Statistical Learning
Introduction to Survival Analysis
Spring 2024

Outline

- **Preliminaries**
- Kaplan–Meier estimator

Survival outcome

- What is the survival outcome? the time to a clinical event of interest: terminal and non-terminal events.
  1. the time from diagnosis of cancer to death
  2. the time between administration of a vaccine and infection date
  3. the time from the initiation of a treatment to the time of the disease progression.
- Let T be a nonnegative random variable denoting the time to the event of interest (survival time/event time/failure time).
- The distribution of T could be discrete, continuous or a mixture of both. We will focus on the continuous distribution.

Survival time T

The distribution of a random variable $T \geq 0$ can be characterized by its probability density function (PDF) and cumulative distribution function (CDF). However, in survival analysis, we often focus on

1. Survival function: $S(t) = \mathrm{pr}(T > t)$. If T is time to death, then $S(t)$ is the probability that a subject survives beyond time t.
2. Hazard function:

$$h(t) = \lim_{\epsilon \downarrow 0} \frac{\mathrm{pr}(T \in (t, t+\epsilon] \mid T \geq t)}{\epsilon}$$

3. Cumulative hazard function: $H(t) = \int_0^t h(u) du$.

Relationships between survival and hazard functions

- Hazard function: $h(t) = f(t)/S(t)$.
- Cumulative hazard function

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

- $f(t) = h(t)S(t) = h(t)\exp\{-H(t)\}$.
- $S(t) = \exp\{-H(t)\}$.

Additional properties of hazard functions

- If $H(t)$ is the cumulative hazard function of T, then $H(T) \sim \mathrm{EXP}(1)$, the unit exponential distribution. (Equivalent to the statement that $F(T) \sim U(0,1)$, where $F(\cdot)$ is the CDF of the random variable T.)
- If T_1 and T_2 are two independent survival times with hazard functions $h_1(t)$ and $h_2(t)$, respectively, then $T = \min(T_1, T_2)$ has a hazard function $h_T(t) = h_1(t) + h_2(t)$. (This statement can be generalized to the case with more than two survival times)

Interpretability

- The hazard function $h(t)$ is NOT the probability that the event (such as death) occurs at time t or before time t.
- $h(t)\gamma$ is approximately the conditional probability that the event occurs within the interval $(t, t + \gamma]$ given that the event has not occurred before time t for small $\gamma > 0$.
- If the hazard function $h(t)$ increases $X\%$ at $[0, \tau]$, the probability of failure before τ in general does not increase $X\%$.

Exponential distribution

- In survival analysis the exponential distribution is the “simplest” parametric distribution for survival time.
- Denote the exponential distribution by $\mathrm{EXP}(\lambda)$.
- $f(t) = \lambda e^{-\lambda t}$.
- $F(t) = 1 - e^{-\lambda t}$.
- $h(t) = \lambda$; constant hazard.
- $H(t) = \lambda t$.

Exponential distribution

- $E(T) = \lambda^{-1}$.
  The higher the hazard, the shorter the expected survival time.
- $\text{Var}(T) = \lambda^{-2}$.
- Memoryless property: $\text{pr}(T > t) = \text{pr}(T > t + s \mid T > s), t, s > 0$.
- $c_0 \times \text{EXP}(\lambda) \sim \text{EXP}\left(\lambda / c_0\right)$ for $c_0 > 0$.
- The log-transformed exponential distribution is the so called extreme value distribution.

Gamma distribution

- Gamma distribution is a generalization of the simple exponential distribution.
- Be careful about the parametrization $G(\alpha, \lambda)$, shape α, rate $\lambda > 0$:
  The density function

$$f(t) = \frac{\lambda^{\alpha} t^{\alpha-1} e^{-\lambda t}}{\Gamma(\alpha)} \propto t^{\alpha-1} e^{-\lambda t}$$

  where

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt$$

  is the Gamma function. For integer $\alpha, \Gamma(\alpha) = (\alpha - 1)\,!$.

