A short note on linear representation of the Cox's profile likelihood estimator

Yen-Chi Chen
University of Washington
June 2, 2020

In survival analysis, the Cox model is a powerful method for understanding the effect of covariates on the time-to-event variable. Consider a typical right-censoring data where we observe IID

$$(Y_1, X_1, \Delta_1), \cdots, (Y_n, X_n, \Delta_n),$$

where $Y_i = \min\{T_i, C_i\}$ is the observed time and $T_i \geq 0$ is the time-to-event of interest and $C_i \geq 0$ is the censoring time and $\Delta_i = I(Y_i = T_i)$ is the censoring indicator and $X_i \in \mathbb{R}^p$ is the covariate. We assume the typical assumption that

$$T \perp C | X.$$

Let $P(t|x) = P(T \leq t | X = x)$ be the CDF of T given $X = x$ and $S(t|x) = 1 - P(t|x)$ is the survival time and $h(t|x) = -\frac{\partial}{\partial t} \log S(t|x)$ be the hazard function and $H(t|x) = \int_0^t h(s|x)ds$ is the cumulative hazard.

The Cox (proportional hazard) model assumes that

$$h(t|x) = h_0(t) \exp(\beta^T x).$$

And the goal is to estimate the coefficients β.

It is known that we can estimate β by solving the following *profile/partial likelihood (score) equation*:

$$\begin{aligned} \widehat{\beta} : 0 &= \sum_{i=1}^n \Delta_i \left(X_i - \frac{S_n^{(1)}(Y_i; \widehat{\beta})}{S_n^{(0)}(Y_i; \widehat{\beta})} \right) \\ S_n^{(1)}(t; \beta) &= \frac{1}{n} \sum_{i=1}^n I(Y_i \geq t) \exp(\beta^T X_i) X_i \\ S_n^{(0)}(t; \beta) &= \frac{1}{n} \sum_{i=1}^n I(Y_i \geq t) \exp(\beta^T X_i) \end{aligned} \tag{1}$$

It is well-known that the estimator $\widehat{\beta}$ is consistent under regularity conditions and has asymptotic normality. Although one may expect this result from the usual theory of estimating equations, equation (1) is not a simple estimating equation that consists of summation of IID terms because $S_n^{(1)}(t; \beta)$ and $S_n^{(0)}(t;]beta)$ depend on every observation.

This fact has made the analysis a bit complicated. Here we will provide a simple way to illustrate the asymptotic linear form (summation of IID random elements) of $\widehat{\beta}$ even if equation (1) does not consists of IID terms. The asymptotic linear form of $\widehat{\beta}$ makes the consistency and asymptotic normality very straight forward.

While there are many ways to derive this, the approach we will be using is based on the following paper:

[LW1989] Lin, D. Y., & Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. Journal of the American statistical Association, 84(408), 1074-1078.

To start with, we describe the population version of equation (1):

$$
\begin{aligned}
\beta^* : 0 &= \mathbb{E}\left[\Delta_i\left(X_i - \frac{s^{(1)}(Y_i;\beta^*)}{s^{(0)}(Y_i;\beta^*)}\right)\right] \\
s^{(1)}(t;\beta) &= \mathbb{E}\left[I(Y_i \geq t)\exp(\beta^T X_i)X_i\right] \\
s^{(0)}(t;\beta) &= \mathbb{E}\left[I(Y_i \geq t)\exp(\beta^T X_i)\right]
\end{aligned} \tag{2}
$$

We will show that

$$
\sqrt{n}(\widehat{\beta} - \beta^*) = \sqrt{n}\sum_{i=1}^{n}\xi_i + o_P(1), \tag{3}
$$

where $\xi_1, \cdots, \xi_n$ are IID random vectors (they are the influence function evaluated at each observation).

1 Derivation of the asymptotic linear form

Step 1: Taylor expansion. Let $U_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\Delta_i\left(X_i - \frac{S_n^{(1)}(Y_i;\beta)}{S_n^{(0)}(Y_i;\beta)}\right)$ and $U_0(\beta) = \mathbb{E}\left[\Delta_i\left(X_i - \frac{s^{(1)}(Y_i;\beta)}{s^{(0)}(Y_i;\beta)}\right)\right]$. Using equation (1) and (2), we can easily decompose

$$
\begin{aligned}
U_n(\beta^*) &= U_n(\beta^*) - U_n(\widehat{\beta}) \\
&= (\beta^* - \widehat{\beta})^T \nabla U_n(\beta^*) + \text{smaller order terms.}
\end{aligned}
$$

Ignoring the smaller order terms, we obtain

$$
\widehat{\beta} - \beta^* \approx [\nabla U_n(\beta^*)]^{-1} U_n(\beta^*).
$$

One can show that under suitable conditions, the matrix

$$
[\nabla U_n(\beta^*)]^{-1} \stackrel{P}{\to} [\nabla U(\beta^*)]^{-1}
$$

and is invertible. Thus, all we need to focus is the term $U_n(\beta^*)$.

We will show that $U_n(\beta^*)$ has an asymptotic linear expansion.

