Survival Analysis

Instructor: Hong-Dar Isaac Wu

Department of Applied Mathematics

National Chung Hsing Uuiversity

February 14, 2012

☆ **Hazard function(hazard rate) :**

$$
\begin{aligned}
h_T(t)\Delta t &= P_r\{T \in [t, t+\Delta t) | T \geq t\} \\
&= \frac{P_r\{t \leq T < t+\Delta t, T \geq t\}}{P_r\{T \geq t\}} \\
&= \frac{P_r\{t \leq T < t+\Delta t\}}{P_r\{T \geq t\}} \\
&= \frac{F(t+\Delta t) - F(t)}{S(t)} \\
&= \frac{f(t)\Delta t}{S(t)}, \ \Delta t \approx 0
\end{aligned}
$$

☆ **Survival function :**

T is the survival time($T > 0$), R.V

$S_T(t) = 1 - F_T(t) = P_r\{T > t\}$ if $F_T(t) = P_r\{T \leq t\}$

✏ **Eg:**

$T \sim exp(\lambda),\ f_T(t) = \lambda e^{-\lambda t}$

$F(t) = 1 - e^{-\lambda t}$

$S(t) = e^{-\lambda t}$

$h(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$, independent of t(constant)

Eg:

$T \sim weibull(a, b)$, $S_T(t) = e^{-(at)^b}$ or $S(t) = e^{-at^b}$, $a > 0$ & $b > 0$,

a : $scale$, b : $shape$

$h(t) = ab(at)^{b-1}$

$f_T(t) = ab(at)^{b-1}e^{-(at)^b}$

☆ Power Generalized Weibull

$S(t) = e^{1-\{1+(at)^b\}^r}$, $r = 1 \Rightarrow$ *Weibull*

$H(t) \equiv$cumulative hazard function$= \int_0^t h(u)du$

$$\begin{cases} H(t) & = -\log S(t) \\ S(t) & = \quad e^{-H(t)} \end{cases} \quad \text{where } h(t) = \frac{f(t)}{S(t)}$$

Therefore

$$\begin{aligned} H(t) &= \int_0^t h(u)du = \int_0^t \frac{f(u)du}{S(u)} = -\int_0^t \frac{dS(u)}{S(u)} \\ &= -\log S(t) + \log S(0) = -\log S(t) \end{aligned}$$

☆ In survival analysis, most of the imposed models are talking about $h(t)$, the hazard rate function.

$$\begin{aligned} h_1 &= \text{hazard of population 1} \\ h_0 &= \text{hazard of population 0} \end{aligned}$$

Therefore

$$\begin{aligned} \frac{h_1}{h_0} &\equiv \text{hazard ratio(rate ratio)} \\ &> 1 \quad h_1 \text{ easy to die than } h_0 \\ &< 1 \quad h_1 \text{ doesn't easy to die than } h_0 \end{aligned}$$

☆ When you have a set of data, $x_1, \cdots, x_n \sim f(x),\ F(x),\ S(x), \cdots$

$$\text{Likelihood} = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \{h(x_i)S(x_i)\}$$

where $h(x_i) = \frac{f(x_i)}{S(x_i)} \Rightarrow f(x_i) = h(x_i)S(x_i)$

☆ **Mean Residual Life**

$T \sim F(\cdot),\ f(\cdot),\ E(T) = \int_0^\infty tf(t)dt,\ E(t) = \int_0^\infty S(t)dt$

(sol):

$$
\begin{aligned}
\int_0^\infty S(t)dt &= S(t)t|_0^\infty - \int tdS(t) \\
&= -\int tdS(t) \\
&= \int tdF(t) \\
&= ET
\end{aligned}
$$

☆ MRL(mrl) of T given x

$$E\{T-x|T>x\} = \int_{t-x}^{\infty}(t-x)\left(\frac{f(t)}{S(x)}\right)dt = \frac{\int_x^{\infty}(t-x)f(t)dt}{S(x)}$$

$$\begin{aligned}
\int_x^{\infty}(t-x)dF(x) &= lim_{c\to\infty}\int_x^{c}(t-x)dF(x) \\
&= lim\{(t-x)F(t)|_x^c - \int_x^c F(t)dt\} \\
&= lim\{(t-c)F(c)-(x-x)F(x) - \int_x^c \{1-S(t)\}dt\} \\
&\Rightarrow -\int_x^c 1dt + \int_x^c S(t)dt \\
&= lim\{(t-c)-c+t+\int_x^c S(t)dt\} \\
&\approx \int_x^{\infty} S(t)dt
\end{aligned}$$

