Background

1.1 Review of Counting Process

Recall that a *counting process* $N(t)$ is a continuous-time stochastic process satisfying $N(0) = 0$, with càdlàg sample paths and increments $dN(t)$ of size 1 at event times. By the Doob-Meyer decomposition, there exists a unique predictable process $\Lambda(t)$ such that the compensated process $M(t) = N(t) - \Lambda(t)$ is a mean-zero martingale. The *compensator*, or cumulative intensity, is expressible as:

$$\Lambda(t) = \int_0^t \lambda(s)ds,$$

where $\lambda(s)$ is a predictable intensity process.

In differential form:

$$dM(t) = dN(t) - \lambda(t)dt. \tag{1.1.1}$$

Since $M(t)$ is a mean-zero martingale:

$$\mathbb{E}\big\{dM(t)|\mathscr{F}(t-)\big\} = 0,$$

and since $\lambda(s)$ is a predictable process:

$$\lambda(t)dt = \mathbb{E}\big\{dN(t)|\mathscr{F}(t-)\big\}.$$

1.1.1 Optional and Predictable Variations

The variance of a mean-zero martingale is the expectation of its optional and predictable variation processes:

$$\mathbb{V}\big\{M(t)\big\} = \mathbb{E}\big[M(t)\big] = \mathbb{E}\big\langle M(t)\big\rangle.$$

The optional variation of the stochastic integral of a predictable process $H(s)$ with respect to a mean-zero counting process martingale is:

$$\left[\int_0^t H(s)dM(s)\right] = \int_0^t H^2(s)d[M(s)] = \int_0^t H^2(s)dN(s).$$

The predictable variation is:

$$\left\langle\int_0^t H(s)dM(s)\right\rangle = \int_0^t H^2(s)d\langle M(s)\rangle = \int_0^t H^2(s)d\Lambda(s).$$

1.1.2 Standard Brownian Motion

Definition 1.1.1. Standard **Brownian motion** is a Gaussian process $W : [0, \infty) \to \mathbb{R}$ with these properties:

- Boundary condition: $W(0) = 0$.
- Moments: $\mathbb{E}\{W(t)\} = 0$ and $\mathbb{C}\{W(s), W(t)\} = \min(s, t)$.
- Continuous sample paths.
- Independent, stationary increments.

■

Discussion 1.1.1. Let $W(t)$ denote standard Brownian motion.

- The independent increments property means that for any $0 = t_0 < t_1 < \cdots < t_n$, the differences:

$$\Delta_1 = W(t_1) - W(t_0), \quad \Delta_2 = W(t_2) - W(t_1), \quad \cdots, \quad \Delta_n = W(t_n) - W(t_{n-1}),$$

  are independent.

- The stationary increments property means that for any $s \leq t$:

$$W(t) - W(s) \stackrel{d}{=} W(t - s).$$

- The optional and predictable variations of $W(t)$ coincide and are equal to t:

$$\big[W(t)\big] = \big\langle W(t) \big\rangle = t.$$

- Both $W(t)$ itself and $W^2(t) - t$ are zero-mean martingales.

♠

1.1.3 Martingale Central Limit Theorem

Theorem 1.1.1. Suppose $M^{(n)}(t)$ is a sequence of mean-zero martingales defined on $[0, \tau]$, and for any $\epsilon > 0$ let $M_{\epsilon}^{(n)}(t)$ denote the martingale containing all jumps of $M^{(n)}(t)$ that are of size greater than ϵ. If the following conditions hold:

i. $\langle M^{(n)} \rangle \xrightarrow{p} \sigma^2(t)$ for all $t \in [0, \tau]$ as $n \to \infty$, where $\sigma^2(t)$ is a strictly increasing continuous function with $\sigma^2(0) = 0$.

ii. $\langle M_{\epsilon}^{(n)} \rangle \xrightarrow{p} 0$ for all $t \in [0, \tau]$ for any $\epsilon > 0$ as $n \to \infty$.

