MATH 659: SURVIVAL ANALYSIS

MID-TERM EXAM
Fall, 2011

(Time allowed: TWO HOURS)

SHOUBHIK MONDAL

INSTRUCTIONS TO STUDENTS:

1. This test contains **FOUR** questions and comprises **EIGHT** printed pages.
2. Answer **ALL** questions for a total of 100 marks.
3. This is a **closed-book** test; only a one-page formula sheet and non-programmable calculators are allowed.
4. Write your name on the front of your answer booklet and on any additional sheets you write on.
5. You do not need to do any complicated calculations. Set things up and show me that you know how to do the calculations. It is OK to leave requested numerical answers as fractions.

98/100.

1. Let X have a uniform distribution on the interval 0 to θ with density function

$$f(x) = \begin{cases} 1/\theta, & \text{for } 0 \le x \le \theta \\ 0, & \text{otherwise} \end{cases}$$

(a) Find the survival function of X.

$$F(x) = \int_0^x \frac{1}{\theta} dt = \frac{x}{\theta}, \quad \theta \geq x \geq 0$$

$$S(x) = 1 - F(x) = 1 - \frac{x}{\theta}, \quad 0 \leq x \leq \theta$$

$$= \frac{\theta - x}{\theta}.$$

$$S(x) = \frac{\theta - x}{\theta}, \quad \text{if } 0 \leq x \leq \theta$$

$$= 0 \quad \text{if } x > \theta$$

(b) Find the hazard function of X.

$$h(x) = \frac{f(x)}{S(x)} = \frac{\frac{1}{\theta}}{1 - \frac{x}{\theta}} = \frac{1}{\theta - x} \quad 0 \leq x < \theta.$$

$$= 0, \quad \text{otherwise}$$

-1

2. A large number of *disease-free* individuals were enrolled in a study beginning January 1, 1970, and were followed for 30 years to assess the age at which they developed breast cancer. Individuals had clinical exams every 3 years after enrollment.

(a) What is the interested event in this study?

Here event is the occurrence of Breast Cancer. (that is if some one developed breast cancer within the study period we count that as one event).

(b) What is the appropriate time scale for this study?

Here time scale is the age at which an individual developed breast cancer and the unit of time will be year.

(c) For two selected individuals described below, what types of censoring and truncation that are represented? And why?

(i) A healthy individual, enrolled in the study at age 25, never developed breast cancer during the study.

It is the type I right censored data, where truncated at left as we don't have any information prior to his age 25.

So $L = 25$.

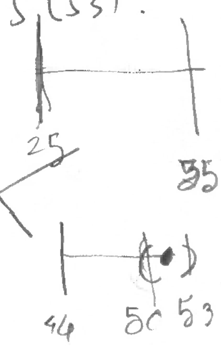
(ii) A healthy individual, enrolled in the study at age 44, was diagnosed with breast cancer at the third exam after enrollment (i.e., the disease started sometime between 6 and 9 years after enrollment).

For this individual as event happened between 6 and 9 years it is interval censoring truncated at left where $L = 44$.

(d) Confining your attention to the two individuals described above, write down the likelihood for this portion of the study. ?

$S(50) - S(53)$.

$$L \propto \frac{S(55)}{S(25)} \cdot \frac{S(53) - S(50)}{S(44)}$$



3. Consider a small study with 8 subjects. The event times were recorded as follows:

$$10, 7, 32+, 23+, 11, 12+, 6+, 27$$

Here "10" means the exact event time is 10. "32+" means the subject is censored at time 32 and its event time is greater than 32.

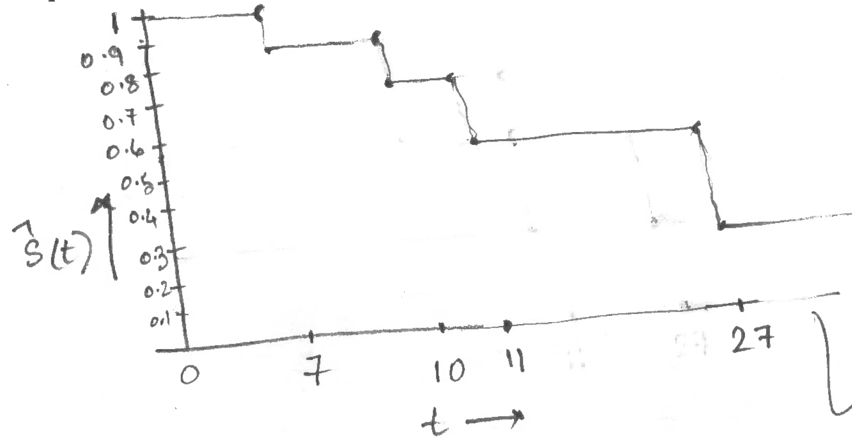
(a) Suppose random censoring is involved in this study. Give a simple example where random censoring is involved.

