Revisiting Cox’s PH Model: Asymptotic Theory

Earlier we sketched how the Martingale CLT can be applied to the score function arising from Cox’s proportional hazards model. This section shows this in greater detail and then sketches why the “MLE” from Cox’s partial likelihood function is consistent and asymptotically normal.

Consider the usual setting in which the survival time and covariate value for the \( i^{th} \) of \( n \) subjects are denoted \( T_i \) and \( Z_i \), where the potential censoring time is denoted \( C_i \), and where we assume noninformative censoring; i.e.,

\[
T_i \perp C_i \mid Z_i.
\]

Assume that the random variables corresponding to different subjects are independent, and that the \( Z_i \) are bounded. The observation for the \( i^{th} \) subject consists of \((U_i, \delta_i, Z_i)\), where

\[
U_i = \min (T_i, C_i), \quad \delta_i = 1(T_i \leq C_i).
\]

The hazard function for a subject with covariate value \( z \) is the usual Cox proportional hazards model

\[
h(t \mid Z = z) = \lambda_0(t) e^{\beta z}.
\]

We assume that there exists a \( \tau \) such that \( C_i \leq \tau \) with probability 1, and that \( P(Y_i(\tau) = 1) > 0 \). Thus, \( \tau \) is the largest possible observed time.

Suppose first that \( Z \) is a scalar which is bounded, and recall that the partial likelihood can be expressed as
\[ L_1(\beta) = \prod_j \frac{e^{\beta Z(j)}}{\sum_{l \in R_j} e^{\beta Z_l}}, \]

where the subscript \( j \) indexes the observed failure times, or equivalently as

\[ L_1(\beta) = \prod_{i=1}^n \int_0^\infty \frac{e^{\beta Z_i}}{\sum_{l=1}^n Y_l(s) e^{\beta Z_l}} dN_i(s). \]

The partial likelihood score function is given by (see Unit 12 page 9)

\[ U(\beta) = \frac{\partial \ln L_1}{\partial \beta} = \sum_{i=1}^n \int_0^\infty \left( Z_i - \sum_{l=1}^n \frac{Z_l e^{\beta Z_l} Y_l(s)}{\sum_{l} e^{\beta Z_l} Y_l(s)} \right) dN_i(s). \]

If we define

\[ S^{(j)}(\beta, t) = \frac{1}{n} \sum_{i=1}^n Y_i(t) Z_i^{j} e^{\beta Z_i} \]

for \( j = 0, 1, 2 \) and

\[ \mathcal{E}(\beta, t) = \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)}, \]

we can write

\[ U(\beta) = \sum_{i=1}^n \int_0^\infty (Z_i - \mathcal{E}(\beta, s)) dN_i(s). \]

Earlier, this led us to consider the process

\[ \overline{U}(\beta, t) = n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t (Z_i - \mathcal{E}(\beta, s)) dN_i(s), \]

which turned out to be the same as (see Unit 12 page 10)
\[ n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_{0}^{t} (Z_i - \mathcal{E}(\beta, s)) dM_i(s), \]

where

\[ M_i(s) = N_i(s) - A_i(s) \]

and

\[ A_i(s) = \int_{0}^{s} \lambda_0(u) e^{\beta Z_i Y_i(u)} du. \]

We then invoked the Martingale Central Limit Theorem to argue that

\[ \overline{U}(\beta, \cdot) \xrightarrow{L} W(\sigma^2(\cdot)) \text{ as } n \to \infty, \]

where \(W(\cdot)\) is a Wiener process (Brownian motion) and \(\sigma^2(t)\) is the probability limit of

\[ \sum_{i=1}^{n} \int_{0}^{t} \left( n^{-\frac{1}{2}} (Z_i - \mathcal{E}(\beta, s)) \right)^2 dA_i(s) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{t} (Z_i - \mathcal{E}(\beta, s))^2 Y_i(s) \lambda_0(s) e^{\beta Z_i} ds. \] (21.1)

