STAT331

1-Sample Problem: General Asymptotic Results

In this unit we will consider the 1-sample problem and prove the consistency and asymptotic normality of the Nelson-Aalen estimator of the cumulative hazard function and the Kaplan-Meier estimator of the survival function. This will justify the approximate tests and pointwise confidence intervals noted when we introduced these estimators in Unit 5. We will use these results to find confidence bands for the integrated hazard and survival functions. In Unit 16 we will use these results to construct confidence intervals for quantiles (e.g., the median) of the survival function.

Setting: We assume the usual 1-sample setting. We have n i.i.d. continuous survival time random variables $T_1, T_2, \cdots, T_n$, with common c.d.f. $F(\cdot)$, and denote the corresponding hazard, cumulative hazard, and survival functions by $\lambda(\cdot)$, $\Lambda(\cdot)$, and $S(\cdot)$.

Further, we assume that the potential censoring times $C_1, C_2, \ldots, C_n$ are i.i.d. with common c.d.f. G, and are independent from $T_1, T_2, \cdots, T_n$, so that censoring occurs noninformatively.

We observe (U_i, δ_i), $i = 1, 2, \ldots, n$, where

$$\begin{aligned} U_i &= min(T_i, C_i) \\ \delta_i &= 1(T_i \leq C_i) \ . \end{aligned}$$

As before, we define the counting and at-risk processes, $N_i(\cdot)$ and $Y_i(\cdot)$, for subject i, as well as the corresponding compensator and martingale processes $A_i(\cdot)$ and $M_i(\cdot)$. Also, denote the sums of these processes over the n subjects with Σ; that is,

$$\Sigma N(\cdot) = \sum_i N_i(\cdot), \quad \Sigma Y(\cdot) = \sum_i Y_i(\cdot), \quad \Sigma A(\cdot) = \sum_i A_i(\cdot),$$

$$\text{and} \qquad \Sigma M(\cdot) = \sum_i M_i(\cdot).$$

Before beginning, we introduce results 15.1 and 15.2 that we will utilize in this and subsequent units. 15.1 is a special case of two more general propositions, 15.1b and 15.1c, which can be found in Andersen, Borgan, Gill, & Keiding (1993, page 85). 15.1 can be used to justify our assumption in Unit 14, when discussing the asymptotic distribution of Cox's partial likelihood score function, that the limit of an integral was the integral of the limit.

Theorem 15.1: Consider a sequence, $X_n(\cdot)$, of stochastic processes for which $\mid X_n(t) \mid \overset{a.s.}{\leq} c$ for all n and all $t \in [0, \tau]$, for some τ and constant c. Then if

$$X_n(t) \xrightarrow{P} h(t) \qquad \text{as } n \to \infty$$

for every $t \in [0, \tau]$, it follows that

$$E\left(\int_0^\tau X_n(t)dt\right) \to \int_0^\tau h(t)dt$$

and

$$\int_0^\tau X_n(t)dt \xrightarrow{P} \int_0^\tau h(t)dt.$$

In some cases, we will need stronger theorems, see Andersen et al. page 85:

Theorem 15.1b: Consider a sequence $X_n(\cdot)$ of stochastic processes and such that for all $\delta > 0$ there exists k_δ with $\int_0^\tau k_\delta < \infty$ such that

$$\liminf_{n\to\infty} P\left(|X_n(t)| \leq k_\delta(t) \text{ for all } t\right) \geq 1 - \delta.$$

Suppose that

$$X_n(t) \xrightarrow{P} h(t) \qquad \text{as } n \to \infty$$

for every $t \in [0, \tau]$ and suppose that

$$\int_0^\tau |h(t)| dt < \infty.$$

Then it follows that

$$\sup_{t\in[0,\tau]} \left| \int_0^t X_n(s) ds - \int_0^t h(s) ds \right| \xrightarrow{P} 0.$$

Theorem 15.1c: Consider a sequence $X_n(\cdot)$ of stochastic processes such that

$$\lim_{C\uparrow\infty} \sup_n E\left(|X_n(t)| \, 1_{\{|X_n(t)|>C\}}\right) = 0$$

for all t (i.e., the sequence $X_n(t)$ is uniformly integrable) and

$$E\,|X_n(t)| \leq k(t)$$

for all t, n, with $\int_0^\tau k(t)dt < \infty$. Suppose moreover that

$$X_n(t) \xrightarrow{p} h(t) \qquad \text{as } n \to \infty$$

for every $t \in [0, \tau]$. Then it follows that

$$E\left(\sup_{t\in[0,\tau]} \left| \int_0^t X_n(s) ds - \int_0^t h(s) ds \right| \right) \xrightarrow{P} 0.$$

