STAT 2261
Orly Olbum

HOMEWORK 2

1. 3.2 p. 87-88

Start Jan 1, 1970, 30 year study
Exams every 3 years

(a) healthy, enrolled at 30, never develops breast cancer during the study
- left truncation because we have no info before age 30, so $L=30$
- right censored (Type I) because no diagnosis before the end of the study $\to C_r = 60$

(b) healthy, age 40, diagnosed at 5^{th} exam
- interval censoring, because we don't know exact time of diagnosis but between 12 & 15 years (52, 55)
- left truncated $L = 40$

(c) healthy, age 50, died from not-bc at 61 (before end of study)
- left truncated $L = 50$
- right censored, $C_r = 61$

(d) 42, moved away at 55, no diagnosis of bc during the study
- Left truncated $L = 42$
- right censored, $C_r = 55$

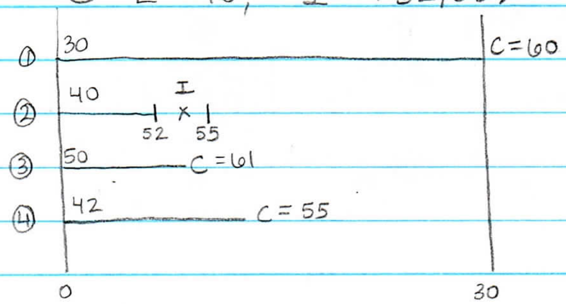
1, continued

ⓔ Likelihood for this portion of the study, considering these 4 subjects

From these four individuals, we have to consider:

① $L=30$, $Cr=60$ ③ $L=50$, $Cr=61$

② $L=40$, $I=(52,55)$ ④ $L=42$, $Cr=55$



$$L \propto \underbrace{\left[\frac{S(60)}{S(30)}\right]}_{①} \cdot \underbrace{\left[\frac{S(52)-S(55)}{S(40)}\right]}_{②} \cdot \underbrace{\left[\frac{S(61)}{S(50)}\right]}_{③} \cdot \underbrace{\left[\frac{S(55)}{S(42)}\right]}_{④}$$

2. 3.8 p. 90

$X=$ time to death $\sim$ exp, hazard rate λ

rc time $C \sim$ exp, hazard rate θ

$T=\min(X,C)$

$\delta=1$, $X \leq C$ $\quad \delta=0$, $X>C$

X & C are independent

$f(x) = \lambda \exp(-\lambda x)$

$h(x) = \lambda$, $S(x) = \exp(-\lambda x)$

$g(c) = \theta \exp(-\theta c)$

$h(c) = \theta$, $S(c) = \exp(-\theta c)$

ⓐ Find $P(\delta=1)$

$$P(\delta=1) = P(X \leq C) = \int_0^\infty Pr(X \leq c \mid C=t)\, Pr(C=t)\, dt$$

$$= \int_0^\infty Pr(X \leq c \mid C=t) \cdot \theta e^{-\theta t}\, dt$$

$$= \int_0^\infty (1-e^{-\lambda t})(\theta e^{-\theta t})\, dt \qquad \text{(because } X \perp C\text{)}$$

$$= \int_0^\infty \theta e^{-\theta t} - \theta e^{(-\lambda-\theta)t}\, dt$$

$$= \int_0^\infty \theta e^{-\theta t}\, dt - \int_0^\infty \theta e^{-(\lambda+\theta)t}\, dt$$

$$= \left[-\tfrac{1}{\theta} \cdot \theta e^{-\theta t}\right]_0^\infty - \left[-\tfrac{\theta}{\lambda+\theta} e^{-(\lambda+\theta)t}\right]_0^\infty$$

$$= [0+1] - \left[0+\frac{\theta}{\lambda+\theta}\right]$$

$$= 1 - \frac{\theta}{\lambda+\theta} =$$

$$= \frac{\lambda}{\lambda+\theta}$$

→

2, continued

(b) Find the distribution of T.

Survival function : $S(t) \overset{\Delta}{=} P(T>t)$

$= P(\min(X,C) > t) = P(X > t, C > t)$

$= P(X>t)\, P(C>t)$ because $X \perp C$

$= S(t)\, S(c)$

$= \exp(-\lambda t)\, \exp(-\theta t)$

$= \exp[-(\lambda+\theta)t]$

$f(t) = -\frac{d}{dt} S(t) = -\{-(\lambda+\theta)\exp[-(\lambda+\theta)t]\}$

$= (\lambda+\theta)\exp\{-(\lambda+\theta)t\}$

(c) Show that $\delta \perp T$.

