An Overview of Estimating Techniques in Length-Biased, Right-Censored Survival Data

Jiajun Zhang
McGill University

Abstract

Survival data often presents characteristic features known as censoring and truncation, where the observer only has partial information about the event times. For right right-censored data, the exact event time is known to exceed a censoring time. On the other hand, length-biased data can be viewed as a left-truncation where only partial data satisfying certain conditions are selected into the study with selection probability proportional to their actual event time. In this article we review different estimating techniques for length-biased and right-censored samples. Several examples with illustrations are introduced at the end of this article.

1 Estimating Equation Under Cox PH Model

1.1 Introduction

In Cox PH model, we assume the covariates vector $\boldsymbol{X}$ acts multiplicity on the baseline hazard $\lambda_0(t)$, where we have

$$\lambda(t|\boldsymbol{X}) = \lambda_0(t)\exp\left(\boldsymbol{\beta}^t\boldsymbol{X}\right) \tag{1}$$

where $\boldsymbol{\beta}^t$ is the coefficients vector we would like to estimate. Without length-biased assumption, the estimating techniques are well studied by Cox (1972), also Klein and Moeschberger (2003) have a detailed discussion on estimating techniques based on different covariates and the presence of censoring. We now consider the data which is length-biased and right censored. Suppose $\tilde{T}$ is the failure time from onset of to the end of the study, A denotes the time from onset to the start of our study (also known as the left truncation time), V denotes the time from the start of our study to failure, and C denotes the censoring time from the start of our study, and assume that C is independent of A and V. Then we are able to observe $\tilde{T}$ if and only if $\tilde{T} > A$. Define $T = \min(A+V, A+C)$ as the observed length-biased lifetime of those satisfying $\tilde{T} > A$, by our length-biased assumption, the probability of selecting a certain individual is proportional to its complete event time. If we let $f(t)$ denotes the true density of $\tilde{T}$, $g(t)$ denotes the density of T, which is the life time conditioning on the fact that $\tilde{T} > A$, we have the following relation:

$$g(t) = \frac{tf(t)}{\displaystyle\int_0^\infty uf(u)du} = \frac{tf(t)}{\mu}$$

where t on the numerator acts as a weight function, indicating the fact that individuals with longer lifetimes are more likely to be selected. We also divide the proportional term by the expected value of X, where $\mu = \mathbb{E}[X]$. Also, the density of A should also follow a pattern proportional to the survival time of an individual. Let $S_U(t)$ be the survival function of $\tilde{T}$, then the density function of A, denote by $f_A(t)$, is given by

$$f_A(t) = \frac{S_U(t)}{\mu} = \frac{S_U(t)}{\displaystyle\int_0^\infty u f(u) du}.$$

In Cox's PH model, let $\boldsymbol{X}$ be the covariates, using a Bayesian argument, the density $g(t|\boldsymbol{X})$ and $f_A(t|\boldsymbol{X})$ given by

$$g(t|\boldsymbol{X}) = \frac{t f(t|\boldsymbol{X})}{\displaystyle\int_0^\infty u f(u|\boldsymbol{X}) du} = \frac{t f(t|\boldsymbol{X})}{\mu(\boldsymbol{X})}, \quad f_A(t|\boldsymbol{X}) = \frac{S_U(t|\boldsymbol{X})}{\displaystyle\int_0^\infty u f(u|\boldsymbol{X}) du} = \frac{S_U(t|\boldsymbol{X})}{\mu(\boldsymbol{X})}.$$

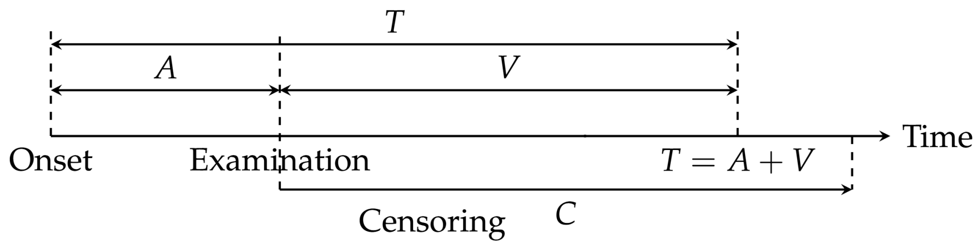

Figure 1: This figure shows the overall relation of those variables we defined, where T is the length-biased time from onset to event (conditioning on $\tilde{T} > A$); V is the observational time to event, or the residual survival time; A is the left truncation time (backward recurrence time); C is the censoring time. $T = \min(A + V, A + C)$.

1.2 Estimating Equation Without Right Censoring

We first derive the estimator without right-censoring. The first approach is given by Qin and Shen (2010). Suppose $h(\boldsymbol{x})$ is the marginal density function of the covariates $\boldsymbol{X}$, then conditioning on the biased density $T = t$ yields

$$h(\boldsymbol{x}|t) = \frac{g(t|\boldsymbol{x})h(\boldsymbol{x})}{\displaystyle\int g(t|\boldsymbol{x})h(\boldsymbol{x})d\boldsymbol{x}} = \frac{t f(t|\boldsymbol{x})h(\boldsymbol{x})/\mu(\boldsymbol{x})}{\displaystyle\int t f(t|\boldsymbol{x})h(\boldsymbol{x})/\mu(\boldsymbol{x})d\boldsymbol{x}}.$$

We use the property that $f(t) = \lambda(t)S(t)$ (see Zhang, Chapter 2.3, (2025)) where $\lambda(t)$ is the hazard function, and applying Cox's PH model described in (1), we have

$$f(t|\boldsymbol{x}) = \lambda_0(t) \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}\right) \cdot S_U(t|\boldsymbol{x})$$

where $\boldsymbol{X} = \boldsymbol{x}$ is the covariates vector. So the conditional density of $\boldsymbol{X}$ given $T = t$ can be written as

$$h(\boldsymbol{x}|t) = \frac{t\lambda_0(t) \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}\right) \cdot S_U(t|\boldsymbol{x})h(\boldsymbol{x})/\mu(\boldsymbol{x})}{\displaystyle\int t\lambda_0(t) \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}\right) \cdot S_U(t|\boldsymbol{x})h(\boldsymbol{x})/\mu(\boldsymbol{x})d\boldsymbol{x}}.$$