Theorem 15.2 (Corollary of Lenglart's Inequality). Let $N(\cdot)$ be a counting process, $A(\cdot)$ denote its continuous compensator, and $M(\cdot) = N(\cdot) - A(\cdot)$ be the corresponding martingale, and assume that $E(M^2(t)) < \infty$ for every t. Suppose that $H(\cdot)$ is a predictable and locally bounded process defined on the same filtration. Then for any stopping time τ, and any $\epsilon > 0$ and $\eta > 0$,

$$P\left(\sup_{t\in[0,\tau]}\left\{\int_0^t H(u)dM(u)\right\}^2 \geq \epsilon\right) < \frac{\eta}{\epsilon} + P\left(\int_0^\tau H^2(t)dA(t) \geq \eta\right).$$

For a proof, see Fleming and Harrington, page 113. One use of this theorem is to prove uniform convergence in probability from pointwise convergence. For example, we will invoke this to prove that $Q(\cdot) \xrightarrow{p} 0$ (see equation (15.1) below) in the proof that the Nelson-Aalen estimator is (uniformly) consistent for the cumulative hazard function.

Large-Sample Behavior of $\hat{\Lambda}(\cdot)$:
We can write the Nelson-Aalen estimator, $\widehat{\Lambda}(\cdot)$ at time t as

$$\widehat{\Lambda}(t) = \int_0^t \frac{1[\Sigma Y(s) > 0]}{\Sigma Y(s)} \, d\Sigma N(s).$$

Note that

- The $M_i(\cdot)$ and $\Sigma M(\cdot)$ are zero-mean martingales
- In the 'old' notation,

$$\widehat{\Lambda}(t) = \sum_{\tau_j \leq t} \frac{d_j}{Y(\tau_j)} \ .$$

  Since we are assuming the T_i are continuous, $d_j = 1$.

- The term '$1[\Sigma Y(s) > 0]$' in the stochastic integral representation of $\widehat{\Lambda}(t)$ seems redundant. However, as is the convention in Fleming & Harrington and Andersen et al., by using this and defining $\frac{0}{0} = 0$, we don't need to worry about the integrand when $\Sigma Y(s) = 0$.

Let $H(s) = \frac{1[\Sigma Y(s)>0]}{\Sigma Y(s)}$ (note: left continuous and adapted, so predictable), so that

$$\widehat{\Lambda}(t) = \int_0^t H(s) d\Sigma N(s),$$

and define the martingale $Q(\cdot)$ by

$$Q(t) = \int_0^t H(s) \, d\Sigma M(s) = \int_0^t H(s) \, d(\Sigma N - \Sigma A)(s).$$

Thus we can write

$$\begin{aligned}
Q(t) &= \widehat{\Lambda}(t) - \int_0^t H(s) d\Sigma A(s) \\
&= \widehat{\Lambda}(t) - \int_0^t H(s)\ \lambda(s)\ \Sigma Y(s)\ ds \\
&= \widehat{\Lambda}(t) - \int_0^t 1[\Sigma Y(s) > 0]\ \lambda(s)\ ds \\
&= \widehat{\Lambda}(t) - \Lambda(t) + \int_0^t \lambda(s)\ \ (1 - 1[\Sigma Y(s) > 0])\ \ ds \\
&= \widehat{\Lambda}(t) - \Lambda(t) + D(t), \qquad (15.1)
\end{aligned}$$

where

$$D(t) \stackrel{def}{=} \int_0^t \lambda(s)\ \ (1 - 1[\Sigma Y(s) > 0])\ \ ds.$$

Since $E(Q(t)) = 0$ and $D(t) \stackrel{a.s.}{\geq} 0$, we have that

$$E[\widehat{\Lambda}(t) - \Lambda(t)] \leq 0$$

or

$$E[\widehat{\Lambda}(t)] \leq \Lambda(t).$$

Thus, the estimator may underestimate $\Lambda(\cdot)$ somewhat.

