Imperial College London

UNIVERSITY OF LONDON

BSc and MSci EXAMINATIONS (MATHEMATICS)

May-June 2005

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

M3S14/M4S14 SURVIVAL ANALYSIS AND ACTUARIAL APPLICATIONS

Date: Tuesday, 31st May 2005 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. (a) Consider a positive real-valued continuous random variable T_x, representing the unknown future lifetime of an individual currently aged x.

(i) Define the survivor function $S_{T_x}(t)$ and the hazard function $\mu(x)$ of T_x.

(ii) Explain what is meant by a realisation of T_x being right-censored at time t, and why the likelihood for such an observation is given by $S_{T_x}(t)$.

(iii) Show that

$$-\frac{d}{dt}S_{T_x}(t) = S_{T_x}(t)\mu(x+t).$$

(Note that $-\frac{d}{dt}S_{T_x}(t)$ is the density of T_x)

(iv) Show using the result of (iii) that the survivor function can be written as

$$S_{T_x}(t) = \exp\{-M_x(t)\}$$

where $M_x(t) = \int_0^t \mu(x+u)du$ is the cumulative hazard function for an individual currently aged x.

(b) Consider the following sequence of survival times for a set of individuals measured from birth:

0.1+, 0.3, 0.3, 0.5+, 0.6, 0.9, 0.9, 1+, 1.4

where numbers followed by a "+" indicate right-censored values.

(i) Calculate the Nelson-Aalen estimate of the cumulative hazard function for these data, presenting the results in a table. Sketch a graph of the estimated function.

(ii) Briefly outline how an alternative estimate of the cumulative hazard function which maximises the likelihood of the data could have been calculated.

(iii) Describe how a graph of either estimate of the cumulative hazard might be used to select a suitable parametric model for the data from a set of possible candidates.

2. In an actuarial investigation we often wish to assess the mortality rates within the integer age groups $[x, x+1)$, $x \in \{0, 1, 2, \ldots\}$. Fixing on a particular integer age x, suppose we have survival data on n individuals observed on intervals $[x + a_i, x + b_i)$, $0 \leq a_i < b_i \leq 1$, $i = 1, \ldots n$, where the data consist of pairs (D_i, V_i) for individual i where $D_i \in \{0, 1\}$ is an indicator equal to 1 if and only if the individual died during the interval and V_i is the individual's time on the study.

(a) Initially assuming $\forall i, a_i = 0, b_i = 1$, describe the Binomial model for estimating the probability q_x of death in one year for a random individual currently aged x, and derive the maximum likelihood estimate under this model.

(b) Explain briefly how the Binomial model can be extended to the general case $0 \leq a_i < b_i \leq 1$ using the Balducci assumption

$$_{1-t}q_{x+t} = (1-t)q_x \qquad 0 \leq t \leq 1$$

where $_tq_x$ is the probability of death within time t for a random individual currently aged x. [*Note* $q_x \equiv {}_1q_x$.]

(c) Show that under the Balducci assumption in (b), the following hyperbolic interpolation rule holds

$$\frac{1}{_tp_x} = \frac{t}{_1p_x} + \frac{1-t}{_0p_x} \qquad 0 \leq t \leq 1$$

where $_tp_x$ is the survivor function for an individual aged x, satisfying $_tp_x = 1 - {}_tq_x$.

(d) Using the result from (c), or otherwise, show that the Balducci assumption leads to a decreasing rate of mortality.

(e) Suppose the following data have been collected for the integer age x:

Individual i	a_i	b_i	d_i
1	0.1	1	0
2	0	0.8	0
3	0.2	1	1
4	0	0.4	1

Find expressions for the likelihood of each observation in terms of q_x under the Balducci assumption.

3. The general N-state Markov model assumes each individual on a study is a realisation of a homogeneous, continuous time Markov process $\{X(t) : t > 0\}$ taking values in $\{1, 2, \ldots, N\}$. Defining the transition probabilities

$$p^{ij}(h) = P(X(t+h) = j | X(t) = i), \qquad h, t \geq 0,$$

for states $i, j \in \{1, 2, \ldots, N\}$, the model assumes

$$p^{ij}(dt) = \mu^{ij} dt + o(dt) \qquad dt \geq 0$$

where μ^{ij} is a constant *transition intensity* between states i and j.

(a) State and prove the Chapman-Kolmogorov equations.

