31

Final exam, Survival Analysis I, 2016 Spring [+35 points]

Name: Lo, Yen-Wen 羅雁文 103225009

- **Not only answer but also derivations**
- **Answers must be simplified**

+4

Q1 [+8] Goodness-of-fit test

1) **[+4]** Let $y_1,\ldots,y_n \overset{iid}{\sim} F(y)=\Pr(Y\le y)$. Consider a goodness-of-fit test for $H_0: F=F_0$ for a specified continuous c.d.f. F_0. Show that the distribution of the Kolmogrov-Smirnov statistics does not depend on F_0 under $H_0: F=F_0$.

+3

Let $Y_1,\ldots,Y_n \overset{iid}{\sim} F(y)=\Pr(Y\le y)$

Let $u_1,\ldots,u_n \overset{iid}{\sim} U(0,1)$, $u_{(1)}\le\cdots\le u_{(n)}$,

The Kolmogrov-Smirnov statistics: $D_n=\sup_y|\hat F_n(y)-F_0(y)|=\sup_y|\hat F_n(y)-F(y)|$

-1 $= \max_{i=1,\ldots,n}\left\{\max\left(\frac{i}{n}-u_{(i)},\, u_{(i)}-\frac{i-1}{n}\right)\right\}$,

need explanation

which does not depend on F_0 under $H_0: F=F_0$ ※

2) **[+4]** Explain how to apply the goodness-of-fit test for testing a log-normal lifetime model under right-censored data (define all necessary mathematical formulas).

+1

T: lifetimes

C: right-censoring

Let $t_1,\ldots,t_n \overset{iid}{\sim} LN(\mu,\sigma^2)$,

$$\delta_i=\begin{cases}1 & \text{if } \Pr(T_i=t_i, C_i>t_i)=f(t_i)\\ 0 & \text{if } \Pr(C_i=t_i, T_i>t_i)=S(t_i)\end{cases}$$

$$L_i=f(t_i)^{\delta_i}\cdot S(t_i)^{1-\delta_i}=\left\{\frac{f(t_i)}{S(t_i)}\right\}^{\delta_i}\cdot S(t_i)=\lambda(t_i)^{\delta_i}S(t_i).$$

Likelihood: $L=\prod_{i=1}^n \lambda(t_i)^{\delta_i}\cdot S(t_i)$

Log-likelihood: $\ell=\sum_{i=1}^n\{\delta_i\cdot\log\lambda(t_i)+\log S(t_i)\}$.

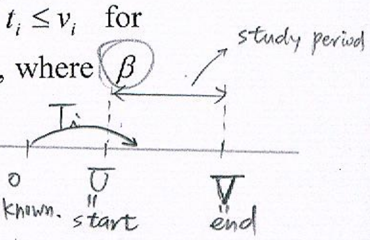
Have to mention about K-M estimator for $\hat S$.

$$L_i = P(T_i = t_i \mid u_i \le T_i \le v_i) = \frac{f(t_i)}{P(u_i \le T_i \le v_i)} = \frac{f(t_i)}{F(v_i) - F(u_i)} = \frac{f(t_i)}{S(u_i) - S(v_i)}.$$

+6

Q2 [+6] Consider doubly-truncated data u_i, t_i, v_i subject to $u_i \le t_i \le v_i$ for $i = 1, \dots, n$. Assume the Weibull lifetime model $\Pr(T \ge t) = \exp\{-(t/\alpha)^\beta\}$, where β is known.

$$f(t) = -\frac{\partial}{\partial t} S(t) = -\left(-\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\beta\right\} = \left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\beta\right\}.$$



+2 1) [+2] Derive the likelihood equation (score function).

Likelihood: $L(\alpha) = \prod_{i=1}^{n} \frac{f(t_i)}{\{F(v_i) - F(u_i)\}} = \prod_{i=1}^{n} \left(\frac{\beta}{\alpha}\right)\left(\frac{t_i}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{t_i}{\alpha}\right)^\beta\right] \Big/ \left[\exp\left\{-\left(\frac{u_i}{\alpha}\right)^\beta\right\} - \exp\left\{-\left(\frac{v_i}{\alpha}\right)^\beta\right\}\right]$, where β is known.

