KENYATTA UNIVERSITY

UNIVERSITY EXAMINATIONS 2011/2012

INSTITUTION BASED PROGRAMME (DECEMBER SESSION)

EXAMINATION FOR THE DEGREE OF MASTER OF SCIENCE

SMA 675: SURVIVAL ANALYSIS

DATE: WEDNESDAY, 2ND MAY 2012 TIME: 9.00 A.M. – 12.00 P.M.

INSTRUCTIONS: Answer Question 1 and any other two questions.

Question 1 (30 marks)

(a) Distinguish between the following terms as used in Survival Analysis

(i) Right censoring and left censoring

(ii) Type I and Type II censoring

(iii) Cumulative distribution function and survivor function.

(iv) Life expectancy and mean residual life

(8 marks)

(b) Let T be a nonnegative random variable with the survivor function

$s(t) = \exp(-\lambda t^{\alpha})$. Find

(i) the probability density function of T

(ii) the hazard function

(iii) the cumulative hazard function

(6 marks)

(c) If the lifetime random variable T has the exponential distribution with the parameter λ then the log likelihood for the parameter λ is

$\ell = d \log \lambda - \lambda \sum t_i$

where d is the total number of failures (deaths) and $\sum x_i$ is the total time at risk.

For this case

(i) Obtain the Maximum Likelihood Estimator $(M.L.E)\hat{\lambda}$ of λ. (2 marks)

(ii) Find an approximate variance of $\hat{\lambda}$ based on observed Fisher Information.

(2 marks)

(iii) The following are remission times (in weeks) for leukaemia patients.

$6,\ 9,\ 13,\ 13^{+},\ 18,\ 23,\ 28,\ 31,\ 34,\ 45^{+},\ 48,\ 56.$

Assuming that these observations are exponentially distributed with parameter λ, obtain the 95% confidence interval for λ. (5 marks)

(d) Define the following models. In each case give an illustration using a survivor graph.

(i) Proportional hazard model (3 marks)

(ii) Accelerated Failure Model (4 marks)

<u>Question 2</u> (20 marks)

(a) Let T be a nonnegative random variable with the pdf $f(t;\phi)$, survivor function $S(t;\phi)$ and hazard function $h(t;\phi)$. Let $u_1 < u_2 < ...$ be the ordered atoms of T. Let d_j and r_j be the number of individuals who fail and who are at risk respectively at u_j and h_j be the hazard function at u_j. Then for n time points $u_1 < u_2 < ... < u_n$ the likelihood function is

$$L = \Pi_j \{h_j(\phi)\}^{d_j} \{1 - h_j(\phi)\}^{r_j - d_j}$$

obtain the log likelihood function, hence find the Maximum Likelihood Estimator (MLE) of h_j and give the Kaplan-Meier estimator of the survivor function s(t).

(5 marks)

(b) The following are survival times of 10 cancer patients after receiving some treatment.

$9,\ 13^{+},\ 23,\ 18,\ 31,\ 28,\ 34\ \ 48,\ 45^{+},\ 50$

Present in a table form the Product-Limit estimator using.

(i) Kaplan-Meier method (5 marks)

(ii) Efron's start-from-the-far left algorithm (5 marks)

(iii) Using a suitable transformation on $\hat{S}(t)$ sketch a graph of the transformed values of $\hat{S}(t)$ against time to check for the suitability of the exponential distribution as a lifetime model for the given data (5 marks)

<u>Question 3</u> (20 marks)

Let there be two treatment groups denoted by 0 and 1, with cumulative distribution functions G_0 (t) and G_1 (t) respectively.

It is desired to test

$H_0 : G_0(t) = G_1(t)$ for all t

Against $H_1 : G_0(t) \neq G_1(t)$ for some t

The following survivor times, in days, are available

Group 0	2,	2,	3*,	4,	6*,	6*,	7,	10,	
Group 1	1*,	3,	4*,	5,	5,	6*,	7,	7*,	8

0 – control
1- treatment

(a) Calculate the value of the log rank statistic. (18 marks)

(b) Assess the p-value of the calculated log rank statistic in part (ii) and give the conclusion in light of the hypothesis being tested. (2 marks)

[You may use $E_j = \dfrac{N_j\left(m_j + n_j\right)}{M_j + N_j}$

and $V_j = \dfrac{M_j N_j \left(m_j + n_j\right)\left(M_j + N_j - m_j - n_j\right)}{\left(M_j + N_j - 1\right)\left(M_j + N_j\right)^2}$]

<u>Question 4</u> (20 marks)

(a) Let Y be a random variable which has a normal distribution with mean μ and Variance σ^2. If $g(Y)$ be a continuous function of Y, then show that $g(Y)$ has an approximate normal distribution with mean $g(\mu)$ and variance $[g'(\mu)]^2\sigma^2$ where $g'(\mu) = \dfrac{d}{d_\mu} g(\mu)$. (6 marks)

(b) Let T be a discrete nonnegative random variable with the pdf $f(t)$ and survivor function *S(t)*. Further let $u_1 < u_2 < .. < u_j <$ be the ordered atoms of *T*. Suppose r_j is the number of individuals at risk just before time u_j and d_j the number who fail at time u_j then Kaplan-Meier estimate of the survivor function is given by

$$\hat{S}(t) = \prod_{j:u_j<t}\left(1-\hat{h}_j\right)$$

where $\hat{h}_j = \dfrac{d_j}{r_j}$

(i) Using the fact that $d_j \sim b\text{ in }\left(r_j, h_j\right)$
where h_j is the hazard rate at time u_j
show that

$$Var\left(\hat{h}_j\right) = \frac{h_j\left(1-h_j\right)}{r_j}$$ (2 marks)

(ii) Using the result of part (a) show that

$$Var\left[\hat{S}(t)\right] = \left[\hat{S}(t)\right]^2 \sum_{j:t_j<t} \frac{d_j}{\left(r_j - d_j\right)r_j}$$ (5 marks)

(iii) Using the result of part (ii) give a $100(1-\alpha)\%$ confidence interval for the unknown survivor function *S(t)* at time t. What is a possible problem with this approach? (3 marks)

(c) Consider the following table:

j	t_j	r_j	d_j	$\hat{s}(t)$
1	1	21	2	0.9048
2	2	19	2	0.8095
3	3	17	1	0.7619
4	4	16	2	0.6667
5	5	14	1	0.6190

Compute $Var\left[\hat{S}(t)\right]$ at each point t_j using result in (b) (ii) (4 marks)

****