BIO 244: Unit 14

Example of Martingale CLT with Cox's Model and Summary of Part 3

In this unit we illustrate the Martingale Central Limit Theorem by applying it to the partial likelihood score function from Cox's model. For simplicity of presentation we assume that we have a scalar covariate Z; however, the same methods apply for vector Z.

Suppose that Z denotes a scalar covariate and assume that the hazard function for someone with covariate value Z is

$$h(t \mid Z) = \lambda_0(t)e^{\beta Z} \ .$$

The observations consist of n triplets (U_i, δ_i, Z_i), for $i = 1, 2, \ldots, n$, arising from i.i.d. (T_i, C_i, Z_i). As usual, we assume that censoring is noninformative $(T_i \perp C_i \mid Z_i)$.

Let

$$U_n = n^{-\frac{1}{2}} \frac{\partial lnL_1(\beta)}{\partial \beta},$$

where L_1 is Cox's partial likelihood. Recall that if L_1 were a 'real' likelihood function based on n i.i.d. observations, then under some regularity conditions we would have that

$$n^{-\frac{1}{2}} \frac{\partial lnL_1(\beta)}{\partial \beta} = \sqrt{n} \quad \left(\frac{1}{n} \partial lnL_1 / \partial \beta \right) \ \xrightarrow{\mathcal{L}} N(0, v) \ ,$$

since the bracketed term is an average of i.i.d. zero-mean random variables (say, with some variance v). This is why we 'standardize' $\partial lnL_1(\beta)/\partial \beta$ by $n^{-1/2}$.

Returning to Cox's partial likelihood, we indicated in Unit 12 that we can express U_n as

$$n^{1/2}U_n = \Sigma U_n(t)\ |_{t=\infty}\ ,$$

where

$$\Sigma U_n(t) = \sum_{i=1}^{n} \int_0^t \left(Z_i - \frac{\sum_{l=1}^{n} Z_l e^{\beta Z_l} Y_l(s)}{\sum_{l=1}^{n} e^{\beta Z_l} Y_l(s)} \right) dM_i(s),$$

and where $N_i(t) = 1[U_i \le t, \delta_i = 1]$, $Y_i(t) = 1(U_i \ge t)$, $M_i(t) = N_i(t) - A_i(t)$, and

$$A_i(t) = \int_0^t \lambda_0(s) e^{\beta Z_l} Y_i(s) ds.$$

Letting

$$H_i(s) \stackrel{\text{def}}{=} n^{-\frac{1}{2}} \left(Z_i - \frac{\sum Z_l e^{\beta Z_l} Y_l(s)}{\sum e^{\beta Z_l} Y_l(s)} \right),$$

we can thus write

$$\Sigma U_n(t) = n^{\frac{1}{2}} \sum_{i=1}^{n} \int_0^t H_i(s) dM_i(s).$$

Setting this equal to zero is the same as setting

$$\sum_{i=1}^{n} \int_0^t H_i(s) dM_i(s)$$

to zero.

Note that $A_i(\cdot)$ is continuous because T_i is continuous, so λ_0 is continuous. To apply the martingale CLT, we need $H_i(\cdot)$ to be predictable and locally bounded. It is clear that $H_i(\cdot)$ is predictable (it is left continuous and adapted). One way to ensure that it is locally bounded is to assume that Z is bounded (Exercise 8 of Unit 12). Also, because the survival times $T_1, T_2, \ldots, T_n$ for the n subjects are assumed to be independent and continuously distributed, they cannot jump at the same time, and so the $M_i(\cdot)$ are orthogonal. Thus, the setting of the martingale central limit theorem, (13.3)-(13.5), has been established.

We now evaluate conditions (a) and (b) of the Martingale CLT.

(a) Does $< \Sigma U_n, \Sigma U_n > (t) \xrightarrow{P} \alpha(t)$ (some deterministic function) as $n \to \infty$?

Because $A_i(\cdot)$ is continuous and M_i and M_j are orthogonal for $i \neq j$,

$$\begin{aligned}
< \Sigma U_n, \Sigma U_n > (t) &= \sum_{i=1}^{n} \sum_{j=1}^{n} \int_0^t H_i(s) H_j(s) d < M_i, M_j > (s) \\
&= \sum_{i=1}^{n} \int_0^t H_i^2(s) dA_i(s) \\
&= \sum_{i=1}^{n} \frac{1}{n} \int_0^t (Z_i - Q_n(s))^2 \lambda_0(s) \; e^{\beta Z_i} Y_i(s) ds \\
&= \int_0^t \lambda_0(s) \left(\frac{1}{n} \sum_{i=1}^{n} (Z_i - Q_n(s))^2 e^{\beta Z_i} Y_i(s) \right) ds \ ,
\end{aligned}$$

where

$$Q_n(s) = \frac{\frac{1}{n} \sum_{l=1}^{n} Z_l e^{\beta Z_l} Y_l(s)}{\frac{1}{n} \sum_{l=1}^{n} e^{\beta Z_l} Y_l(s)} \ .$$