Log-likelihood: $\ell(\alpha) = \sum_{i=1}^{n} \left[\log\beta - \beta \cdot \log\alpha + (\beta - 1)\log t_i - \left(\frac{t_i}{\alpha}\right)^\beta - \log\left[\exp\left\{-\left(\frac{u_i}{\alpha}\right)^\beta\right\} - \exp\left\{-\left(\frac{v_i}{\alpha}\right)^\beta\right\}\right]\right] = n\log\beta - n\beta\cdot\log\alpha + \beta\sum_{i=1}^{n}\log t_i - \sum_{i=1}^{n}\log t_i - \sum_{i=1}^{n}\left(\frac{t_i}{\alpha}\right)^\beta - \sum_{i=1}^{n}\log\left[\exp\left\{-\left(\frac{u_i}{\alpha}\right)^\beta\right\} - \exp\left\{-\left(\frac{v_i}{\alpha}\right)^\beta\right\}\right]$

Score function: $U(\alpha) = \frac{\partial}{\partial\alpha}\ell(\alpha) = \frac{-n\beta}{\alpha} - \left(\sum_{i=1}^{n} t_i^\beta\right)\cdot(-\beta)\cdot\alpha^{-\beta-1} - \sum_{i=1}^{n} \frac{-\beta\left(\frac{u_i}{\alpha}\right)^{\beta-1}\left(\frac{-u_i}{\alpha^2}\right)\exp\left\{-\left(\frac{u_i}{\alpha}\right)^\beta\right\} + \beta\left(\frac{v_i}{\alpha}\right)^{\beta-1}\left(\frac{-v_i}{\alpha^2}\right)\exp\left\{-\left(\frac{v_i}{\alpha}\right)^\beta\right\}}{\exp\left\{-\left(\frac{u_i}{\alpha}\right)^\beta\right\} - \exp\left\{-\left(\frac{v_i}{\alpha}\right)^\beta\right\}}$

$$= \frac{-n\beta}{\alpha} + \frac{\beta}{\alpha}\cdot\sum_{i=1}^{n}\left(\frac{t_i}{\alpha}\right)^\beta - \sum_{i=1}^{n} \frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{u_i}{\alpha}\right)^\beta \exp\left\{-\left(\frac{u_i}{\alpha}\right)^\beta\right\} - \left(\frac{\beta}{\alpha}\right)\left(\frac{v_i}{\alpha}\right)^\beta \exp\left\{-\left(\frac{v_i}{\alpha}\right)^\beta\right\}}{\exp\left\{-\left(\frac{u_i}{\alpha}\right)^\beta\right\} - \exp\left\{-\left(\frac{v_i}{\alpha}\right)^\beta\right\}}$$

※

+2 2) [+2] Obtain the MLE if $v_i = \infty$ (no right-truncation)

If $v_i = \infty$, $\exp\left\{-\left(\frac{v_i}{\alpha}\right)^\beta\right\} = 0$, and $\left(\frac{v_i}{\alpha}\right)^\beta \cdot \exp\left\{-\left(\frac{v_i}{\alpha}\right)^\beta\right\} = 0$.

MLE: Set $U(\alpha) = 0$, $U(\alpha) = \frac{-n\beta}{\alpha} + \frac{\beta}{\alpha}\sum_{i=1}^{n}\left(\frac{t_i}{\alpha}\right)^\beta - \sum_{i=1}^{n} \frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{u_i}{\alpha}\right)^\beta \cdot \cancel{\exp\left\{-\left(\frac{u_i}{\alpha}\right)^\beta\right\}} - \cancel{\left(\frac{\beta}{\alpha}\right)\left(\frac{v_i}{\alpha}\right)^\beta \exp\left\{-\left(\frac{v_i}{\alpha}\right)^\beta\right\}}^{\to 0}}{\cancel{\exp\left\{-\left(\frac{u_i}{\alpha}\right)^\beta\right\}} - \cancel{\exp\left\{-\left(\frac{v_i}{\alpha}\right)^\beta\right\}}^{\to 0}} \overset{\text{set}}{=} 0.$

