Self-Consistency, Product-Limit and IPCW Estimator

Yi-Cheng Tai

National Yang Ming Chiao Tung University

February 22, 2023

Introduction

- Strawderman (2023)[2] proposed a unified method to show the equivalence among the solution to self-consistency equation, product-limit estimator and inverse-probability-censoring-weight (IPCW) estimator for the survival function with right-censored data.
- The method is based on the Volterra integral equation and the relationship between the survival functions for failure time and censoring time.

Notations

- T: failure time of interest with survival function S and corresponding cumulative hazard, A
- C: censoring time with survival function G
- $(Y = T \wedge C, \delta = I(T \leq C))$: observed data
- $N(t) = \sum_{i=1}^{n} I(Y_i \leq t, \delta_i = 1)$
- $C(t) = \sum_{i=1}^{n} I(Y_i \leq t, \delta_i = 0)$
- $R(t) = \sum_{i=1}^{n} I(Y_i \geq t)$

Volterra Integral Equation

We provides a motivation to memorize Volterra integral equation and its solution.

We start with the definition of hazard function:

$$\begin{aligned} dA(t) &= -\frac{dS(t)}{S(t-)} \\ \Rightarrow dS(t) &= -S(t-)dA(t) \\ \Rightarrow S(t) &= -\int_0^t S(u-)dA(u) + Const \end{aligned}$$

Since $S(0) = 1$, we have $Const = 1$. Then the Volterra integral equation is obtained:

$$S(t) = 1 - \int_0^t S(u-)dA(u) \tag{1}$$

Solution to Volterra Integral Equation

The survival function can be written as the product-limit form:

$$S(t) = \prod_{0 \leq u \leq t} [1 - dA(u)]$$

Hence, the product-limit is a solution to (1). In fact, it is the unique solution.

Product-Limit Estimator

The product-limit (Kaplan-Meier) estimator for survival function is known as

$$\hat{S}(t) = \prod_{0 \le u \le t} \left[1 - \frac{dN(u)}{R(u)} \right]$$

Assume that there are no ties between observed failure and censoring events. i.e $[dN(u) > 0 \Rightarrow dC(u) = 0]$. Then the product-limit estimator can be re-written as

$$\begin{aligned}
\hat{S}(t) &= \prod_{0 \le u \le t} \left[1 - \frac{dN(u)}{R(u)} \right] \\
&= \prod_{0 \le u \le t} \left[1 - \frac{R(u) - R(u+) - dC(u)}{R(u) - dC(u)} \right] \\
&= \prod_{0 \le u \le t} \left[\frac{R(u+)}{R(u) - dC(u)} \right]
\end{aligned}$$

Product-Limit Estimator (Cont.)

$$
\begin{aligned}
\left[\hat{S}(t)\right]^{-1} &= \prod_{0 \le u \le t} \left[\frac{R(u) - dC(u)}{R(u+)}\right] \\
&= \prod_{0 \le u \le t} \left[\frac{R(u) - dC(u)}{R(u)}\right] \prod_{0 \le u \le t} \frac{R(u)}{R(u+)} \\
&= \prod_{0 \le u \le t} \left[1 - \frac{dC(u)}{R(u)}\right] \frac{R(0)}{R(t+)} \\
&= \prod_{0 \le u \le t} \left[1 - \frac{dC(u)}{R(u)}\right] \frac{n}{R(t+)}
\end{aligned}
$$

Note that $\hat{G}(t) = \prod_{0 \le u \le t}[1 - \frac{dC(u)}{R(u)}]$ and $\hat{S}_0(t) = n^{-1}R(t+)$ is the empirical estimator of $Pr(Y > t)$.

Self-Consistency Equation

Since $\hat{G}$ is of the product-limit form, it satisfies the corresponding Volterra integral equation.

$$\hat{G}(t) = 1 - \int_0^t \hat{G}(u-)\frac{dC(u)}{R(u)}$$

Based on the relationship derived above, $\hat{S}_0(t) = \hat{S}(t)\hat{G}(t)$, we may replace $\hat{G}$. Then,

$$\frac{\hat{S}_0(t)}{\hat{S}(t)} = 1 - \int_0^t \frac{\hat{S}_0(u-)}{\hat{S}(u-)}\frac{dC(u)}{R(u)}$$

$$\Rightarrow \hat{S}(t) = \hat{S}_0(t) + n^{-1}\int_0^t \frac{\hat{S}(t)}{\hat{S}(u-)}dC(u)$$

Hence, the product-limit estimator satisfies the self-consistency equation proposed by Efron (1967)[1].

IPCW Estimator

Again, $\hat{S}$ satisfies the Volterra integral equation and we may replace $\hat{S}(u-)$ in RHS.

$$\begin{aligned}
\hat{S}(t) &= 1 - \int_0^t \hat{S}(u-)\frac{dN(u)}{R(u)} \\
&= 1 - n^{-1}\int_0^t \frac{dN(u)}{\hat{G}(u-)} \\
&= 1 - n^{-1}\sum_{i=1}^n \frac{I(Y_i \le t, \delta_i = 1)}{\hat{G}(Y_i-)} \\
&= 1 - n^{-1}\sum_{i=1}^n \frac{I(\delta_i = 1)}{\hat{G}(Y_i-)} + n^{-1}\sum_{i=1}^n \frac{I(Y_i > t, \delta_i = 1)}{\hat{G}(Y_i-)}
\end{aligned}$$

IPCW Estimator (Cont.)

Let $Y_{(n)} = max(Y_i, 1 \leq i \leq n)$. Then,

$$\begin{aligned}
\hat{S}(Y_{(n)}) &= 1 - n^{-1} \sum_{i=1}^{n} \frac{I(\delta_i = 1)}{\hat{G}(Y_i-)} + n^{-1} \sum_{i=1}^{n} \frac{I(Y_i > Y_{(n)}, \delta_i = 1)}{\hat{G}(Y_i-)} \\
&= 1 - n^{-1} \sum_{i=1}^{n} \frac{I(\delta_i = 1)}{\hat{G}(Y_i-)}
\end{aligned}$$

It can be treated as estimated cured rate if there are cured fraction in the population.

If $\hat{S}(t) \to 0$ as $t \to \infty$, then the product-limit estimator can be written as IPCW estimator.

$$\hat{S}(t) = n^{-1} \sum_{i=1}^{n} \frac{I(Y_i > t, \delta_i = 1)}{\hat{G}(Y_i-)}$$

References

B. Efron.
The two sample problem with censored data.
In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 4, pages 831–853, 1967.

R. L. Strawderman.
On the solutions to efron's self-consistency equation, 2023.