Variance of m(), questions in the paper

2019-11-05

We denote $T_i, i = 1, ..., N$ are the independent, identically, distributed (iid) lifetimes, whose corresponding cumulative distriubtion function (CDF) is F, probability distribution function (PDF) is f; the censoring time is defined as $C_i, i = 1, ..., N$. C_is are also iid, with CDF denoted as G and PDF denoted as g. We set the censors happen on the right and the ovserved time is $Z_i = T_i \wedge C_i$, whose CDF is H and PDF is h. The $\delta_i = I_{[T_i \leq C_i]}$ is the status indicator, which shows whether subject i is censored ($\delta_i = 0$) or not ($\delta_i = 0$). The corresponding hazard function of lifetime is λ_F and cumulative hazard function is Λ_F.

The variance comparison goal

We denote the estimate of $F(t)$:

- with Kaplan Meier as $F_n^{km}(t)$
- with Dikta's method: $\hat{F}_n^D(t)$
- with the new method: $\hat{F}_n^N(t)$

We denote the asymptiotic variance of $n^{1/2}(F_n^{km}(t) - F(t))$ by $v^{km}(t)$, that of $n^{1/2}(F_n^D(t) - F(t))$ by $v^D(t)$, that of $n^{1/2}(F_n^N(t) - F(t))$ by $v^N(t)$.

We have know that $v^{km}(t) > v^D(t)$ from the Dikta's paper. Our goal is to compare $v^{km}(t)$ with $v^N(t)$, and $v^D(t)$ with $v^N(t)$

We have that

$$v^{km}(t) = (1 - F(t))^2 \int_0^t \frac{1}{(1 - H(x))^2} H^1(dx)$$

$$v^{km}(t) - v^D(t) = (1 - F(t))^2 r(t) \geq 0$$

$$v^D(t) = v^{km}(t) - (1 - F(t))^2 r(t) = (1 - F(t))^2 \Big[\int_0^t \frac{1}{(1 - H(x))^2} H^1(dx) - r(t) \Big]$$

where

$$r(t) = \int_0^t \frac{1 - m(x, \theta_0)}{(1 - H(x))^2} H^1 d(x) - \int_0^t \int_0^t \frac{\alpha(x, y)}{(1 - H(x))(1 - H(y))} H(dy) H(dx)$$

$$\alpha(x, y) = \langle \text{Grad}(m(x, \theta_0)), \Big(I^{-1}(\theta_0) \text{Grad}(m(x, \theta_0)) \Big) \rangle$$

Our goal is to find

$$n^{1/2}(F_n^N(t) - F(t))$$

Since there is $n^{1/2}(\Lambda_n(t) - \Lambda(t))$, could we use a Delta method to calculate $n^{1/2}(F_n^N(t) - F(t))$?

New assumption:

Instead of the strong assumption of independent between T_i and C_i, we proposed that $T \perp\!\!\!\perp C$ at a small neighborhood, where $T = C$. That is, we have

$$\lim_{dt\to 0} P(C > t, T \geq t + dt) = P(C > t)P(T \geq t + dt) \tag{1}$$

As well as

$$P(C > t, T \geq t) = P(C > t)P(T \geq t) \tag{2}$$

With this assumption, we can show:

$$\begin{aligned} P(C > t|T = t) &= \lim_{dt\to 0} P(C > t|t \leq T < t + dt) \\ &= \lim_{dt\to 0} \frac{P(C > t, t \leq T < t + dt)}{P(t \leq T < t + dt)} \\ &= \lim_{dt\to 0} \frac{P(C > t, T \geq t) - P(C > t, T > t + dt)}{P(T \geq t) - P(T > t + dt)} \\ &= \lim_{dt\to 0} \frac{P(C > t)\big(P(T \geq t) - P(T > t + dt)\big)}{P(T \geq t) - P(T > t + dt)} \\ &= P(C > t) \end{aligned} \tag{3}$$

And since indpendent,

$$P(C > t|T > t) = \frac{P(C > t, T > t)}{P(T > t)} = \frac{P(C > t)P(T > t)}{P(T > t)} = P(C > t)$$

Therefore,

$$P(C > t|T > t) = P(C > t|T = t) \tag{4}$$

Given (Eq 1), we could derive that

$$P(\delta = 1|X = t) = \frac{P(C > t, T = t)}{P(X = t)} = \frac{P(T = t)}{P(X = t)} \frac{P(C > t, T > t)}{P(T > t)} = \frac{f(t)S_x(t)}{h(t)S(t)} = \frac{\lambda_F(t)}{\lambda_H(t)}$$

where $\lambda_H(t)$ is the hazard function corresponding to Z, which is known as crude hazard rate as well.