Since \(L_1(\beta)\) is not necessarily a likelihood function, it doesn’t necessarily follow that the probability limit of the sample information obtained from \(L_1(\beta)\), which we use to standardize the partial likelihood score function, is equal to the variance of the asymptotic distribution of \(\overline{U}(\beta, \tau)\). To assess this, let the probability limit of \(S^{(j)}(\beta, t)\) be denoted \(s^{(j)}(\beta, t)\) and define

\[ V(\beta, t) = \frac{S^{(2)}(\beta, t)}{S^{(0)}(\beta, t)} - \left( \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)} \right)^2 \]

\[ = \ldots = \frac{1}{n} \sum_{i=1}^{n} (Z_i - \mathcal{E}(\beta, t))^2 Y_i(t) e^{\beta Z_i} \]

\[ = \ldots = \frac{1}{S^{(0)}(\beta, t)} \] .
(21.1) can be rewritten as
\[ \int_0^t S^{(0)}(\beta, s)V(\beta, s)\lambda_0(s)ds. \]

If the probability limit of \( V(\beta, t) \) is denoted \( v(\beta, t) \), \( \sigma^2 \) can be rewritten as,
\[ \sigma^2(t) = \int_0^t s^{(0)}(\beta, s) v(\beta, s) \lambda_0(s) ds. \] (21.2)

(notice that everything here is bounded since \( Z \) is bounded so pointwise convergence gives convergence of the integral (Theorem 15.1)).

Let the value of this variance function when \( t = \tau \) be denoted \( i(\beta) \), and assume that it is positive. It follows that
\[ n^{-\frac{1}{2}} U(\beta) = n^{-1/2} U(\beta, \tau) \xrightarrow{L} \mathcal{N}(0, i(\beta)) \quad \text{as} \quad n \to \infty. \]

Note also that the sample information at \( \beta \) can be written as
\[ \hat{I}(\beta) = -\frac{\partial^2 \ln L_1(\beta)}{\partial \beta^2} \]
\[ = \ldots = \sum_{i=1}^n \int_0^\tau \left( \frac{S^{(2)}(\beta, s)}{S^{(0)}(\beta, s)} - \mathcal{E}^2(\beta, s) \right) dN_i(s) \]
\[ = \sum_{i=1}^n \int_0^\tau V(\beta, s) dN_i(s) \]
\[ = \sum_{i=1}^n \int_0^\tau V(\beta, s) dM_i(s) + \sum_{i=1}^n \int_0^\tau V(\beta, s) dA_i(s). \]
Thus,

\[ \frac{1}{n} \hat{I}(\beta) = \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} V(\beta, s) dM_i(s) + \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} V(\beta, s) dA_i(s). \]

Because the first term has expectation 0 and variance going to 0, the first term here converges in probability to 0. Thus, \( \frac{1}{n} \hat{I}(\beta) \) has the same probability limit as

\[
\frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} V(\beta, s) dA_i(s)
\]

\[
= \int_0^{\tau} \lambda_0(s) \frac{1}{n} \sum_i Y_i(s) e^{\beta Z_i} \cdot V(\beta, s) ds
\]

\[
= \int_0^{\tau} \lambda_0(s) \cdot V(\beta, s) S^{(0)}(\beta, s) ds.
\]

From (21.2), the probability limit of this is \( \sigma^2(\tau) = i(\beta) \). Thus, it follows from Slutsky’s theorem that as \( n \to \infty \)

\[
\frac{U(\beta)}{\sqrt{\hat{I}(\beta)}} = \frac{n^{-1/2}U(\beta)}{\sqrt{\hat{I}(\beta)/n}} \xrightarrow{L} \mathcal{N}(0, 1).
\]

It can be shown that the limiting distribution is unchanged if we replace \( \hat{I}(\beta) \) by \( \hat{I}(\hat{\beta}) \).
We now sketch the proofs of the consistency and asymptotic normality of $\hat{\beta}$. For details, see Fleming and Harrington, Chapter 8, Section 4.

