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 (SOLUTIONS) SURVIVAL ANALYSIS AND ACTUARIAL APPLICATIONS

Date: Tuesday, 31st May 2005 Time: 2 pm – 4 pm

1. (a) (i) The survivor function

$$S_{T_x}(t) = P(T_x > t)$$

is the probability of an individual currently aged x surviving beyond time t.
The hazard function at age x

$$\mu(x) = \lim_{h \to 0^+} \frac{P(T_x \le h)}{h}.$$

(ii) A realisation of T_x is right-censored at time t if we do not observe the exact value of T_x and learn only that t provides a lower bound for T_x.
The probability of such an observation is given by

$$P(T_x > t) = S_{T_x}(t).$$

(iii)

$$\begin{aligned}
-\frac{d}{dt} S_{T_x}(t) &= -\lim_{h \to 0^+} \frac{P(T_x > t+h) - P(T_x > t)}{h} \\
&= \lim_{h \to 0^+} \frac{P(T_x > t) - P(T_x > t+h)}{h} \\
&= P(T_x > t) \lim_{h \to 0^+} \frac{1 - P(T_x > t+h | T_x > t)}{h} \\
&= P(T_x > t) \lim_{h \to 0^+} \frac{1 - P(T_{x+t} > h)}{h} \\
&= P(T_x > t) \lim_{h \to 0^+} \frac{P(T_{x+t} \le h)}{h} \\
&= S_{T_x}(t)\mu(x+t).
\end{aligned}$$

(iv)

$$\begin{aligned}
\frac{d}{dt} S_{T_x}(t) &= -S_{T_x}(t)\mu(x+t) \\
\Rightarrow S_{T_x}(t) &= \exp\left\{ -\int_0^t \mu(x+u)du \right\} + C
\end{aligned}$$

for some positive constant C. Using the boundary condition at $t = 0$, $S_{T_x}(0) = 1$ (the individual will survive a further time $t = 0$ with probability 1), we get $C = 0$ and

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

(b) (i) The *Nelson-Aalen* estimator is given by

$$\tilde{M}(t) = \sum_{t_j \le t} \frac{d_j}{n_j}.$$

Death Time	Censoring Time	n_j	d_j	$\tilde{M}(t)$
	0.1	9	0	0
0.3		8	2	0.25
	0.5	6	0	0.25
0.6		5	1	0.45
0.9		4	2	0.95
	1	2	0	0.95
1.4		1	1	1.95

Table 1: Nelson Aalen estimate.

Figure 1: Nelson Aalen estimate.

(ii) The maximum likelihood estimate of the survivor function is given by the *Kaplan-Meier* estimator

$$\hat{S}(t) \quad = \quad \prod_{t_j \leq t} \left(1 - \frac{d_j}{n_j}\right).$$

By the result of (a)(iv) this can be used to provide the maximum likelihood estimate of the cumulative hazard function

$$\hat{M}(t) = \exp\left\{-\hat{S}(t)\right\}.$$

(iii) A plot of $\hat{M}(t)$ or $\log(\hat{M}(t))$ against t or $\log(t)$ can be compared with the known theoretical properties of $M(t)$ for a particular parametric model. For example,

- $M(t)$ vs. t is linear for the exponential distribution.
- $\log M(t)$ vs. $\log t$ is linear for the Weibull distribution.
- $\log M(t)$ vs. t is linear for the Gompertz-Makeham distribution.

2. (a) The simple Binomial model assumes that each individual has the same probability q_x of death, so the ith life is a Bernoulli trial

$$P[D_i = d_i] = {q_x}^{d_i}(1-q_x)^{(1-d_i)}.$$

The lifetimes of the individuals are assumed independent and so defining $D = \sum_{i=1}^{n} D_i$ to be the number of individuals who died on the study, the Binomial model implies $D \sim$ Binomial(n, q_x).
That is,

$$P(D = d) = \binom{n}{d} q_x^d (1-q_x)^{(n-d)}.$$

To find the maximum likelihood estimator of q_x, we differentiate the log-likelihood

$$\begin{aligned}\log(P(D=d)) &= \log\left\{\binom{n}{d}\right\} + d\log(q_x) + (n-d)\log(1-q_x)\\ \frac{d}{dq_x}\log(P(D=d)) &= \frac{d}{q_x} - \frac{n-d}{1-q_x}.\end{aligned}$$