(b) Using the Chapman-Kolmogorov equations, verify the Kolmogorov backward equations

$$\frac{d}{dt} p^{ij}(t) = \sum_{k=1}^{N} \mu^{ik} p^{kj}(t)$$

(c) A life assurance company models the lifetimes of its policy holders as a three state Markov model with transitions indicated by the diagram

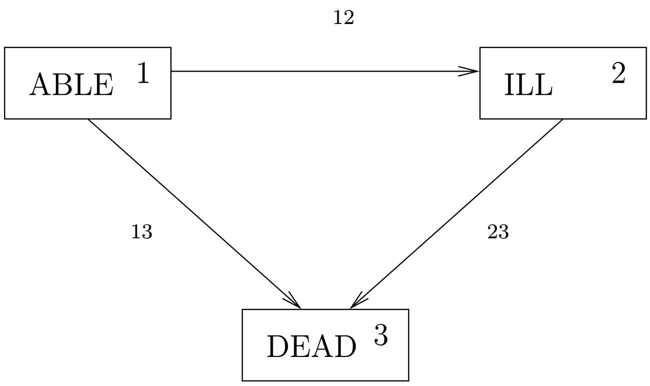

Figure 1: One-way Able-ill-dead model.

where state 2 represents a chronic illness from which transitions back to the able state 1 are not expected (such as contracting the HIV virus), and thus differs from the standard 'Able-ill-dead' Markov model.

(continued next page)

(i) Derive the following equations for the transition probabilities

$$\begin{aligned}
\frac{d}{dt}p^{12}(t) &= \mu^{12}\exp(-\mu^{23}t) - (\mu^{12}+\mu^{13})p^{12}(t) \\
\frac{d}{dt}p^{13}(t) &= -\mu^{12}\exp(-\mu^{23}t) + (\mu^{12}+\mu^{13})(1-p^{13}(t)) \\
p^{23}(t) &= 1-\exp(-\mu^{23}t)
\end{aligned}$$

Suppose the following state change data have been gathered from this adjusted "Able-ill-Dead" Markov model: 10 transitions from state 1 to state 2; 1 death from state 1; 5 deaths from state 2. Additionally, suppose we have a total time spent in states 1 and 2 for the individuals on our study as 20 years and 5 years respectively.

(ii) Write down the likelihood function for these data.

(iii) Find the maximum likelihood estimates for the three transition intensities, and give an expression for an estimate of the probability of an individual currently in state 2 surviving for at least one more year.

4. (a) State the assumptions of the *proportional hazards* (PH) and *accelerated failure time* (AFT) models in the presence of explanatory covariates z, assuming a baseline hazard function μ_0 and baseline survivor function S_0. Also explain the partial likelihood method for inference on the covariate coefficients under PH.

(b) Show that any positive scalar multiple of a Weibull random variable is also a Weibull random variable.

(c) Using the result from (b) show that if the baseline hazard function is that of a Weibull random variable then there is an equivalence between the assumptions of the PH and AFT models.

(d) Suppose the following survival time data have been gathered on patients following a heart bypass operation:

$4.22, 6.2+, 6.23, 6.6, 10.44+, 15.99$

where numbers followed by a "+" indicate right-censored values. Suppose also that we have recorded as a potentially explanatory variable the corresponding number of cigarettes smoked on average per week by each individual:

$60, 0, 40, 100, 0, 0$

(i) Find the partial likelihood function for the covariate regression coefficient under the proportional hazards model.

(ii) Numerical maximisation of the expression found in (i) yields a maximum likelihood estimate of 0.017 with corresponding log likelihood -3.439.
What is the interpretation of this value on the effect of smoking on future lifetime for heart bypass patients?

(iii) By considering the partial likelihood of the data under a null model where the patients are homogeneous with no covariate effects (the regression coefficients fixed at zero), perform a hypothesis test to investigate whether the number of cigarettes smoked is a statistically significant predictor of future lifetime for these individuals. Comment on your findings. [*Note you may use the fact that the 90th percentile point of* χ^2_1 *is 2.71, and that* $\log(72) \approx 4.277$]

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

(a) Censoring in survival data.

(b) The Poisson model as a survival process.

(c) The Kaplan-Meier estimate and Greenwood's formula.

(d) Comparison of the Binomial and 2-state Markov models.

(e) Choosing a parametric model for survival data.