UNIVERSITY OF LONDON

IMPERIAL COLLEGE LONDON

BSc and MSci EXAMINATIONS (MATHEMATICS)
MAY–JUNE 2003

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

M3S14 (SOLUTIONS) SURVIVAL ANALYSIS AND ACTUARIAL APPLICATIONS

DATE: Thursday, 5th June 2003 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)* The hazard function, $\mu(t)$, records the instantaneous event rate, or force of mortality given by

seen ⇓

$$\mu(t) \equiv \lim_{h \to 0^+} \frac{1}{h} Pr(T \leq t + h | T > t)$$

2

b) The hazard function can be written as

$$\mu(t) = \frac{f_T(t)}{\int_t^\infty f_T(s) ds},$$

where $f_T(t)$ is the probability density function of the random event time T. 1

c) Some reasons for considering the hazard function include,

(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: for example Exponential, Weibull etc.

2

d) i) The Kaplan-Meier estimate of the survivor function $\widehat{S}(t)$ is given by

unseen ⇓

$$\widehat{S}(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i}$$

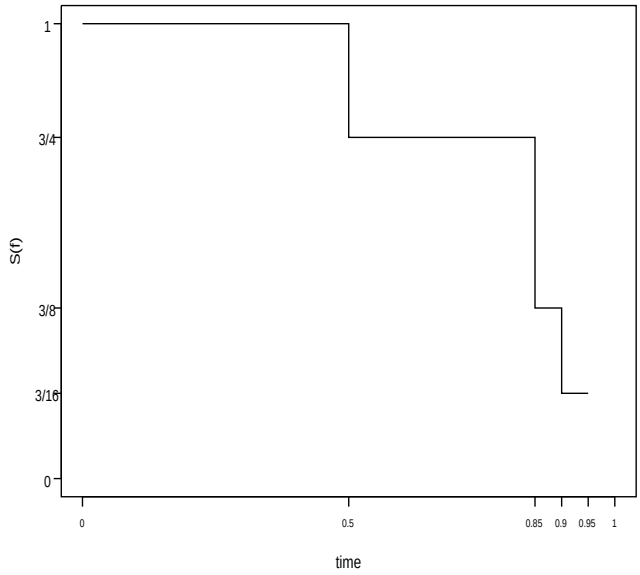
for unique death times t_i and the variance, by Greenwood's formula is

$$\widehat{\text{Var}}\widehat{S}(t) = [\widehat{S}(t)]^2 \sum_{t_i < t} \frac{d_i}{n_i(n_i - d_i)}$$

8

Death time	Censored time	n_j	d_j	$\widehat{S}(t)$	$\widehat{\text{Var}}\widehat{S}(t)$
-	0.4	-	-	1	0
0.5	-	8	2	$\frac{3}{4}$	$\left(\frac{3}{4}\right)^2 \frac{1}{24}$
-	0.6	-	-	$\frac{3}{4}$	$\left(\frac{3}{4}\right)^2 \frac{1}{24}$
-	0.8	-	-	$\frac{3}{4}$	$\left(\frac{3}{4}\right)^2 \frac{1}{24}$
0.85	-	4	2	$\frac{3}{8}$	$\left(\frac{3}{8}\right)^2 \frac{7}{24}$
0.9	-	2	1	$\frac{3}{16}$	$\left(\frac{3}{16}\right)^2 \frac{19}{24}$
-	0.95	-	-	$\frac{3}{16}$	$\left(\frac{3}{16}\right)^2 \frac{19}{24}$

ii) Note from the sketch below that the survivor function is undefined past the last observation, $t = 0.95$, as this is censored.



3

iii) Some of the assumptions made by the model include

seen ⇓

- that the lives are independent and identically distributed.
- that the censoring mechanism is independent to the mortality process.

2

iv) The term "product limit" refers to the derivation of the Kaplan-Meier as the function implied by the maximum likelihood estimate of the cumulative hazard function defined on a finer and finer partition of the positive half-line, $\{[0, dt), [dt, 2dt), \ldots\}$ as $dt \to 0^+$.

2

2. *a) i)* The study is a Type II censoring procedure, where individuals are observed until a specified number are seen to fail, the remaining individuals being right censored. Individuals lost to follow up during the study are right censored. The patients who have failed prior to the start of the study are left censored.

seen ⇓

3

ii) **A.** The left censored observations (at 2 months) contribute terms, $F_T(2)$, to the likelihood - where $F_T(t)$ denotes the distribution function.

B. The right censored observations contribute terms $1 - F_T(t) \equiv S_T(t)$, where $S_T(t)$ is the survivor function and t is the censoring times.