Let's examine $D(t)$ as $n \to \infty$. First note that the bracketed term in the integrand of D(t) is just $1[SumY(s) = 0]$. We will now show that this converges to zero in probability.

To see this, suppose $\epsilon \in$(0,1). Then

$$P[1[\Sigma Y(s) = 0] > \epsilon] \ = \ P[\Sigma Y(s) = 0]$$

$$
\begin{aligned}
&= P[Y_1(s)=0, Y_2(s)=0, \cdots, Y_n(s)=0] \\
&= \Pi_{i=1}^n P(Y_i(s)=0) = \Pi_{i=1}^n (1-P(Y_i(s)=1)) \\
&= \Pi_{i=1}^n \{1-(1-F(s))(1-G(s))\} \\
&= \{1-(1-F(s))(1-G(s))\}^n.
\end{aligned}
$$

In what follows, we will assume that τ lies in the interior of the support of $F(\cdot)$ and $G(\cdot)$; that is, that $1-F(\tau)>0$ and $1-G(\tau)>0$, and hence that

$$1-F(t)>0 \quad \text{and} \quad 1-G(t)>0 \quad \text{for } 0 \le t \le \tau. \tag{15.2}$$

Then the bracketed term above is strictly less than one, and hence the expression goes to zero as $n \to \infty$. This proves that the integrand in D(t) converges in probability to zero. It follows from Theorem 15.1 (λ is continuous on the compact interval $[0,\tau]$, so bounded there) that

$$E(D(t)) \to 0$$

and

$$D(t) \xrightarrow{P} 0.$$

Thus, from (15.1) we have that $E\widehat{\Lambda}(t) \to \Lambda(t)$ as $n \to \infty$.

Also,

$$
\begin{aligned}
\text{Var }(Q(t)) &= E\int_0^t H^2(s) d\Sigma A(s) \\
&= E\int_0^t \frac{1[\Sigma Y(s)>0]\lambda(s)}{\Sigma Y(s)} ds.
\end{aligned}
$$

Equation (15.2) ensures that $\Sigma Y(s) \xrightarrow{p} \infty$, so that the integrand above converges in probability to zero. Thus, applying Theorem 15.1 again, we have that as $n \to \infty$

$$\text{var}\ (Q(t)) = E \int_0^t \frac{1[\Sigma Y(s) > 0]\lambda(s)}{\Sigma Y(s)} ds \to 0.$$

Since $E(Q(t)) = 0$ (it is a martingale 0 at time 0), it follows that $Q(t) \xrightarrow{P} 0$ (Exercise 4). We already had that $D(t) \xrightarrow{P} 0$, so (15.1) guarentees that

$$\widehat{\Lambda}(t) \xrightarrow{P} \Lambda(t).$$

Thus, <u>$\widehat{\Lambda}(t)$ is consistent</u> provided $1 - F(t) > 0$ and $1 - G(t) > 0$.

Note that if $1 - G(t) = 0$ or $1 - F(t) = 0$ for some t, we could never observe a U_i greater than t. Thus, $\widehat{\Lambda}(t)$ would not be estimable, even as $n \to \infty$. This is the reason for (15.2).

The above result shows the pointwise consistency of $\Lambda(t)$. However, since $D(s)$ is non-negative and nondecreasing for $0 \leq t \leq \tau$,

$$sup_{0 \leq t \leq \tau} \mid D(t) \mid = D(\tau)\ ,$$

so that

$$P(sup_{0 \leq t \leq \tau} \mid D(t) \mid > \epsilon) = P(D(\tau) > \epsilon) \to 0.$$

Thus $D(\cdot) \xrightarrow{P} 0$ over $[0, \tau]$. Also, from Theorem 15.2, $Q(\cdot) \xrightarrow{P} 0$ (Exercise 5). Hence from Equation (15.1), <u>$\hat{\Lambda}(\cdot)$ converges uniformly</u>; i.e.,

$$\sup_{0 \leq t \leq \tau} \mid \widehat{\Lambda}(t) - \Lambda(t) \mid \leq \sup_{0 \leq t \leq \tau} \mid Q(t) \mid + \sup_{0 \leq t \leq \tau} \mid D(t) \mid \xrightarrow{P} 0$$

because of Slutzky's lemma for random variables.