- **median survival:**

  $x_p = inf\{x : F(x) \leq p\}$

  if $p = 50\% \Rightarrow$ *median survival*

- **Statistical issue:**

  $\hat{x}_p(p = \frac{1}{2})$, confident interval of $x_p(p = \frac{1}{2})$

☆ **log-normal density:**

$Y = \log X \sim Normal$, $|J| = \frac{1}{X}$

Therefore $f(y) = \frac{1}{\sqrt{2\pi}\sigma} exp\{-\frac{(y-\mu)^2}{2\sigma^2}\}$

$$
\begin{aligned}
\Rightarrow f_X(x) &= f_Y(y(x))|J| = \frac{1}{\sigma\sqrt{2\pi}} exp\left\{ - \frac{(\log x - \mu)^2}{2\sigma^2} \right\} \frac{1}{x} \\
&= \phi\left(\frac{\log x - \mu}{\sigma}\right)/x \\
S(x) &= 1 - \Phi\left(\frac{\log x - \mu}{\sigma}\right)
\end{aligned}
$$

☆ $MLE\widehat{\theta} = argmax_\theta L_\theta,\ \theta = (\theta_1, \cdots, \theta_p)$

$H_0 : \theta = \theta_0(\theta_1 = \theta_{10}, \theta_2 = \theta_{20}, \cdots, \theta_p = \theta_{p0})$

Score:

$$\frac{\partial}{\partial\theta}(\log L_\theta) = \frac{\partial}{\partial\theta}(l_\theta)$$

$$(\frac{\partial}{\partial\theta} l_\theta)' Var(\frac{\partial}{\partial\theta} l_\theta)(\frac{\partial}{\partial\theta} l_\theta)|_{\theta=\widehat{\theta}}$$

Wald:

$$\widehat{\theta}'(Var\widehat{\theta})^{-1}\widehat{\theta} = \widehat{\theta}'\mathfrak{I}_\theta\widehat{\theta}$$

To put something on parametric model to do regression analysis.

☆ **Conclusion(TEMP):**

We can do regression analysis through a parametric model, say weibull.

✏ **Example:**

Can you fit a weibull regression model on a set of survival data? Further, can you test $\beta_1 = 0$?

$$Z_1 = \begin{cases} 1 & \leftarrow \text{treatment} \quad \text{group} \\ 0 & \leftarrow \text{control} \quad \text{group} \end{cases}$$

☆ **Weibull distribution fullfils:**

1. proportional hazards(Cox 1972; Semi-parametric model)
2. accelerated failure time $S_T(t) = S_{T^*}(\phi(r'z)t)$

☆ $S(t)=\exp(-at^b),\ S_0(t)=\exp(-a_0t^b)$

$$\begin{aligned} H(t) &= at^b \\ &= (\frac{a}{a_0})(a_0t^b) = a^*H_0(t) \\ a^* &= \exp(\beta_*'z) = \exp(r'z) \end{aligned}$$

$$\begin{aligned} H(t;z) &= H_0(t)\exp(r'z) \Longleftarrow \textit{proportionality} = \exp(r'z) \\ \Big\downarrow \quad & \quad \frac{\exp(-H(t)) = S(t)}{H(t) = -\log S(t)} \\ S(t;z) &= \exp(-H(t;z)) = \exp(-H_0(t)e^{r'z}) = (\exp(-H_0(t)))^{\mu(r,z)} \\ &= \{S_0(t)\}^{\mu(r,z)},\ \mu(r,z)=\exp(r'z) \end{aligned}$$

☆ **Weibull(a,b)**

$$S(t) = \exp(-at^b) \begin{cases} \textit{Porportional hazards Lehmann family} \\ AFT \end{cases}$$

$S(t)$ rejaramterized as $S_T(t) = \exp(-(at)^b) = P_r\{T > t\} \equiv S_0(t)$

If there is another RV $T^* = aT(t^* = at)$

$$\Rightarrow P_r\{T > t\} = P_r\{aT > at\} = P_r\{T^* > at\} \equiv S_{T^*}(at)$$

Which means the failure of T^* is "a"-feld accelerated to the failure of T.