We now take the expectation of $h(\boldsymbol{x}|T=t)$, then after simplifications we will get

$$\mathbb{E}[\boldsymbol{X}|T=t] = \frac{\displaystyle\int \boldsymbol{x} \cdot \exp\Big(\boldsymbol{\beta}^t \boldsymbol{X}\Big) S_U(t|\boldsymbol{x}) h(\boldsymbol{x})/\mu(\boldsymbol{x}) d\boldsymbol{x}}{\displaystyle\int \exp\Big(\boldsymbol{\beta}^t \boldsymbol{X}\Big) S_U(t|\boldsymbol{x}) h(\boldsymbol{x})/\mu(\boldsymbol{x}) d\boldsymbol{x}} = \frac{\mathbb{E}\Big[\boldsymbol{X} \cdot \exp\Big(\boldsymbol{\beta}^t \boldsymbol{X}\Big) S_U(t|\boldsymbol{x})/\mu(\boldsymbol{x})\Big]}{\mathbb{E}\Big[\exp\Big(\boldsymbol{\beta}^t \boldsymbol{X}\Big) S_U(t|\boldsymbol{x})/\mu(\boldsymbol{x})\Big]}. \quad (2)$$

The term $S_U(t|\boldsymbol{x})/\mu(\boldsymbol{x})$ satisfies the following:

$$\mathbb{E}\left[\frac{1}{T} \cdot \mathbf{1}\{T \geq t\} \middle| \boldsymbol{X} = \boldsymbol{x}\right] = \int_t^\infty \frac{1}{t} \cdot g(t|\boldsymbol{x}) dt = \frac{1 - F(t|\boldsymbol{x})}{\mu(\boldsymbol{x})} = \frac{S_U(t|\boldsymbol{x})}{\mu(\boldsymbol{x})}$$

let $\delta(t)$ be the indicator of $T \geq t$, so the expectation of $\boldsymbol{X}$ can be written as

$$\mathbb{E}[\boldsymbol{X}|T=t] = \frac{\mathbb{E}\Big[\boldsymbol{X} \cdot \exp\Big(\boldsymbol{\beta}^t \boldsymbol{X}\Big) \cdot \mathbb{E}[T^{-1}\delta(t)|\boldsymbol{X}]\Big]}{\mathbb{E}\Big[\exp\Big(\boldsymbol{\beta}^t \boldsymbol{X}\Big) \cdot \mathbb{E}[T^{-1}\delta(t)|\boldsymbol{X}]\Big]} = \frac{\mathbb{E}\Big[\boldsymbol{X} \cdot \exp\Big(\boldsymbol{\beta}^t \boldsymbol{X}\Big) \cdot T^{-1}\delta(t)\Big]}{\mathbb{E}\Big[\exp\Big(\boldsymbol{\beta}^t \boldsymbol{X}\Big) \cdot T^{-1}\delta(t)\Big]}$$

using the law of total expectation. Now suppose T_i is the biased event time for individual X_i with covariates $\boldsymbol{X}_i$, we use the empirical estimate of expectation

$$\widehat{\mathbb{E}}\Big[\boldsymbol{X} \cdot \exp\Big(\boldsymbol{\beta}^t \boldsymbol{X}\Big) \cdot T^{-1}\delta(t)\Big] = \frac{1}{n} \sum_{i=1}^n \boldsymbol{X}_j \cdot \exp\Big(\boldsymbol{\beta}^t \boldsymbol{X}_j\Big) \cdot T_j^{-1} \delta_j(t)$$

and

$$\widehat{\mathbb{E}}\Big[\exp\Big(\boldsymbol{\beta}^t \boldsymbol{X}\Big) \cdot T^{-1}\delta(t)\Big] = \frac{1}{n} \sum_{i=1}^n \exp\Big(\boldsymbol{\beta}^t \boldsymbol{X}_j\Big) \cdot T_j^{-1} \delta_j(t).$$

We use the fact that $\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}|T=t]$ is unbiased, and we sum over the entire sample with $T_1, \cdots, T_n$ to get the following estimating equation for the length-biased sample:

$$\sum_{i=1}^n \left(\boldsymbol{X}_i - \frac{\displaystyle\sum_{j=1}^n \boldsymbol{X}_j \cdot \exp\Big(\boldsymbol{\beta}^t \boldsymbol{X}_j\Big) \cdot T_j^{-1} \delta_j(T_i)}{\displaystyle\sum_{j=1}^n \exp\Big(\boldsymbol{\beta}^t \boldsymbol{X}_j\Big) \cdot T_j^{-1} \delta_j(T_i)} \right) = 0. \quad (3)$$

The second approach to obtain (3) is done by Wang (1996) using a modified risk set from Cox's partial likelihood function (D.R Cox, 1972). We will derive the estimating equation and show those two are conceptually the same. Without length-biased data, Cox's partial likelihood function is given by

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \frac{\exp\Big(\boldsymbol{\beta}^t \boldsymbol{X}_i\Big)}{\displaystyle\sum_{j \in R(t_i)} \exp\Big(\boldsymbol{\beta}^t \boldsymbol{X}_j\Big)}$$

where $\boldsymbol{X}_i$ is the covariate vector, and $R(t_i)$ is the risk set at any failure time t_i defined by $R(t_i) = \{j : t_j \geq t_i\}$. An adjusted risk set $R^*(t_i)$ is defined by

$$R^*(t_i) = \{j : t_j \geq t_i, \Delta_j(t_i) = 1\}$$

with the indicator $\Delta_j(t_i) = 1$ with probability $\frac{t_i}{t_j}$ and $\Delta_j(t_i) = 0$ with probability $1 - \frac{t_i}{t_j}$, representing the probability of an individual being selected is proportional to its lifetime. Wang (1996) also showed that individuals in adjusted risk set also presents the population risk structure, and we have the modified partial likelihood:

$$L^*(\boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{\exp\left(\boldsymbol{\beta}^t \boldsymbol{X}_i\right)}{\sum_{j \in R^*(t_i)} \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}_j\right)} \tag{4}$$

The score U of the partial likelihood (4) is given by

$$\begin{aligned}
U &= \frac{\partial}{\partial \boldsymbol{\beta}} \log L^*(\boldsymbol{\beta}) \\
&= \frac{\partial}{\partial \beta} \left[\sum_{i=1}^{n} \boldsymbol{\beta}^t \boldsymbol{X}_i - \sum_{i=1}^{n} \log \left(\sum_{j \in R^*(t_i)} \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}_j\right) \right) \right] \\
&= \sum_{i=1}^{n} \left(\boldsymbol{X}_i - \frac{\sum_{j \in R^*(t_i)} \boldsymbol{X}_j \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}_j\right)}{\sum_{j \in R^*(t_i)} \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}_j\right)} \right) \\
&= \sum_{i=1}^{n} \left(\boldsymbol{X}_i - \frac{\sum_{j \in R(t_i)} \Delta_j(t_i) \boldsymbol{X}_j \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}_j\right)}{\sum_{j \in R(t_i)} \Delta_j(t_i) \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}_j\right)} \right).
\end{aligned}$$