Now, consider the asymptotic distribution of the process

$$\sqrt{n}\left(\widehat{\Lambda}(\cdot) - \Lambda(\cdot)\right).$$

Using (15.1), we have

$$\sqrt{n}\left(\widehat{\Lambda}(t) - \Lambda(t)\right) = \int_0^t \frac{\sqrt{n}\ 1[\Sigma Y(s) > 0]}{\Sigma Y(s)}\, d\Sigma M(s) - \sqrt{n}D(t).$$

It is easily shown (see Exercise 2) that $\sqrt{n}D(t) \xrightarrow{P} 0$, and thus that $\sqrt{n}D(\cdot) \xrightarrow{P} 0$ on $[0, \tau]$. Thus, by Slutsky's theorem for stochastic processes, $\sqrt{n}(\widehat{\Lambda}(\cdot) - \Lambda(\cdot))$ converges weakly on $[0, \tau]$ to the same limit as

$$\int H^*(s) d\Sigma M(s),$$

where

$$H^*(s) = \sqrt{n}\ \frac{1[\Sigma Y(s) > 0]}{\Sigma Y(s)}.$$

Conditions (a) and (b) of the Martingale CLT, when applied to $\int H^*(s) d\Sigma M(s)$, are satisfied (Exercise 3). Thus, as $n \to \infty$,

$$\sqrt{n}\left(\widehat{\Lambda}(\cdot) - \Lambda(\cdot)\right) \xrightarrow{w} X(\cdot)\ ,$$

where $X(\cdot)$ is a zero-mean Gaussian process with independent increments and

$$\text{var}\ (X(t)) = \sigma^2(t),$$

where $\sigma^2(t)$ is the probability limit of $\int_0^t [H^*(s)]^2 d\Sigma A(s)$. This can be shown to equal

$$\sigma^2(t) = \int_0^t \frac{\lambda(s)}{(1 - F(s))(1 - G(s))} ds.$$

Since (1-F(s))(1-G(s))=P(U≥s) is estimated by $\Sigma Y(s)/n$, we can estimate $\sigma^2(t)$ by

$$\begin{aligned}\widehat{\sigma}^2(t) &= \int_0^t \frac{1[\Sigma Y(s)>0]}{\Sigma Y(s)/n} d\widehat{\Lambda}(s) \\ &= n\int_0^t \frac{1[\Sigma Y(s)>0]}{\Sigma Y(s)^2} d\Sigma N(s) \qquad \left(\begin{array}{l}\text{using} \\ d\hat{\Lambda}(s) = \frac{1[\Sigma Y(s)>0]}{\Sigma Y(s)} d\Sigma N(s)\end{array}\right).\end{aligned}$$

Using the 'old' notation, this is just

$$\widehat{\sigma}^2(t) = n\sum_{\tau_j \le t} \frac{d_j}{Y(\tau_j)^2}\ , \tag{15.3}$$

where d_j and $Y(\tau_j)$ are the numbers of failures and subjects at risk at time τ_j, respectively. It can be shown (but not really easily, Exercise 7) that the consistency of this estimator for $\sigma^2(t)$ follows from Theorems 15.1b and 15.1c.

Comment: These results can be used to make inferences about $\Lambda(t)$. Specifically,

- An approximate 95% (pointwise) CI for $\Lambda(t)$ is given by

  $$\widehat{\Lambda}(t) \pm 1.96\ \widehat{\sigma}(t)/\sqrt{n}.$$

  This is based on taking

  $$\widehat{\Lambda}(t) \approx \mathcal{N}\left(\Lambda(t),\ \sigma^2(t)/n\right).$$

- An approximate 95% confidence <u>band</u> for

  $$\Lambda(t),\ 0 \le t \le \tau$$

can be obtained by first choosing c such that

$$P\left(\sup_{0\le t\le\tau} \mid X(t) \mid \le c\right) = .95,$$

where $X(\cdot)$ is a zero-mean Gaussian process with independent increments and variance $\sigma^2(t)$. Then, because of the Continuous Mapping Theorem (recall that taking the supremum is continuous with respect to the Skorohod topology),

$$\sup_{t\in[0,\tau]} \left|\sqrt{n}\left(\hat{\Lambda}(t) - \Lambda(t)\right)\right| \xrightarrow{\mathcal{L}} \sup_{t\in[0,\tau]} |X(t)|\,.$$