Then, $M^{(n)}(t)$ converges weakly to a Gaussian martingale:

$$M^{(n)}(t) \rightsquigarrow W\big\{\sigma^2(t)\big\}.$$

□

1.1.4 Functional Delta Method

Theorem 1.1.2. Suppose g is a continuously differentiable, and that for $n \to \infty$:

$$\sqrt{n}\big\{\hat{\mu}_n(\cdot) - \mu(\cdot)\big\} \rightsquigarrow Z(\cdot), \tag{1.1.2}$$

where $Z(\cdot)$ has continuous sample paths. Then:

$$\sqrt{n}\big\{g \circ \hat{\mu}_n(\cdot) - g \circ \mu(\cdot)\big\} \rightsquigarrow \{g' \circ \mu(\cdot)\}Z(\cdot)$$

and:

$$\sqrt{n}\big\{g \circ \hat{\mu}_n(\cdot) - g \circ \mu(\cdot)\big\} = \{g' \circ \mu(\cdot)\} \cdot \sqrt{n}\big\{g \circ \hat{\mu}_n(\cdot) - g \circ \mu(\cdot)\big\} + o_p(1).$$

□

1.2 Data

Let $\{(U_i, \delta_i)\}_{i=1}^n$ denote IID observations, where:

$$U_i = \min(C_i, T_i), \qquad \delta_i = \mathbb{I}(T_i \leq C_i).$$

Define the individual-level *event* and *at-risk* processes:

$$N_i(t) = \mathbb{I}\big(U_i \leq t, \delta_i = 1\big), \qquad Y_i(t) = \mathbb{I}\big(U_i \geq t\big).$$

The intensity of $N_i(t)$ will take the form:

$$\lambda_i(t) = \alpha(t)Y_i(t),$$

where $\alpha(t)$ is the hazard of the event-time distribution.

Denote the aggregated, sample-level processes by:

$$N(t) = \sum_{i=1}^n N_i(t), \qquad Y(t) = \sum_{i=1}^n Y_i(t), \qquad \lambda(t) = \sum_{i=1}^n \lambda_i(t).$$

The aggregated intensity process satisfies the multiplicative property:

$$\lambda(t) = \alpha(t)Y(t). \tag{1.2.3}$$

Nelson-Aalen

2.1 Cumulative Hazard

Definition 2.1.1. The **cumulative hazard** may be defined as:

$$A(t) = -\int_0^t \frac{dS(u)}{S(u-)}. \tag{2.1.4}$$

■

Discussion 2.1.1. The survival increment is interpretable as:

$$-dS(t) = S(t-) - S(t) = \mathbb{P}(t \leq T < t + dt).$$

If the distribution of T is absolutely continuous, then $-dS(t) = f(t)dt$ and $S(t-) = S(t)$, such that (2.1.4) reduces to:

$$A(t) = \int_0^t \frac{f(u)}{S(u)} du = \int_0^t \alpha(u) du,$$

where $\alpha(u)$ is the continuous hazard.

If the distribution is discrete, then $-dS(t) = S(t-) - S(t) = \mathbb{P}(T = t)$ and:

$$\frac{-dS(t)}{S(t-)} = \frac{\mathbb{P}(T = t)}{\mathbb{P}(T \geq t)} = \alpha(t).$$

Thus, in the discrete case,

$$A(t) = \sum_{u \leq t} \alpha(u),$$

where $\alpha(u)$ is the discrete hazard.

The differential form of (2.1.4) is:

$$dS(t) = -S(t-)dA(t). \tag{2.1.5}$$

♠

2.2 Estimator

Definition 2.2.1. The **Nelson-Aalen** estimator of the cumulative hazard is:

$$\hat{A}(t) = \int_0^t \frac{J(s)}{Y(s)} dN(s).$$

where $J(t) = \mathbb{I}\{Y(t) > 0\}$, and $J(t)/Y(t) = 0$ if $Y(t) = 0$. ■

Proposition 2.2.1. Define the *modified cumulative hazard*:

$$A^*(t) = \int_0^t J(s)\alpha(s)ds, \tag{2.2.6}$$

then:

$$\hat{A}(t) - A^*(t) = \int_0^t \frac{J(s)}{Y(s)} dM(s). \tag{2.2.7}$$

where $M(s)$ is the counting process martingale corresponding to $N(s)$. ♦

Proof. From (1.1.1) and (1.2.3):

$$dN(s) - \alpha(s)Y(s)ds = dM(s).$$

Multiplying by $J(s)/Y(s)$ and integrating over $[0, t]$:

$$\int_0^t \frac{J(s)}{Y(s)} dN(s) - \int_0^t J(s)\alpha(s)ds = \int_0^t \frac{J(s)}{Y(s)} dM(s).$$

