Imperial College London

UNIVERSITY OF LONDON

BSc and MSci EXAMINATIONS (MATHEMATICS)

May-June 2007

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

M3S14/M4S14

Survival Models and Actuarial Applications

Date: Monday, 4th June 2007 Time: 10 am – 12 noon

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.

1. Let T be a continuous random variable with probability density function $f(t)$, where T represents the unknown lifetime of a newborn individual. Suppose $f(t)$ is defined in such a way that T only has positive probability density across the interval $[0, \omega]$ so that ω acts as a limiting age for T.

(i) (a) Define what is my meant by the survivor function of T, and list the properties required by a continuous function S in order for it to act as the survivor function for T.

(b) Consider the random variable $S(T)$, the survivor function evaluated at the unknown lifetime value T. Using carefully the properties of S you gave in part (i)(a), show that $S(T) \sim \text{Uniform}[0, 1]$.

(ii) Suppose T has probability density function

$$f(t) = ce^{-\lambda t}, \quad t \in [0, \omega],$$

where $\lambda > 0$ and c is a normalising constant.

(a) Find an expression for the value of the normalising constant c.

(b) Derive expressions for the survivor, hazard and cumulative hazard functions for T.

(c) How does the hazard function behave here as a function of time, and how is this affected as ω is increased? Briefly comment.

2. A machine has two independent, identical components organised in parallel, such that only one of the two components needs to be working in order for the machine to still work. Each component lifetime exhibits the 'lack of memory property', meaning that for times $s, t > 0$,

$$\mathsf{P}(\text{Component broken at } s + t | \text{Component not broken at } s) = \mathsf{P}(\text{Component broken at } t).$$

The status of the machine, which always begins life with both components working, can thus be viewed as 3-state Markov model schematically represented by the following diagram.

We are interested in the time to failure of the machine.

For a machine currently in state i, let $p^{ij}(t)$ be the transition probability of being in state j after t units of time have passed.

(i) Write down the generator matrix for this process.

(ii) Using the 'lack of memory property' stated above, show that the component lifetimes have constant hazard λ and hence follow an exponential distribution.

(iii) Starting from the result from part (ii), prove that the holding times in the non-absorbing states follow exponential distributions. Hence give formulae for the transition intensities μ^{12} and μ^{23} in terms of the component hazard λ, and identify a formula for $p^{23}(t)$.

(iv) Again starting from the result from part (ii), derive the cumulative distribution function for the machine lifetime, and hence find $p^{13}(t)$.
Verify your equation for $p^{13}(t)$ using the Kolmogorov backward equations.

3. Consider a mortality study of individuals falling between the ages x and $x+1$ for some $x > 0$, where individual i can only be observed across the sub-interval $[x+a_i, x+b_i)$, $0 \leq a_i < b_i \leq 1$, $= 1, \ldots, n$. Let ${}_tq_x$ be the probability that a life currently aged x dies within time t, and D be the n-vector of death indicators for the individuals on the study.

(i) (a) State the Binomial model of mortality, and explain the need for a smoothing method in order to perform maximum likelihood estimation under this model.

(b) Show how an assumption of uniform distribution of deaths within $[x, x+1]$ implies an increasing force of mortality.

(c) Under the assumption of uniform distribution of deaths within $[x, x+1]$, show that for $s, t > 0, s + t \leq 1$,

$$ {}_tq_{x+s} = \frac{tq_x}{1 - sq_x}, $$

where $q_x \equiv {}_1q_x$.

(ii) Consider the following data gathered for individuals aged between 119 and 120 years old.

Individual (i)	Age on entry into study	Age on exit from study	Death indicator d_i
1	119	120	0
2	119	120	1
3	119	119.9	1
4	119	119.8	1

(a) Derive the maximum likelihood estimate for q_{119} for these data under the assumption of uniform distribution of deaths.

(b) Calculate the actuarial estimate of q_{119}.

4. Suppose we have an inhomogeneous population of individuals whose differences in lifetime distribution are attributable to a vector of measurable covariates z.

Let T_0 be the lifetime random variable of a possibly hypothetical 'baseline' individual with covariates $z = 0$. Write $\mu_0(t)$ for the hazard function and $S_0(t)$ for the survivor function of T_0. We seek to relate the distribution of the lifetime T_z of an individual with covariates z to that of T_0, via a suitably chosen positive-valued function $\psi(z)$.

(i) (a) Define the proportional hazards (PH) and accelerated failure time (AFT) models which include covariate effects in survival analysis, in terms of T_z, T_0, $\psi(z)$ and $\mu_0(t)$. For both models state equations showing the form of the generalised survivor function $S(t; z)$ in terms of the baseline survivor function for T_0, $S_0(t)$.

(b) Suppose that the baseline hazard is of the form

$$\mu_0(t) = \begin{cases} 1, & t \leq 2 \\ \frac{1}{2}, & t > 2 \end{cases} .$$

Sketch the hazard function for two groups with either $\psi(z) = 1$ (baseline) or $\psi(z) = 2$ under both the PH and AFT models. Comment on the difference between the hazards of the two models.

(ii) Now suppose it is known that $T_0 \sim$ Weibull(α, η) for known parameters α and η. Hence T_0 has survivor and hazard functions

$$S_0(t) = \exp(-(t/\alpha)^\eta),$$
$$\mu_0(t) = \eta\alpha^{-\eta}t^{\eta-1}.$$

Suppose also that the covariate z is a scalar binary variable taking value 0 or 1, where $\psi(0) = 1$ and $\psi(1) = \psi$ for some $\psi > 0$.

(a) Show that under the Weibull baseline model given above, the PH and AFT models both imply Weibull survival models for an individual with covariate z.

(b) Under this scenario where the baseline model is fully understood, suppose we then observe a sample of data from the covariate $z = 1$ group (there is nothing to learn about the $z = 0$ group as this distribution is known). Assume that there is no censoring and that the sample is of size n.
Without performing the maximization, find a simplified equation for the likelihood ratio statistic that could be used in a hypothesis test to investigate whether the single covariate z has a significant bearing on the lifetime of an individual under the AFT model. State the null distribution against which the test statistic should be compared.

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

(i) Censoring mechanisms.

(ii) Comparison of Markov, Poisson and Binomial models for mortality data.

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

(iv) Nonparametric estimation of the cumulative hazard function and applications.

(v) Counting processes for survival data.