Hence

$$\begin{aligned} .95 &\approx P[\sup_{0\le t\le\tau} \mid \sqrt{n}[\widehat{\Lambda}(t) - \Lambda(t)] \mid \le c] \\ &= P\left(\sqrt{n} \mid \widehat{\Lambda}(t) - \Lambda(t) \mid \le c, \quad 0 \le t \le \tau\right) \\ &= P\left(\widehat{\Lambda}(t) - \frac{c}{\sqrt{n}} \le \Lambda(t) \le \widehat{\Lambda}(t) + \frac{c}{\sqrt{n}}, \quad 0 \le t \le \tau\right). \end{aligned}$$

These bounds will include the function $(\Lambda(t),\ 0 \le t \le \tau)$ with probability of approximately 0.95.

Comment: One problem with the above is finding (computing) the desired c. Another is that $\widehat{\Lambda}(t) - c/\sqrt{n}$ can be negative, which is undesirable. Similar concerns arise when finding confidence bands for $S(\cdot)$. We will return to these points later.

Large-Sample Behavior of $\tilde{S}(\cdot)$ and $\hat{S}(\cdot)$:

Next consider estimation of the survival function $S(t) = P(T \geq t)$. Since $S(t) = e^{-\Lambda(t)}$, one possible estimate for $S(t)$ is

$$\tilde{S}(t) = e^{-\hat{\Lambda}(t)}.$$

As we saw in Unit 5, this is algebraically similar to the Kaplan-Meier estimator of $S(t)$. We'll focus on the Kaplan-Meier estimator below, but first examine the asymptotic properties of $\tilde{S}(\cdot)$ because (1) they derive easily from those of $\hat{\Lambda}(\cdot)$, and (2) because $\tilde{S}(\cdot)$ is asymptotically equivalent to the Kaplan-Meier estimator, and hence the two have the same asymptotic properties.

Since

$$\widehat{\Lambda}(\cdot) \xrightarrow{p} \Lambda(\cdot),$$

it follows (why is this?) that

$$\tilde{S}(\cdot) \xrightarrow{p} S(\cdot).$$

Furthermore, we can apply the functional delta method (Unit 4) to the function $g(x) = e^{-x}$. Thus, we get

$$\sqrt{n}(\tilde{S}(\cdot) - S(\cdot)) = \sqrt{n}\left(g(\widehat{\Lambda}(\cdot)) - g(\Lambda(\cdot))\right) \xrightarrow{w} -S(\cdot)X(\cdot). \quad (15.5)$$

Note that $-S(\cdot)X(\cdot)$ has the same distribution as $S(\cdot)X(\cdot)$, which is a zero-mean Gaussian process with variance function $S^2(\cdot)\sigma^2(\cdot)$. That is, $\sqrt{n}(\tilde{S}(\cdot) - S(\cdot))$ converges weakly to a zero-mean Gaussian process with variance function $S^2(\cdot)\sigma^2(\cdot)$. From (15.3) and the consistency of $\tilde{S}(\cdot)$, a pointwise consistent estimator of this variance function is $\tilde{S}(\cdot)^2\hat{\sigma}^2(\cdot)$.

Now consider the Kaplan-Meier estimator of $S(\cdot)$, denoted $\hat{S}(\cdot)$. It can be shown that

$$\sqrt{n}\left(\hat{S}(\cdot) - \tilde{S}(\cdot)\right) \xrightarrow{P} 0$$

as $n \to \infty$. From this it follows immediately that the Kaplan-Meier estimator is consistent for $S(\cdot)$. Furthermore, since

$$\sqrt{n}\left(\hat{S}(\cdot) - S(\cdot)\right) \;=\; \sqrt{n}\left(\tilde{S}(\cdot) - S(\cdot)\right) \;+\; \sqrt{n}\left(\hat{S}(\cdot) - \tilde{S}(\cdot)\right),$$

it follows from Slutsky's theorem that (15.5) also holds for $\hat{S}$, so that $\sqrt{n}\left(\hat{S}(\cdot) - S(\cdot)\right)$ converges weakly to a zero-mean Gaussian process with variance function $S^2(\cdot)\sigma^2(\cdot)$.