■

Corollary 2.2.1. Since the RHS of (2.2.7) is the stochastic integral of a predictable process with respect to a mean-zero martingale, the Nelson-Aalen estimator $\hat{A}(t)$ is unbiased for $A^*(t)$:

$$\mathbb{E}\big\{\hat{A}(t) - A^*(t)\big\} = 0.$$

♣

Corollary 2.2.2. An estimate for $\mathbb{V}\{\hat{A}(t) - A^*(t)\}$ is given by:

$$\hat{\sigma}^2_{\text{NA}}(t) = \int_0^t \frac{J(s)}{Y^2(s)} dN(s). \tag{2.2.8}$$

♣

Proof. The optional variation of (2.2.7) is:

$$\big[\hat{A}(t) - A^*(t)\big] = \int_0^t \left\{\frac{J(s)}{Y(s)}\right\}^2 d[M(s)] = \int_0^t \frac{J(s)}{Y^2(s)} dN(s).$$

■

2.3 Asymptotics

Proposition 2.3.2 (Consistency). If $\inf_{s\in[0,\tau]} Y(s) \xrightarrow{p} \infty$ as $n \to \infty$, then:

$$\sup_{t\in[0,\tau]} |\hat{A}(t) - A(t)| \xrightarrow{p} 0. \tag{2.3.9}$$

See [1] (IV.1.1). ♦

Proposition 2.3.3 (Asymptotic Normality). Let $A^*(t) = \int_0^t J(s)dA(s)$ denote the modified cumulative hazard, and $\hat{A}(t)$ the Nelson-Aalen estimator. Suppose there exists a deterministic function $y(s) = \operatorname*{plim}_{n\to\infty} n^{-1}Y(s)$ strictly positive on $[0,\tau]$. The normalized process $\sqrt{n}\big\{\hat{A}(t) - A^*(t)\big\}$ converges weakly to a Gaussian martingale:

$$\sqrt{n}\big\{\hat{A}(t) - A^*(t)\big\} \rightsquigarrow W\big\{\sigma^2_{\text{NA}}(t)\big\}, \tag{2.3.10}$$

with variance function:

$$\sigma^2_{\text{NA}}(t) = \int_0^t \frac{\alpha(s)}{y(s)}ds.$$

Moreover,

$$\sup_{s\in[0,\tau]} \big|n \cdot \hat{\sigma}^2_{\text{NA}}(s) - \sigma^2_{\text{NA}}(s)\big| \xrightarrow{p} 0,$$

where $\hat{\sigma}^2_{\text{NA}}(t)$ is the optional variation estimator (2.2.8). See [1] (IV.1.2). ♦

Proof (*Sketch*)**.** Consider the normalized difference:

$$\sqrt{n}\big\{\hat{A}(t) - A^*(t)\big\} \equiv \int_0^t H(s)dM(s),$$

where $H(s)$ is the predictable process:

$$H(s) = \sqrt{n}\frac{J(t)}{Y(t)}$$

The predictable variation is:

$$\begin{aligned}
\big\langle\sqrt{n}\big\{\hat{A}(t) - A^*(t)\big\}\big\rangle &= \int_0^t \big\{H(s)\big\}^2 d\langle M(s)\rangle \\
&= \int_0^t \frac{J(s)}{n^{-1}Y^2(s)}d\Lambda(s) \\
&= \int_0^t \frac{J(s)}{n^{-1}Y^2(s)}Y(s)\alpha(s)ds \\
&= \int_0^t \frac{J(s)}{n^{-1}Y(s)}\alpha(s)ds.
\end{aligned}$$

By hypothesis $n^{-1}Y(s) \xrightarrow{p} y(s)$, which is strictly on $[0,\tau]$, hence $J(s) \xrightarrow{p} 1$ and:

$$\operatorname*{plim}_{n\to\infty}\big\langle\sqrt{n}\big\{\hat{A}(t) - A^*(t)\big\}\big\rangle = \int_0^t \operatorname*{plim}_{n\to\infty} \frac{J(s)}{n^{-1}Y(s)}\alpha(s)ds = \int_0^t \frac{\alpha(s)}{y(s)}ds.$$