Suppose we are intrusted in event of breast cancer as like in question 2. Now one individual never developed breast cancer but died due to causes other than breast cancer before the study end. This is an example of right random censoring.

(b) Calculate the Kaplan-Meier estimates for the survival function $S(t)$. The Kaplan-Meier estimator is $\hat{S}(t) = \prod_{t_i \le t} \left(1 - \frac{d_i}{Y_i}\right)$.

i	t_i	d_i	Y_i	$\prod_{t_i \le t} \left(1 - \frac{d_i}{Y_i}\right) = \hat{S}(t)$.
1	7	1	7 (as already one censored data 6+)	$\left(1 - \frac{1}{7}\right) = 0.857$
2	10	1	6	$0.857\left(1 - \frac{1}{6}\right) = 0.714$
3	11	1	5	$0.74\left(1 - \frac{1}{5}\right) = 0.571$
4	27	1	2	$0.571\left(1 - \frac{1}{2}\right) = 0.286$

(c) Make a plot of the survival function based on the estimates obtained in part (b).



(d) Suppose we have $\hat{S}(t_0)$ and the variance of $\hat{S}(t_0)$ is $\hat{V}[\hat{S}(t_0)] = \hat{S}(t_0)^2\sigma^2(t_0)$. Using the delta method with a log transformation of the survival function to derive a $100(1-\alpha)\%$ confidence interval for $\hat{S}(t_0)$. That is, use the transformation $\log[\hat{S}(t_0)]$.

$$g(t) = \ln(t) \qquad g^{-1}(t) = e^t$$

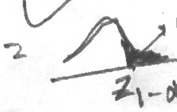
$$g'(t) = \frac{1}{t} \qquad g'(\hat{S}_0(t)) = \frac{1}{\hat{S}_0(t)}$$

$$\text{Var}[g(t)] = V(t)\,(g'(t))^2 = \frac{V(t)}{t^2}.$$

$$\therefore \text{var}[\ln[\hat{S}(t_0)]] = \frac{V(\hat{S}(t_0))}{\hat{S}(t_0)^2} = \frac{\hat{S}(t_0)^2\sigma^2(t_0)}{\hat{S}(t_0)^2} = \sigma^2(t_0)$$

$100(1-\alpha)\%$ C.I. for $\log[\hat{S}(t_0)]$ is

$$\log[\hat{S}(t_0)] \pm Z_{1-\alpha/2}\,\sigma(t_0)$$ – where $Z_{1-\alpha/2}$ = upper $1-\alpha/2$ % point of $N(0,1)$



$\therefore$ $100(1-\alpha)\%$ C.I. for $\hat{S}(t_0)$ is

$$g^{-1}\left[\log[\hat{S}(t_0)] \pm Z_{1-\alpha/2}\,\sigma(t_0)\right]$$

$$\Rightarrow \exp\left[\log[\hat{S}(t_0)] \pm Z_{1-\alpha/2}\,\sigma(t_0)\right]$$

$100(1-\alpha)\%$ C.I. is $\left(\exp\left[\log[\hat{S}(t_0)] - Z_{1-\alpha/2}\,\sigma(t_0)\right], \exp\left[\log[\hat{S}(t_0)] + Z_{1-\alpha/2}\,\sigma(t_0)\right]\right)$

i.e. $= \left[\frac{\hat{S}(t_0)}{\theta}, \hat{S}(t_0)\cdot\theta\right]$ where $\theta = \exp\left[Z_{1-\alpha/2}\,\sigma(t_0)\right]$

(e) Using the procedure you developed in part (d) to obtain a 95% confidence interval for $\hat{S}(10.5)$. Suppose $\hat{V}[\hat{S}(10.5)] = 0.2$.

$\hat{S}(10.5) = 0.714$

$\hat{V}[\hat{S}(10.5)] = 0.2$

$\sigma(t_0) = \sqrt{\frac{V(\hat{S}(10.5))}{\hat{S}(10.5)^2}} = \frac{\sqrt{0.2}}{0.714}$

$= 0.626$

log confidence interval

$= \left[\frac{0.714}{3.4}, 0.714 \times 3.4\right]$

$= [0.2, 2.428]$

$\theta = \exp[1.96 \times 0.625]$

$= e^{1.225} = 3.40$

(f) How do you interpret the confidence interval you obtained in part (e)?