Gamma distribution

- $E(T) = \alpha\lambda^{-1}$.
- $\mathrm{Var}(T) = \alpha\lambda^{-2}$.
- If $T_i \sim G(\alpha_i, \lambda), i = 1, \cdots, K$ and $T_i, i = 1, \cdots, K$ are independent, then

$$\sum_{i=1}^{K} T_i \sim G\left(\sum_{i=1}^{K} \alpha_i, \lambda\right).$$

- $G(1, \lambda) \sim \mathrm{EXP}(\lambda)$. (A generalization of the exponential distribution)
- Increasing hazard $\alpha > 1$; constant hazard $\alpha = 1$; decreasing hazard $0 < \alpha < 1$.

Weibull distribution

- Weibull distribution is also a generalization of the simple exponential distribution.
- Be careful about the parametrization $W(p, \lambda), \lambda > 0$ (rate parameter) and $p > 0$ (shape parameter).

1. $S(t) = e^{-(\lambda t)^p}$.
2. $f(t) = p\lambda(\lambda t)^{p-1} e^{-(\lambda t)^p} \propto t^{p-1} e^{-(\lambda t)^p}$.
3. $h(t) = p\lambda(\lambda t)^{p-1} \propto t^{p-1}$.
4. $H(t) = (\lambda t)^p$.

Weibull distribution

- $E(T) = \lambda^{-1}\Gamma(1 + 1/p)$.
- $\mathrm{Var}(T) = \lambda^{-2}\left[\Gamma(1 + 2/p) - \{\Gamma(1 + 1/p)\}^2\right]$.
- $W(1, \lambda) \sim \mathrm{EXP}(\lambda)$.
- $W(p, \lambda) \sim \{\mathrm{EXP}\left(\lambda^p\right)\}^{1/p}$.

Log-normal distribution

- The log-normal distribution is another commonly used parametric distribution for characterizing the survival time.
- $LN\left(\mu, \sigma^2\right) \sim \exp\left\{N\left(\mu, \sigma^2\right)\right\}$.
- $E(T) = e^{\mu+\sigma^2/2}$.
- $\mathrm{Var}(T) = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)$.

Generalized gamma distribution

- The generalized gamma distribution becomes popular due to its flexibility.
- Again be careful about its parametrization $GG(\alpha, p, \lambda)$.
- $f(t) = p\lambda(\lambda t)^{\alpha-1}e^{-(\lambda t)^p}/\Gamma(\alpha/p) \propto t^{\alpha-1}e^{-(\lambda t)^p}$.
- $S(t) = 1 - \gamma\left\{\alpha/p, (\lambda t)^p\right\}/\Gamma(\alpha/p)$, where

$$\gamma(s, x) = \int_0^x t^{s-1}e^{-t}dt$$

is the incomplete gamma function.

Generalized gamma distribution

- For $k = 1, 2, \cdots$

$$E\left(T^k\right) = \frac{\Gamma\{(\alpha + k)/p\}}{\lambda^k \Gamma(\alpha/p)}.$$

- If $p = 1, GG(\alpha, 1, \lambda) \sim G(\alpha, \lambda)$.
- if $\alpha = p, GG(p, p, \lambda) \sim W(p, \lambda)$.
- if $\alpha = p = 1, GG(1, 1, \lambda) \sim \mathrm{EXP}(\lambda)$.
- The generalized gamma distribution can be used to test the adequacy of commonly used Gamma, Weibull and Exponential distributions, since they are all nested within the generalized gamma distribution family.

Homogeneous Poisson Process

- $N(t) = \#$ events occurring in $(0, t)$.
- T_1 denotes the time to the first event
  T_2 denotes the time from the first to the second event
  T_3 denotes the time from the second to the third event ...
- If the gap times $T_1, T_2, \cdots$ are i.i.d. $\mathrm{EXP}(\lambda)$, then

  $$N(t+s) - N(t) \sim \mathrm{Poisson}(\lambda s).$$

  The process $N(t)$ is called the homogeneous Poisson process.
- The interpretation of the intensity function (similar to hazard function)

  $$\lim_{\epsilon \downarrow 0} \frac{\mathrm{pr}\{N(t+\epsilon) - N(t) > 0\}}{\epsilon} = \lambda.$$

Censoring

- A common feature of survival data is the presence of censoring.
- There are different types of censoring. Suppose that $T_1, T_2, \cdots, T_n$ are i.i.d. survival times.