■

2.4 Confidence Bands

Proposition 2.4.4 (Gill Band). A simultaneous level $(1-\alpha)$ confidence band for $A^*(t)$ and $t \in [0, \tau]$ is given by:

$$\hat{A}(t) - \frac{\gamma_{1-\alpha}\hat{\sigma}_{\text{NA}}(\tau)}{\sqrt{n}} \leq A^*(t) \leq \hat{A}(t) + \frac{\gamma_{1-\alpha}\hat{\sigma}_{\text{NA}}(\tau)}{\sqrt{n}},$$

where $\gamma_{1-\alpha}$ is a critical value such that:

$$\mathbb{P}\left\{\sup_{u\in[0,1]} \big|W(u)\big| \leq \gamma_{1-\alpha}\right\} = 1 - \alpha.$$

♦

Proof. Recall that $\sqrt{n}\big\{\hat{A}(t) - A^*(\tau)\big\}$ converges weakly to a Gaussian martingale:

$$\sqrt{n}\big\{\hat{A}(t) - A^*(t)\big\} \rightsquigarrow W\big\{\sigma^2_{\text{NA}}(t)\big\}.$$

Since $\sigma^2_{\text{NA}}(t)$ is monotone increasing:

$$\sup_{t\in[0,\tau]}\left\{\frac{\sqrt{n}}{\hat{\sigma}_{\text{NA}}(\tau)}\Big|\hat{A}(t) - A^*(t)\Big|\right\} \xrightarrow{\mathcal{L}} \sup_{t\in[0,\tau]}\left|W\left\{\frac{\sigma^2_{\text{NA}}(t)}{\sigma^2_{\text{NA}}(\tau)}\right\}\right| = \sup_{u\in[0,1]}\big|W(u)\big|.$$

Let $\gamma_{1-\alpha}$ denote a critical value such that:

$$\mathbb{P}\left\{\sup_{u\in[0,1]} \big|W(u)\big| \leq \gamma_{1-\alpha}\right\} = 1 - \alpha,$$

then:

$$\begin{aligned}
1 - \alpha &\doteq \mathbb{P}\left\{\sup_{t\in[0,\tau]} \frac{\sqrt{n}}{\hat{\sigma}_{\text{NA}}(\tau)}\Big|\hat{A}(t) - A^*(t)\Big| \leq \gamma_{1-\alpha}\right\} \\
&= \mathbb{P}\left\{\hat{A}(t) - \frac{\gamma_{1-\alpha}\hat{\sigma}_{\text{NA}}(\tau)}{\sqrt{n}} \leq A^*(t) \leq \hat{A}(t) + \frac{\gamma_{1-\alpha}\hat{\sigma}_{\text{NA}}(\tau)}{\sqrt{n}} \text{ for } \forall t \in [0, \tau]\right\}
\end{aligned}$$

■

Definition 2.4.1. The standard **Brownian bridge** is a Gaussian process $B : [0, 1] \to \mathbb{R}$ with these properties:

- Boundary conditions: $B(0) = B(1) = 0$.
- Moments: $\mathbb{E}\{B(t)\} = 0$ and $\mathbb{C}\{B(s), B(t)\} = \min(s, t) - st$.
- Continuous sample paths.

■

Proposition 2.4.5. Let $W(t)$ denote standard Brownian motion, and define:

$$B(t) = (1-t)W\left(\frac{t}{1-t}\right), \tag{2.4.11}$$

with $B(1) \equiv 0$. Then $B(t)$ is the standard Brownian bridge. ♦

Proof. The boundary conditions are satisfied since $B(0) = W(0) = 0$ and $B(1) = 0$. The mean of $B(t)$ is:

$$\mathbb{E}\{B(t)\} = \mathbb{E}\left\{(1-t)W\left(\frac{t}{1-t}\right)\right\} = 0,$$

since $\mathbb{E}\{W(\cdot)\} = 0$. Noting that $s/(1-s) < t/(1-t)$:

$$\begin{aligned}
\mathbb{C}\{B(s), B(t)\} &= \mathbb{C}\left\{(1-s)W\left(\frac{s}{1-s}\right), (1-t)W\left(\frac{t}{1-t}\right)\right\} \\
&= (1-s)(1-t)\mathbb{C}\left\{W\left(\frac{s}{1-s}\right), W\left(\frac{t}{1-t}\right)\right\} \\
&= (1-s)(1-t)\frac{s}{1-s} = s(1-t) \\
&= \min(s,t) - st.
\end{aligned}$$