Finding the root of the equation

$$\begin{aligned}0 &= \frac{d}{q_x} - \frac{n-d}{1-q_x}\\ \Rightarrow d - dq_x &= (n-d)q_x\\ \Rightarrow d &= nq_x\end{aligned}$$

yielding the maximum likelihood estimator of q_x

$$\hat{q_x} = \frac{D}{n}.$$

(b) Under the general assumption $0 \le a_i < b_i \le 1$ the individual death probabilities will depend on (a_i, b_i); individual i will die with probability ${}_{b_i - a_i}q_{x+a_i}$
The ith life is therefore a Bernoulli trial

$$P[D_i = d_i] = ({}_{b_i-a_i}q_{x+a_i})^{d_i}(1 - {}_{b_i-a_i}q_{x+a_i})^{(1-d_i)}.$$

To relate these individual probabilities to the unit interval probability q_x, the equation for the Balducci assumption can be used to derive formulae for the probabilities ${}_{b_i-a_i}q_{x+a_i}$ in terms of q_x. This enables us to proceed with estimation of the single parameter q_x.

(c) We have ${}_tp_x = 1 - {}_tq_x$ (and hence $p_x = 1 - q_x$), which when substituted into the equation for the Balducci assumption gives

$$\begin{aligned} 1 - {}_{1-t}p_{x+t} &= (1-t)(1-p_x) \qquad 0 \le t \le 1 \\ \iff {}_{1-t}p_{x+t} &= t + (1-t)p_x \end{aligned}$$

We have the general identity ${}_{s+t}p_x = {}_tp_x {}_sp_{x+t}$, so ${}_sp_{x+t} = \dfrac{{}_{s+t}p_x}{{}_tp_x}$. Letting $s = 1-t$, we get ${}_{1-t}p_{x+t} = \dfrac{p_x}{{}_tp_x}$. Substituting into the Balducci equation

$$\begin{aligned} \frac{p_x}{{}_tp_x} &= t + (1-t)p_x \\ \iff \frac{1}{{}_tp_x} &= \frac{t}{p_x} + (1-t) \\ \iff \frac{1}{{}_tp_x} &= \frac{t}{{}_1p_x} + \frac{1-t}{{}_0p_x}, \qquad 0 \le t \le 1. \end{aligned}$$

(d) Recall the hazard function is given by $\dfrac{d}{dt}\{-\log({}_tp_x)\}$. Then

$$\begin{aligned} \frac{1}{{}_tp_x} &= \frac{t}{p_x} + (1-t) \\ &= 1 + t\left(\frac{1}{p_x} - 1\right) \\ \iff -\log({}_tp_x) &= \log\left(1 + t\left(\frac{1}{p_x} - 1\right)\right) \\ \Rightarrow \frac{d}{dt}\{-\log({}_tp_x)\} &= \frac{\frac{1}{p_x} - 1}{1 + t\left(\frac{1}{p_x} - 1\right)}, \end{aligned}$$

a decreasing function of t as p_x is a probability and thus lies between 0 and 1.

(e) From the Balducci assumption we have

$$\begin{aligned} {}_{1-t}p_{x+t} &= 1-(1-t)q_x \\ {}_{1-t}q_{x+t} &= (1-t)q_x \end{aligned}$$

and from part (c) we have

$$\begin{aligned} {}_tp_x &= \left(\frac{t}{1-q_x}+(1-t)\right)^{-1} \\ {}_tq_x &= 1-\left(\frac{t}{1-q_x}+(1-t)\right)^{-1} \end{aligned}$$

So the likelihood contributions for each individual are

1. $1-0.9q_x$
2. $\left(\frac{0.8}{1-q_x}+0.2\right)^{-1}$
3. $0.8q_x$
4. $1-\left(\frac{0.4}{1-q_x}+0.6\right)^{-1}$

3. (a) The Chapman-Kolmogorov equations state that for $s, t > 0$

$$p^{ij}(t+s) = \sum_{k=1}^{N} p^{kj}(t)p^{ik}(s)$$

To verify these, we have

$$\begin{aligned}
p^{ij}(t+s) &= P(X(t+s)=j|X(0)=i) \\
&= \sum_{k=1}^{N} P(X(t+s)=j|X(s)=k)P(X(s)=k|X(0)=i) \\
&\qquad \text{(by the law of total probability)} \\
&= \sum_{k=1}^{N} P(X(t)=j|X(0)=k)P(X(s)=k|X(0)=i) \\
&\qquad \text{(by the homogeneity of the process)} \\
&= \sum_{k=1}^{N} p^{kj}(t)p^{ik}(s).
\end{aligned}$$

