Survival Analysis: Weeks 2-3

Lu Tian and Richard Olshen
Stanford University
Kaplan-Meier (KM) Estimator

- Nonparametric estimation of the survival function
  \[ S(t) = \text{pr}(T > t) \]

- 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

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

- $\hat{S}$ 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 $f_1 = \Pr(T = T_1), f_2 = \Pr(T = T_2), \cdots$.
- 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 $\delta_i = I(T_i \leq C_i)$.

- **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 $f_1 = \Pr(T = v_1), f_2 = \Pr(T = v_2), \cdots$.
- Objective: estimate $f_1, f_2, \cdots$. 
- 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 $(f_1, f_2, \cdots)$:
  \[
  L(F) = f_1^2(f_3 + f_4 + f_5 + f_6 + f_7 + f_8) f_3(f_4 + f_5 + f_6 + f_7 + f_8) f_4 f_5 f_6 f_8,
  \]
  where $f_1 + f_2 + f_3 + f_4 + f_5 + f_6 + f_7 + f_8 = 1$
Reparametrization tricks

• The discrete hazard function: \( h_1 = \text{pr}(T = v_1) \) and
  \[ h_j = \text{pr}(T = v_j | T > v_{j-1}), j > 2 \]

• For \( t \in [v_j, v_{j+1}) \)

  \[ S(t) = \text{pr}(T > t) = \text{pr}(T > v_j) = \prod_{i=1}^{j} (1 - h_i). \]

• For \( t = v_j \)

  \[ f_j = f(t) = \text{pr}(T = t) = h_j \prod_{i=1}^{j-1} (1 - h_i). \]
Reparametrization tricks

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

\[
L(F) = h_1^2 \times \{(1 - h_1)(1 - h_2)\} \times \{(1 - h_1)(1 - h_2)h_3\} \\
\times \{(1 - h_1)(1 - h_2)(1 - h_3)\} \times \{(1 - h_1)(1 - h_2)(1 - h_3)h_4\} \\
\times \{(1 - h_1)(1 - h_2)(1 - h_3)(1 - h_4)h_5\} \\
\times \{(1 - h_1)(1 - h_2)(1 - h_3)(1 - h_4)(1 - h_5)h_6\}^2 \\
\times \{(1 - h_1)(1 - h_2)(1 - h_3)(1 - h_4)(1 - h_5)(1 - h_6)(1 - h_7)\}
\]

\[
= h_1^2(1 - h_1)^8 \times (1 - h_2)^8 \times h_3(1 - h_3)^6 \\
h_4(1 - h_4)^4 \times h_5(1 - h_5)^3 \times h_6^2(1 - h_6) \times (1 - h_7)
\]
KM estimation

- The likelihood function

\[ L(F) = \prod_j h_j^{d_j} (1 - h_j)^{Y(v_j) - d_j} \]

where

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

\[ Y(v_j) = \sum_{i=1}^n I(U_i \geq v_j) = \# \text{“at risk” at } v_j. \]
\[
\hat{h}_j = d_j / Y(v_j)
\]

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

which is the Kaplan-Meier Estimator.
### Example

<table>
<thead>
<tr>
<th>$v_j$</th>
<th>$Y(v_j)$</th>
<th>$d_j$</th>
<th>$\hat{h}_j$</th>
<th>$\hat{S}(v_j) = \prod_{i=1}^{j} (1 - \hat{h}_i) = \hat{P}(T &gt; v_j)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>10</td>
<td>2</td>
<td>2/10</td>
<td>.8</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>1</td>
<td>1/7</td>
<td>.69 ($= .8 \times \frac{6}{7}$)</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>1</td>
<td>1/5</td>
<td>.55 ($= .69 \times \frac{4}{5}$)</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>1</td>
<td>1/4</td>
<td>.41 ($= .55 \times \frac{3}{4}$)</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>2</td>
<td>2/3</td>
<td>.14 ($= .41 \times \frac{1}{3}$)</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>.14</td>
</tr>
</tbody>
</table>
Suppose that \( v_g \) denotes the largest \( v_j \) for which \( Y(v_j) > 0 \).

1. If \( d_g = Y(v_j) \), then \( \hat{S}(t) = 0 \) for \( t \geq v_g \).
2. If \( d_g < Y(v_j) \), then \( \hat{S}(t) > 0 \) but not defined for \( t > v_g \). (Not identifiable beyond \( v_g \)).

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 estimation

- KM estimator is a special MLE

\[ 1 - \hat{S}(t) = \text{argmax}_F L(F) \]

where \( F \) is the CDF for all discrete random variables (nonparametric MLE).
Self-Consistency

- No censoring: \( \hat{S}(t) = n^{-1} \sum_{i=1}^{n} I(T_i > t) \)

- Right censoring: \( \hat{S}(t) = n^{-1} \sum_{i=1}^{n} E(I(T_i > t)|U_i, \delta_i) \)
  1. \( E(I(T_i > t)|U_i, \delta_i = 1) = I(U_i > t) \)
  2. \( E(I(T_i > t)|U_i, \delta_i = 0) = S(t)/S(U_i)I(t \geq U_i) + I(U_i > t) \)

- 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.
### Redistribution of Mass