The risk set interpretation $j \in R(t_i)$ can be written as $\mathbf{1}\{t_j \geq t_i\} = \delta_j(t_i)$ as in (3), followed by a similar argument, we construct the unbiased estimator $\boldsymbol{X} - \mathbb{E}[U]$, the expected value of the second term above can be estimated by their empirical form:

$$\widehat{\mathbb{E}} \left[\sum_{j \in R(t_i)} \Delta_j(t_i) \boldsymbol{X}_j \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}_j\right) \right] = \frac{1}{n} \sum_{j=1}^{n} \frac{t_i}{t_j} \boldsymbol{X}_j \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}_j\right) \delta_j(t_i)$$

and

$$\widehat{\mathbb{E}} \left[\sum_{j \in R(t_i)} \Delta_j(t_i) \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}_j\right) \right] = \frac{1}{n} \sum_{j=1}^{n} \frac{t_i}{t_j} \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}_j\right) \delta_j(t_i).$$

Since the terms t_i, $\frac{1}{n}$ are canceled when we take the ratio of those two empirical estimates, and by summing over all t_i, our estimating equation would result in the same as (3).

1.3 Modifications for Right Censored Data

With potential right censoring, the estimating equation (3) needs to be adjusted. We will first consider the uncensored, joint density of A and V. Under stationarity, the conditional distribution of A given $T = t$ is $Uniform(0, t)$ (Asgharian and Wolfson, 2005), meaning the onset of the event among those selected sample happens equally between 0 and t, hence a conditioning argument yields

$$f_{A,T}(a,t) = f_{A|T=t}(a|t) \cdot f_T(t) = \frac{1}{t} \cdot \frac{tf(t)}{\mu} = \frac{f(t)}{\mu}$$

where $\mu = \displaystyle\int_0^\infty uf(u)du$. Since $T = V + A$, we also have the joint density of A and V:

$$f_{A,V}(a,v) = \frac{f(a+v)}{\mu}.$$

When we have potential right censoring, it is reasonable to assume C is independent from A, V, and we denote f_C as the density of C, $T = \min\{A+V, A+C\}$, when $T = A+V$, we have

$$\begin{aligned}\mathbb{P}(T = A+V|\boldsymbol{X} = \boldsymbol{x}) &= \mathbb{P}(A = a, V = t-a, C \geq t-a|\boldsymbol{X} = \boldsymbol{x}) \\ &= \mathbb{P}(A = a, V = t-a|\boldsymbol{X} = \boldsymbol{x}) \cdot \mathbb{P}(C \geq t-a) \\ &= \frac{f(t|\boldsymbol{x})}{\mu} \cdot S_C(t-a). \end{aligned} \tag{5}$$

When $T = A + C$, we have

$$\begin{aligned}\mathbb{P}(T = A+C|\boldsymbol{X} = \boldsymbol{x}) &= \mathbb{P}(A = a, C = t-a, V \geq t-a|\boldsymbol{X} = \boldsymbol{x}) \\ &= \mathbb{P}(C = t-a) \cdot \mathbb{P}(A = a, V \geq t-a|\boldsymbol{X} = \boldsymbol{x}) \\ &= \mathbb{P}(C = t-a) \cdot \frac{\int_t^\infty f(u)du}{\mu} \\ &= \mathbb{P}(C = t-a) \cdot \frac{S_U(t)}{\mu}. \end{aligned} \tag{6}$$

Denote δ as the censoring indicator, $\delta = \mathbf{1}\{V \geq C\}$. Consider the conditional expectation $\boldsymbol{X}$ (with marginal density $h(\boldsymbol{X})$) given $T = t, A = a, \delta$, it is the same as the one we obtained in (2)

$$\mathbb{E}[\boldsymbol{X}|T = t, A = a, \delta] = \frac{\mathbb{E}\left[\boldsymbol{X} \cdot \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}\right) S_U(t|\boldsymbol{X})/\mu(\boldsymbol{X})\right]}{\mathbb{E}\left[\exp\left(\boldsymbol{\beta}^t \boldsymbol{X}\right) S_U(t|\boldsymbol{X})/\mu(\boldsymbol{X})\right]} \tag{7}$$

Different unbiased estimator of the term $S_U(t|\boldsymbol{X})/\mu(\boldsymbol{X})$ can be obtained, one is that (Qin and Shen, 2010)

$$\begin{aligned}\mathbb{E}\left[\frac{\delta \cdot \mathbf{1}\{T \geq t\}}{TS_C(T-A)}\middle|\boldsymbol{X} = \boldsymbol{x}\right] &= \int_t^\infty \int_0^t \cdot \frac{1}{t} \cdot \frac{1}{S_C(t-a)} \cdot f(t|\boldsymbol{x}) S_C(t-a) \cdot \frac{1}{\mu(\boldsymbol{x})} dadt \\ &= \int_t^\infty \frac{f(t|\boldsymbol{x})}{\mu(\boldsymbol{x})} dt \\ &= \frac{S_U(t|\boldsymbol{x})}{\mu(\boldsymbol{x})}. \end{aligned} \tag{8}$$

Combine (7), (8) and the empirical estimate technique we used earlier, the adjusted estimating equation with right-censored data with T_i is now

$$U_1(\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \cdot \left[\boldsymbol{X}_i - \frac{\sum_{j=1}^{n} \mathbf{1}\{T_j \geq T_i\}\delta_j \boldsymbol{X}_j \cdot \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}_j\right) \cdot \left(T_j S_C(T_j - A_j)\right)^{-1}}{\sum_{j=1}^{n} \mathbf{1}\{T_j \geq T_i\}\delta_j \boldsymbol{X}_j \cdot \exp\left(\boldsymbol{\beta}^t\right) \cdot \left(T_j S_C(T_j - A_j)\right)^{-1}} \right] = 0. \tag{9}$$

In the above estimating equation, $T_j - A_j = V_j$, which is the time from examination to event. We may replace S_C by its consistent Kaplan-Meier estimator for the residual censoring time.