$B(\cdot)$ has continuous sample paths since $W(\cdot)$ has continuous sample paths and $(1-t)$ is continuous with $(1-t) \to 0$ as $t \to 1$. ■

Proposition 2.4.6. Define:

$$K(t) = \frac{\sigma^2_{\text{NA}}(t)}{1+\sigma^2_{\text{NA}}(t)},$$

and let $q(u)$ denote a continuous, non-negative function on $[0,1]$. Then,

$$q\{K(t)\} \cdot \{1-K(t)\} \cdot \sqrt{n}\{\hat{A}(t) - A^*(t)\} \rightsquigarrow q\{K(t)\} \cdot B\{K(t)\}.$$

♦

Proof. Observe that:

$$\sigma^2_{\text{NA}}(t) = \frac{K(t)}{1-K(t)}.$$

From (2.3.10):

$$\sqrt{n}\{\hat{A}(t) - A^*(t)\} \rightsquigarrow W\left\{\frac{K(t)}{1-K(t)}\right\},$$

therefore:

$$q\{K(t)\} \cdot \{1 - K(t)\} \cdot \sqrt{n}\{\hat{A}(t) - A^*(t)\} \rightsquigarrow g\{K(t)\} \cdot \{1 - K(t)\} \cdot W\left\{\frac{K(t)}{1 - K(t)}\right\}.$$

By (2.4.11):

$$q\{K(t)\} \cdot \{1 - K(t)\} \cdot W\left\{\frac{K(t)}{1 - K(t)}\right\} \overset{d}{=} q\{K(t)\} \cdot B\{K(t)\}.$$

■

Proposition 2.4.7.

$$\sup_{t \in [0,\tau]} \left| q\{K(t)\} \cdot \{1 - K(t)\} \cdot \sqrt{n}\{\hat{A}(t) - A^*(t)\} \right| \xrightarrow{\mathcal{L}} \sup_{u \in [0, K(\tau)]} q\{K(u)\} \cdot |B(u)| \tag{2.4.12}$$

where $B(\cdot)$ is the standard Brownian bridge. See [2]. ♦

Proposition 2.4.8 (Hall-Wellner Band). A simultaneous level $(1 - \alpha)$ confidence band for $A^*(t)$ and $t \in [0, \tau]$ is given by:

$$\hat{A}(t) - \frac{\gamma_{1-\alpha}\{1 + \sigma_{\text{NA}}^2(t)\}}{\sqrt{n}} \leq A^*(t) \leq \hat{A}(t) + \frac{\gamma_{1-\alpha}\{1 + \sigma_{\text{NA}}^2(t)\}}{\sqrt{n}},$$

where $\gamma_{1-\alpha}$ is a critical value such that:

$$\mathbb{P}\left\{\sup_{u \in [0, K(\tau)]} |B(u)| \leq \gamma_{1-\alpha}\right\} = 1 - \alpha,$$

♦

Proof. From (2.4.12) with $q(\cdot) \equiv 1$:

$$\begin{aligned}
1 - \alpha &\doteq \mathbb{P}\left\{\sup_{t \in [0,\tau]} \left|\{1 - K(t)\} \cdot \sqrt{n}\{\hat{A}(t) - A^*(t)\}\right| \leq \gamma_{1-\alpha}\right\} \\
&= \mathbb{P}\left\{\sup_{t \in [0,\tau]} \left|\frac{1}{1 + \sigma_{\text{NA}}^2(t)} \cdot \sqrt{n}\{\hat{A}(t) - A^*(t)\}\right| \leq \gamma_{1-\alpha}\right\} \\
&= \mathbb{P}\left\{\hat{A}(t) - \frac{\gamma_{1-\alpha}\{1 + \sigma_{\text{NA}}^2(t)\}}{\sqrt{n}} \leq A^*(t) \leq \hat{A}(t) + \frac{\gamma_{1-\alpha}\{1 + \sigma_{\text{NA}}^2(t)\}}{\sqrt{n}} \text{ for } \forall t \in [0, \tau]\right\}
\end{aligned}$$