It is 95% chance that the value of Survival function at $t = 10.5$ will lie in the interval

$[0.2, 2.428]$

4. Suppose that the time to death X has an exponential distribution with hazard rate λ and is left-censored by the censor time C_l. Let $T = \max(X, C_l)$ and $\delta = 1$ if $X > C_l$ and $\delta = 0$ if $X \le C_l$. Let $(T_1, \delta_1), \ldots, (T_n, \delta_n)$ be a random sample from this model.

(a) Determine the likelihood function based on the random sample. $X \sim \exp(\lambda)$ left censored $= C_L$

when $\delta = 1$ contribution is $f(x)$

when $\delta = 0$ contribution is $(1 - S(x))$

$$L \propto \prod_{i=1}^{n} f(T_i)^{\delta_i} (1 - S(T_i))^{1-\delta_i} = \prod_{i=1}^{n} \left(\lambda e^{-\lambda T_i}\right)^{\delta_i} \left(1 - e^{-\lambda T_i}\right)^{1-\delta_i}$$

$$= \prod_{i=1}^{n} \left(\frac{\lambda e^{-\lambda T_i}}{1 - e^{-\lambda T_i}}\right)^{\delta_i} \left(1 - e^{-\lambda T_i}\right) \quad \left[\prod_{i=1}^{n} \left(\lambda e^{-\lambda T_i}\right)^{\delta_i} \left(1 - e^{-\lambda C_l}\right)^{1-\delta_i} \right.$$

as when $\delta_i = 0$

$T_i = C_l$]

(b) Calculate the maximum likelihood estimator of λ.

$$\ln L = \sum_{i=1}^{n} \left[\delta_i \ln\left(\lambda e^{-\lambda T_i}\right) + (1 - \delta_i) \ln\left(1 - e^{-\lambda T_i}\right)\right]$$

$$= \sum_{i=1}^{n} \delta_i (\ln \lambda - \lambda T_i) + \sum_{i=1}^{n} (1 - \delta_i) \ln\left(1 - e^{-\lambda T_i}\right)$$

$$\frac{\partial \ln L}{\partial \lambda} = 0 \Rightarrow \sum_{i=1}^{n} \delta_i \left(\frac{1}{\lambda} - T_i\right) + \sum_{i=1}^{n} (1 - \delta_i) \frac{T_i e^{-\lambda T_i}}{1 - e^{-\lambda T_i}} = 0$$

$$\Rightarrow \frac{1}{\lambda} \sum_{i=1}^{n} \delta_i - \sum_{i=1}^{n} \delta_i T_i - \sum_{i=1}^{n} (1 - \delta_i) T_i \left[1 - \frac{1}{1 - e^{-\lambda T_i}}\right] = 0$$

$$\Rightarrow \frac{1}{\lambda} \sum_{i=1}^{n} \delta_i = \sum_{i=1}^{n} \delta_i T_i + \sum_{i=1}^{n} (1 - \delta_i) T_i - \sum_{i=1}^{n} \frac{(1 - \delta_i) T_i}{1 - e^{-\lambda T_i}}$$

$$\Rightarrow \hat{\lambda} = \frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} \delta_i T_i + \sum_{i=1}^{n} (1 - \delta_i) T_i - \sum_{i=1}^{n} \frac{(1 - \delta_i) T_i}{(1 - e^{-\lambda T_i})}} = \frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} T_i - \sum_{i=1}^{n} \frac{(1-\delta_i) T_i}{(1 - e^{-\lambda T_i})}}$$

8

from this eqn we can calculate $\hat{\lambda}$ iteratively with some initial choice of λ.

(P.T.O)

$$\therefore \hat{\lambda} = \frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} T_i - \sum_{i=1}^{n} \frac{(1-\delta_i) T_i}{(1-e^{-\hat{\lambda} T_i})}}$$

$$\left[\text{or } \hat{\lambda} = \frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} T_i - \sum_{i=1}^{n} (1-\delta_i) C_\ell / 1 - e^{-\lambda C_\ell})} \right]$$

as when $\delta_i = 0$, $T_i = C_\ell$

$$\text{or, } \hat{\lambda} = \frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} T_i - \frac{C_\ell}{(1-e^{-\lambda C_\ell})} \sum_{i=1}^{n} (1-\delta_i)}$$