The second choice (Qin and Shen, 2010) is based on a modification of (5), where we showed that

$$\mathbb{P}(T = A + V | \boldsymbol{X} = \boldsymbol{x}) = \frac{f(t|\boldsymbol{x}) S_C(t - a)}{\mu}$$

by integrating a from 0 to t, we have

$$\mathbb{P}(T = t | \boldsymbol{X} = \boldsymbol{x}) = \frac{f(t|\boldsymbol{x})}{\mu} \int_0^t S_C(u) du \equiv \frac{f(t|\boldsymbol{x})}{\mu} w_C(t),$$

and hence we have the expression

$$\begin{aligned}
\mathbb{E}\left[\frac{\mathbf{1}\{T > t, \delta = 1\}}{w_C(t)} \middle| \boldsymbol{X} = \boldsymbol{x} \right] &= \int_t^\infty \frac{f(u|\boldsymbol{x}) du}{\mu} \cdot \int_0^t \frac{S_C(v)}{w_C(v)} dv \\
&= \frac{S_U(t|\boldsymbol{x})}{\mu(\boldsymbol{x})}.
\end{aligned}$$

Another estimating equation follows:

$$U_2(\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \cdot \left[\boldsymbol{X}_i - \frac{\sum_{j=1}^{n} \mathbf{1}\{T_j \geq T_i\}\delta_j \boldsymbol{X}_j \cdot \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}_j\right) \cdot \left(w_C(T_j)\right)^{-1}}{\sum_{j=1}^{n} \mathbf{1}\{T_j \geq T_i\}\delta_j \boldsymbol{X}_j \cdot \exp\left(\boldsymbol{\beta}^t\right) \cdot \left(w_C(T_j)\right)^{-1}} \right] = 0. \tag{10}$$

Estimating equations (9),(10) requires the knowledge of the density of censoring variable C, a final approach using delayed-entry/left-truncated data (Qin and Shen, 2010) does not require the knowledge of the density of C. By assuming C is independent of $\boldsymbol{X}$, We have

$$\begin{aligned}
\mathbb{E}[\mathbf{1}\{T \geq t, A \leq t, \delta = 1\} | \boldsymbol{X} = \boldsymbol{x}] &= \mathbb{E}[\mathbf{1}\{T \geq t, A + C \geq t, A \leq t\} | \boldsymbol{X} = \boldsymbol{x}] \\
&= \int_t^\infty \int_0^t \frac{f(t|\boldsymbol{x}) S_C(t - a)}{\mu(\boldsymbol{x})} da dt \\
&= \frac{S_U(t|\boldsymbol{x})}{\mu(\boldsymbol{x})} \cdot w_C(t).
\end{aligned}$$

Using the same method in (7) and (2), we have

$$\mathbb{E}[\boldsymbol{X} | T \geq t, A \leq t, \delta] = \frac{\mathbb{E}\left[\boldsymbol{X} \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}\right) S_U(t|\boldsymbol{x}) / \mu(\boldsymbol{x})\right]}{\mathbb{E}\left[\exp\left(\boldsymbol{\beta}^t \boldsymbol{X}\right) S_U(t|\boldsymbol{x}) / \mu(\boldsymbol{x})\right]} \tag{11}$$

where $w_C(t)$ is canceled out from the numerator and denominator, and we have another estimating equation

$$U_L(\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \left[\boldsymbol{X}_i - \frac{\sum_{j=1}^{n} \mathbf{1}\{T_j \geq T_i, A_j \leq T_i\} \boldsymbol{X}_j \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}_j\right)}{\sum_{j=1}^{n} \mathbf{1}\{T_j \geq T_i, A_j \leq T_i\} \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}_j\right)} \right] = 0 \tag{12}$$

It is important to know that (12) can also derived from a modified partial likelihood function, Under prevalent sampling and quasi-stationarity conditions discussed by Wang et al. (1993), the likelihood function of the event time $T = \min\{A + V, A + C\}$ of the sample is proportional to

$$L(\boldsymbol{\beta}) \propto \prod_{i=1}^{n} \frac{f(t_i|\boldsymbol{X})^{\delta_i} \cdot S_U(t_i|\boldsymbol{X})^{1-\delta_i}}{S_U(a_i|\boldsymbol{X})}$$

where $\delta_i = \mathbf{1}\{V \geq C\}$ is the censoring indicator, a_i is the left truncation time. Then adapting Cox PH model, a partial likelihood can be written as

$$L_p(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left[\frac{\exp\left(\boldsymbol{\beta}^t \boldsymbol{X}_i\right)}{\sum_{j \in R(y_i)} \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}\right)} \right]^{\delta_i} \tag{13}$$

with modified risk set $R(y) \equiv \{j : a_j \leq y \leq y_j\}$, and Wang et al. (1993) showed that the likelihood $L(\boldsymbol{\beta})$ is proportional to the partial likelihood times a residual likelihood. Then the estimating equation (12) is just the score of the partial likelihood (13) by simple calculation:

$$\begin{aligned}
U &= \frac{\partial}{\partial \boldsymbol{\beta}} \log L_p(\boldsymbol{\beta}) \\
&= \frac{\partial}{\partial \boldsymbol{\beta}} \sum_{i=1}^{n} \delta_i \left[\boldsymbol{\beta}^t \boldsymbol{X}_i - \log \left(\sum_{j \in R(t_i)} \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}_j\right) \right) \right] \\
&= \sum_{i=1}^{n} \delta_i \left[\boldsymbol{X}_i - \frac{\sum_{j \in R(t_i)} \boldsymbol{X}_j \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}_j\right)}{\sum_{j \in R(t_i)} \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}_j\right)} \right] \\
&= \sum_{i=1}^{n} \delta_i \left[\boldsymbol{X}_i - \frac{\sum_{j=1}^{n} \boldsymbol{X}_i \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}_j\right) \mathbf{1}\{A_j \leq T_j \leq T_i\}}{\sum_{j=1}^{n} \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}_j\right) \mathbf{1}\{A_j \leq T_j \leq T_i\}} \right].
\end{aligned}$$

since U has mean zero, we will then have the same estimating equation.

1.4 Asymptotic Properties of Estimating Equation

In this section, we will briefly investigate the asymptotic properties of the estimating equations (9),(10) we derived in a length-biased, right-censored sample. We shall denote them by $U_1(\boldsymbol{\beta}), U_2(\boldsymbol{\beta})$ separately. For individual i, define the risk set $R_i(t) = \mathbf{1}\{T_i \geq t\}\delta_i$ and $N_i(t) = \mathbf{1}\{T_i \leq t, C_i \geq T_i - A_i\}$ and

$$S_k^{(l)}(\boldsymbol{\beta}, t) = \frac{1}{n}\sum_{i=1}^{n} w_C(t) R_i(t) W_{ki} \boldsymbol{X}_i^{\otimes l} \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}_i\right)$$

where in (9) $k = 1$ and in (10) $k = 2$, $l = 0, 1, 2$, W_{ki} represents the different weight function we derived, and so $W_{1i} = (T_i S_C(T_i - A_i))^{-1}$ and $W_{2i} = (w_C(T_i))^{-1}$. We furthermore define $E_k(\boldsymbol{\beta}, t) = S_k^{(1)}(\boldsymbol{\beta}, t)/S_k^{(0)}(\boldsymbol{\beta}, t)$, $e_k(\boldsymbol{\beta}, t)$ as the expectation of $E_k(\boldsymbol{\beta}, t)$, $s_k^{(l)}(\boldsymbol{\beta}, t)$ as the expectation of $S_k^{(l)}(\boldsymbol{\beta}, t)$. The theoretical approach relies on the counting processes (Wang, 1996) and we construct

$$M_{ki}(t) = N_i(t) - \int_0^t w_C(u) R_i(u) W_{ki} \exp\left(\boldsymbol{\beta}^t \boldsymbol{X}_i\right) \lambda(u) du, \quad k = 1, 2; i = 1, \cdots, n. \tag{14}$$