$\delta \perp T$ if $P(\delta=1, T>t) = P(\delta=1)\, P(T>t)$

and $P(\delta=0, T>t) = P(\delta=0)\, P(T>t)$

$P(\delta=1, T>t) = P(\delta=1, \min(X,C) > t)$

$= P(X \leq C, X>t, C>t)$

$= P(X \leq C, X>t \mid C=d)\, P(C=y)$

$= \int_t^\infty P(X \leq C, X>t \mid C=y) \cdot \theta e^{-\theta y}\, dy$

$= \int_t^\infty P(X \leq y, X>t) \cdot \theta e^{-\theta y}\, dy$

$= \int_t^\infty P(t < X \leq y) \cdot \theta e^{-\theta y}\, dy$

$= \int_t^\infty [F(y) - F(t)] \cdot \theta e^{-\theta y}\, dy$

$= \int_t^\infty [-e^{-\lambda y} + e^{-\lambda t}] \cdot \theta \cdot e^{-\theta y}\, dy$

$= \int_t^\infty \theta \cdot e^{-(\lambda+\theta)y} + \theta \cdot e^{-\lambda t - \theta y}\, dy$

$= \int_t^\infty \theta e^{-(\lambda+\theta)y}\, dy + \int_t^\infty \theta \cdot e^{-\lambda t}\, e^{-\theta y}\, dy$

$= \frac{-\theta}{\lambda+\theta} e^{-(\lambda+\theta)t} + \cancel{\theta} e^{-\lambda t} \cdot -\frac{1}{\cancel{\theta}} e^{-\theta t}$

$= \frac{-\theta}{\lambda+\theta} e^{-(\lambda+\theta)t} + e^{-(\lambda+\theta)t}$

$= (e^{-(\lambda+\theta)t}) \cdot \left[-\frac{\theta}{\lambda+\theta} + 1\right]$

$= \underbrace{\frac{\lambda}{\lambda+\theta}}_{P(\delta=1)} \cdot \underbrace{e^{-(\lambda+\theta)t}}_{P(T>t)}$

→

2, continued

$P(\delta=0, T>t) = P(X>C, \min(X,C)>t)$

$= P(X>C,\ X>t,\ C>t)$

$= \int_t^\infty P(X>y, X>t \mid C=y)\, P(C=y)\, dy$

$= \int_t^\infty P(X>y, X>t) \cdot \theta \cdot e^{-\theta y}\, dy$

$= \int_t^\infty -e^{-\lambda y} \cdot -e^{-\lambda t} \cdot \theta e^{-\theta y}\, dy$

$= \int_t^\infty \theta \cdot e^{-\lambda y - \lambda t - \theta y}\, dy$

$= \int_t^\infty \theta e^{-\lambda t} \cdot e^{-(\lambda+\theta)y}\, dy$

$= \theta e^{-\lambda t} \cdot \frac{1}{\lambda+\theta} e^{-(\lambda+\theta)t}$

$= \frac{\theta}{\lambda+\theta} \cdot e^{-(\lambda+\theta)t}$

$\downarrow P(\delta=0) \quad \downarrow P(T>t)$

so, $\delta \perp T$

ⓓ $(T_1, \delta_1), \ldots, (T_n, \delta_n)$ → random sample from this model. Show that **MLE** of λ is $\sum_{i=1}^n \delta_i / \sum_{i=1}^n T_i$. Use parts a-c to find mean and variance of $\hat{\lambda}$.

we have event times & RC times with f_x, S_x, f_c, S_c

$$L \propto \prod_{i=1}^n f_x(T_i)^{\delta_i} f_c(T_i)^{1-\delta_i} S_x(T_i)^{1-\delta_i} S_c(T_i)^{\delta_i}$$

$$\propto \prod_{i=1}^n (\lambda e^{-\lambda T_i})^{\delta_i} (\theta e^{-\theta T_i})^{1-\delta_i} (e^{-\lambda T_i})^{1-\delta_i} (e^{-\theta T_i})^{\delta_i}$$

$$\propto \prod_{i=1}^n (\lambda e^{-\lambda T_i} e^{-\theta T_i})^{\delta_i} (\theta e^{-\theta T_i} e^{-\lambda T_i})^{1-\delta_i}$$

(keep λ terms)

$$\propto \prod_{i=1}^n (\lambda e^{-\lambda T_i})^{\delta_i} (e^{-\lambda T_i})^{1-\delta_i} \underbrace{(e^{-\theta T_i})^{\delta_i} (\theta e^{-\theta T_i})^{1-\delta_i}}_{\text{no } \lambda}$$

$$\log L \propto \sum \delta_i \log(\lambda e^{-\lambda T_i}) + (1-\delta_i)\log(e^{-\lambda T_i})$$

$$\propto \sum \delta_i \log\lambda + \delta_i \cdot -\lambda T_i + (1-\delta_i)(-\lambda T_i)$$

$$\propto \sum \delta_i \log\lambda - \cancel{\delta_i \lambda} T_i - \lambda T_i + \cancel{\delta_i \lambda} T_i$$

$$\propto \sum \delta_i \log\lambda - \lambda T_i$$

$$\partial \log L / \partial\lambda = \sum_{i=1}^n (\delta_i/\lambda - T_i) = 0$$

$$\hat{\lambda} = \sum_{i=1}^n (\delta_i / T_i)$$

→

2, continued

(ⓓ continued)

$$\hat{\lambda} = \sum_{i=1}^{n}(\delta_i / T_i)$$

Find mean & variance of $\hat{\lambda}$

From ⓑ, $f(t) = (\lambda+\theta)\exp\{-(\lambda+\theta)t\}$

and $\hat{\lambda} = \sum_{i=1}^{n}(\delta_i / T_i)$

$E(\hat{\lambda}) = \frac{1}{\hat{\lambda}} = \sum_{i=1}^{n}(T_i / \delta_i)$

→ exponential because both X & C are exponential and $T = \min(x, c)$, and δ_i is just our indicator