Consider the term in large brackets for fixed s. This has the same probability limit as

$$\frac{1}{n} \sum_{i=1}^{n} \left\{ (Z_i - \mu(s))^2 e^{\beta Z_i} Y_i(s) \right\}, \tag{14.1}$$

where

$$\mu(s) = \frac{E\left(Z_l e^{\beta Z_l} Y_l(s)\right)}{E\left(e^{\beta Z_l} Y_l(s)\right)}.$$

To see this, note that their difference equals

$$\frac{1}{n} \sum_{i=1}^{n} (Z_i - Q_n(s))^2 e^{\beta Z_i} Y_i(s) \quad - \quad \frac{1}{n} \sum_{i=1}^{n} (Z_i - \mu(s))^2 e^{\beta Z_i} Y_i(s)$$

$$= \frac{1}{n} \sum_{i=1}^{n} Y_i(s) e^{\beta Z_i} \left(Q_n^2(s) - \mu^2(s) + 2 Z_i (\mu(s) - Q_n(s)) \right)$$

$$= \left(Q_n^2(s) - \mu^2(s)\right) \frac{1}{n} \sum_{i=1}^{n} Y_i(s) e^{\beta Z_i} - 2\left(Q_n(s) - \mu(s)\right) \frac{1}{n} \sum_{i=1}^{n} Y_i(s) e^{\beta Z_i} Z_i \,. \quad (14.2)$$

By the Law of Large Numbers,

$$\frac{1}{n} \sum_{i=1}^{n} Y_i(s) e^{\beta Z_i} \xrightarrow{P} E(Y_i(s) e^{\beta Z_i})$$

and

$$\frac{1}{n} \sum_{i=1}^{n} Y_i(s) e^{\beta Z_i} Z_i \xrightarrow{P} E(Y_i(s) e^{\beta Z_i} Z_i).$$

Since $Q_n(s)$ converges in probability to $\mu(s)$ (Law of Large Numbers and Slutzky), $Q_n^2(s) - \mu^2(s)$ converges in probability to zero (Continuous Mapping Theorem), and hence (14.2) converges to zero (repeatedly apply Slutzky's lemma).

Now let's consider (14.1) for any arbitrary s. Since it is an average of i.i.d. random variables, it converges in probability to

$$m(s) \stackrel{\text{def}}{=} E\left((Z_i - \mu(s))^2 e^{\beta Z_i} Y_i(s) \right).$$

It follows from this that

$$< \Sigma U_n, \Sigma U_n > (t) = \int_0^t \lambda_0(s) \left(\frac{1}{n} \sum_{i=1}^{n} (Z_i - Q_n(s))^2 e^{\beta Z_i} Y_i(s) \right) ds$$

$$\xrightarrow{P} \int_0^t \lambda_0(s) m(s) ds. \quad (14.3)$$

That is, if the limit of the integral is the integral of the limit (we will show this more rigorously in Unit 15). Thus, in the notation of the Martingale CLT,

$$\alpha(t) = \int_0^t \lambda_0(s) m(s) ds.$$

Now consider condition (b) of the Martingale Central Limit Theorem:

Does $< \Sigma U_{n,\epsilon}, \Sigma U_{n,\epsilon} > (t) \xrightarrow{P} 0$ for all $\epsilon > 0$?
We have

$$< \Sigma U_{n,\epsilon} \ , \Sigma U_{n,\epsilon} > (t) = \sum_{i=1}^{n} \int_0^t H_i^2(s) \cdot 1[\mid H_i(s) \mid \geq \epsilon] dA_i(s) =$$

$$= \int_0^t \lambda_0(s) \{ \frac{1}{n} \sum_{i=1}^{n} \left(Z_i - \frac{\frac{1}{n}\sum_{j=1}^n Z_j(s) e^{\beta Z_j} Y_j(s)}{\frac{1}{n}\sum_{j=1}^n e^{\beta Z_j} Y_j(s)} \right)^2 \cdot 1 \left[\mid Z_i - \frac{\frac{1}{n}\sum_{j=1}^n Z_j e^{\beta Z_j} Y_j(s)}{\frac{1}{n}\sum_{j=1}^n e^{\beta Z_j} Y_j(s)} \mid \geq \sqrt{n}\epsilon \right] e^{\beta Z_i} Y_i(s) \} ds.$$