<table>
<thead>
<tr>
<th>Step 1</th>
<th>2</th>
<th>2</th>
<th>3⁺</th>
<th>5</th>
<th>5⁺</th>
<th>7</th>
<th>9</th>
<th>16</th>
<th>16</th>
<th>18⁺</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td>1/10</td>
<td>1/10</td>
<td>1/10</td>
<td>1/10</td>
<td>1/10</td>
<td>1/10</td>
<td>1/10</td>
<td>1/10</td>
</tr>
<tr>
<td>Step 3</td>
<td></td>
<td></td>
<td>1/70</td>
<td>1/70</td>
<td>1/70</td>
<td>1/70</td>
<td>1/70</td>
<td>1/70</td>
<td>1/70</td>
<td>1/70</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1/5(8/70)</td>
<td>1/5(8/70)</td>
<td>1/5(8/70)</td>
<td>1/5(8/70)</td>
<td>1/5(8/70)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>2/10</td>
<td>0</td>
<td>8/70</td>
<td>0</td>
<td>24/175</td>
<td>24/175</td>
<td>48/175</td>
<td></td>
</tr>
<tr>
<td>Mass</td>
<td></td>
<td></td>
<td>Assume this is somewhere &gt; 18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Nelson-Aalen Estimator

- How to estimate the cumulative hazard function?

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

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

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

for \( v_j \leq t < v_{j+1} \).
Asymptotical 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 \( N(0, \sigma^2(t)) \) in distribution.
Asymptotical 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

  \[
  \frac{\hat{h}_i(1 - \hat{h}_i)}{Y(v_i)} = \frac{d_i(Y(v_i) - d_i)}{Y(v_i)^3}
  \]

- $\hat{h}_i$ and $\hat{h}_j$ are asymptotically independent (why?).
Asymptotical variance of KM estimator

For $v_j \leq t < v_{j+1}$: $\text{Var} \left( \ln \hat{S}(t) \right) \approx \sum_{i=1}^{j} \text{Var} \left( \ln(1 - \hat{h}_i) \right)$

$\delta$-method

$= \sum_{i=1}^{j} \text{Var} \left( \hat{h}_i \right) \cdot \frac{1}{(1 - \hat{h}_i)^2}$

$= \sum_{i=1}^{j} \frac{d_i}{Y(v_i)(Y(v_i) - d_i)}$.
Asymptotical variance of KM estimator

Using the $\delta$-method

$$\text{Var} \left( \hat{S}(t) \right) \approx \text{Var} \left( \ln \hat{S}(t) \right) \left( e^{\ln \hat{S}(t)} \right)^2$$

$$= \hat{S}(t)^2 \text{Var} \left( \ln \hat{S}(t) \right)$$

$$\approx \hat{S}(t)^2 \sum_{i=1}^{j} \frac{d_i}{Y(v_i)(Y(v_i) - d_i)} \quad (v_j \leq t < v_{j+1})$$

$$= \hat{\sigma}^2(t)$$

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

$$\text{var}(\hat{H}(t)) = \sum_{i=1}^{j} \frac{d_{i}}{Y(v_{i})(Y(v_{i}) - d_{i})}, \quad v_{j} \leq t < v_{j+1}$$
Confidence Interval

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

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

- How to estimate the median survival time

- Solving $\hat{S}(\hat{t}_M) = 1/2$, not always solvable!

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

\[ 0.975 = \Pr(\hat{S}_L(t_M) < S(t_M)) = \Pr(\hat{S}_L(t_M) < 0.5) \]
\[ = \Pr(\hat{S}_L(t_M) < \hat{S}_L(\hat{t}_{ML})) = \Pr(t_M \geq \hat{t}_{ML}) \]

\[ 0.975 = \Pr(\hat{S}_U(t_M) > S(t_M)) = \Pr(\hat{S}_U(t_M) > 0.5) \]
\[ = \Pr(\hat{S}_U(t_M) > \hat{S}_U(\hat{t}_{MU})) = \Pr(t_M \leq \hat{t}_{MU}) \]
Restricted mean survival time

- The area under the survival curve is a nice summary for the curve.
- The AUC $\mu = \int_0^\tau S(t)dt$:
  $$
  \mu = tS(t)\bigg|_0^\tau + \int_0^\tau tf(t)dt
  = \tau S(\tau) + \int_0^\tau tf(t)dt
  = \int_0^\infty \min(t, \tau)f(t)dt = E\{\min(t, \tau)\}
  $$
- $\mu$ 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

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

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

- Rational

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

- This type of estimator is called the inverse probability weighting estimator
Restricted mean survival time

- \( \hat{\mu} \) and \( \hat{\mu}_{IPW} \) are equivalent!

- The variance of the estimated area under the survival curve is complicated (the derivation will be given later).

\[
\frac{1}{n} \int_0^\tau \left\{ \int_t^\tau S(u)du \right\}^2 \frac{h(t)dt}{P(U \geq t)}.
\]
Area under the KM curve
Model Checking

- Under the exponential distribution: \( \log(S(t)) = -\lambda t \)
- Plot \( t \) vs. \( \log(\hat{S}(t)) \) to visually check the expected linear pattern.
- Under the Weibull distribution
  \[
  \log(-\log(S(t))) = p \log(\lambda) + p \log(t)
  \]
- Plot \( \log(t) \) vs. \( \log(-\log(\hat{S}(t))) \) to visually check the expected linear pattern.
Model Checking

\[ \log(-\log(S(t))) \]
The distribution of $\hat{S}(t)$ as a process.