☆ Therefore, if $T \sim Weibull(a, b)$

Let $Y = \log T$ & $w = \frac{Y-\mu}{\sigma}$

(Where $Y = \mu + \sigma w \Rightarrow \log t = \mu + \sigma w \quad \therefore t = \exp(\mu + \sigma w)$)

$$f_W = f_T |\frac{dw}{dT}|^{-1} = \exp(-a(e^{\mu+\sigma w})^b ab(e^{\mu+\sigma w})^{b-1})\sigma^{\mu+\sigma w}\sigma$$

$(|\frac{dw}{dT}|^{-1} = |\frac{dt}{dw}| = \exp(\mu + \sigma w)\sigma,\ S_T = \exp(-at^b),$

$f_T = \exp(-at^b)abt^{b-1})$

$$\begin{aligned} &= \exp(-e^w)\exp(w) \\ &= \exp(w - e^w) \leftarrow \text{extreme value distribution} \end{aligned}$$

$F_W(w) = \int_0^w f_W(u)du = 1 - \exp(-e^w)$

Therefore $S_W(w) = \exp(-e^w)$

☆ If $Y = \mu + \sigma w (\equiv \beta' z + \sigma w)$

$$
\begin{aligned}
S_Y(y) &= P_r\{Y > y\} \\
&= P_r\{\frac{Y-\mu}{\sigma} > \frac{y-\mu}{\sigma}\} \\
&= P_r\{w > \frac{y-\mu}{\sigma}\} \text{ (the "survival" of } w) \\
&= \exp(-e^{\frac{y-\mu}{\sigma}}) \\
&= \exp(-e^{\frac{y}{\sigma}} e^{-\frac{\mu}{\sigma}}) \; (\frac{1}{\sigma} = b) \\
&= \exp(-(te^{-\mu})^b) = S_T(\phi(\cdot)t) \; (\phi(\cdot) = e^{-\mu}, \; \mu = \beta' z) \\
&= S_T(e^{-\beta' z} t)
\end{aligned}
$$

☆ That means: $S_Y(y) = S_T(e^{-\beta' z} t)$

☆ **R Miller:**

Bias-corrected estimator

✏ **Eg:**(exponential distribution)

$T_1, \cdots, T_n \sim exp(\lambda)$, $S_T(t) = e^{-\lambda t}$ estimation $\widehat{S}_T(t) = e^{-\widehat{\lambda} t}$

$E\overline{T} = \frac{1}{\lambda}$, so $\widehat{\lambda} = \frac{1}{\overline{T}}$, therefore $\widehat{S}_T(t) = e^{-\frac{t}{\overline{T}}}$

(a reasonable estimator, but it is abriously biased,

$\because E\{e^{-\frac{t}{\overline{T}}}\} \neq e^{-\lambda t}$)

☆ **Bias-corrected estimator:**

Let $E\{\overline{T}\} = \theta = \frac{1}{\lambda}$

$$
\begin{aligned}
\Rightarrow e^{-\frac{t}{\overline{T}}} &= e^{-\frac{t}{\theta}} + (\overline{T} - \theta) e^{-\frac{t}{\theta}} \frac{t}{\theta^2} + \frac{1}{2}(\overline{T} - \theta)^2 e^{-\frac{t}{\theta}} \left\{ (\frac{t}{\theta^2})^2 - \frac{2t}{\theta^3} \right\} \\
&\quad + \cdots \\
\Rightarrow E\{e^{-\frac{t}{\overline{T}}}\} &= e^{-\frac{t}{\theta}} + 0 + \frac{1}{2} E\{(\overline{T} - \theta)^2\} e^{-\frac{t}{\theta}} \left\{ \frac{t^2}{\theta^4} - \frac{2t}{\theta^3} \right\} \\
&= e^{-\frac{t}{\theta}} \left\{ 1 + \frac{1}{2}\frac{\theta^2}{n} (\frac{t^2}{\theta^4} - \frac{2t}{\theta^3}) \right\}
\end{aligned}
$$

$$
\widetilde{S}(t) = \frac{e^{-\widehat{\lambda} t}}{1 + \frac{1}{2n}(t^2 \widehat{\lambda}^2 - 2t\widehat{\lambda})}
$$

- **TFR , DFR**
- **IFRA , DFRA**

  F or f has an increasing failure if $h(T)$ is increasing

$$H(t) = \int_0^t h(u)du \Rightarrow \frac{H(t)}{t} \text{ increasing} \Rightarrow \text{IFRA}$$

Survival Analysis

Instructor: Hong-Dar Isaac Wu

Department of Applied Mathematics

National Chung Hsing Uuiversity

February 14, 2012

- Censoring $\begin{cases} \text{type I censoring} \\ \text{type II censoring} \end{cases}$

- Censoring $\begin{cases} \text{right censoring} \\ \text{left censoring} \\ \text{interval censoring} \end{cases}$

☆ T_1 is the survival time of "1"(imagine there is a "right-censoring" time C_1)

☞ $C_1 > T_1$

you observe T_1 if $T_1 \leq C_1$

you observe C_1 if $C_1 < T_1$

☞ $T_2 > C_2$

you observe C_2 rather than T_2(you only know $T_2 > C_2$)

