

UNIVERSITY OF LONDON  
BSc and MSci EXAMINATIONS (MATHEMATICS)  
May 2006

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

**M3S14/M4S14**

**Survival Models and Actuarial Applications**

Date: Tuesday, 23rd May 2006

Time: 2 pm – 4 pm

Credit will be given for all questions attempted but extra credit will be given for complete or nearly complete answers.

Calculators may not be used.

Statistical tables will not be available.

1. Let  $T_x$  be the unknown future lifetime of an individual currently aged  $x$ .

- (i) (a) For  $t > 0$ , define the cumulative distribution and survivor functions of  $T_x$ ,  ${}_tq_x$  and  ${}_tp_x$ . Give an expression for the force of mortality  $\mu(x+t)$  in terms of  ${}_tp_x$ .
- (b) Show that for  $s, t > 0$ ,  ${}_tp_{x+s} = \frac{{}_{s+t}p_x}{{}_sp_x}$ .
- (c) If the force of mortality is assumed to be a constant function over  $[x, x+1)$ , show that  ${}_tp_x = (p_x)^t$  for  $0 \leq t < 1$ , where  $p_x \equiv {}_1p_x$ .
- (ii) Consider a study to investigate mortality of individuals between the ages  $x$  and  $x+1$ , where individual  $i$  is observed over the interval  $[x+a_i, x+b_i)$ ,  $0 \leq a_i < b_i \leq 1$ ,  $i = 1, \dots, n$ . Suppose the following data have been obtained from five individuals inside the age range  $[x, x+1)$ :

Individual ( $i$ )	$a_i$	$b_i$	$d_i$
1	0	1	0
2	0	1	1
3	0	0.7	0
4	0.2	1	0
5	0.3	0.8	0

where  $d_i = 1$  if individual  $i$  died on the study, otherwise  $d_i = 0$ .

- (a) Assuming a binomial model for the data, find the joint likelihood function in terms of the survivor functions  ${}_tp_s$ .
- (b) By assuming a constant force of mortality, simplify the likelihood to a function of the single parameter  $p_x$ .
- (c) Find the maximum likelihood estimate for  $p_x$ .
- (d) Suppose a further three individuals were to enter the study, one aged  $x$  and the other two both aged  $x + \frac{1}{2}$ , and all could be observed until the end of the unit interval  $[x, x+1)$ . Estimate the probability that all three individuals will be alive at age  $x+1$ .

2. A clinical trial is conducted to investigate the efficacy of a new drug for angina, which is to be used to accelerate pain relief in the period immediately following an attack. To measure improvement against existing recovery rates, some of the patients in the trial are randomly allocated to a control group where an existing treatment is given; the rest are given the new drug. Suppose the following recovery times in minutes are observed for the two groups:

New drug: 1.5 1.8 2.2 5.0

Control: 2.4 2.8 3.3

- (i) For each treatment group, sketch the nonparametric maximum likelihood estimate of the survivor function.
- (ii) (a) State the *proportional hazards* model for time to event data in the presence of a vector of explanatory covariates  $z$ , assuming a baseline hazard function  $\mu_0(t)$ . By considering individuals in the control group as baseline individuals, briefly explain how this model can be applied to the data above.
- (b) For a set of ordered event times  $t_1 < t_2 < \dots < t_n$  with no censoring, and corresponding covariate vectors  $z_1, z_2, \dots, z_n$ , write down the partial likelihood function under the proportional hazards model for inference on covariate effects. Why is this only a partial likelihood?
- (c) Write down an expression for the partial likelihood for the clinical trial data above in terms of a suitably chosen parameter.
- (d) Maximising the partial likelihood gives an estimate of 1.37 for the multiplicative effect on the hazard function when receiving the new drug. Interpret this result, and also suggest how this inference may have been affected by the fact that the magnitude of the highest recovery time for the new drug has not been explicitly taken into account.
- (e) Using an estimate of the effect of the new drug, state a revised estimate for the baseline survivor function which makes use of all of the data. Suggest one advantage and one disadvantage of using this function as an estimate for the control group survivor function in comparison with the corresponding plot in (i).