C. The observations seen to fail contribute, $\frac{d}{dt}F_T(t) \equiv f_T(t)$, where $f_T(t)$ is the probability density function evaluated at the observed failure time $T = t$.

3

b) i) Breslow's approximation to the partial-likelihood function is given by

$$L(\beta) = \prod_{i \in U} \frac{\exp(\beta s_i')}{[\sum_{j \in R_i} \exp(\beta z_j')]^{d_i}}$$

where U is the set of k unique event times $U = \{t_1, \ldots, t_k\}$, d_i denotes the number of events at time t_i, s_i is the sum of covariates z_j for events at time t_i and R_i denotes the "risk set" of those observations still in view at time t_i.

3

ii) Hence,

unseen ⇓

$$L(\beta) = \frac{e^\beta}{2(e^\beta + 1)} \frac{1}{(e^\beta + 2)^2} \frac{e^\beta}{e^\beta}$$

with log-likelihood

$$l(\beta) = \beta - \log(1 + e^\beta) - 2\log(2 + e^\beta) - \log(2)$$

To find the maximum we note

$$\frac{dl}{d\beta} = 1 - \frac{e^\beta}{1 + e^\beta} - \frac{2e^\beta}{2 + e^\beta}$$

Let $x = e^{\beta}$ then at maximum we have

$$0 = 1 - \frac{x}{1+x} - \frac{2x}{2+x} = \frac{1}{1+x} + \frac{4}{2+x} - 2$$

So,

$$\frac{d^2 l}{dx^2} = -\frac{1}{(1+x)^2} - \frac{4}{(2+x)^2} < 0$$

Hence this is a maximum. So,

$$\begin{aligned} (1+x)(2+x) - x(2+x) - 2x(1+x) &= 0 \\ 2x^2 + x - 2 &= 0 \end{aligned}$$

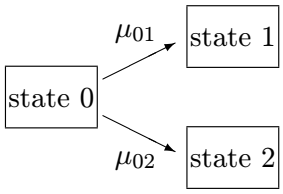
Hence, due to positivity constraints, $x = e^{\beta}$, we find

$$\widehat{\beta} = \log\left(\frac{-1+\sqrt{17}}{4}\right)$$

The estimtate $\widehat{\beta} < 0$ implies that smoking has a relatively decreasing effect on the hazard function. 8

iii) To test for the significance of an extra term we would fit two models: one with the single parameter and one with both parameters recording the log-likelihood of both models. Under the null hypothesis that the extra parameter has value zero the deviance (twice the difference in the log-likelihoods) follows a chi-squared distribution on one degree of freedom. If the deviance is large we would reject the null hypothesis at some level of significance. seen ⇓ 3

3. *a)* The 3-state competing risk model is, seen ⇓



2

b) The three assumptions are that

Assumption 1 - *Markov assumption*: The probabilities that a life at any given age will be found in any state at any subsequent age depends only on the ages involved and on the state currently occupied.

$$P(X(t+s)|X(0), X(1), \ldots, X(t)) = P(X(t+s)|X(t))$$

where $X(t)$ denotes the state of the process at time t.

Assumption 2 : For $X(t) = i$ various things can happen in the small time interval $(t, t+dt)$.

- The chain can remain in state 0 with probability $p_{00}(t, dt) + o(dt)$
- The chain may move to another state j with probability $p_{0j}(t, dt) + o(dt)$

where $\frac{o(dt)}{dt} \to 0$ as $dt \to 0^+$. For small dt, the transition probability $p_{0j}(t, dt)$, $j \in 1, 2$ is approximately linear in dt with constant of proportionality μ_{0j},

Assumption 3 : The model is an homogeneous chain so that

$$\begin{aligned} p_{0j}(t, dt) &= \mu_{0j}dt + o(dt) \qquad j = 1, 2 \\ p_{00}(t, dt) &= 1 + \mu_{00}dt + o(dt) \end{aligned} \tag{1}$$

where $\mu_{00} = -\mu_{01} - \mu_{02}$.