Next, we estimate the variance of the Kaplan Meier estimator $\widehat{S}(t)$. It can be shown that (but not easily, see Fleming and Harrington page 98 and further):

$$\begin{aligned}
&\text{Var}\left(\sqrt{n}(\widehat{S}(t) - S(t))\right) = n \;\cdot\; \text{Var}(\widehat{S}(t)) \\
&\approx\; S^2(t)\left(n\; E\int_0^t \frac{\widehat{S}^2(s-)1(\Sigma Y(s) > 0)\lambda(s)}{S^2(s)\;\Sigma Y(s)} ds\right).
\end{aligned}$$

Thus, we can estimate var $[\widehat{S}(t)]$ by

$$\widehat{\text{var}}\left(\widehat{S}(t)\right) = \widehat{S}^2(t)\int_0^t \frac{\widehat{S}^2(s-)1[\Sigma Y(s) > 0]}{\hat{S}^2(s)(\Sigma Y(s))^2} d\Sigma N(s), \tag{15.6}$$

where we have replaced $d\Lambda(s) = \lambda(s)ds$ by $d\widehat{\Lambda}(s) = \frac{d\Sigma N(s)}{\Sigma Y(s)}$. This variance estimator is often close to, but not identical to Greenwood's formula.

Another way of estimating the variance of $\widehat{S}(t)$ can be based on (15.5). That is,

$$\begin{aligned} \text{var}\ (\widehat{S}(t)) &\approx \tfrac{S^2(t)\sigma^2(t)}{n} \\ \Rightarrow\ \widetilde{var}\ (\widehat{S}(t)) &= \tfrac{\widehat{S}^2(t)\widehat{\sigma}^2(t)}{n}. \end{aligned}$$

Replacing $\hat{\sigma}^2(t)$ by its estimator in (15.3) yields

$$\widetilde{var}(\widehat{S}(t)) = \widehat{S}^2(t) \int_0^t \frac{1(\Sigma Y(s) > 0)}{(\Sigma Y(s))^2} d\Sigma N(s). \tag{15.7}$$

This differs from the previous estimator of $var\ (\widehat{S}(t))$ in that it omits the factor $\frac{\widehat{S}^2(s-)}{\widehat{S}^2(s)}$ in the integrand. Since $\frac{\widehat{S}^2(s-)}{\widehat{S}^2(s)} \geq 1$, the previous estimator is somewhat larger, leading to more conservative inferences.

Exercises

1. Show that $P(\Sigma Y(t) > 0) \to 1$ if and only if $1 - 1(\Sigma Y(t) > 0) \xrightarrow{P} 0$.

2. Verify that $\sqrt{n}\ D(\cdot) \xrightarrow{P} 0$ on $[0, \tau]$, where D(t) is defined below equation (15.1).

3. In the proof of the weak convergence of $\hat{\Lambda}(\cdot)$, verify conditions (a) and (b) of the Martingale CLT and find the expression for $\sigma^2(t)$, the variance of the limiting process at time t. If you are imprecise (e.g., use Theorem 15.1 instead of 15.1b or c but the conditions of Theorem 15.1 are not satisfied), please indicate which condition(s) are not satisfied and why.

4. Show that if $EQ(t) = 0$ and $\mathrm{Var}(Q(t)) \to 0$, then $Q(t) \xrightarrow{P} 0$.

5. Show that $Q(\cdot) \xrightarrow{P} 0$ on page 8. Hint: one has to prove that $\sup_{t\in[0,\tau]} |Q(t) - 0|$ converges in probability to 0, or equivalently that for any $\epsilon > 0$

$$P\left(\sup_{t\in[0,\tau]} |Q(t) - 0| > \epsilon\right) < \epsilon$$

   for n large enough. You may want to use Theorem 15.2.

6. Show consistency of $\hat{\sigma}^2$ on page 9 (difficult to do in a precise way).

7. Consider the estimated variance function given in equation (15.6). Using the notation we used when we first introduced the Kaplan-Meier estimator in Unit 5, show how this estimator in (15.6) compares to Greenwood's formula for the variance of the Kaplan-Meier estimator. Indicate why they should often be close.

References

Fleming TR and Harrington DP (1991). Counting Processes and Survival Analysis, Wiley, New York

Anderson PK, Borgan O, Gill RD, and Keiding N. (1993). Statistical Models Based on Counting Processes, Springer-Verlag