1. Type I censoring: observe only

$$(U_i, \delta_i) = \{\min(T_i, c), I(T_i \leq c)\}, i = 1, \cdots, n,$$

   i.e., we only have the survival information up to a fixed time c.

2. Type II censoring: observe only

$$T_{(1,n)}, T_{(2,n)}, \cdots, T_{(r,n)}$$

   where $T_{(i,n)}$ is the i th smallest survival time, i.e., we only observe the first r smallest survival times.

Censoring

3. Random censoring (The most common type of censoring): $C_1, C_2, \cdots, C_n$ are potential censoring times for n subjects, observe only

$$(U_i, \delta_i) = \{\min(T_i, C_i), I(T_i \leq C_i)\}, i = 1, \cdots, n$$

We often treat the censoring time C_i as i.i.d. random variables in statistical inferences.

4. Interval censoring: observe only $(L_i, U_i), i = 1, \cdots, n$ such that $T_i \in [L_i, U_i)$.

Non-informative censoring

- If T_i and C_i are independent, then censoring is non-informative.
- Examples of non-informative censoring.
  1. administrative censoring
  2. random drop off

Non-informative censoring

- Noninformative censoring condition:

$$h(t) = \lim_{\epsilon \downarrow 0} \frac{\mathrm{pr}(T \in [t, t+\epsilon] \mid T \geq t, C \geq t)}{\epsilon}$$

- It is slightly weaker than the independence between T and C.
- Consequences of informative censoring:
  1. There are more than one distribution for (T, C) with different marginal distribution of T correspond to the same distribution of $(U, \delta) = \{\min(T, C), I(T \leq C)\}$.
  2. Based on the distribution (U, δ) alone, it is impossible to determine the distribution of T.

Likelihood construction

- In the presence of right censoring, we only observe $(U_i, \delta_i), i = 1, \cdots, n$.
- The likelihood construction must be with respect to the bivariate random variable $(U_i, \delta_i), i = 1, \cdots, n$.
    1. If $(U_i, \delta_i) = (u_i, 1)$, then $T_i = u_i, C_i \geq u_i$.
    2. If $(U_i, \delta_i) = (u_i, 0)$, then $T_i > u_i, C_i = u_i$.

Likelihood construction

- Assuming $C_i, 1 \leq i \leq n$ are i.i.d. random variables with a CDF $G(\cdot)$.

$$L_i(F,G) = \begin{cases} f(u_i)\,(1 - G(u_i))\,, \text{ if } \delta_i = 1 \\ S(u_i)\,g(u_i)\,, \text{ if } \delta_i = 0 \end{cases}$$

$$\Rightarrow L(F,G) = \prod_{i=1}^{n} L_i(F,G) = \prod_{i=1}^{n} \left[\{f(u_i)\,(1 - G(u_i))\}^{\delta_i} \{S(u_i)\,g(u_i)\}^{1-\delta_i} \right]$$

$$= \left\{ \prod_{i=1}^{n} f(u_i)^{\delta_i} S(u_i)^{1-\delta_i} \right\} \left\{ \prod_{i=1}^{n} g(u_i)^{1-\delta_i} (1 - G(u_i))^{\delta_i} \right\}.$$

Likelihood construction

- We have used the noninformative censoring assumption in the likelihood construction.
- $L(F,G) = L(F) \times L(G)$ and therefore the likelihood-based inference for F can be made based on

$$L(F) = \prod_{i=1}^{n} \left\{ f\left(u_i\right)^{\delta_i} S\left(u_i\right)^{1-\delta_i} \right\} = \prod_{i=1}^{n} h\left(u_i\right)^{\delta_i} S\left(u_i\right)$$

  only.