5

c) i) For our homogeneous competing risk model, the likelihood function factor for life i can be written as

$$l(\mu_{01}, \mu_{02}|V_i = v_i, D_i = d_i) = f_V(v_i)p_{0d_i}(dt)$$

where the first term on the right hand side is simply the probability density for the waiting time in state 0 denoted $p_{\bar{0}\bar{0}}(t)$. From assumption 2 above, for small dt

$$p_{\bar{0}\bar{0}}(t + dt) = p_{\bar{0}\bar{0}}(t)(1 + \mu_{00}dt) + o(dt)$$

and hence,

$$\frac{d}{dt}p_{\bar{0}\bar{0}}(t) = p_{\bar{0}\bar{0}}(t)\mu_{00}$$

which leads to

$$p_{\bar{0}\bar{0}}(t) = \exp[-t(\mu_{01} + \mu_{02})]$$

so that the waiting time is exponentially distributed. Hence the likelihood function for $(V_i = v_i, D_i = 0)$ is,

$$l(\mu_{01}, \mu_{02}|V_i = v_i, D_i = 0) = \exp[-v_i(\mu_{01} + \mu_{02})]$$

and,

$$\begin{aligned} l(\mu_{01}, \mu_{02}|V_i = v_i, D_i = 1) &= p_{\bar{0}\bar{0}}(v_i)p_{01}(dt) \\ l(\mu_{01}, \mu_{02}|V_i = v_i, D_i = 2) &= p_{\bar{0}\bar{0}}(v_i)p_{02}(dt) \end{aligned}$$

from assumption 3,

$$\begin{aligned} l(\mu_{01}, \mu_{02}|V_i = v_i, D_i = 1) &= \exp[-v_i(\mu_{01} + \mu_{02})]\mu_{01} \\ l(\mu_{01}, \mu_{02}|V_i = v_i, D_i = 2) &= \exp[-v_i(\mu_{01} + \mu_{02})]\mu_{02} \end{aligned}$$

The likelihood for the full data set is then,

$$l(\mu_{01}, \mu_{02}) = exp[-v^+(\mu_{01} + \mu_{02})](\mu_{01})^{\#01}(\mu_{02})^{\#02}$$

where $\#0j$ denotes the number of transitions between state 0 and state j and v^+ denotes the total waiting time in state 0.
This yields the maximum likelihood estimates as,

$$\widehat{\mu}_{01} = \frac{\#01}{v^+}, \qquad \widehat{\mu}_{02} = \frac{\#02}{v^+}.$$

7

ii) The probability $p_{0j}(t)$ is given by the product

part seen ⇓

$$p_{0j}(t) = \Pr(\text{jumps to } j \mid \text{jump})\Pr(\text{jumps during } t)$$

where Pr(jumps to j given a jump) is

$$\begin{aligned} \Pr(\text{jumps to } j \mid \text{jump}) &= \frac{\mu_{0j}dt}{1 - (1 + \mu_{00}dt)} \\ &= \frac{\mu_{0j}}{\mu_{01} + \mu_{02}} \end{aligned}$$

and Pr(jumps during t) = 1 - Pr(no jump during t),

$$\Pr(\text{jumps during } t) = 1 - \exp[-(\mu_{01} + \mu_{02})t]$$

hence,

$$p_{0j}(t) = \frac{\mu_{0j}}{\mu_{01} + \mu_{02}}[1 - \exp(-\mu_{01} + \mu_{02})t)$$

4

iii) From the formula derived in (c)(i) we find, $\widehat{\mu}_{01} = 0.8$ and $\widehat{\mu}_{02} = 1.2$. From (c)(ii) we see,

unseen ⇓

$$\begin{aligned} \widehat{p}_{01}(\log 2) &= \frac{2}{5} \times \left[1 - \frac{1}{2^2}\right] = 0.3 \\ \widehat{p}_{02}(\log 2) &= \frac{3}{5} \times \left[1 - \frac{1}{2^2}\right] = 0.45 \end{aligned}$$

2

seen ⇓

4. *a)* The three assumptions for the Binomial model are

a. The uniform distribution of deaths:

$$\theta_x(t) = t\theta_x(1) \qquad 0 \le t \le 1$$

b. The Balducci assumption

$$\theta_{x+t}(1-t) = (1-t)\theta_x(1) \qquad 0 \le t \le 1$$

c. The constant force of mortality

$$\theta_x(t) = 1 - \exp(-\mu t)$$

5

b) To begin we recall the identity relating to the survival functions $(1 - \theta_x(t))$, unseen ⇓

$$(1-\theta_x(1)) = (1-\theta_x(a))(1-\theta_{x+a}(b))(1-\theta_{x+c}(1-c))$$

where $c = a + b$ and $0 \le a \le b < 1$. Hence,

$$\theta_{x+a}(b) = 1 - \frac{1-\theta_x(1)}{(1-\theta_x(a))(1-\theta_{x+c}(1-c))}$$

Under the assumption of *uniform distribution of deaths* we have

$$\theta_x(a) = a\theta_x(1)$$

and to find $\theta_{x+c}(1-c)$ we note from the identity above

$$\begin{aligned} 1-\theta_{x+c}(1-c) &= \frac{1-\theta_x(1)}{1-\theta_x(c)} \\ &= \frac{1-\theta_x(1)}{1-c\theta_x(1)} \end{aligned}$$