We make the following claim (Qin and Shen, 2010):

Claim 1. *$M_{ki}(t)$ is a mean zero stochastic process.*

Proof. When $k = 1$, (equation $U_1(\boldsymbol{\beta})$), we first notice that the expectation of an indicator is just the probability that its values takes 1, so for the first term, we have

$$\mathbb{E}[N_i(t)] = \mathbb{P}(N_i(t) = 1) = \mathbb{P}(T_i \leq t, C_i \geq T_i - A_i).$$

Since C is independent of A, T, hence the joint density of A, T, C is equal to the joint density of A, T we derived in the beginning of section 1.3 times the density of C, with bounds carefully chosen.

$$\begin{aligned}
\mathbb{P}(T_i \leq t, C_i \geq T_i - A_i) &= \int_0^t \int_0^t \int_{t_i - a_i}^{\infty} f_{A,T}(a_i, t_i) \cdot f_C(c_i) dc_i da_i dt_i \\
&= \int_0^t \int_0^t \frac{f(t_i|\boldsymbol{x}_i)}{\mu(\boldsymbol{x}_i)} \cdot S_C(t_i - a_i) da_i dt_i \\
&= \int_0^t \frac{f(t_i|\boldsymbol{x}_i)}{\mu(\boldsymbol{x}_i)} \cdot w_C(t_i) dti \\
&= \int_0^t \frac{S_U(t_i|\boldsymbol{x}_i) \cdot \lambda(t_i|\boldsymbol{x}_i)}{\mu(\boldsymbol{x}_i)} \cdot w_C(t_i) dt_i.
\end{aligned} \tag{15}$$

On the other hand, for the second term on the right hand side of (14) (denote by $K_i(t)$), using (8) its expectation is given by

$$\mathbb{E}[K_i(t)] = \mathbb{E}\left[\int_0^t \mathbb{E}\left[\frac{\delta_i \mathbf{1}\{T_i \geq u\}}{T_i S_C(T_i - A_i)}\middle| \boldsymbol{X}_i = \boldsymbol{x}_i\right] \cdot R_i(u) w_C(u) \lambda(u|\boldsymbol{X}_i) du\right]$$

with further simplification, we have

$$\mathbb{E}[K_i(t)] = \int_0^t \frac{S_U(u|\boldsymbol{x}_i)}{\mu(\boldsymbol{x}_i)} \cdot w_C(u) \lambda(u|\boldsymbol{x}_i) du. \tag{16}$$

From (15),(16), we see that $\mathbb{E}[M_{1i}(t)] = 0$. The same idea can be applied to $U_2(\boldsymbol{\beta})$ and $U_L(\boldsymbol{\beta})$. □

Under this condition, the estimating equation can be written as

$$U_k(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^\tau (\boldsymbol{X}_i - e_k(\boldsymbol{\beta}, t)) dM_i(t) + o_p(1)$$

where τ is a predetermined constant with $\mathbb{P}(A_i + C_i \geq \tau) > 0$. If S_C is a unknown function, we may replace S_C by its Kaplan-Meier estimator $\widehat{S}_C$, and due to the consistency of Kaplan-Meier estimator, we have

$$\widehat{U}_k(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^\tau (\boldsymbol{X}_i - \widehat{E}_k(\boldsymbol{\beta}, t)) dN_i(t). \tag{17}$$

Set $\widehat{U}_k(\boldsymbol{\beta}) = 0$, we can solve for the estimate $\widehat{\boldsymbol{\beta}}$. We now study each equation separately.First let $k = 1$, a second claim follows.

Claim 2. *Under regularity conditions (see appendix), there is a unique solution to* (17) *such that* $\widehat{U}_1(\boldsymbol{\beta}) = 0$. *Denote* $\widehat{\boldsymbol{\beta}}_1$ *be the solution, it is also consistent.*

Moreover, $\widehat{U}_1(\boldsymbol{\beta})/\sqrt{n}$ converges weakly to a mean zero Gaussian process (Qin and Shen, 2010).

Some other asymptotic properties will be added, I need some extra reading to understand the proofs.

2 Non-parametric Model

2.1 Introduction

When right-censoring is absent, Vardi (1982) derived a non-parametric estimating model using non-parametric maximum likelihood estimator (NPMLE) from a pooled sample $\{X_1, \cdots, X_m\} \cup \{Y_1, \cdots, Y_n\}$ where X_is are unbiased (complete) observations while Y_js are length-biased observations. Vardi (1982) showed that the survival function $\widehat{S}_U(t)$ obtained follows the asymptotic property that $\sqrt{m+n}(\widehat{S}_U(t) - S_U(t))$ converges weakly to a pinned Gaussian process when $m, n \to \infty$ and $m/n \to$ constant. With the presence of right-censoring, Vardi (1989) derived the NPMLE of a length-biased and uninformatively censored data; Asgharian et al. (2002) used an unconditional approach to obtain an estimate with asymptotics. We again begin by introducing the machinery of estimating equations without right censoring; we then bring in the discussion of potential censoring; we shall last study the asymptotic properties of those estimators.