☆ Right-censoring:

T is the true survival time, C is the censoring time. We observe:

① $T \cap C \equiv \min(T, C) \equiv x$

② $\delta = 1_{\{T \leq C\}}$

☆ For type II censoring,

$$x_{(1)} < x_{(2)} < \cdots < x_{(r)} < x_{(r+1)} < \cdots < x_{(n)}$$

☆ Likelihood:

if $x_1, ..., x_n \sim f_\theta(x)$,

$$l_\theta = \binom{n}{r} \left(\prod_{i=1}^{r} f_\theta(x_{(i)}) \right) \left(S_\theta(x_{(r)}) \right)^{n-r}$$

☆ For right-censored data

$T_1, \ldots T_n \sim f_\theta(t),\ F_\theta(t),\ S_\theta(t) = 1 - F_\theta(t),\ C_1, \ldots, C_n \sim g(\cdot),\ G(\cdot)$

$$l_\theta = \prod_{i=1}^{n} f_\theta(x_i)^{\delta_i} \left(1 - G(x_i)\right)^{\delta_i} g(x_i)^{1-\delta_i} S_\theta(x_i)^{1-\delta_i}$$

where $x_i = T_i \cap C_i$, $\delta_i = 1_{\{T_i \le C_i\}}$, $i = 1, 2, \ldots, n$.

☆ Because G and g do not contain the information of θ, the likelihood can be re-written as

$$l_\theta = \prod_{i=1}^{n} f_\theta(x_i)^{\delta_i} S_\theta(x_i)^{1-\delta_i} = \prod_{i=1}^{n} h_\theta(x_i)^{\delta_i} S_\theta(x_i)$$

☆ Statistical information can be made based on the constructed likelihood. (Parametric!)

☆ In some non-or semi-parametric model(s)(say, Cox's PH model), you have $h(\cdot)$ & $S(\cdot)$.

§ 3.6 Counting Process

☆ $N(t)(t \geq 0)$ is a stochastic process

(i) $N(0) = 0$

(ii) $N(t) < \infty \Rightarrow P_r\{w; N(t) < \infty\} = 1$

(iii) The sample path of $N(t)$ are

① right continuous;

② piecewise constant;

③ jump size=+1;

☆ Right-censoring:

$$N_i(t) = 1_{\{T_i \le t,\ \delta_i = 1\}}$$

☆ $N_i(t)$ is the counting process associated with individual i, $\sum N_i(t) \equiv N(t)$ is also a counting process. It counts the number of "deaths"(or "events") at and prior to time t.

☆ For right censored data, the information contained in $\{N_i(t)\}_{i=1}^n$ includes knowledge of who has been censored prior to t and who had died before or at t.

$Z(t) \longrightarrow$

☆ **Def : Data history(or filtration)**

The accumulated knowledge about a set of covariates(possibly be time-dependent), censoring variables, counting processes up to time t.

$$\begin{aligned} F_t &= \sigma(\{Z_i(t),\ T_i,\ \delta_i\}_{i=1}^n,\ 0 < t) \\ &\equiv \text{sigma algebra generated by } (\{Z_i(t),\ T_i,\ \delta_i\}_{i=1}^n,\ 0 < t) \end{aligned}$$

$F_s \subseteq F_t$ if $s < t$ and F_t is a filtration (increasing σ-algebra)

☆ For a given $N(t)$

- $dN(t) = N(t + dt-) - N(t-)$
- assuming no "fies"
- For right-censored data,

$$dN(t) = \begin{cases} 1, & \text{if there is "one event" accurred at "}t\text{"} \\ 0, & \text{o.w.} \end{cases}$$

- at-risk indicator process

$$\begin{cases} Y_i(t) = 1_{\{x_i \geq t\}} = 1, & \text{if } t \leq T_i \text{ and } t \leq C_i \\ X_i \equiv T_i \cap C_i = 0, & \text{o.w.} \end{cases}$$

It is "known" at t(constant) predictable process.

$Y(t) = \sum_{i=1}^{n} Y_i(t)$ =total# of "at risk" individuals at time t.