Thus, 5

$$\begin{aligned} \theta_{x+a}(b) &= 1 - \frac{1-\theta_x(1)}{(1-\theta_x(a))(1-\theta_{x+c}(1-c))} \\ &= 1 - \frac{1-c\theta_x(1)}{1-a\theta_x(1)} \end{aligned}$$

which for $a = 0.2, b = 0.2, c = 0.4$ gives

$$\theta_{x+0.2}(0.2) = \frac{0.02}{0.98}$$

and for $a = 0.4, b = 0.2, c = 0.6$ gives

$$\theta_{x+0.4}(0.2) = \frac{0.02}{0.96}$$

Under the Balducci assumption we have

$$\theta_{x+c}(1-c) = (1-c)\theta_x(1)$$

and to find $\theta_x(a)$ we note

$$\begin{aligned}\theta_x(a) &= 1 - \frac{1-\theta_x(1)}{1-\theta_{x+a}(1-a)} \\ &= 1 - \frac{1-\theta_x(1)}{1-(1-a)\theta_x(1)}\end{aligned}$$

Thus,

$$\begin{aligned}\theta_{x+a}(b) &= 1 - \frac{1-\theta_x(1)}{(1-\theta_x(a))(1-\theta_{x+c}(1-c))} \\ &= 1 - \frac{1-(1-a)\theta_x(1)}{1-(1-c)\theta_x(1)}\end{aligned}$$

which for $a = 0.2, b = 0.2, c = 0.4$ gives

$$\theta_{x+0.2}(0.2) = \frac{0.02}{0.94}$$

and for $a = 0.4, b = 0.2, c = 0.6$ gives

$$\theta_{x+0.4}(0.2) = \frac{0.02}{0.96}$$

4

c) The results show that the uniform distribution of death implies an increasing force of mortality (hazard rate) while the opposite is true for the Balducci assumption.

seen ⇓

2

d) The actuarial estimate is given by

unseen ⇓

$$\widehat{\theta}_x(1) = \frac{d}{v + 0.5d}$$

where d is the number of deaths and v the total observation time on the population. Hence, given $\theta_x(1) = 0.1$, $d = 80$ we obtain $v = 760$. For the Poisson model of deaths we have

$$Pr(D = d) = \exp(-\mu v)\frac{(\mu v)^d}{d!}$$

where μ is the intensity rate and v the total observation time. This leads to the maximum likelihood estimate as

$$\widehat{\mu} = \frac{d}{v}$$

which from above is

$$\widehat{\mu} = \frac{80}{760} \quad = \frac{2}{19}$$

4

5. I would hope to see some or all of the following issues mentioned.

For graduation and adherence to standard tables.

- The reasons for graduation (smoothing) of interval estimated mortality.
- The models used for graduation, for example, parametric models and parametric fitting to a standard table.
- The range of statistical tests to compare observed data to those published in standard tables. Chi-squared tests, groupings of signs tests, bias detection.
- The type of features that each test is designed to detect.

The Binomial and 2-state Markov models.

- Both methods are used by actuaries to model mortality in the time interval, $[x, x+1)$.
- Binomial model estimates the mortality rate, the Markov model estimates the force of mortality (hazard).
- The Binomial requires approximations to allow for inference when observations are made on the sub-interval.
- The Markov model is easily extended to more complex scenarios involving multiple decrements and increments while the Binomial model is not easily extended.
- The Binomial model uses an estimator crudely based on a method of moments to estimate the mortality rate using the "Actuarial Estimate". The Markov model uses a probabilistic likelihood based approach.

The Nelson-Aarlen estimate.

- The Nelson-Aarlen estimate (NAE) is a non-parametric estimate of the integrated hazard function.
- The NAE is an isotonic (monotone increasing) step function with jumps at the observed death times.
- The form of the NAE should be given.
- The NAE is derived as the maximum likelihood estimate of the integrated hazard function.

Greenwood's formula

- The use of Greenwood's formula as an estimate of the variance in the estimated product-limit nonparametric survivor function.
- An expression of the formula should be given.
- The approximations made in the derivation of the formula, using the delta rule.
- Either a derivation or some mention of the estimate based on the variance of a binomial random variable.

A comparison of proportional hazards (PH) and accelerated time to failure models.

- Expressions for the models should be provided.
- The implications of both models for the survivor function should be given.
- The estimation procedures for parameters in the model should be discussed. In particular the semi-parametric feature of the PH model should be highlighted.
- Methods for model checking should be discussed.
- Interpretation of the parameter values in terms of relative survival times for two individuals with different covariates.