3. Let  $\{X(t) : t \geq 0\}$  be a continuous-time  $N$ -state homogeneous Markov process, and let  $\{\mathbf{P}_t : t \geq 0\}$  be the transition probability matrices with  $ij$ th entry  $p^{ij}(t) = \Pr(X(s+t) = j | X(s) = i), \forall s \geq 0$ .

- (i) (a) Define the intensity functions and generator matrix for  $\{X(t) : t \geq 0\}$ . What property does each row of the generator necessarily possess?
  - (b) Explain in terms of the generator what it means for a state  $i$  to be an *absorbing state*.
  - (c) Prove the Chapman Kolmogorov equations  $\mathbf{P}_{s+t} = \mathbf{P}_s \mathbf{P}_t$  for all  $s, t \geq 0$ .
- (ii) Suppose we have the following generator matrix for a three state Markov process:

$$\mathbf{G} = \begin{pmatrix} \mu^{11} & \mu^{12} & 0 \\ \mu^{21} & \mu^{22} & \mu^{23} \\ 0 & 0 & 0 \end{pmatrix},$$

where the values  $\mu^{ij}$  represent non-zero entries of  $G$ . We will label the three states as “Safe”, “At Risk” and “Dead” respectively, and refer to this process as the *Safe–At Risk–Dead* model. Suppose that any realisation of the process begins life in state 1.

- (a) Draw a diagram of the Safe–At Risk–Dead model, indicating the possible transitions. Write a sentence to describe the nature of each of the states.
- (b) Given the process is in state 2, give, without proof, the respective jump probabilities of moving to states 1 or 3 when the next transition occurs.
- (c) Suppose we observe a single individual from the Safe–At Risk–Dead model for a fixed period and observe  $m > 0$  transitions. By considering the two cases where  $m$  is odd or even, write down expressions for the conditional probability, given  $m$  transitions occurred, that the individual entered the absorbing state.
- (d) A study observes 20 lives following this model, and in total 27 transitions of state occur; five of the individuals eventually enter the absorbing state, and of the remaining 15, five do not experience any transitions, seven experience one transition and three experience two transitions. The total times spent by individuals in states 1 and 2 are 10 and 20 years respectively. Estimate the transition intensities of the model using these data.

4. The lifetime of a particular electronic component is believed to follow an exponential distribution with unknown rate parameter.

Recording the time until failure of five such components yielded the following data, where a number followed by a '+' indicates a right-censoring time:

0.7, 2.2+, 4.4, 2.6, 0.1.

- (i) (a) Explain what is meant by right-censoring, and describe the contribution to the likelihood function from a right-censored survival time.
- (b) Starting with the assumption of a constant hazard rate  $\mu(t) = \mu$ , derive the probability density function and the survivor function for an exponentially distributed random variable.
- (c) Find an expression for the maximum likelihood estimator for  $\mu$  and evaluate the estimate for these data.
- (ii) We could consider the exponentially distributed failure times of the components as being the event times of a counting process  $\{N(t) : t \geq 0\}$  with intensity function  $\lambda(t)$ . Let  $Y(t)$  be the number of components which have not failed or been censored at time  $t$ .
- (a) State without proof an equation for the failure intensity in this case, in terms of  $Y(t)$  and  $\mu$ .
- (b) Sketch estimates of  $\lambda(t)$  and the cumulative intensity  $\Lambda(t) = \int_0^t \lambda(s)ds$  for the data.
- (c) Briefly explain why this model is different from the homogeneous Poisson process.

5. Write brief essays on *three* of the following topics:

- (i) Counting processes for survival data.
- (ii) Comparison of Markov, Poisson and Binomial models for mortality data.
- (iii) Hypothesis tests for choosing between nested models in survival analysis.
- (iv) Nonparametric estimation of the cumulative hazard function and its uses.
- (v) Censoring.