(b) Considering a small positive increment dt, from (a) we have

$$\begin{aligned}
p^{ij}(t+dt) &= \sum_{k=1}^{N} p^{ik}(dt)p^{kj}(t) \\
p^{ij}(t+dt) &= (1+\mu^{ii}dt)p^{ij}(t) + \sum_{k\neq j} \mu^{ij}p^{jk}(t)dt + o(dt) \\
&= p^{ij}(t) + dt\sum_{k} \mu^{ik}p^{kj}(t) + o(dt).
\end{aligned}$$

Rearranging and taking the limit as $dt \to 0^+$,

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

(c) (i) Use the Kolmogorov backward equations

$$\frac{d}{dt}p^{ij}(t) = \sum_{k=1}^{3} \mu^{ik}p^{kj}(t).$$

$$\begin{aligned}
\frac{d}{dt}p^{12}(t) &= \sum_{k=1}^{3} \mu^{1k}p^{k2}(t) \\
&= \mu^{11}p^{12}(t) + \mu^{12}p^{22}(t) + \mu^{13}p^{32}(t) \\
&= \mu^{11}p^{12}(t) + \mu^{12}p^{22}(t) \\
&\qquad \text{(since } \forall t,\ p^{32}(t)=0) \\
&= -(\mu^{12}+\mu^{13})p^{12}(t) + \mu^{12}p^{22}(t) \\
&\qquad \text{(since } \mu^{11}+\mu^{12}+\mu^{13}=0).
\end{aligned}$$

Similarly

$$\frac{d}{dt}p^{13}(t) = -(\mu^{12}+\mu^{13})p^{13}(t) + \mu^{12}p^{23}(t) + \mu^{13}$$

since $\forall t$, $p^{33}(t) = 1$ (the state dead is *absorbing*), and

$$\begin{aligned}\frac{d}{dt}p^{23}(t) &= \mu^{22}p^{23}(t) + \mu^{23} \\ &= -\mu^{23}p^{23}(t) + \mu^{23}.\end{aligned}$$

Solving the equation for $p^{23}(t)$ is straightforward

$$p^{23}(t) = C\exp(-\mu^{23}t) + 1$$

and using the boundary condition $p^{23}(0) = 0$ we get $C = -1$ and

$$p^{23}(t) = 1 - \exp(-\mu^{23}t)$$

Noting that $p^{22}(t) + p^{23}(t) = 1$ (can only go to states 2 or 3 from state 2), we can substitute this result in to the differential equations for $p^{12}(t)$ and $p^{13}(t)$:

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

(ii)

$$(\mu^{12})^{10}(\mu^{13})(\mu^{23})^5\exp\{-20(\mu^{12}+\mu^{13})\}\exp\{-5\mu^{23}\}.$$

(iii) The log likelihood is given by

$$10\log(\mu^{12}) + \log(\mu^{13}) + 5\log(\mu^{23}) - 20(\mu^{12}+\mu^{13}) - 5\mu^{23}.$$

Differentiating this equation with respect to one of the transition intensities and setting equal to zero yields the corresponding maximum likelihood estimate. We find

$$\begin{aligned}\frac{10}{\hat\mu^{12}} - 20 &= 0 \\ \frac{1}{\hat\mu^{13}} - 20 &= 0 \\ \frac{5}{\hat\mu^{23}} - 5 &= 0\end{aligned}$$

Hence the maximum likelihood estimates are

$$\begin{aligned}\hat\mu^{12} &= 0.5 \\ \hat\mu^{13} &= 0.05 \\ \hat\mu^{23} &= 1\end{aligned}$$

The holding time in state 2 follows an exponential distribution with parameter μ^{23}, which has been estimated to be 1. Therefore the probability of an individual in state 2 surviving beyond one year is $\approx e^{-1}$.

4. (a) Defining

$$\psi(z;\beta) = \exp\{z\beta\},$$

the proportional hazards model states $\mu(t;z) = \mu_0(t)\psi(z;\beta)$ and the accelerated failure time model states $S(t;z) = S_0(t\psi(z;\beta))$.