☆ **Property :**

$$
\begin{aligned}
E\{dN(t)|F(t-)\} &= E\{\#\text{ of obs. } w/, t \le x_i \le t+dt,\\
&\quad c_i > t+dt|F(t-)\}\\
&= \text{"\# of total at-risk"} \times P_r\{T \in [t, t+dt)|F(t-)\}\\
&= \left(\sum Y_i(t)\right) \times h_i(t)dt\\
&= \sum Y_i(t)h_i(t)dt
\end{aligned}
$$

- $Y(t)h(t)$ is called the intensity process
- $\lambda_i(t) = Y_i(t)h_i(t)$
  $\Lambda_i(t) = \int_0^t \lambda_i(u)du$ cumulative intensity process
- $E\{dN(t)|F(t-)\} = Y(t)h(t)dt = d\Lambda(t) \Rightarrow E\{N(t)|F(t-)\} = \Lambda(t)$

  $d\Lambda(t)$ is called a compensator of $dN(t)$, $\Lambda(t)$ is the compensator of $N(t)$.

☆ **Def:**

$M(t) = N(t) - \Lambda(t)$ or $dM(t) = dN(t) - d\Lambda(t)$

$$\begin{aligned} E\{M(t)|F(t-)\} &= M(t-) \\ \text{or } E\{M(t)|F(s),\ s<t\} &= M(s) \\ \text{or } E\{dM(t)|F(s),\ s<t\} &= 0 \end{aligned}$$

- Predictable process:

$$E\{\Lambda(t)|F(t-)\} = \Lambda(t)$$

where $\Lambda(t)$, $Y(t)$, and $Z(t)$ are predictable.

- Predictable covariation process of
$M(t) \equiv < M > (t) \equiv$ compensator of $M^2(t)$

$$\begin{aligned} dM^2(t) &= \{M((t+dt)-)\}^2 - \{M(t-)\}^2 \\ &= \{M(t-) + dM(t)\}^2 - (M(t-))^2 \\ &= 2M(t-)dM(t) + (dM(t))^2 \end{aligned}$$

☆ **Property :**

$$\mathrm{Var}\{dM(t)|F(t-)\} = d < M > (t) = dM^2(t)$$

✏ **Eg :**

Non-parametric estimate of $H(t)$, the cumulative hazard $\hat{H}(t)$

$$\begin{aligned} \Lambda(t) &= Y(t)H(t) \\ d\Lambda(t) &= Y(t)dH(t) \end{aligned}$$

➢ *Sol* :

$$
\begin{aligned}
N(t) &= \Lambda(t) + M(t) \\
M(t) &= N(t) - \Lambda(t) \\
dN(t) &= d\Lambda(t) + dM(t)
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow \quad & \frac{dN(t)}{Y(t)} = \frac{d\Lambda(t)}{Y(t)} + \frac{dM(t)}{Y(t)} \\
\Rightarrow \quad & E\left\{ \frac{dN(t)}{Y(t)} \middle| F(t-) \right\} = E\left\{ \frac{d\Lambda(t)}{Y(t)} \middle| F(t-) \right\} + E\left\{ \frac{dM(t)}{Y(t)} \middle| F(t-) \right\} \\
& \text{Let } J(t) = 1_{\{Y(t)>0\}} \\
\Rightarrow \quad & \int_0^t \frac{J(u)dN(u)}{Y(u)} = \int_0^t \frac{J(u)}{Y(u)} Y(t)dH(t) + \int_0^t \frac{J(u)dM(u)}{Y(u)} \\
& = \int_0^t J(u)dH(u) \equiv H^*(t)
\end{aligned}
$$

☆ Non-parametric estimator of $H(t)$

$$\begin{aligned} N(t) &= \Lambda(t) + M(t) \\ M &= N - \Lambda \end{aligned}$$

$$\begin{aligned} \int_0^t \frac{J(u)dN(u)}{Y(u)} &= \sum_i \frac{J(t_i)}{Y(t_i)} \cdot 1 \\ &= \int_0^t J(u)dH(u) + \int_0^t \frac{J(u)}{Y(u)} dM(u) \end{aligned}$$

☆ Martingale Transformation Theorem(1984) R.Gill (JASA)
$\int p(t)dM(t) \equiv M^*(t)$ is still a martingale process.

$$\begin{aligned} Y &= X\beta + \varepsilon \\ \frac{1}{n}X'Y &= \frac{1}{n}X'X\beta + \frac{1}{n}X'\varepsilon \\ \hat{\beta} &\approx (X'X)^{-1}X'Y \xrightarrow{pr.} 0 \end{aligned}$$