$$\Rightarrow n = \sum_{i=1}^{n}\left(\frac{t_i}{\alpha}\right)^\beta - \sum_{i=1}^{n}\left(\frac{u_i}{\alpha}\right)^\beta = \alpha^{-\beta}\left(\sum_{i=1}^{n} t_i^\beta - \sum_{i=1}^{n} u_i^\beta\right)$$

(Given $\alpha = \hat{\alpha}$)

$$\Rightarrow \hat{\alpha}^\beta = \frac{\sum_{i=1}^{n}(t_i^\beta - u_i^\beta)}{n} \Rightarrow \boxed{\hat{\alpha} = \left\{\frac{\sum_{i=1}^{n}(t_i^\beta - u_i^\beta)}{n}\right\}^{\frac{1}{\beta}}},$$

and $\frac{\partial^2}{\partial\alpha^2}\ell(\alpha) < 0$. ※.

+2 3) [+2] Explain how to obtain the MLE if $u_i = 0$ (no left-truncation)

If $u_i = 0$, $\exp\left\{-\left(\frac{u_i}{\alpha}\right)^\beta\right\} = 1$, and $\left(\frac{u_i}{\alpha}\right)^\beta \cdot \exp\left\{-\left(\frac{u_i}{\alpha}\right)^\beta\right\} = 0$.

MLE: Set $U(\alpha) = 0$, $U(\alpha) = \frac{-n\beta}{\alpha} + \frac{\beta}{\alpha}\sum_{i=1}^{n}\left(\frac{t_i}{\alpha}\right)^\beta - \sum_{i=1}^{n} \frac{\cancel{\left(\frac{\beta}{\alpha}\right)\left(\frac{u_i}{\alpha}\right)^\beta \exp\left\{-\left(\frac{u_i}{\alpha}\right)^\beta\right\}}^{\to 0} - \left(\frac{\beta}{\alpha}\right)\left(\frac{v_i}{\alpha}\right)^\beta \cdot \exp\left\{-\left(\frac{v_i}{\alpha}\right)^\beta\right\}}{\underbrace{\cancel{\exp\left\{-\left(\frac{u_i}{\alpha}\right)^\beta\right\}}}_{1} - \exp\left\{-\left(\frac{v_i}{\alpha}\right)^\beta\right\}} \overset{\text{set}}{=} 0$

$$\Rightarrow n = \sum_{i=1}^{n}\left(\frac{t_i}{\alpha}\right)^\beta + \sum_{i=1}^{n}\left[\frac{\left(\frac{v_i}{\alpha}\right)^\beta \cdot \exp\left\{-\left(\frac{v_i}{\alpha}\right)^\beta\right\}}{1 - \exp\left\{-\left(\frac{v_i}{\alpha}\right)^\beta\right\}}\right]$$

$$\Rightarrow \alpha^\beta \cdot n = \sum_{i=1}^{n} t_i^\beta + \sum_{i=1}^{n}\left[\frac{v_i^\beta \cdot \exp\left\{-\left(\frac{v_i}{\alpha}\right)^\beta\right\}}{1 - \exp\left\{-\left(\frac{v_i}{\alpha}\right)^\beta\right\}}\right]$$

Given $\alpha = \hat{\alpha}$

$$\Rightarrow \hat{\alpha}^\beta = \frac{1}{n}\left[\sum_{i=1}^{n}\left\{t_i^\beta + \frac{v_i^\beta \cdot \exp\left\{-\left(\frac{v_i}{\alpha}\right)^\beta\right\}}{1 - \exp\left\{-\left(\frac{v_i}{\alpha}\right)^\beta\right\}}\right\}\right]$$

$$\Rightarrow \hat{\alpha} = \left[\frac{\sum_{i=1}^{n}\left[t_i^\beta + \left\{\frac{v_i^\beta \cdot \exp\left\{-\left(\frac{v_i}{\alpha}\right)^\beta\right\}}{1 - \exp\left\{-\left(\frac{v_i}{\alpha}\right)^\beta\right\}}\right\}\right]}{n}\right]^{\frac{1}{\beta}}$$

(Suppose some value $\alpha = \alpha_0$.)