2.2 NPMLE of Length-Biased Data

We again denote by f as the true density of the sample and g as the length-biased density. We take two independent samples $X_1, \cdots, X_m \overset{i.i.d}{\sim} f$ and $Y_1, \cdots, Y_n \overset{i.i.d}{\sim} g$ and study the pooled sample $\{X_i\} \cup \{Y_i\}$. Denote the ordered failure times in the pooled sample to be $t_1 < t_2 < \cdots < t_h$ where $t_h \leq m + n$ allowing potential ties. We also use ξ_i, η_i as the multiplicity of X_i's and Y_i's at time t_i. Then the likelihood function of the pooled sample takes the form

$$L = \prod_{i=1}^{h} f(t_i)^{\xi_i} \cdot \left(\frac{t_i f(t_i)}{\displaystyle\int_0^\infty u f(u) du} \right)^{\eta_i} = \prod_{i=1}^{h} f(t_i)^{\xi_i} \cdot \left(\frac{t_i f(t_i)}{\mu} \right)^{\eta_i}. \tag{18}$$

Vardi's approach to maximize (18) is to introduce a class of step functions that has positive jumps at each observed failure times $t_1, \cdots, t_h$ only. Denote by p_j as the jump at t_i, $p_j = dF(t_j)$, then the discrete case likelihood function with p_j as variables can be written as

$$L(p_1, \cdots, p_h) = \prod_{i=1}^{h} p_i^{\xi_i} \cdot \left(\frac{t_i p_i}{\sum_{j=1}^{h} t_j p_j} \right)^{\eta_i}$$

where $\sum_{j=1}^{h} p_j = 1$ and $p_j > 0$. Hence we now have an optimization problem maximizing $L(p_1, \cdots, p_h)$ under the constraint $\sum_{j=1}^{h} p_j = 1$. We first derive the log-likelihood function as

$$\ell(p_1, \cdots, p_h) = \sum_{i=1}^{h} \left[\xi_i \log p_i + \eta_i \log(t_i p_i) - \eta_i \log \left(\sum_{j=1}^{h} t_j p_j \right) \right]$$

by definition, $m = \sum_{i=1}^{h} \xi_i, n = \sum_{i=1}^{h} \eta_i$ and we additionally define $n_i = \xi_i + \eta_i$, so the log-likelihood function is now

$$\ell(p_1, \cdots, p_h) = \left(\sum_{i=1}^{h} n_i \log p_i \right) - n \log \left(\sum_{j=1}^{h} t_j p_j \right) + \sum_{i=1}^{h} \eta_i \log t_i \tag{19}$$

where the last term in equation (19) is a constant with respect to p_i's. So now our goal is to maximize this equation under the constraint $\sum_{j=1}^{h} p_j = 1$, where we shall use Lagrange multiplier method. Define

$$L(\lambda) = \left(\sum_{i=1}^{h} n_i \log p_i \right) - n \log \left(\sum_{j=1}^{n} t_j p_j \right) + \lambda \left(1 - \sum_{j=1}^{n} p_j \right)$$

and the partial derivatives are given by

$$\frac{\partial L(\lambda)}{\partial p_i} = \frac{n_i}{p_i} - \frac{n t_j}{\sum_{j=1}^{n} t_j p_j} - \lambda$$

we solve for $\frac{\partial L(\lambda)}{\partial p_i} = 0$ and it yields

$$p_i = \frac{n_i \mu}{\lambda \mu + n t_i} \tag{20}$$

where $\mu = \sum_{j=1}^{h} t_j p_j$, we apply (20) to the constraint and we have

$$\sum_{i=1}^{h} \frac{n_i \mu}{\lambda \mu + n t_i} = 1.$$

Here we specifically set $\lambda = m$, so the denominator contains both the weight from the biased data $n t_i$ and the mean of unbiased data λm. By solving p_i under this constraint, the final solution takes the form

$$\widehat{p}_i = \frac{n_i \widehat{\mu}}{n t_i + m \widehat{\mu}} \tag{21}$$

where $\sum_{i=1}^{h} \widehat{p}_i = 1$ and $\widehat{\mu} = \sum_{i=1}^{h} \widehat{p}_i t_i$. Under these two constraints, $\widehat{\mu}$ solves the equation

$$\sum_{j=1}^{h} \frac{n_j t_j}{n t_j + m\widehat{\mu}} = 1.$$

We call $\widehat{p}_i$ as the non-parametric maximum likelihood estimator (NPMLE) of the sample. Then we may estimate the cumulative distribution function (cdf) of the biased sample as

$$\widehat{G}(x) = \sum_{t_i \leq x} \widehat{g}_i = \sum_{t_i \leq x} \frac{t_i \widehat{p}_i}{\widehat{\mu}} = \sum_{t_i \leq x} \frac{n_i t_i}{n t_i + m\widehat{\mu}}.$$

By the invariance of MLE, $\widehat{G}(x)$ is the NPMLE of the length-biased cdf of $G(x)$. We now study some properties of the estimator (21) proposed by Vardi (1982). First is that this estimator is vaild for the case when $m = 0$ or $n = 0$, representing a sample of fully length-biased or fully unbiased data. If $n = 0$, then $\widehat{\mu}$ is simply the mean of the unbiased sample, and $\widehat{G}$ is the empirical distribution of the unbiased sample; If $m = 0$, then $\widehat{\mu}$ is the harmonic mean of the biased sample, and $\widehat{p}_k \propto \frac{n_k}{t_k}$, which corresponds to the estimator proposed by Cox (1969) when estimating length-biased sample (Vardi, 1982).

2.2.1 Asymptotic Properties

The asymptotic properties of the estimator are studied as follows (Vardi, 1982): We let F be the true cdf of the sample, and assume it is absolutely continuous with respect to Lebesgue measure, with density f and mean μ, let $N = m + n$ (sample size) and $\lambda = \frac{m}{N}$ (proportion of unbiased observations), then we are able to get the estimates $\widehat{F}, \widehat{G}, \widehat{\mu}$ using the method proposed. Under the condition that $N \to \infty$ while λ being fixed, define

$$K(x) = \int_0^x \frac{y}{\lambda\mu + (1-\lambda)y} \cdot f(y)dy$$

and $K = K(\infty)$. One remark is that if we assume F is absolutely continuous with respect to Lebesgue measure, then we have

$$\mathbb{P}(X_i = X_j) = 0, \text{almost surely}$$

this is because for a fixed x, the set $\mathcal{S} = \{(x, y) : x = y\}$ has measure 0 in $\mathbb{R}^2$ hence

$$\mathbb{P}(X_i = x, X_j = x) = \iint_{\mathcal{S}} f(x)f(y)dxdy = 0$$

and similarly we can show $\mathbb{P}(X_i = Y_j) = 0$, almost surely so in this case there are no ties and hence $h = m + n = N$, $\xi_i + \eta_i = 1; \xi_i, \eta_i \in \{0, 1\}$. Then we have the following theorem:

Theorem 1. *Under the conditions stated above, $\widehat{\mu} - \mu \to 0$ almost surely.*

Our proof follows the same outline by Vardi (1982), however we gave the full detail of the proof:

Proof. In our original estimate (21), we replace $n_i = \xi_i + \eta_i = 1$, and divide the numerator and denominator by N, and use the fact that $m = \frac{\lambda}{N}$, $\frac{n}{N} = 1 - \lambda$, we have a modified estimate

$$\widehat{p}_k = \frac{1}{N} \cdot \frac{\widehat{\mu}}{(1-\lambda)t_k + \lambda\widehat{\mu}}$$

such that $\widehat{\mu}$ solves the equation

$$Q_N(\widehat{\mu}) = \frac{1}{N} \sum_{k=1}^{N} \frac{t_k}{(1-\lambda)t_k + \lambda\widehat{\mu}} = 1.$$

the equation above can be viewed as the sample mean of the random variable $\dfrac{T}{(1-\lambda)T + \lambda\widehat{\mu}}$. Since our observations are taken from a pooled sample $\{X_i\} \cup \{Y_j\}$, hence we may estimate the mean as the sum of two separate sample means multiplied by their population proportion:

$$\begin{aligned}
\mathbb{E}\left[\frac{T_k}{(1-\lambda)T_k + \lambda\widehat{\mu}}\right] &= \lambda\mathbb{E}\left[\frac{X}{(1-\lambda)X + \lambda\widehat{\mu}}\right] + (1-\lambda)\mathbb{E}\left[\frac{Y}{(1-\lambda)Y + \lambda\widehat{\mu}}\right] \\
&= \lambda \int_0^\infty \frac{t}{(1-\lambda)t + \lambda\widehat{\mu}} \cdot f(t)dt + (1-\lambda) \int_0^\infty \frac{t}{(1-\lambda)t + \lambda\widehat{\mu}} \cdot \frac{tf(t)}{\mu} dt \\
&= \frac{1}{\mu} \int_0^\infty \left(\frac{\lambda\mu + (1-\lambda)t}{(1-\lambda)t + \lambda\widehat{\mu}}\right) \cdot tf(t)dt.
\end{aligned}$$

By Strong Law of Large Numbers (SLLN), we have

$$Q_N(\widehat{\mu}) \xrightarrow{a.s} Q(\widehat{\mu}) = \frac{1}{\mu} \int_0^\infty \left(\frac{(1-\lambda)t + \lambda\mu}{(1-\lambda)t + \lambda\widehat{\mu}}\right) \cdot tf(t)dt$$

which equals 1 if and only if $\mu = \widehat{\mu}$. Hence by Slutsky's theorem we have $\widehat{\mu} \xrightarrow{a.s} \mu$. □

Furthermore Vardi showed the weak convergence of the estimator:

Theorem 2. $\sqrt{N}\left(\dfrac{\widehat{\mu}}{\mu} - 1\right) \to N\left(0, \dfrac{1-K}{K\lambda(1-\lambda)}\right)$ *weakly.*

Proof. First we note that Q_N is monotone in $\widehat{\mu}$, also continuously differentiable, hence we may perform a Taylor expansion to its inverse function:

$$Q_N^{-1}(Q_N(a) + \epsilon_N(a)) = a + \frac{\epsilon_N(a)}{Q'(a)} + O(\epsilon^2(a))$$

where $\epsilon_N(a) = Q_N(a) - Q(a)$. By definition $Q(\mu) = 1$, $Q_N(\widehat{\mu}) = 1$ so letting $a = \mu$ we have $Q_N^{-1}(Q_N(\mu) + \epsilon_N(\mu)) = Q_N^{-1}(Q_N(\mu) - Q_N(\mu) + Q(\mu)) = Q_N^{-1}(1) = \widehat{\mu}$, and hence we have

$$\widehat{\mu} = \mu + \frac{\epsilon_N(\mu)}{Q'_N(\mu)} + O(\epsilon_N^2(\mu)).$$

also note that

$$Q'(\mu) = \frac{d}{da}\left(\frac{1}{\mu}\int_0^\infty \frac{(1-\lambda)t + \lambda\mu}{(1-\lambda)t + \lambda a} tf(t)dt\right)_{a=\mu} = -\frac{1}{\mu}\int_0^\infty \frac{((1-\lambda)t + \lambda\mu)\lambda}{((1-\lambda)t + \lambda\mu)^2} tf(t)dt = -\frac{\lambda}{\mu}K$$

and we now have

$$\widehat{\mu} - \mu = -\frac{\mu}{\lambda K}\epsilon_N(\mu) + O(\epsilon_N^2(\mu)).$$

Finally we use central limit theorem to approximate $\epsilon_N(\mu)$. We have that

$$\sqrt{N}(Q_N(\mu) - Q(\mu)) \xrightarrow{d} N\Big(0, \sigma^2\Big)$$

where $\sigma^2 = \mathbf{Var}\left(\dfrac{T}{(1-\lambda)T + \lambda\mu}\right)$. □

As for the asymptotic properties of the cdf $\widehat{F}$ and $\widehat{G}$, Vardi (1982) also showed that they converge in distribution to a pinned Gaussian processes, where

$$\sqrt{N}(\widehat{F} - F) \to V_F \text{ weakly}$$

with

$$\mathbf{Cov}(V_F(s), V_F(t)) = \frac{1}{\lambda}(F(s)(1-F(t)) - (1-\lambda)K(s)(1-K(t)/K)), 0 \le s \le t$$

and

$$\sqrt{N}(\widehat{G} - G) \to V_G \text{ weakly}$$

with

$$\mathbf{Cov}(V_G(s), V_G(t)) = \frac{1}{1-\lambda}(G(s)(1-G(t)) - \lambda K(s)(1-K(t)/K)), 0 \le s \le t.$$

2.3 An Adjusted NPMLE with Right Censoring

In this section we will derive the NPMLE of a length-biased, right-censored sample. Adapting the notion from section 1.1, let $f(t)$ be the true density of the sample, $g(t) = \frac{tf(t)}{\mu}$ be the length-biased density where $\mu = \int_0^\infty uf(u)du$. For each individual X_i belonging to the length-biased observation, we introduce a covariate vector $(X_i, A_i, V_i, C_i, T_i, \delta_i)$: A_i as the left truncation time (length-biased); V_i as the time from the start of the observation to death (event); C_i as the right censoring time from the start of the observation; $T_i = \min\{A_i + V_i, A_i + C_i\}$ as the complete length-biased, right censored event time; $\delta_i = \mathbf{1}\{V_i \le C_i\}$ as the censoring indicator. For length-biased, right-censored sample, we have derived in section 1.3 that if $T = A + V$ yielding the uncensored observation, we have

$$\mathbb{P}(T = A + V) = \frac{f(t)}{\mu} \cdot S_C(t-a)$$

while when $T = A + C$ yielding the censored observation,

$$\mathbb{P}(T = A + C) = \mathbb{P}(C = t - a) \cdot \frac{S_U(t)}{\mu}.$$

Assume that C_i is independent from A_i, V_i and additionally let η_i be the multiplicity at t_i, so we will have a set of observations $\{t_1, \cdots, t_h\}$ where $h \le n$ and $\sum_{i=1}^h \eta_i = n$. The likelihood function of the sample is proportional to

$$L = \prod_{i=1}^h \left(\frac{f(t_i)}{\mu}\right)^{\delta_i \eta_i} \cdot \left(\frac{S_U(t_i)}{\mu}\right)^{(1-\delta_i)\eta_i}$$