☆ Observe that

$$\int_0^t J(u)dH(u) = H(t)1_{\{0<t<\tau\}} = \text{cumulative hazard}$$

$$\begin{aligned} \therefore \hat{H}(t) &= \int_0^\tau \frac{dN(u)}{Y(u)} \\ N(t) &= \sum N_i(t) \\ Y(t) &= \sum Y_i(t) \end{aligned}$$

$$
\begin{aligned}
\hat{H}(t) &= 0 \; if \;\; t < x_1 \\
&= \frac{1}{9} = \frac{1}{1+1+\cdots+1} \quad x_1 \le t < x_2 \\
&= \frac{1}{9} + \frac{1}{0+1+\cdots+1} \quad x_2 \le t < x_4 \\
&= \frac{1}{9} = \frac{1}{8} + \frac{1}{6} \quad x_4 \le t \\
&\vdots
\end{aligned}
$$

☆ $Q : (HW)\ \mathrm{Var}(\hat{H}(t)) = ?$

$\hat{H}_{N-A}(t)$, $\exp\{-H(t)\} = S(t)$

So you can estimate $S(t)$ by $\exp\{-\hat{H}_{N-A}(t)\} = \hat{S}(t)$

☆ If you actually estimate $S(\cdot)$ in the previous manner

$$\Rightarrow d\hat{S}(t) = \exp\{-d\hat{H}(t)\} \approx 1 - d\hat{H}(t)$$

(if dH is considered as being "very" small)

☆ For discrete case and by the predict of consecutive conditional probability.(see later for $K - M$ estimator)

$$
\begin{aligned}
\prod_{\{t\}} d\hat{s}(\text{discrete case}) &= \hat{S}_{K-M}(t) \\
&= \prod_{u=0}^{t} (1 - d\hat{H}(u)) \\
&= \prod_{u=0}^{t} \{1 - \frac{dN(u)}{Y(u)}\} \\
&\equiv \text{Kaplan-Meier(1958) estimator} \\
&\equiv \hat{S}_{K-M}(t)
\end{aligned}
$$

(More details)(Fleming and Harringtor, 1991, P96∼P97)

$$H(t) = \int_0^t h(u)du = (-1)\int_0^t \frac{dS(u)}{S(u-)}$$

$$S(t) = \int_0^t dS(u) = -\int_0^t S(u-)dH(u)$$

which means S(·) is uniqualy determined by $H(\cdot)$

$\therefore$ By plugging in $\hat{H}$,

$$\hat{S}(t) = -\int_0^t \hat{S}(u-)d\hat{H}(u) = -\int_0^t \hat{S}(u-)\cdot\frac{dN(u)}{Y(u)}$$

$$\hat{S}(t-) - \hat{S}(t) = (-1)\Delta\hat{S}(t) = \hat{S}(t-)\frac{\Delta N(t)}{Y(t)}$$

$$\therefore \hat{S}(t) = \hat{S}(t-)\left\{1 - \frac{\Delta N(t)}{Y(t)}\right\}$$

☆ consider all jump points,

$$\hat{S}_{K-M}(t) = \prod_{S<t} \left\{ 1 - \frac{\Delta N(S)}{Y(S)} \right\}$$

$$\sup |F_n(x) - F_0(x)| \xrightarrow{a.s} 0$$

§ Chapter 4

☆ T_i may be a failure time or censoring time

We have n observation, $T_1 < T_2 < \cdots < T_n$(ordered)

☆ Let $T_1 < T_2 < \cdots < T_n$ be time points where there is exactly one-failure at time $t_j \Rightarrow$

$$
\begin{aligned}
P_r\{T \geq t_j\} &= S_T(t_j) \\
&= P_r\{T \geq t_j | T \geq t_{j-1}\} \cdot P_r\{T \geq t_{j-1}\} \\
&= \cdots \\
&= P_r\{T \geq t_j | T \geq t_{j-1}\} \cdot P_r\{T \geq t_{j-1} | T \geq t_{j-2}\} \cdots \\
& \quad P_r\{T \geq t_2 | T \geq t_1\} \cdot P_r\{T \geq t_1 | T \geq t_0\} \cdot \\
& \quad P_r\{T \geq t_0\}
\end{aligned}
$$

☆ or

$$1 \cdot (1 - \frac{d_1}{Y_1}) \cdot (1 - \frac{d_2}{Y_2}) \cdot \cdots \cdot (1 - \frac{d_j}{Y_j})$$

$$d_j = \# \text{ of failures occured at } t_j \text{ } (if\ d_j \geq 2 \Rightarrow \text{ ties})$$

$$Y_j = \# \text{ of at-risk at } t_j = \sum_{l=1}^{n} Y_l(t_j) = \#R(t_j)$$

☆ Therefore

$$\hat{S}_{KM}(t_j) = \prod_{i=1}^{j} \left\{ 1 - \frac{d_i}{Y_i} \right\}$$

$$= \prod_{s<t} \left\{ 1 - \frac{dN(S)}{Y(S)} \right\} \text{ in terms of " counting process "}$$