**Consistency of $\hat{\beta}$:** For this argument, denote the true value of $\beta$ by $\beta_0$. Define

$$X_n(\beta) = \frac{1}{n} \ln \left( \frac{L_1(\beta)}{L_1(\beta_0)} \right)$$

and regard this as a function of $\beta$ in some neighborhood, say $N_0$, of $\beta_0$. It can be shown that $X_n(\beta)$ is concave, and it is clear that $X_n(\beta)$ is maximized at $\beta = \hat{\beta}$. Now consider

$$A_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{T} \left( (\beta - \beta_0) Z_i - \ln \left( \frac{S^{(0)}(\beta, s)}{S^{(0)}(\beta_0, s)} \right) \right) Y_i(s)e^{\beta_0 Z_i \lambda_0(s)} ds.$$

By writing $X_n(\beta) - A_n(\beta)$ as a stochastic integral with respect to $M_i(\cdot)$, which has mean 0 and variance converging to 0, it follows that $X_n(\beta)$ and $A_n(\beta)$ have the same probability limit as $n \to \infty$. But $A_n(\beta)$ converges in probability to (Theorem 15.1)

$$A(\beta) = \int_{0}^{T} \left( (\beta - \beta_0) s^{(1)}(\beta_0, s) - \ln \left[ \frac{s^{(0)}(\beta, s)}{s^{(0)}(\beta_0, s)} \right] s^{(0)}(\beta_0, s) \right) \lambda_0(s) ds.$$

Thus, $X_n(\beta) \xrightarrow{P} A(\beta)$ as $n \to \infty$. It can also be shown that this convergence is uniform in $\beta$ over $N_0$.

Since $X_n(\beta)$ is maximized at $\hat{\beta}$, and $A(\beta)$ is maximized at the true value $\beta_0$ (Exercise 1), and $X_n(\beta)$ converges (uniformly) in probability to $A(\beta)$, it follows that $\hat{\beta} \xrightarrow{P} \beta$ (see e.g. Van der Vaart Theorem 5.7).
Asymptotic Normality of $\hat{\beta}$: The asymptotic normality of $\hat{\beta}$ follows from that of $n^{-\frac{1}{2}} U(\beta)$ using standard techniques: Expanding $U(\hat{\beta})$ about the true value $\beta$, we get

$$U(\hat{\beta}) = U(\beta) - \mathcal{I}(\beta_*) \cdot (\hat{\beta} - \beta),$$

where $|\beta - \beta_*| \leq |\beta - \hat{\beta}|$. Thus, since $U(\hat{\beta}) = 0$,

$$n^{-\frac{1}{2}} U(\beta) = \left( \frac{1}{n} \mathcal{I}(\beta_*) \right) \cdot \sqrt{n} (\hat{\beta} - \beta).$$

Since $\hat{\beta} \xrightarrow{P} \beta$ implies $\beta_* \xrightarrow{P} \beta$, it can be shown that $\frac{1}{n} \mathcal{I}(\beta_*) \xrightarrow{P} i(\beta)$.

Thus, since

$$n^{-\frac{1}{2}} U(\beta) \xrightarrow{L} \mathcal{N}(0, i(\beta)) \quad \text{as } n \to \infty,$$

it follows from Slutsky’s theorem that as $n \to \infty$,

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{L} \mathcal{N}(0, i^{-1}(\beta)).$$

Since $\hat{\mathcal{I}}(\beta)/n$ is a consistent estimator of $i(\beta)$, it also follows that
\[
\frac{\hat{\beta} - \beta}{\sqrt{\hat{I}^{-1}(\hat{\beta})}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad \text{as } n \to \infty.
\]

Thus, for scalar $Z$, the asymptotic normality of Cox’s partial likelihood score function and maximum partial likelihood estimator have been established. These results are the basis for approximate CIs and tests concerning the regression parameter $\beta$. 
Vector $Z$