Outline

- Preliminaries
- **Kaplan–Meier estimator**

Kaplan-Meier (KM) Estimator

- Nonparametric estimation of the survival function $S(t) = \mathrm{pr}(T > t)$.
- The nonparametric estimation is more robust and does not depend on any parametric assumption.

If there is no censoring

- $S(t)$ can be consistently estimated by

$$\hat{S}(t) = n^{-1} \sum_{i=1}^{n} I\left(T_i > t\right).$$

- $\hat{S}$ is a discrete distribution with mass probability of n^{-1} at observed times $T_1, \cdots, T_n$.
- $\hat{S}(t)$ is the nonparametric maximum likelihood estimator (NPMLE) for $S(t)$.

CDF as NPMLE

- Assuming that $F(\cdot)$ is discrete with mass probability at $T_1 < T_2 < \cdots < T_n$, where $\{T_1, T_2, \cdots\}$ are observed times.
- Let $f_1 = \text{pr}\left(T = T_1\right), f_2 = \text{pr}\left(T = T_2\right), \cdots$.
- Objective: estimate $f_1, f_2, \cdots$.
- Method: maximize $\prod_{i=1}^{n} f_i$ subject to $\sum_{i=1}^{n} f_i = 1$.
- Solution: $\hat{f}_1 = \hat{f}_2 = \cdots = \hat{f}_n = n^{-1}$.

Data and Assumptions

- Data: $\{(U_i, \delta_i), i = 1, \cdots, n\}$ where $U_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \le C_i)$.
- Assumptions:
  1. $T_1, \cdots, T_n$ i.i.d. $\sim F(\cdot) = 1 - S(\cdot)$
  2. $C_1, \cdots, C_n$ i.i.d. $\sim G(\cdot)$
  3. $T_i \perp C_i, i = 1, \cdots, n$. Noninformative censoring!

If there is censoring

- Assuming that $F(\cdot)$ is discrete with mass probability at $v_1 < v_2 < \cdots$, where $\{v_1, v_2, \cdots\}$ are observed times.
- Let $f_1 = \mathrm{pr}\left(T = v_1\right), f_2 = \mathrm{pr}\left(T = v_2\right), \cdots$.
- Objective: estimate $f_1, f_2, \cdots$.

Example

- Obs: $2, 2, 3^+, 5, 5^+, 7, 9, 16, 16, 18^+$, where $^+$means censored
- $v_1 = 2; v_2 = 3, v_3 = 5, v_4 = 7, v_5 = 9, v_6 = 16, v_7 = 18, v_8 = 18^+$
- The likelihood function in terms of $(f_1, f_2, \cdots)$:
  $L(F) =$
  $f_1^2 (f_3 + f_4 + f_5 + f_6 + f_7 + f_8) f_3 (f_4 + f_5 + f_6 + f_7 + f_8) f_4 f_5 f_6^2 f_8,$
  where $f_1 + f_2 + f_3 + f_4 + f_5 + f_6 + f_7 + f_8 = 1$

Reparametrization tricks

- The discrete hazard function: $h_1 = \mathrm{pr}\left(T = v_1\right)$ and $h_j = \mathrm{pr}\left(T = v_j \mid T > v_{j-1}\right), j > 2$
- For $t \in \left[v_j, v_{j+1}\right)$

$$S(t) = \mathrm{pr}(T > t) = \mathrm{pr}\left(T > v_j\right) = \prod_{i=1}^{j}(1 - h_i)$$