Note that the limit, in probability, of

$$1 \left[\mid Z_i - \frac{\frac{1}{n}\sum_{j=1}^n Z_j e^{\beta Z_j} Y_j(s)}{\frac{1}{n}\sum_{j=1}^n e^{\beta Z_j} Y_j(s)} \mid \ \geq \ \sqrt{n} \ \epsilon \right]$$

is zero because the term in absolute value is bounded (still assume Z is bounded) whereas $\sqrt{n}\epsilon \to \infty$ (see Exercises). It follows that the entire integrand in the above integral converges to zero in probability, and hence that $< \Sigma U_{n,\epsilon}, \Sigma U_{n,\epsilon} > (t) \xrightarrow{P} 0$ for every t and ϵ as $n \to \infty$ (implicit in this statement is that the limit of the integral equals the integral of the limit. This requires some assumptions that we will return to in Unit 15).

Thus, the conditions of Martingale CLT hold, so that

$$\Sigma U_n(\cdot) \xrightarrow{w} Q(\cdot),$$

as $n \to \infty$, where $Q(\cdot)$ is a zero-mean Gaussian process with independent increments and var $(Q(t)) = \alpha(t)$. By taking $t = \infty$, it follows that

$$U_n = \Sigma U_n(\infty) \xrightarrow{\mathcal{L}} N(0, \sigma^2) \qquad \text{as } n \to \infty \ ,$$

where $\sigma^2 = \alpha(\infty) = \int_0^\infty \lambda_0(s) m(s) ds.$

Note: not clear why we can evaluate these processes at infinity. Usually, one assumes that the support of U does not extend beyond some time τ, so that the process evaluated at time τ is same as at infinity.

The expression (14.3) for $\alpha(t)$ cannot be simplified much more without being more specific about Z and the censoring distribution. For example, consider the 2-sample problem, where

$$Z_i = \begin{cases} 0 & \text{for group } 0 \\ 1 & \text{for group } 1 \end{cases}$$

and suppose that $P(Z_i = 0) = \frac{1}{2}$, and that the distribution of C_i does not depend on (T_i, Z_i), and has c.d.f. $G(\cdot)$. Then it is not difficult to show that under $H_0 : \beta = 0$,

$$E\left((Z_l e^{\beta Z_l} Y_l(s)\right) = \frac{1}{2}(1 - F_0(s))(1 - G(s))$$

and

$$E\left(e^{\beta Z_l} Y_l(s)\right) = (1 - F_0(s))(1 - G(s)),$$

where

$$F_0(t) = 1 - e^{-\int_0^t \lambda_0(s)ds}.$$

Thus, $\mu(s) = 1/2$. It can also be easily shown that, under H_0,

$$\begin{aligned} m(s) &= E\left((Z_i - \tfrac{1}{2})^2 Y_i(s)\right) \\ &= \tfrac{1}{4}(1 - F_0(s))(1 - G(s)). \end{aligned}$$

Thus

$$\begin{aligned} \alpha(t) &= \tfrac{1}{4}\int_0^t \lambda_0(s)(1 - F_0(s))(1 - G(s))ds \\ &= \tfrac{1}{4}\int_0^t f_0(s)(1 - G(s))ds, \end{aligned}$$

where $f_0(s) = \lambda_0(s)(1 - F_0(s))$.

When $t = \infty$,

$$\sigma^2 = \alpha(\infty) = \frac{1}{4}\int_0^\infty f_0(s)(1 - G(s))ds = \frac{p}{4},$$

where $p = Pr(\delta_i = 1)$.

Hence we have shown that under $H_0 : \beta = 0$,

$$U_n = n^{-1/2} \partial ln L_1(\beta)/\partial\beta$$

converges in distribution to a normal distribution with mean zero and variance σ^2. Therefore, the partial likelihood score test of H_0, obtained by evaluating this at $\beta = 0$ and standardizing it by σ, is asymptotically $N(0, 1)$ under H_0. We later use very similar techniques to find the non-null distribution of Cox's partial likelihood score test when $\beta \neq 0$.

Note: this result shows that logrank test is asymptotically N(0,1) under the null: recall that it can be viewed as arising from a Cox proportional hazards model as a score test!