These results extend to vector $Z$ in a straightforward way, although the notation gets a bit messy. Suppose we have $p$ covariates and we define

$$Z \overset{\text{def}}{=} (Z_1, \cdots, Z_p)^T \quad \text{and} \quad \beta \overset{\text{def}}{=} (\beta_1, \cdots, \beta_p).$$

Let $U(\beta)$ denote the $p \times 1$ score function having $j^{th}$ component

$$\frac{\partial \ln L_1(\beta)}{\partial \beta_j},$$

let $Z_i$ denote the value of $Z$ for subject $i$, and let $S^{(1)}(\beta, t)$ be the $p \times 1$ vector

$$\frac{1}{n} \sum_i Z_i Y_i(t) e^{\beta Z_i}. $$

Note that this is a vector because $Z_i$ is $p \times 1$. Also define $S^{(2)}(\beta, t)$ be the $p \times p$ matrix

$$\frac{1}{n} \sum_i Y_i(t) e^{\beta Z_i} Z_i \otimes Z_i,$$

where $Z \otimes Z \overset{\text{def}}{=} ZZ^T$. Then with the $p \times p$ matrix $V(\beta, t)$ defined as

$$V(\beta, t) = \frac{S^{(2)}(\beta, t)}{S^{(0)}(\beta, t)} - \left( \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)} \right)^{\otimes 2}$$

and $i(\beta)$ defined as the probability limit of

$$\int_0^\tau V(\beta, t) S^{(0)}(\beta, t) \lambda_0(t) dt,$$

similar arguments to those for scalar $Z$ can be used to show that
where “0” denotes the px1 vector of zeros. These results justify applying the usual kinds of asymptotic approximations for tests and CIs for the vector $\beta$ (or components of it) as we do for regular parametric likelihood theory.

Fleming and Harrington show how the results can be generalized to time-varying $Z$, more general assumptions about censoring, and other multiplicative intensity models.
Underlying Hazard Function.

It can also be shown that the estimated cumulative underlying hazard function (Breslow)

\[ \hat{\Lambda}_0(t) = \int_0^t \frac{\hat{N}(s)}{\sum_{i=1}^{n} Y_i(s) e^{\beta Z_i}} \]

is consistent and that

\[ \sqrt{n}(\hat{\Lambda}_0(\cdot) - \Lambda(\cdot)) \]

converges to a Gaussian process with zero mean and independent increments. The proof is a bit detailed, and outlined in Fleming and Harrington, section 8.3.
Exercises

1. Suppose that the covariates \( Z_i \) take values in a bounded space. Show that \( X_n(\beta) - A_n(\beta) \) on page 6 of Unit 21 converges in probability to 0. Hint 1: use the first formula on page 2 of Unit 21 (the formula for the Cox partial likelihood in terms of the \( Z_{(j)} \) and \( \tau_j \)) to express \( X_n(\beta) \) as some integral with respect to the \( N_i \) \( X_n(\beta) = \sum_{i=1}^n \int_0^{\tau_i} h_i(s) dN_i(s) \) for some \( h_i(s) \); what should you hope \( h_i \) to be in order to solve this problem?). Hint 2: the definition of \( S^{(0)} \) is also on page 2 of Unit 21. Hint 3: show that \( \int_0^t \ldots \) is a martingale, and therefore \( EM^2(t) = E < M > (t) \). What is its limit?

2. (33 points). In Unit 21 page 6 in the proof of consistency of \( \hat{\beta} \) in the Cox model, we needed that

\[
A(\beta) = \int_0^\tau \left( (\beta - \beta_0) s^{(1)}(\beta_0, s) - \ln \left( \frac{s^{(0)}(\beta, s)}{s^{(0)}(\beta_0, s)} \right) s^{(0)}(\beta_0, s) \right) \lambda_0(s) ds
\]

is maximized at the true value \( \beta_0 \). Show that this is true, including also that it is a maximum and not a minimum.

References