We may define $m(t) = P(\delta = 1|X = t) = E(\delta|X = t)$. Then

$$m(t) = \frac{\lambda_F(t)}{\lambda_H(t)} \tag{5}$$

which is the same parameter defined in Dikta's papers. Therefore, the independence between T and C is not the necessory condition for equation (2).

A question about the $H^1, m(t)$ relationship

$$m(x) = E(\delta = 1|Z = x) = P(\delta = 1|Z = x) = \frac{\lambda_F(x)}{\lambda_H(x)}$$

And there is a relationship in Dikta's paper:

$$\begin{aligned} H_1(x) =& P(\delta = 1, Z \leq x) = P(T \leq x, T < C) \\ =& \int_0^x \int_t^\infty f_{ts}(t, s) ds dt \end{aligned} \tag{6}$$

There is another relationship in the paper

$$\begin{aligned} H_1(x) =& P(\delta = 1, Z \leq x) = \int_0^x \bar{G}(t) F(dt) \\ =& \int_0^x \int_t^\infty g(s) ds f(t) dt \\ =& \int_0^x \int_t^\infty f(t) g(s) ds dt \end{aligned} \tag{7}$$

And eq (t) = eq (7), is that means that $f_{t,s}(t, s) = f(t)g(s)$ every where?

$m(t)$ function, $H_1(t)$ function

$$H_1(x) = P(\delta = 1, Z \leq x) = \int_0^x m(z) H(dz)$$

Since $m(x) = \frac{\lambda_F(x)}{\lambda_H(x)}$, then $\int_0^x m(z)H(dz) = \int_0^x \lambda_F(z)dz$, then $H_1(x) = P(\delta = 1, Z \leq x) = \int_0^x \lambda_F(z)dz$? is that correct?

$$H^1(t) = \int_0^t m(x) H(dx)$$

$$\begin{aligned} \Lambda(t) = \int_0^t \frac{1}{1 - F(x)} F(dx) = \int_0^t \frac{1}{1 - H(x)} H^1(dx) \\ \rightarrow \Lambda(t) = \int_0^t \frac{m(x)}{1 - H(x)} H(dx) \end{aligned}$$

Formulas

The relationship

$$m(x) = P(\delta = 1|Z = x) = E(\delta|Z = x)$$

$$m(x) = \frac{\lambda_F}{\lambda_H} = \frac{f}{S}\frac{S_h}{f_h}$$

$$\begin{aligned}
H^1(t) =& P(\delta = 1, Z \le x) = \int_0^x m(z)h(z)dz \\
=& \int_0^x \frac{f(z)}{S(z)}\frac{S_h(z)}{d}z, \\
& \text{since } S_h(x) = P(Z > x) = P(T > x)P(C > x) = S(x)S_c(x) \\
=& \int_0^x f(z)S_c(z)dz = \int_0^x f(z)G(z)dz, \text{ (we denote } G(z) = S_c(z))
\end{aligned}$$

$$H^1(dt) = m(z)h(z)dz$$

$$\Lambda(t) = \int_0^t \frac{f(x)}{S(x)}dx = \int_0^t \lambda_f(x)dx = \int_0^t m(x)\lambda_H(x)dx = \int_0^t \frac{m(x)f_H(x)}{S_H(x)}dx = \int_0^t \frac{m(x)}{1 - H(x)}H(dx)$$

Since $H^1(t) = \int_0^x m(z)h(z)dz$,

$$\Lambda(t) = \int_0^t \frac{m(x)}{1 - H(x)}H(dx) = \int_0^t \frac{1}{1 - H(x)}H^1(dx)$$