Brief Review of Some Key Martingale Theory Results

- $X(\cdot)$ is a martingale if it is adapted to $\mathcal{F}_t$ and:

$$\begin{cases} \cdot \quad X(\cdot) \text{ is right continuous with left-hand limits} \\ \cdot \quad E|X(t)| < \infty \quad\quad \forall t \\ \cdot \quad E[X(t+s)|\mathcal{F}_t] \overset{a.s.}{=} X(t) \quad\quad \forall s \geq 0, t \geq 0. \end{cases}$$

- $X(\cdot)$ is a sub-martingale if the '=' is replaced by '$\geq$'.
- If $X(\cdot)$ is a martingale, then $E(X(t))$ is constant in t; without loss of generality we can take $E(X(t)) = 0$.
- $N(\cdot)$ is a counting process if

$$\begin{cases} \cdot \quad N(0) = 0, \quad N(t) < \infty \\ \cdot \quad N(\cdot) \text{ is right-continuous} \\ \cdot \quad N(\cdot) \text{ is a step function with jumps of size } +1. \end{cases}$$

- $H(\cdot)$ is a predictable process if its value at t determined by $\mathcal{F}_{t-}$ (e.g., left continuous and adapted).
- Doob-Meyer decomposition: $X(\cdot)$ = non-negative sub-martingale $\Rightarrow \exists$ right-continuous, $\nearrow$, predictable process $A(\cdot)$ s.t. $E(A(t)) < \infty \;\; \forall t$, and $M(\cdot) = X(\cdot) - A(\cdot)$ is a martingale. $A(\cdot)$ is called the compensator for $X(\cdot)$. If $A(0) \overset{a.s.}{=} 0$, $A(\cdot)$ is almost surely unique.
- Any counting process $N(\cdot)$ such that $E(N(t)) < \infty$ is a sub-martingale, $\Rightarrow \exists\; A(\cdot)$ such that $N(\cdot) - A(\cdot)$ is a martingale.
- If $M(\cdot)$ is any martingale and $EM^2(t) < \infty, M^2(\cdot)$ is a sub-martingale $\Rightarrow \exists\;!$ predictable process $< M, M > (\cdot)$ such that $M^2(\cdot) - < M,\; M > (\cdot)$ is a martingale. The process $< M, M > (\cdot)$ is called the predictable quadratic variation.

Note: If $E(M(t)) = 0$, var $(M(t)) = E(M^2(t)) = E(< M, M > (t))$.

- If the zero mean martingale $M = N - A$ satisfies $E(M^2(t)) < \infty$ and $A(\cdot)$ is <u>continuous</u>, then $< M, M > (\cdot) \overset{as}{=} A(\cdot)$. Thus, $Var(M(t)) = E(A(t))$.

Next suppose that $M_1(\cdot), M_2(\cdot), \cdots$ are zero-mean martingales defined on the same filtration. Then

- $a\ M_1(\cdot) + bM_2(\cdot)$ is a martingale (for any constants a, b).
- If $E(M_j^2(t)) < \infty$ for j = 1, 2, $\exists$ a right-continuous predictable process $< M_1, M_2 > (\cdot)$ such that $< M_1, M_2 > (0) = 0$, $E\,|< M_1, M_2 > (t)| < \infty$, and $M_1(\cdot) \cdot M_2(\cdot) - < M_1, M_2 > (\cdot)$ is a martingale.
- If $< M_1, M_2 > \overset{as}{=} 0$, $\qquad M_1(\cdot)$ and $M_2(\cdot)$ are called orthogonal.
- If $M_j = N_j - A_j$, where N_j is a counting process and A_j is its continuous compensator, then if N_i and N_j cannot jump at the same time, $< M_i, M_j > (\cdot) \overset{as}{=} 0$.

Next suppose that $N(\cdot)$ is a bounded counting process, that $A(\cdot)$ is the compensator for $N(\cdot)$, and $E(N - A)^2(t) < \infty$ for every t. Moreover, $H(\cdot)$ is a bounded, predictable process. Define $Q(\cdot)$ by

$$Q(t) = \int_0^t H(s)dM(s) \qquad \text{where } M(\cdot) = N(\cdot) - A(\cdot).$$

Then

- $Q(\cdot)$ is a zero-mean martingale.
- $Var(Q(t)) = E(Q^2(t))$.
- $Q^2(\cdot)$ is a sub-martingale, and thus $\exists$ predictable $< Q, Q > (\cdot)$ such that $Q^2(\cdot) - < Q, Q > (\cdot)$ is a martingale. Hence, var $(Q(t)) = E(< Q, Q > (t))$.
- If $A(\cdot)$ is continuous, $< Q, Q > (t) \overset{a.s.}{=} \int_0^t H^2(s)dA(s)$, and thus $Var(Q(t)) = E(\int_0^t H^2(s)dA(s))$.
- If M_1, M_2 are 2 martingales defined similarly to $M(\cdot)$, H_1 and H_2 are bounded predictable processes, and $Q_1(\cdot)$ and $Q_2(\cdot)$ are defined in the same way as $Q(\cdot)$, the predictable quadratic covariance process $< Q_1, Q_2 >$ satisfies $< Q_1, Q_2 > (t) = \int_0^t H_1(s)\ H_2(\cdot)d < M_1, M_2 > (s)$.