$$\begin{cases} N(t) = \sum_{i=1}^{n} N_i(t) \\ Y(t) = \sum Y_i(t) \end{cases}$$

$$\hat{S}_{KM}(t) = \prod_{\{i:t_i \leqslant t\}} \left\{1 - \frac{d_i}{Y_i}\right\} = \prod_{\{i:t_i \leqslant t\}} \left\{1 - \frac{Y_i - d_i}{Y_i}\right\}$$

$$\therefore \ \log \hat{S}_{KM} = \sum \log \hat{P}_i \qquad \text{where } \hat{P}_i = \frac{Y_i - d_i}{Y_i}$$

$$Y_i \hat{P}_i \sim \text{Bin}(Y_i, P_i)$$

$$\text{Var}(\log \hat{P}_i) \overset{\text{Taylor exp.}}{\approx} \frac{1}{P_i^2}\text{Var}(\hat{P}_i) = \frac{1}{P_i^2}\frac{P_i(1-P_i)}{Y_i} = \frac{1-P_i}{Y_i P_i}$$

$$\left(* \ \log \hat{P}_i \simeq \log P_i + (\hat{P}_i - P_i)\frac{1}{P_i} + \text{remainder}\right)$$

☆ assume $\left\{ \log \hat{P}_i \right\}_{i=1}^{n}$ are "mutually indep"

$$\Rightarrow \text{Var}\{\log \hat{S}_{KM}\} = \sum \frac{1-P_i}{Y_i P_i} \approx \sum \frac{1-\hat{P}_i}{Y_i \hat{P}_i} = \sum \frac{\frac{d_i}{Y_i}}{Y_i \frac{Y_i - d_i}{Y_i}} = \sum \frac{d_i}{Y_i(Y_i - d_i)}$$

- This is called Greenwood's formula derived from "actuarial method".
- Again by $\delta-$method,

$$\begin{aligned} \text{Var}\hat{S} &\approx \hat{S}^2\text{Var}(\log \hat{S}) \\ \therefore \text{Var}\hat{S}_{KM}(t) &= \hat{S}^2_{KM}(t) \sum_{\{i:t_i \le t\}} \frac{d_i}{Y_i(Y_i - d_i)} \end{aligned}$$

- The "independence" assumption. Is a problem!

(sol):

$$\begin{aligned} \log \hat{S} &\approx \log S + (\hat{S} - S)\frac{1}{S} \\ &\Rightarrow \text{Var}(\log \hat{S}) = \frac{1}{S^2}\text{Var}(\hat{S}) \\ &\Rightarrow \text{Var}(\hat{S}) = S^2\text{Var}(\log \hat{S}) \end{aligned}$$

☆ 95% CI of $S(t)$, at every t

$$\begin{aligned} g(\hat{S}(t)) &= g(S(t)) + (\hat{S}(t) - S(t))g'(S(t)) + \\ &\quad \frac{1}{2}(\hat{S}(t) - S(t))^2 g''(S(t)) + \cdots \\ \mathrm{Var}(g(\hat{S})) &\approx \left(g'(S(t))\right)^2 \mathrm{Var}(\hat{S}(t)) \end{aligned}$$

☆ In general, we can meet the following large-sample property:

$$\frac{g(\hat{S}(t)) - g(S(t))}{\sqrt{\hat{\mathrm{Var}} g(\hat{S}(t))}} = \pm Z_{\alpha/2}$$

☆ The 100(1-α)% CI of $g\big(S(t)\big)$, at t is

$$g\big(\hat{S}(t)\big) \pm Z_{\alpha/2} g'\big(S(t)\big)\sqrt{\text{Var}\hat{S}(t)}$$

Now, take $g(x) = \log\{-\log x\} \quad 0 < x < 1$

i.e. $g(\cdot)$ is the so-called complementary $\log - \log$ link

$$\text{Therefore } g'(x) = \frac{1}{-\log x}(-1)\frac{1}{x} = \frac{1}{x\log x}$$

So

$$\log\Big\{-\log\hat{S}(t)\Big\} \pm Z_{\alpha/2}\frac{1}{\hat{S}(t)\log\hat{S}(t)}\sqrt{\text{Var}\hat{S}(t)}$$

☆ Let $\frac{\sqrt{\text{Var}\hat{S}}}{\hat{S}} \equiv \sigma_{\hat{S}}$ or $\frac{\text{Var}\hat{S}}{\hat{S}^2} \equiv \sigma^2_{\hat{S}}$