- For $t = v_j$

$$f_j = f(t) = \mathrm{pr}(T = t) = h_j \prod_{i=1}^{j-1}(1 - h_i)$$

Reparametrization tricks

- The likelihood function in terms of $(h_1, h_2, \cdots)$:

$$
\begin{aligned}
L(F) =& h_1^2 \times \{(1-h_1)(1-h_2)\} \times \{(1-h_1)(1-h_2)h_3\} \\
& \times \{(1-h_1)(1-h_2)(1-h_3)\} \times \{(1-h_1)(1-h_2)(1-h_3)h_4\} \\
& \times \{(1-h_1)(1-h_2)(1-h_3)(1-h_4)h_5\} \\
& \times \{(1-h_1)(1-h_2)(1-h_3)(1-h_4)(1-h_5)h_6\}^2 \\
& \times \{(1-h_1)(1-h_2)(1-h_3)(1-h_4)(1-h_5)(1-h_6)(1-h_7)\} \\
=& h_1^2 (1-h_1)^8 \times (1-h_2)^8 \times h_3 (1-h_3)^6 \\
& h_4 (1-h_4)^4 \times h_5 (1-h_5)^3 \times h_6^2 (1-h_6) \times (1-h_7)
\end{aligned}
$$

KM estimation

- The likelihood function

$$L(F) = \prod_j h_j^{d_j} \left(1 - h_j\right)^{Y\left(v_j\right) - d_j}$$

where

$$d_j = \sum_{i=1}^{n} \delta_i I\left(U_i = v_j\right) = \#\text{ failures at } v_j$$

$$Y\left(v_j\right) = \sum_{i=1}^{n} I\left(U_i \geq v_j\right) = \#\text{ "at risk" at } v_j.$$

KM estimation

- $\hat{h}_j = d_j / Y\left(v_j\right)$

$$\hat{S}(t) = \begin{cases} 1 & t < v_1 \\ \prod_{i=1}^{j}\left(1 - \hat{h}_i\right) & v_j \leq t < v_{j+1} \end{cases}$$

which is the Kaplan-Meier estimator.

Example

v_j	$Y(v_j)$	d_j	$\hat{h}_j$	$\hat{S}(v_j) = \prod_{i=1}^{j}(1-\hat{h}_i) = \hat{P}(T > v_j)$
2	10	2	$2/10$	$.8$
5	7	1	$1/7$	$.69 \quad (= .8 \times \frac{6}{7})$
7	5	1	$1/5$	$.55 \quad (= .69 \times \frac{4}{5})$
9	4	1	$1/4$	$.41 \quad (= .55 \times \frac{3}{4})$
16	3	2	$2/3$	$.14 \quad (= .41 \times \frac{1}{3})$
18	1	0	0	$.14$

Remarks

- Suppose that v_g denotes the largest v_j for which $Y\left(v_j\right) > 0$.
  1. if $d_g = Y\left(v_g\right)$, then $\hat{S}(t) = 0$ for $t \geq v_g$
  2. if $d_g < Y\left(v_g\right)$, then $\hat{S}(t) > 0$ but not defined for $t > v_g$.
- The survival distribution may not be estimable with right-censored data. Implicit extrapolation is sometimes used.
- The KM estimator can also be used to estimate the survival function for the censoring distribution.
- KM estimator is a special MLE

$$1 - \hat{S}(t) = \operatorname{argmax}_F L(F)$$

where F is the CDF for all discrete random variables (nonparametric MLE).

Redistribution of Mass

Step 1	$\underline{2}$	$\underline{2}$	$\underline{3^+}$	$\underline{5}$	$\underline{5^+}$	$\underline{7}$	$\underline{9}$	$\underline{16}$	$\underline{16}$	$\underline{18^+}$
Step 2	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$
Step 3	$\downarrow$	$\downarrow$	$\hookrightarrow$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$
	$\downarrow$	$\downarrow$		$\downarrow$	$\hookrightarrow$	$\frac{1}{5}\left(\frac{8}{70}\right)$	$\frac{1}{5}\left(\frac{8}{70}\right)$	$\frac{1}{5}\left(\frac{8}{70}\right)$	$\frac{1}{5}\left(\frac{8}{70}\right)$	$\frac{1}{5}\left(\frac{8}{70}\right)$
Total Mass	$\underbrace{\qquad\qquad}_{\frac{2}{10}}$		0	$\frac{8}{70}$	0	$\frac{24}{175}$	$\frac{24}{175}$	$\underbrace{\qquad\qquad}_{\frac{48}{175}}$		$\underbrace{\qquad}$ Assume this is somewhere > 18