$\mathrm{var}(\hat{\lambda}) = \left[E\left(\frac{\partial^2}{\partial \lambda^2} L\right)\right]^{-1}$

by large sample assumption

$= \frac{\lambda^2}{\sum_{i=1}^{n} \delta_i}$ → number of deaths

→

3. X = age of onset of diabetes
$X \sim$ continuous $f(x)$ & $S(x)$

ⓐ Describe censoring/truncation for full dataset.

Patient	Censoring / Truncation
1	RT; RC - no diagnosis yet, as of survey age
2	RT; LC - we know $X <$ age, but we don't know exact
3	RT; $X = 19$
4	RT; LC - we know $X <$ age, but we don't know exact
5	RT; RC - no diagnosis yet, as of survey age
6	RT; $X = 23$

All patients are subject to right truncation because they are asked about time leading up to their age at the survey, and any information afterwards is not observable

ⓑ Construct likelihood for X with individuals 1-6

Patient	Values
1	$R = 29$, $C_r = 29$
2	$R = 30$, $C_l = 30$
3	$R = 35$, $X = 19$
4	$R = 38$, $C_l = 38$
5	$R = 42$, $C_r = 42$
6	$R = 43$, $X = 23$

$$L \propto \underbrace{\left[\frac{S(29)}{1-S(29)}\right]}_{①} \times \underbrace{\left[\frac{1-S(30)}{1-S(30)}\right]}_{②} \times \underbrace{\left[\frac{f(19)}{1-S(35)}\right]}_{③}$$
$$\times \underbrace{\left[\frac{1-S(38)}{1-S(38)}\right]}_{④} \times \underbrace{\left[\frac{S(42)}{1-S(42)}\right]}_{⑤} \times \underbrace{\left[\frac{f(23)}{1-S(43)}\right]}_{⑥}$$

(c) Restrict analysis to patients with prior diagnosis (ie, only individuals with diabetes surveyed).

Without 1&5, there is no right censoring, but we still have right truncation on all patients and left censoring on patients who don't have exact age of diagnosis.

(d) Likelihood for only individuals with prior diagnosis

$$L \propto \left[\frac{1-S(30)}{1-S(30)}\right] \times \left[\frac{f(19)}{1-S(35)}\right] \times \left[\frac{1-S(38)}{1-S(38)}\right] \times \left[\frac{f(23)}{1-S(43)}\right]$$

(e) Individuals 2&4 are both right truncated and left censored, so they do not provide any information to the likelihood (their calculations cancel out to 1).

(f) This additional information allows us to interval censor these individuals, so it does change the likelihood in (d) to L':

$$L' \propto \left[\frac{S(16)-S(30)}{1-S(30)}\right] \times \left[\frac{f(19)}{1-S(35)}\right] \times \left[\frac{S(21)-S(38)}{1-S(38)}\right] \times \left[\frac{f(23)}{1-S(43)}\right]$$

4.(a) 3.4 p.88-89

X = time from treatment to leukemia relapse

~ exp with $h(x)=\lambda$

construct likelihood function, find MLE of λ by maximizing likelihood function

$f(x) = \lambda \exp(-\lambda x)$; $h(x)=\lambda$; $S(x) = \exp(-\lambda x)$

Pair	L_i	Pair	L_i	
1	$f(10)$	12	$S(20)$	$f \to 9$
2	$f(7)$	13	$S(19)$	$S \to 12$
3	$S(32)$	14	$f(6)$	
4	$f(23)$	15	$S(17)$	
5	$f(22)$	16	$S(35)$	
6	$f(6)$	17	$f(6)$	
7	$f(16)$	18	$f(13)$	
8	$S(34)$	19	$S(9)$	
9	$S(32)$	20	$S(6)$	
10	$S(25)$	21	$S(10)$	
11	$S(11)$			

$L \propto$ product of L_i above:

$\propto \lambda^9 \exp\{(10+7+23+22+6+16+6+6+13)\lambda\}$

$\times \exp\{-(32+34+32+25+11+20+19+17+35+9+6+10)\lambda\}$

$\propto \lambda^9 \cdot \exp\{-(109+250)\lambda\} \propto \boxed{\lambda^9 \exp\{-359\lambda\}}$

$L(\lambda) = \log L = 9\log\lambda - 359\lambda$

$L'(\lambda) = \frac{9}{\lambda} - 359 = 0$

$\frac{9}{\lambda} = 359 \to \boxed{\hat{\lambda} = .0251}$