Step 2: linking U_n to empirical process. We will derive an alternative representation of U_n to make a clear link to the empirical process. Define $N_i(t) = I(Y_i \leq t, \Delta_i = 1)$. Then $U_n(\beta)$ can be represented as

$$
\begin{aligned}
U_n(\beta) &= \frac{1}{n}\sum_{i=1}^{n} \Delta_i \left(X_i - \frac{S_n^{(1)}(Y_i;\beta)}{S_n^{(0)}(Y_i;\beta)} \right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \int \left(X_i - \frac{S_n^{(1)}(t;\beta)}{S_n^{(0)}(t;\beta)} \right) dN_i(t) \\
&= \frac{1}{n}\sum_{i=1}^{n} \int X_i dN_i(t) - \underbrace{\frac{1}{n}\sum_{i=1}^{n} \int \frac{S_n^{(1)}(t;\beta)}{S_n^{(0)}(t;\beta)} dN_i(t)}_{(*)}
\end{aligned}
\tag{4}
$$

We will now focus on the term $(*)$. Let $G_n(t) = \frac{1}{n}\sum_{i=1}^{n} N_i(t) = \frac{1}{n}\sum_{i=1}^{n} I(Y_i \leq t, \Delta_i = 1)$. This quantity behaves like the observed event empirical distribution function but note that the denominator is n. Let $\bar{G}(t) = \mathbb{E}[G_n(t)]$. We then decompose $(*)$ by the following

$$
\begin{aligned}
(*) &= \int \frac{S_n^{(1)}(t;\beta)}{S_n^{(0)}(t;\beta)} dG_n(t) \\
&= \underbrace{\int \frac{s^{(1)}(t;\beta)}{s^{(0)}(t;\beta)} d(G_n(t) - \bar{G}(t))}_{(I)} + \underbrace{\int \frac{S_n^{(1)}(t;\beta)}{S_n^{(0)}(t;\beta)} d\bar{G}(t)}_{(II)} \\
&\quad + \underbrace{\int \left(\frac{S_n^{(1)}(t;\beta)}{S_n^{(0)}(t;\beta)} - \frac{s^{(1)}(t;\beta)}{s^{(0)}(t;\beta)} \right) d(G_n(t) - \bar{G}(t))}_{(III)}.
\end{aligned}
\tag{5}
$$

Step 3: controlling (III). First, we want to note that by the usual empirical process theory, $\sqrt{n}(G_n(t) - \bar{G}(t))$ converges to to a Gaussian process uniformly. Also, one can easily see that

$$
S_n^{(1)}(t;\beta) = \frac{1}{n}\sum_{i=1}^{n} I(Y_i \geq t) \exp(\beta^T X_i) X_i
$$

is essentially an average of IID random element and $\mathbb{E}[S_n^{(1)}(t;\beta)] = s^{(1)}(t;\beta)$. Assuming that both T and C are bounded from the above and the parameter β is restricted to a compact set $\mathcal{B}$, one can easily show that the function

$$
\{\eta_{t,\beta}(y,x) = I(y \leq t)\exp(\beta^T x) : t \in [0,\bar{T}], \beta \in \mathcal{B}\}
$$

forms a GC class so $\sup_{t,\beta} |S_n^{(1)}(t;\beta) - s^{(1)}(t;\beta)| = o_P(1)$ and similarly $\sup_{t,\beta} |S_n^{(0)}(t;\beta) - s^{(0)}(t;\beta)| = o_P(1)$. As a result, one can see that $(III) = o_P(1/\sqrt{n})$.

Step 4: controlling (I) and (II). Now we control the second term (II). We denote

$$
\varepsilon_j = S_n^{(j)}(t;\beta) - s^{(j)}(t;\beta)
$$

for $j=0,1$. Due to the uniform convergence property, ε_j is approaching 0. So we can rewrite the ratio using the Taylor's theorem as

$$\begin{aligned}
\frac{S_n^{(1)}(t;\beta)}{S_n^{(0)}(t;\beta)} &= \frac{S_n^{(1)}(t;\beta)}{s^{(0)}(t;\beta)+\varepsilon_0} \\
&= \frac{S_n^{(1)}(t;\beta)}{s^{(0)}(t;\beta)(1+\frac{\varepsilon_0}{s^{(0)}(t;\beta)})} \\
&\approx \frac{S_n^{(1)}(t;\beta)}{s^{(0)}(t;\beta)}\left(1-\frac{\varepsilon_0}{s^{(0)}(t;\beta)}\right) \\
&\approx \frac{1}{s^{(0)}(t;\beta)}\left(S_n^{(1)}(t;\beta)-\frac{s^{(1)}(t;\beta)}{s^{(0)}(t;\beta)}\varepsilon_0\right) \\
&= \frac{1}{s^{(0)}(t;\beta)}\left(S_n^{(1)}(t;\beta)-\frac{s^{(1)}(t;\beta)}{s^{(0)}(t;\beta)}S_n^{(0)}(t;\beta)+s^{(1)}(t;\beta)\right).
\end{aligned}$$