Can use Fixed-point iteration or Newton-Raphson to obtain $\hat{\alpha}$.

※.

+12 **Q3 [+12]** Consider left-truncated and right-censored data $(u_i,\ t_i,\ \delta_i,\ x_i)$ subject to $u_i \le t_i$ for $i=1,\ldots,n$. Assume the proportional hazard model $h(t\,|\,x_i)=h_0(t)e^{\beta x_i}$ for $x_i=1$ (group 1) and $x_i=0$ (group 2). Define ordered times at death, $t_{(1)} \le t_{(2)} \le \cdots \le t_{(k)}$, where $k=\sum_{i=1}^{n}\delta_i$. Also, define $d_i=\delta_i$ and $d_{1i}=\delta_i x_i$.

$x_i=\begin{cases}1, & \text{group 1}\\ 0, & \text{group 2}\end{cases}$

$\{C=j\}=\{T=T_j\}=\{T_j<T_\ell,\ \forall \ell\neq j\}$

Define $R_i=\{j : u_i \le t_{(i)} \le t_j\}$, $j=1,2$.

+2 1) [+2] Define Cox's partial likelihood for β.

Likelihood: $L(\beta)=\prod_{i=1}^{n}\left\{\frac{e^{\beta x_i}}{\sum_{\ell\in R_i} e^{\beta x_\ell}}\right\}^{\delta_i}=\prod_{i=1}^{k}\left\{\frac{e^{\beta x_{(i)}}}{\sum_{\ell\in R_i} e^{\beta x_\ell}}\right\}^{1}$ ※

$n_{1i}\equiv\sum_{i=1}^{n}x_i$

$n_{2i}\equiv\sum_{i=1}^{n}(1-x_i)$

+2 2) [+2] Derive the score function

Log-likelihood: $\ell(\beta)=\sum_{i=1}^{n}\delta_i\left[\beta\cdot x_i-\log\left\{\sum_{\ell\in R_i}e^{\beta x_\ell}\right\}\right]$

Score function: $U(\beta)=\frac{\partial}{\partial\beta}\ell(\beta)=\sum_{i=1}^{n}\delta_i\left[x_i-\bar{x}(t_i,\beta)\right]$, where $\bar{x}(t_i,\beta)=\frac{\sum_{\ell\in R_i}x_\ell e^{\beta x_\ell}}{\sum_{\ell\in R_i}e^{\beta x_\ell}}=\frac{n_{1i}\cdot e^{\beta}}{\underbrace{n_{1i}e^{\beta}}_{x_\ell=1}+\underbrace{n_{2i}}_{x_\ell=0}}$

(Let $d_i=\delta_i$, $d_{1i}=\delta_i x_i$) $=\sum_{i=1}^{n}\left(d_{1i}-\frac{d_i n_{1i}e^{\beta}}{n_{1i}e^{\beta}+n_{2i}}\right)$ ※

+2 3) [+2] Derive the Hessian

$I(\beta)=-\frac{\partial^2}{\partial\beta^2}\ell(\beta)=\sum_{i=1}^{n}\frac{d_i n_{1i}e^{\beta}(n_{1i}e^{\beta}+n_{2i})-(d_i n_{1i}e^{\beta})\cdot(n_{1i}e^{\beta})}{(n_{1i}e^{\beta}+n_{2i})^2}=\sum_{i=1}^{n}\frac{d_i n_{1i}n_{2i}e^{\beta}}{(n_{1i}e^{\beta}+n_{2i})^2}$

$H(\beta)=-I(\beta)=-\sum_{i=1}^{n}\frac{d_i n_{1i}n_{2i}e^{\beta}}{(n_{1i}e^{\beta}+n_{2i})^2}$ ※

+2 4) [+2] Derive the fixed-point iteration algorithm to obtain $\hat{\beta}$.