The maximizing idea is the same as Vardi (1982): Consider a class of continuous step functions that has positive jumps p_i at each t_i only, so $p_i = dF(t_i)$ and $\sum_{i=1}^{n} p_i = 1$. It now suffices to maximize

$$L(p_1, \cdots, p_n) = \prod_{i=1}^{n} \left(\frac{p_i}{\sum_{i=1}^{n} t_i p_i} \right)^{\delta_i \eta_i} \cdot \left(\frac{\sum_{t_j \geq t_i} p_j}{\sum_{i=1}^{n} t_i p_i} \right)^{(1-\delta_i)\eta_i} \tag{22}$$

under the constrain $\sum_{i=1}^{n} p_i = 1$, which can be obtained by Lagrange multiplier method. Denote $\ell(p_1, \cdots, p_n) = \log L(p_1, \cdots, p_n)$, define

$$\mathcal{L}(\lambda) = \ell(p_1, \cdots, p_n) - \lambda \sum_{i=1}^{n} p_i$$

then we have

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial p_i} &= \frac{\partial}{\partial p_i} \sum_{i=1}^{n} \left(\delta_i \eta_i \left(\log(p_i) - \log \sum_{i=1}^{n} t_i p_i \right) + (1-\delta_i)\eta_i \left(\log \sum_{t_j \geq t_i} p_j - \log \sum_{i=1}^{n} t_i p_i \right) - \lambda \sum_{i=1}^{n} p_i \right) \\
&= \left(\frac{\delta_i \eta_i}{p_i} + \frac{(1-\delta_i)\eta_i}{\sum_{t_i \geq t_j} p_j} - \frac{n t_i}{\mu} - \lambda \right).
\end{aligned}$$

Set $\frac{\partial \mathcal{L}}{\partial p_i} = 0$ and we have

$$\begin{cases} p_i = \dfrac{\eta_i \mu}{(\lambda \mu + n t_i)} & \text{if } \delta_i = 1 \\ \displaystyle\sum_{t_j \geq t_i} p_j = \dfrac{\eta_i \mu}{\lambda \mu + n t_i} & \text{if } \delta_i = 0 \end{cases}. \tag{23}$$

At each uncensored point, its jump p_i can be estimated by $\widehat{p}_i = \frac{\eta_i \mu}{\lambda \mu + n t_i}$ where λ is chosen such that it solves the equation $\sum_{i=1}^{n} p_i = 1$ and $\sum_{i=1}^{n} p_i t_i = \mu$; On the other hand, at each censored point t_i, its estimate depends on the survival function, and the following algorithm will compute each separately:

Initiation: Let U_j be the set of all uncensored observations after t_j, C_j be the set of all censored observations after t_j.

Step 1: Identify the latest censored point, denote by t_{c_1}, and its estimate is given by

$$\widehat{p}_{c_1} = \sum_{t_j \geq t_{c_1}} p_j - \sum_{t_j > t_{c_1}} p_j = \frac{\eta_{c_1} \mu}{\lambda \mu + n t_{c_1}} - \sum_{t_j > t_{c_1}} \frac{\eta_j \mu}{\lambda \mu + n t_j}$$

Step 2: Iterate backwards and find the second latest censored point, denote by t_{c_2}, and we compute its estimate by

$$\widehat{p}_{c_2} = \sum_{t_j \geq t_{c_2}} p_j - \sum_{t_j > t_{c_2}} p_j = \frac{\eta_{c_2} \mu}{\lambda \mu + n t_{c_2}} - \left(\sum_{t_j \in C_{c_2}} \widehat{p}_{t_j} + \sum_{t_j \in U_{c_2}} \widehat{p}_{t_j} \right)$$

Add t_{c_2} to the set C.

Step 3: Continue this iteration, in general the kth latest censored point t_{c_k} has the form

$$\widehat{p}_{c_k} = \sum_{t_j \geq t_{c_k}} p_j - \sum_{t_j > t_{c_k}} p_j = \frac{\eta_{c_k}\mu}{\lambda\mu + nt_{c_k}} - \left(\sum_{t_j \in C_{c_k}} \widehat{p}_{t_j} + \sum_{t_j \in U_{c_k}} \widehat{p}_{t_j} \right)$$

until we have locate all the censored observations, and output all censored estimates $\widehat{p}_{c_1}, \cdots, \widehat{p}_{c_l}$.

We now discuss some properties of the estimator (23) we proposed. First notice that without censoring, we will have

$$\widehat{p}_i = \frac{\eta_i\mu}{\lambda\mu + nt_i}$$

where λ solves

$$\sum_{i=1}^{h} \frac{\eta_i t_i}{\lambda\mu + nt_i} = 1.$$

It is straightforward to let $\lambda = 0$ and the summation will add up to one by definition. In this case we have $\widehat{p}_i = \frac{\eta_i\mu}{nt_i} \propto \frac{\eta_i}{t_i}$ which coincide with the one obtained by Vardi (1982).

References

[1] Asgharian, Masoud and Wolfson, David B. *Asymptotic behavior of the unconditional NPMLE of the length-biased survivor function from right censored prevalent cohort data*. The Annals of Statistics. 2005.

[2] Klein, John P. and Moeschberger, Melvin L. *Survival Analysis: Techniques for censored and truncated data*. Second Edition. 2003.

[3] Qin, Jing and Shen, Yu. *Statistical Methods for Analyzing Right-censored Length-biased Data under Cox Model*. Biometrics. 2010 June ; 66(2): 382–392.

[4] Vardi, Y. *Nonparametric Estimation in the Presence of Length-Bias*. The Annuals of Statistics, 1982.

[5] Wang, Mei-Cheng et al. *Statistical Models for Prevalent Cohort Data*. Biometrics. 1993.

[6] Wang, Mei-Cheng. *Hazards Regression Analysis for Length-Biased Data*. Biometrics. 1996.

[7] Zhang, Jiajun. *Analyzing Incident and Prevalent Cohort Survival Data: An Introduction to Survival Analysis*. McGill University. 2025.