in part (a) and (b). Only a few candidates achieved high marks for part (c) - many students failed to adequately compare the uncertainty of the two beta estimates (using the provided standard errors) and, whilst a reasonable number gave a passing reference to the difference in baseline hazards, a much smaller subset of these candidates discussed this in the context of the data being fitted; such a discussion would have obtained full marks.

Survival Models_MATH60048 MATH97075 MATH70048	4	Question 4 was reasonably well done. Part (a) was executed correctly by most students, with few issues; many students got full marks here. Amongst those who lost marks in part (a), a common mistake was failing to provide correct, or sufficiently clear, working in part (iv). Part (b) provided the greatest conceptual challenge, though a reasonable number of students (perhaps half) managed to establish that S was a sum of two holding times, each of which was exponentially distributed; of those who got this far, some missed out on the final mark by failing to name the resulting distribution of S as a Gamma distribution. Part (c) was generally quite well done. A good proportion of students proceeded to check the CK equations; of those that followed this approach, those who lost marks generally failed to state values of s,t for which the CK equations do not hold (they do hold for the trivial s=t=0).
Survival Models_MATH60048 MATH97075 MATH70048	5	Some of the candidates did very well, whilst others struggled a bit with this question. Some reoccuring errors/mistakes a) many got this right. Some said that the variance of x of those at risk increases over time - it does not (but this did not cost any points if the rest of the answer was correct). b) Sometimes the specifics of the Gamma distribution was not plugged in / used. c) Sometimes this was not adapted enough from the general frailty model to the specific situation. d) Sometimes the gap time convention missed the third tuple and sometimes the tuples were not explained at all. e) Sometimes not attempted at all, probably due to a lack of time.