Estimate Kaplan Meier curve When Patients Following Certain Treatment Decisions

ISO's draft, please comments and contribute ***

November 15, 2018

Abstract

Key words and Phrases: *** ***. **Short title**: ***

1 Notations

- There are N subjects from a large population.
- A_i is the treatment assignment (actions), where $i = 1, , N$.
- T_i is the time until an event of interest takes place.
- C_i is the censoring time for subject i.
- $Y_i = T_i \wedge C_i$ is the observed time.
- $\Delta_i = I(T_i \leq C_i)$ is an indicator for the occurrence of the event at censoring time.
- X_i is a vector of covariates.
- (T, A, X) is the generic random variable of $\{(T_i, A_i, X_i)\}$.
- $\mathcal{P}$ is the distribution of (T, A, X).
- E is the expectation with respect to P.
- Population space $\mathcal{X}$, i.e. $X_i \in \mathcal{X}$.
- $\mathcal{D}(\cdot)$ is a treatment recommendation based on covariates, i.e. $\mathcal{D}(\cdot) : \mathcal{X} \rightarrow \mathcal{A}$.
- $\mathcal{P}^{\mathcal{D}}$ is the distribution of (T, A, X) given that $A = \mathcal{D}(X)$.
- $E^{\mathcal{D}}$ is the expectation with respect to $\mathcal{P}^{\mathcal{D}}$.

2 Survival function under treatment rule of $\mathcal{D}$

Let $S^{\mathcal{D}}(t) = P(T > t | A = \mathcal{D}(X))$ be the survival function under treatment rule of $\mathcal{D}$. At time t,

$$S^{\mathcal{D}}(t) = E^{\mathcal{D}}\left[I(T > t)\right] = \int I(T > t) d\mathcal{P}^{\mathcal{D}} = \int I(T > t) \frac{d\mathcal{P}^{\mathcal{D}}}{d\mathcal{P}} d\mathcal{P}. \tag{2.1}$$

Since

$$\frac{d\mathcal{P}^{\mathcal{D}}}{d\mathcal{P}} = \frac{p(T|X, A) I[A = \mathcal{D}(X)] p(X)}{p(Y|X, A) p(A|X) p(X)} = \frac{I[A = \mathcal{D}(X)]}{p(A|X)}, \tag{2.2}$$

we have

$$S^{\mathcal{D}}(t) = \int I(T > t) \frac{I[A = \mathcal{D}(X)]}{p(A|X)} d\mathcal{P} = E\left[\frac{I[A = \mathcal{D}(X)]}{p(A|X)} I(T > t)\right]. \tag{2.3}$$

3 Generate Kaplan Meier curve

The following steps describe an algorithm to generate the Kaplan Meier curve if the patients following treatment rule $\mathcal{D}(\cdot)$ with two treatment situation.

(1) We have the observed data as (Y, X, A, C), where A is 1 if patients take one treatment and is 0 if patients following another treatment. We can use the observed data and a logistic regression to estimate propensity scores $e = P(A = 1|X)$.

(2) Given time t, define $J(t) = \{1 \leq i \leq n : C_i \geq t\}$, i.e. the events for which the outcome was not censored before time t. $Q(t) = \{1 \leq i \leq n : C_i \geq t, Y_i \geq t\}$, i.e. the events for which the outcome was not censored and not happened before time t.

(3) We calculate our KM estimator,

$$\widehat{S}^{\mathcal{D}}(t-1) = \frac{1}{|J(t)|} \sum_{i=1}^{N} \frac{I[i \in Q(t)] I[A_i = \mathcal{D}(X_i)]}{A_i e_i + (1 - A_i)(1 - e_i)}$$