Suppose we have observed event times $t_1, \ldots, t_n$ some of which are censored. Each individual has associated covariates $z_1, \ldots, z_n$. To make inference about β when the baseline hazard function is unknown, we restrict attention to the ordering of the survival/censoring times and ignore the exact event time values.

Assuming unique survival times, the partial likelihood for this reduced data set is given by

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

(b) If $T \sim$ Weibull(α, η) then the survivor function for T is $S(t) = \exp\{-(t/\alpha)^\eta\}$.

Let $c > 0$ be constant w.r.t T. Then the random variable $T' = Tc$ has survivor function $\exp\{-(t/(c\alpha))^\eta\}$. This is the survivor function of a Weibull$(c\alpha, \eta)$ random variable.

(c) Let the baseline model be Weibull$(c\alpha, \eta)$.

Consider the lifetime T of a random individual with covariates z. It follows immediately from (b) that under the AFT model $T \sim$ Weibull$(\alpha/\psi(z;\beta), \eta)$.

Under PH, we have

$$\begin{aligned} S(t;z) &= S_0(t)^{\psi(z;\beta)} \\ &= (\exp\{-(t/\alpha)^\eta\})^{\psi(z;\beta)} \\ &= \exp\{-(t/\alpha)^\eta \psi(z;\beta)\} \end{aligned}$$

and hence $T \sim$ Weibull$(\alpha/\psi(z;\beta)^{1/\eta}, \eta)$.

Since $\psi(z;\beta) = \exp\{z\beta\}$, $\psi(z;\beta)^{1/\eta} = \exp\{z\beta/\eta\}$ and thus we have an equivalence with the AFT model with the regression coefficients simply rescaled by the Weibull shape parameter η.

(d) (i)

$$\left(\frac{e^{60\beta}}{3 + e^{40\beta} + e^{60\beta} + e^{100\beta}}\right)\left(\frac{e^{40\beta}}{2 + e^{40\beta} + e^{100\beta}}\right)\left(\frac{e^{100\beta}}{2 + e^{100\beta}}\right)$$

(ii) That the estimated value for β is positive suggests that an individuals risk of death (hazard) increases with the number of cigarettes smoked.

(iii) Under the null model we have a partial likelihood of

$$\left(\frac{1}{6}\right)\left(\frac{1}{4}\right)\left(\frac{1}{3}\right) = \frac{1}{72}$$

Hence the log partial likelihood without covariates is $-\log(72) = -4.227$.
The likelihood ratio statistic is twice the difference in the maximised log partial likelihoods, so $W = 2(-3.439 + 4.227) = 1.675$.
Since there is one covariate parameter to estimate we compare with χ_1^2 which has a 90th percentile point of 2.71, and thus we do not find the effect of the number of cigarettes smoked by an individual significant at the 10% level.
Notice that the data we have collected appears fairly supportive a different conclusion, that smoking cigarettes increases the hazard. (Looking at the death times, these are strictly higher for the non smokers than for the smokers). We could therefore conclude that the amount of data gathered was insufficient to gain the statistical significance required, or that a simpler binary smoker/non-smoker covariate would have been more effective.

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

(a) Censoring in survival data.

- The mechanisms: Left, Right, Interval censoring.
- Types/study designs: Type I, Generalised Type I, Type II.
- Form of likelihood for observations under different censoring mechanisms.

(b) The Poisson model as a survival process.

- Models data over a unit time interval, $[x; x + 1)$, with individual entry and exit times from the study. An approximation to the two state Markov model assuming a fixed total waiting time.
- Poisson formula for distribution of the number of deaths.
- Definition of a Poisson process and stating the link with the Poisson model.
- Exponentially distributed inter-arrival times.
- Maximum likelihood estimation of the intensity parameter.

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

- The Kaplan-Meier estimate (KM) is a non-parametric estimate of the survivor function.
- The KM is an isotonic (monotone decreasing) step function with jumps at the observed death times.
- The formula for the KM should be given.
- The KM is derived as the maximum likelihood estimate of the survivor function over the space of all distributions.

(d) Comparison of 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.

(e) Choosing a parametric model for survival data.

* $*$ Ad hoc approach of comparing the empirical cumulative hazard with known distributional forms.
* $*$ Hypothesis testing approach between nested parametric models.
* $*$ Description of the likelihood ratio test.
* $*$ Description of the Wald test.
* $*$ Some comparison of Wald and likelihood ratio tests.