Self-Consistency

- No censoring $\hat{S}(t) = n^{-1} \sum_{i=1}^{n} I(T_i > t)$
- Right censoring: $\hat{S}(t) = n^{-1} \sum_{i=1}^{n} E\left(I(T_i > t) \mid U_i, \delta_i\right)$
  1. $E\left(I(T_i > t) \mid U_i, \delta_i = 1\right) = I(U_i > t)$
  2. $E\left(I(T_i > t) \mid U_i, \delta_i = 0\right) = S(t)/S(U_i)\, I(t \geq U_i) + I(U_i > t)$
- Self-consistency iteration:

$$\hat{S}_{new}(t) = n^{-1} \sum_{i=1}^{n} \left\{ I(U_i > t) + (1 - \delta_i) \frac{\hat{S}_{old}(t)}{\hat{S}_{old}(U_i)} I(U_i \leq t) \right\}$$

- The solution is still the KM estimator.

Nelson-Aalen Estimator

- How to estimate the cumulative hazard function?

$$\hat{H}(t) = \sum_{i=1}^{j} \hat{h}_i \text{ for } v_j \le t < v_{j+1}$$

- $H(t) = -\log\{S(t)\}$

$$-\log\{\hat{S}(t)\} = \sum_{i=1}^{j} \left\{-\log\left(1-\hat{h}_i\right)\right\} \approx \sum_{i=1}^{j} \hat{h}_i = \hat{H}(t)$$

for $v_j \le t < v_{j+1}$.

Asymptotic properties of KM estimator

- As $n \to \infty, \hat{S}(t) \to S(t)$ in probability.
- As $n \to \infty, n^{1/2}\{\hat{S}(t) - S(t)\}$ converges to $N\left(0, \sigma^2(t)\right)$ in distribution.

Asymptotic properties of KM estimator

- How to estimate the variance of $\hat{S}(t)$
- $\hat{h}_i$ is an estimated probability.
- The variance of $\hat{h}_i$ can be approximated by

$$\frac{\hat{h}_i\left(1-\hat{h}_i\right)}{Y\left(v_i\right)}=\frac{d_i\left(Y\left(v_i\right)-d_i\right)}{Y\left(v_i\right)^3}$$

- $\hat{h}_i$ and $\hat{h}_j$ are asymptotically independent.

Asymptotic variance of KM estimator

$$
\begin{aligned}
\text{For } v_j \le t < v_{j+1} : \operatorname{Var}(\ln \hat{S}(t)) &\approx \sum_{i=1}^{j} \operatorname{Var}\left(\ln\left(1-\hat{h}_i\right)\right) \\
&\approx \sum_{i=1}^{j} \operatorname{Var}\left(\hat{h}_i\right) \cdot \frac{1}{\left(1-\hat{h}_i\right)^2} \\
&= \sum_{i=1}^{j} \frac{d_i}{Y\left(v_i\right)\left(Y\left(v_i\right)-d_i\right)}
\end{aligned}
$$

Asymptotic variance of KM estimator

$$
\begin{aligned}
\operatorname{Var}(\hat{S}(t)) &\approx \operatorname{Var}(\ln \hat{S}(t)) \left(e^{\ln \hat{S}(t)}\right)^2 \\
&= \hat{S}(t)^2 \operatorname{Var}(\ln \hat{S}(t)) \\
&\approx \hat{S}(t)^2 \sum_{i=1}^{j} \frac{d_i}{Y(v_i)\,(Y(v_i) - d_i)} \qquad \left(v_j \le t < v_{j+1}\right) \\
&= \hat{\sigma}^2(t)
\end{aligned}
$$

Greenwood's formula

By-Product

The by-product of the Greenwood's formula is the variance estimator for Nelson-Aalen Estimator:

$$\mathrm{var}(\hat{H}(t)) = \sum_{i=1}^{j} \frac{d_i}{Y(v_i)\,(Y(v_i) - d_i)}, \quad v_j \le t < v_{j+1}$$