Thus,

$$(II)\approx\int\left(\frac{S_n^{(1)}(t;\beta^*)}{s^{(0)}(t;\beta^*)}-\frac{s^{(1)}(t;\beta^*)}{s^{(0)}(t;\beta^*)}\frac{S_n^{(0)}(t;\beta^*)}{s^{(0)}(t;\beta^*)}+\frac{s^{(1)}(t;\beta^*)}{s^{(0)}(t;\beta^*)}\right)d\bar{G}(t).$$

Notice that the third term of (II) also appears in (I). So

$$(I)+(II)+(III)=\int\frac{s^{(1)}(t;\beta^*)}{s^{(0)}(t;\beta^*)}dG_n(t)+\int\left(\frac{S_n^{(1)}(t;\beta^*)}{s^{(0)}(t;\beta^*)}-\frac{s^{(1)}(t;\beta^*)}{s^{(0)}(t;\beta^*)}\frac{S_n^{(0)}(t;\beta^*)}{s^{(0)}(t;\beta^*)}\right)d\bar{G}(t)+o_P(1/\sqrt{n}).$$

Step 5: Final expression. Using the fact that $G_n(t), S_n^{(j)}(t;\beta^*)$ are both summation of IID terms, we can rewrite (*) as

$$\begin{aligned}
(*) &= \frac{1}{n}\sum_{i=1}^n W_i+o_P(1/\sqrt{n}), \\
W_i &= \int\frac{s^{(1)}(t;\beta^*)}{s^{(0)}(t;\beta^*)}dN_i(t)+\int\frac{I(Y_i\geq t)\exp(\beta^{*T}X_i)}{s^{(0)}(t;\beta^*)}\left(X_i-\frac{s^{(1)}(t;\beta^*)}{s^{(0)}(t;\beta^*)}\right)d\bar{G}(t).
\end{aligned}\tag{6}$$

Note that $W_1,\cdots,W_n$ are IID random variables. Putting this into equation (4), we conclude that

$$\begin{aligned}
U_n(\beta^*) &= \frac{1}{n}\sum_{i=1}^n\Gamma_i+o_P(1/\sqrt{n}), \\
\Gamma_i &= \int X_i dN_i(t)+W_i=X_i\Delta_i+W_i
\end{aligned}\tag{7}$$

and $W_1,\cdots,W_n$ are IID random elements. As a result, the estimator $\widehat{\beta}$ can be written as

$$\sqrt{n}(\widehat{\beta}-\beta^*)=[\nabla U(\beta^*)]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n(X_i\Delta_i+W_i)+o_P(1),$$

which provides an asymptotic linear form of the estimator in equation (3) with $\xi_i=[\nabla U(\beta^*)]^{-1}(X_i\Delta_i+W_i)$. The asymptotic normality can be derived easily under this form (note: it is not hard to see that $E[X_i\Delta_i+W_i]=0$).

2 Remarks

- In [LW1989], they derive the asymptotic linear form in terms of time-varying covariates. In this case, the covariate $X_i = X_i(t)$ and we observe the covariate $\{X_t(t) : t \in [0, Y_i]\}$ until the observed time point Y_i. The score equation remains very similar; here is the score equation (c.f. equation (1)):

$$\begin{aligned} \widehat{\beta} : 0 &= \sum_{i=1}^{n} \Delta_i \left(X_i(Y_i) - \frac{S_n^{(1)}(Y_i; \widehat{\beta})}{S_n^{(0)}(Y_i; \widehat{\beta})} \right) \\ S_n^{(1)}(t; \beta) &= \frac{1}{n} \sum_{i=1}^{n} I(Y_i \geq t) \exp(\beta^T X_i(t)) X_i(t) \\ S_n^{(0)}(t; \beta) &= \frac{1}{n} \sum_{i=1}^{n} I(Y_i \geq t) \exp(\beta^T X_i(t)) \end{aligned} \tag{8}$$

  and one can modify the population version in equation (2) accordingly. The asymptotic linear form remains very similar and we only need to modify

$$\begin{aligned} \Gamma_i &= \int X_i(t) dN_i(t) + W_i \\ W_i &= \int \frac{s^{(1)}(t; \beta^*)}{s^{(0)}(t; \beta^*)} dN_i(t) + \int \frac{I(Y_i \geq t) \exp(\beta^{*T} X_i(t))}{s^{(0)}(t; \beta^*)} \left(X_i(t) - \frac{s^{(1)}(t; \beta^*)}{s^{(0)}(t; \beta^*)} \right) d\bar{G}(t) \end{aligned}$$

  and $s^{(j)}(t; \beta^*)$ is the modified version of the population quantity.

- This idea can be combined with IPW estimators under complex design. In particular, the following paper discussed the idea of generalizing it into a survey sample problem:

  Lin, D. Y. (2000). On fitting Cox's proportional hazards models to survey data. Biometrika, 87(1), 37-47.

  And the following paper considered the problem of missing covariates:

  Lin, D. Y., & Ying, Z. (1993). Cox regression with incomplete covariate measurements. Journal of the American Statistical Association, 88(424), 1341-1349.

  Note that both of the above papers are working on time-varying covariates as well.