Set $U(\beta)\big|_{\beta=\hat\beta}=0 \Rightarrow \left(\sum_{i=1}^{n}d_{1i}\right)=e^{\beta}\cdot\left(\sum_{i=1}^{n}\frac{d_i n_{1i}}{n_{1i}e^{\beta}+n_{2i}}\right)\Bigg|_{\beta=\hat\beta} \Rightarrow e^{\hat\beta}=\frac{\sum_{i=1}^{n}d_{1i}}{\sum_{i=1}^{n}\left(\frac{d_i n_{1i}}{n_{1i}e^{\beta}+n_{2i}}\right)}$

suppose some value $e^{\beta}=e^{\beta_0}$

Fixed-point iteration algorithm:

Step 1: Set β_0

Step 2: $e^{\beta_J}=\frac{\sum_{i=1}^{n}d_{1i}}{\sum_{i=1}^{n}\left(\frac{d_i n_{1i}}{n_{1i}e^{\beta_{J-1}}+n_{2i}}\right)}$, $J=1,2,\ldots$

Step 3: $\hat\beta=\beta_J$, if $e^{\beta_J}-e^{\beta_{J-1}}\approx 0$ ※

+2 5) [+2] Derive the Newton-Raphson algorithm to obtain $\hat{\beta}$.

Newton-Raphson algorithm:

Step 1: Set β_0.

Step 2: $\beta_J=\beta_{J-1}-H(\beta_{J-1})\cdot U(\beta_{J-1})^{-1}$, $J=1,2,\ldots$

Step 3: $\hat\beta=\beta_J$, if $e^{\beta_J}-e^{\beta_{J-1}}\approx 0$. ※

+2 6) [+2] Derive the Wald test for $H_0:\beta=0$ at level $\alpha=0.05$.

If $\beta=0$ is true, $\hat\beta \,\dot\sim\, N(0, I(\hat\beta)^{-1})$, $se(\hat\beta)=I(\hat\beta)^{-\frac{1}{2}}$

$Z=\frac{\hat\beta}{se(\hat\beta)} \,\dot\sim\, N(0,1)$.

If $|Z|>1.96$, reject H_0 at $\alpha=0.05$ ※

+9

Q4 [+9] Let (T_1, T_2) be bivariate lifetimes, and U be the frailty variable.

+3

1) **[+3]** State all assumptions required to define the gamma frailty model

(i) Gamma frailty : $U \sim f_U(u) = \frac{1}{\Gamma(\frac{1}{\phi})\phi^{\frac{1}{\phi}}} u^{\frac{1}{\phi}-1} e^{-\frac{u}{\phi}}, \phi \geq 0$, (Gamma $(\alpha = \frac{1}{\phi}, \beta = \phi)$).

$E(U) = 1$, $Var(U) = \phi$.

(ii) Conditional hazard : $\lambda_j(t|u) = u_j \cdot \lambda_j(t)$, $j = 1, 2$.

(iii) Conditional independence : $T_1 \perp T_2 | U$.

#

+3

2) **[+3]** The Clayton model for bivariate survival function is written as $c(u_1, u_2) = S(t_1, t_2) = \Pr(T_1 \geq t_1, T_2 \geq t_2) = C(S_1(t_1), S_2(t_2))$, where $C(u_1, u_2)$ is called copula. Derive the formula $C(u_1, u_2)$ from the shared gamma frailty model. (By using above assumption in 1))