Further, $\log S(t) = -H(t)$ or $\log \hat{S}(t) = -\hat{H}(t)$

$$\frac{Z_{\alpha/2}\sigma_{\hat{S}}}{\log \hat{S}} \equiv \log \hat{\theta}$$

$$\log\left(\hat{H}(t)\right) \pm \log(\hat{\theta}) = -\hat{\theta}\log \hat{S},\ \frac{-1}{\hat{\theta}}\log \hat{S}$$

- CI of $H(t)$: $\exp\{\log \hat{H} \pm \log \hat{\theta}\} = \hat{H}\hat{\theta},\ \hat{H}\frac{1}{\hat{\theta}}$
- CI of S:

$$\exp\{-H\} = \left(\exp\{-\hat{H}\frac{1}{\hat{\theta}}\}, \exp\{-\hat{H}\hat{\theta}\}\right) = \left(\left(\hat{S}(t)\right)^{\frac{1}{\hat{\theta}}}, \left(\hat{S}(t)\right)^{\hat{\theta}}\right)$$

☆ Two-sample comparison Test $H_0 : \mu_1 = \mu_2$ $T - test$

$H_0 : m_1 = m_2$?

$$m_1 = \inf\{x : F(x) \geqslant 1/2\}$$

$$m_2 = \inf\{x : G(x) \geqslant 1/2\}$$

☆ $H_0 : S_1(t) = S_2(t) \;\; \forall t \;\; (\text{or } h_1(t) = h_2(t))$

Group1: treatment Group2: control

$$P_r\{\text{the death } \in G1 | \text{one death at time } t_1\} = \frac{n_1}{n_1 + m_1}$$

$$P_r\{\text{the death } \in G2 | \text{one death at time } t_1\} = \frac{m_1}{n_1 + m_1}$$

☆ a is the random variable (fixed marginal total)

$a = 1,\ P = \frac{n1}{n1+m1}$

$a = 0,\ p = \frac{m1}{n1+m1}$

$\mathrm{E}(a) = p,\ \mathrm{Var}(a) = p(1-p)$

	death	survival	
G1	a		n1
G2	1-a		m1
	1	$N_1 - 1$	

✍ Ex: $n1 = 5,\ m1 = 5$

at t_1	death	survival	
G1	1	4	5
G2	0	5	5
	1	9	10

at t_2	death	survival	
G1	0	3	3
G2	1	4	5
	1	7	8

$n_j = \sum_{k=1}^{n1} Y_{1k}(t_j),\ m_j = \sum_{k=1}^{m1} Y_{2k}(t_j)$

$P_r\{a_2 = 1 | \text{one death at } t_2\} = \frac{n_2}{N_2}$

$P_r\{a_2 = 0 | \text{one death at } t_2\} = \frac{m_2}{N_2}$

☆ k-number 2×2 table

a_1	b_1	n_1
c_1	d_1	m_1
s_1	t_1	N_1

. . .

a_k	b_k	n_k
c_k	d_k	m_k
s_k	t_k	N_k

$\hat{\varphi}_1$ =odds ratio of table 1$=\frac{a_1 d_1}{b_1 c_1}$ . . . $\hat{\varphi}_k = \frac{a_k d_k}{b_k c_k}$

Sum:

$$\frac{\sum_{i=1}^k a_i - \mathrm{E}(\sum_{i=1}^k a_i)}{\sqrt{\mathrm{Var}(\sum_{i=1}^k a_i)}} \sim N(0,1)$$

✍ $H_0 : \varphi_1 = \cdots = \varphi_k \equiv \varphi_0$ (common odds ratio homogeneity)

$\hat{\varphi}_0 =$? estimation!

☆ $a_k = 1$,

$$P_r\{a_k = 1 | \text{one death at } t_k\} = \frac{n_k}{n_k + m_k}$$

$$\mathrm{E}(a_k) = \frac{n_k}{n_k + m_k} \text{ if } h_1(t) = h_2(t)$$

$$\left(\frac{\sum_{i=1}^{k} a_i + \sum_{i=1}^{k} (\frac{n_i}{n_i+m_i})}{\sqrt{\sum_{i=1}^{k} (N_i \frac{n_i}{n_i+m_i} \frac{m_i}{n_i+m_i})}} \right)^2 \overset{H_0}{\sim} \mathcal{X}_1^2$$

☆ Note:

a		n
		m
d	N-d	N

$$\mathrm{E}(a) = N\frac{n}{N}\frac{d}{N} = \frac{nd}{N} \stackrel{d=1}{=} \frac{n}{N}$$

$$\mathrm{Var}(a) = \frac{nmd(N-d)}{N^2(N-1)} \stackrel{d=1}{=} \frac{nm}{N^2}$$