Confidence Interval

- $\hat{S}(t) \pm 1.96\hat{\sigma}(t)$, drawbacks?
- By δ-method

$$\text{var}(\log(-\log(\hat{S}(t)))) = \frac{\hat{\sigma}^2(t)}{(\log(\hat{S}(t)))^2\hat{S}(t)^2}$$

- The confidence interval for $\hat{S}(t)$

$$\left[\exp\left\{-e^{\log(-\log(\hat{S}(t)))-\frac{1.96\hat{\sigma}(t)}{\log(\hat{S}(t))\hat{S}(t)}}\right\}, \exp\left\{-e^{\log(-\log(\hat{S}(t)))+\frac{1.96\hat{\sigma}(t)}{\log(\hat{S}(t))\hat{S}(t)}}\right\}\right]$$

Median survival time

- How to estimate the median survival time
- Solving $\hat{S}\left(\hat{t}_M\right) = 1/2$
- How to construct the CI for the median survival time? Suppose that
  1. $\mathrm{pr}\left(\hat{S}_L(t) < S(t)\right) = \mathrm{pr}\left(\hat{S}_U(t) > S(t)\right) = 0.975.$
  2. $\hat{S}_L\left(\hat{t}_{ML}\right) = 0.5$
  3. $\hat{S}_U\left(\hat{t}_{MU}\right) = 0.5$
  4. The confidence interval for t_M is $\left[\hat{t}_{ML}, \hat{t}_{MU}\right]$.

Median survival time

$$
\begin{aligned}
0.975 &= \text{pr}\left(\hat{S}_L\left(\hat{t}_M\right) < S\left(\hat{t}_M\right)\right) = \text{pr}\left(\hat{S}_L\left(\hat{t}_M\right) < 0.5\right) \\
&= \text{pr}\left(\hat{S}_L\left(\hat{t}_M\right) < \hat{S}_L\left(\hat{t}_{ML}\right)\right) = \text{pr}\left(\hat{t}_M \geq \hat{t}_{ML}\right) \\
0.975 &= \text{pr}\left(\hat{S}_U\left(\hat{t}_M\right) > S\left(\hat{t}_M\right)\right) = \text{pr}\left(\hat{S}_U\left(\hat{t}_M\right) > 0.5\right) \\
&= \text{pr}\left(\hat{S}_U\left(\hat{t}_M\right) > \hat{S}_U\left(\hat{t}_{MU}\right)\right) = \text{pr}\left(\hat{t}_M \leq \hat{t}_{MU}\right)
\end{aligned}
$$

Restricted mean survival time

- The area under the survival curve is a nice summary for the curve
- The AUC $\mu = \int_0^\tau S(t)dt$:

$$\mu = tS(t)|_0^\tau + \int_0^\tau tf(t)dt$$
$$= \tau S(\tau) + \int_0^\tau tf(t)dt = \int_0^\infty \min(t,\tau)f(t)dt = E\{\min(T,\tau)\}$$

- μ can be estimated as

$$\int_0^\tau \hat{S}(t)dt$$

Restricted mean survival time

- The restricted mean survival time $E\{\min(T,\tau)\}$ can also be estimated as

$$\hat{\mu}_{IPW} = n^{-1} \sum_{i=1}^{n} \frac{\delta_i + (1-\delta_i)\, I\,(U_i \geq \tau)}{\widehat{S}_C\,(T_i \wedge \tau)} T_i \wedge \tau$$

  where $\widehat{S}_C(\cdot)$ is a consistent estimator of the survival function of the censoring time C.

- Rational

$$E\left[\frac{I\,(C_i \geq \tau \wedge T_i)}{\widehat{S}_C\,(T_i \wedge \tau)} T_i \wedge \tau \mid T_i\right] \approx (T_i \wedge \tau)\,\frac{P\,(C_i \geq \tau \wedge T_i \mid T_i)}{S_C\,(T_i \wedge \tau)} = T_i \wedge \tau$$

- This type of estimator is called the inverse probability weighting estimator