$U_1 \equiv S_1(t_1) \sim U(0,1)$

$U_2 \equiv S_2(t_2) \sim U(0,1)$

$(\because P(A) = E[P(A|X)])$

$$S(t_1, t_2) = \Pr(T_1 \geq t_1, T_2 \geq t_2) = E[\Pr(T_1 \geq t_1, T_2 \geq t_2 | U)] = E[\Pr(T_1 \geq t_1 | U) \cdot \Pr(T_2 \geq t_2 | U)]$$

$$= E[\exp\{-U\Lambda_1(t_1)\} \cdot \exp\{-U\Lambda_2(t_2)\}], \text{ where } \Lambda_j(t) = \int_0^t \lambda_j(u)\,du, \; j = 1, 2.$$

(Set $\Lambda = \Lambda_1(t_1) + \Lambda_2(t_2)$)

$$= E[\exp\{-U(\Lambda_1(t_1) + \Lambda_2(t_2))\}] = \int_0^\infty e^{-u(\Lambda_1(t_1)+\Lambda_2(t_2))} f_U(u)\,du$$

$$= \int_0^\infty e^{-u\Lambda} \frac{1}{\Gamma(\frac{1}{\phi}) \cdot \phi^{\frac{1}{\phi}}} u^{\frac{1}{\phi}-1} e^{-\frac{u}{\phi}}\,du = \frac{1}{\Gamma(\frac{1}{\phi})\phi^{\frac{1}{\phi}}} \cdot \int_0^\infty u^{\frac{1}{\phi}-1} e^{-(\Lambda+\frac{1}{\phi})u}\,du = \frac{1}{\Gamma(\frac{1}{\phi})\phi^{\frac{1}{\phi}}} \int_0^\infty \left(\frac{s}{\Lambda+\frac{1}{\phi}}\right)^{\frac{1}{\phi}-1} \cdot e^{-s} \frac{ds}{(\Lambda+\frac{1}{\phi})}$$

Gamma function

$$= \frac{\int_0^\infty s^{\frac{1}{\phi}-1} e^{-s}\,ds}{\Gamma(\frac{1}{\phi}) \cdot \phi^{\frac{1}{\phi}} (\Lambda+\frac{1}{\phi})^{\frac{1}{\phi}}} = \frac{\Gamma(\frac{1}{\phi})}{\Gamma(\frac{1}{\phi})(\phi\Lambda+1)^{\frac{1}{\phi}}} = (\phi\Lambda+1)^{-\frac{1}{\phi}}$$

Let $s = (\Lambda + \frac{1}{\phi})u \Rightarrow u = \frac{s}{\Lambda+\frac{1}{\phi}}$, $\frac{ds}{du} = \Lambda + \frac{1}{\phi}$

$$= \{1 + \phi\Lambda_1(t_1) + \phi\Lambda_2(t_2)\}^{-\frac{1}{\phi}} = \{1 + (u_1^{-\phi}-1) + (u_2^{-\phi}-1)\}^{-\frac{1}{\phi}} = \{u_1^{-\phi} + u_2^{-\phi} - 1\}^{-\frac{1}{\phi}} = \{S_1(t_1)^{-\phi} + S_2(t_2)^{-\phi} - 1\}^{-\frac{1}{\phi}} : \text{Clayton model}$$

$\because u_1 = S_1(t_1) = \{1 + \phi\Lambda_1(t_1)\}^{-\frac{1}{\phi}} \Rightarrow u_1^{-\phi} = 1 + \phi\Lambda_1(t_1) \Rightarrow \boxed{\phi\Lambda_1(t_1) = u_1^{-\phi} - 1}$

similarly, $\boxed{\phi\Lambda_2(t_2) = u_2^{-\phi} - 1}$.

#

+3

3) **[+3]** Discuss if $T_1 \perp T_2$ holds or not

a) T_1=time to death,
T_2=time to marriage

$T_1 \not\perp T_2$, T_2 is smaller or equal to T_1, $T_2 \leq T_1$.

Since the time to marriage must earlier or equal to the time to death of a person.

#

b) T_1=time to breakdown of computer display,
T_2=time to breakdown of computer battery

$T_1 \perp T_2$. Since the display and battery are the different components of the computer. Then, the time to breakdown of computer display and battery are not dependent.

#