Martingale CLT:
For i=1,2,...,n, let

$$\begin{aligned} N_{in}(\cdot) &= \text{ counting processes that can't jump at the same time} \\ A_{in}(\cdot) &= \text{ continuous compensators} \\ H_{in}(\cdot) &= \text{ locally bounded predictable processes} \end{aligned}$$

Define

$$\begin{aligned} M_{in(\cdot)} &= N_{in}(\cdot) - A_{in}(\cdot) \\ U_{in}(t) &= \int_0^t H_{in}(s)dM_{in}(s) \\ H^*_{in}(s) &= H_{in}(s)1[|H_{in}(s)| \geq \epsilon] \\ U_{in,\epsilon}(t) &= \int_0^t H^*_{in}(s)dM_{in}(s) \\ \Sigma U_n(\cdot) &= \sum_i U_{in}(\cdot), and \\ \Sigma U_{in,\epsilon}(\cdot) &= \sum_i U_{in,\epsilon}(\cdot). \end{aligned}$$

Suppose that

(a) $< \Sigma U_n, \Sigma U_n > (t) \xrightarrow{P} \alpha(t)$ for all t, for some function $\alpha(\cdot)$.
(b) $< \Sigma U_{n,\epsilon}, \Sigma U_{n,\epsilon} >$(t) $\xrightarrow{P} 0$ for all t and all $\epsilon > 0$.

Then

$$\Sigma U_n(\cdot) \xrightarrow{\mathcal{L}} U(\cdot),$$

where $U(\cdot)$ is a zero-mean Gaussian process with independent increments and variance function $Var(U(t)) = \alpha(t)$.

Note: By the definitions of $N_{in}(\cdot)$, $A_{in}(\cdot)$, and $H_{in}(\cdot)$:

$$\begin{aligned} < \Sigma U_n, \Sigma U_n > (t) &= \sum_{i=1}^n \int_0^t H^2_{in}(s)dA_{in}(s) \\ < \Sigma U_{n,\epsilon}, \Sigma U_{n,\epsilon} > (t) &= \sum_{i=1}^n \int_0^t H^{*2}_{in}(s)dA_{in}(s), \end{aligned}$$

where $H^*_{in}(s) = H_{in}(s) \cdot 1[|H_{in}(s)| \geq \epsilon]$.

Intuition: While the proof of the Martingale CLT is rather detailed, the result is not surprising. One forms a sum, $U_n(\cdot) = \sum U_{in}(\cdot)$, of n orthogonal martingales, where each has zero mean, and where the variance of the sum converges in probability to some constant function $\alpha(\cdot)$. In the classical CLT we are used to seeing a sum of i.i.d. random variables multiplied by $n^{-1/2}$. Thus, the variance of the standardized sum is (and converges to) $\frac{1}{n}n\sigma^2 = \sigma^2$, where σ^2 is the variance of each term in the sum. In the way we have set up the Martingale CLT, the multiplier $n^{-1/2}$ is buried in the integrands, $H_{in}(\cdot)$ of the martingales Q_{in}, so the analogy is really the same as in the classical CLT. If we consider the asymptotic distribution of $U_n(t)$ for some fixed t, it is not surprising that the result is a normal distribution with mean 0. From this one would expect the finite-dimensional distributions of $U_n(\cdot)$ to converge to zero mean multivariate normal distributions. Thus, if tightness holds, we would expect $U_n(\cdot)$ to converge to a zero mean Gaussian process. The independent increments of this limiting process are assured from the uncorrelated increment property of the individual martingales $U_{in}(\cdot)$.

Exercises

1. Prove that (see page 5)

$$Q_n(s) \xrightarrow{P} 0 \qquad \text{as} \quad n \to \infty,$$

where

$$Q_n(s) = 1\left\{\left(Z_i - \frac{\frac{1}{n}\sum_{j=1}^{n} Z_j e^{\beta Z_j} Y_j(s)}{\frac{1}{n}\sum_{j=1}^{n} e^{\beta Z_j} Y_j(s)}\right) \geq \sqrt{n}\epsilon\right\}.$$

2. Prove the expressions for μ, m and σ^2 on page 6. Do not take any of the expressions on page 6 for granted.