■

Proposition 2.4.9 (Equi-Precision Band). A simultaneous level $(1 - \alpha)$ confidence band for $A^*(t)$ and $t \in [0, \tau]$ is given by:

$$\hat{A}(t) - \frac{\gamma_{1-\alpha}\sigma_{\text{NA}}(t)}{\sqrt{n}} \leq A^*(t) \leq \hat{A}(t) + \frac{\gamma_{1-\alpha}\sigma_{\text{NA}}(t)}{\sqrt{n}},$$

where $\gamma_{1-\alpha}$ is a critical value such that:

$$\mathbb{P}\left\{\sup_{u \in [0, K(\tau)]} |B(u)| \cdot \{u(1 - u)\}^{-1/2} \leq \gamma_{1-\alpha}\right\} = 1 - \alpha,$$

♦

Proof. Let $q(u) = \{u(1-u)\}^{-1/2}$, then:

$$q\{K(t)\} \cdot \{1 - K(t)\} = \left\{\frac{\sigma_{\text{NA}}^2}{(1+\sigma_{\text{NA}}^2)^2}\right\}^{-1/2} \cdot \frac{1}{1+\sigma_{\text{NA}}^2} = \frac{1}{\sigma_{\text{NA}}}.$$

Now, from (2.4.12):

$$\begin{aligned}
1 - \alpha &\doteq \mathbb{P}\left\{\sup_{t\in[0,\tau]} \left|q\{K(t)\} \cdot \{1 - K(t)\} \cdot \sqrt{n}\big\{\hat{A}(t) - A^*(t)\big\}\right| \leq \gamma_{1-\alpha}\right\} \\
&= \mathbb{P}\left\{\sup_{t\in[0,\tau]} \left|\frac{1}{\sigma_{\text{NA}}(t)} \cdot \sqrt{n}\big\{\hat{A}(t) - A^*(t)\big\}\right| \leq \gamma_{1-\alpha}\right\} \\
&= \mathbb{P}\left\{\hat{A}(t) - \frac{\gamma_{1-\alpha}\sigma_{\text{NA}}(t)}{\sqrt{n}} \leq A^*(t) \leq \hat{A}(t) + \frac{\gamma_{1-\alpha}\sigma_{\text{NA}}(t)}{\sqrt{n}} \text{ for } \forall t \in [0,\tau]\right\}
\end{aligned}$$

■

Kaplan-Meier

3.1 Survival Function

Proposition 3.1.1. Let $0 = t_0 < t_1 < \cdots < t_K = t$ partition the interval $(0, t]$. Then:

$$S(t) = \prod_{k=1}^{K} \mathbb{P}(T > t_k | T > t_{k-1}). \tag{3.1.13}$$

♦

Proof. By successive conditioning:

$$\begin{aligned}
\mathbb{P}(T > t) &= \mathbb{P}(T > t_K) = \mathbb{P}\big\{(T > t_K) \cap (T > t_{K-1})\big\} \\
&= \mathbb{P}(T > t_K | T > t_{K-1})\mathbb{P}(T > t_{K-1}) \\
&= \mathbb{P}(T > t_K | T > t_{K-1})\mathbb{P}(T > t_{K-1} | T > t_{K-2})\mathbb{P}(T > t_{K-2}) \\
&= \cdots \\
&= \prod_{k=1}^{K} \mathbb{P}(T > t_k | T > t_{k-1})
\end{aligned}$$

Note that $\mathbb{P}(T > t_1 | T > t_0) = \mathbb{P}(T > t_1)$ since $\mathbb{P}(T > 0) = 1$. ■

Proposition 3.1.2. Suppose $S(t)$ is the survival function $S(t) = \mathbb{P}(T > t)$ of a positive random variable, and that the *cumulative hazard* is defined as in (2.1.4). Then,

$$S(t) = \prod_{u \leq t} \big\{1 - dA(u)\big\} \equiv \lim_{M\to 0} \prod_{k=1}^{K} \big\{1 - \Delta A(t_k)\big\},$$

where the limit is over finite partitions of $[0, t]$ as $M = \max_k |t_k - t_{k-1}| \to 0$. See [1] (II.6.6) for details. ♦

Discussion 3.1.1. For an absolutely continuous distribution:

$$\mathbb{P}(T > t) = \prod_{u \le t} \{1 - \alpha(u)du\} = \exp\left\{-\int_{u \le t} \alpha(u)du\right\} = e^{-A(t)}.$$

For a discrete distribution:

$$\mathbb{P}(T > t) = \prod_{u \le t} \{1 - dA(u)\} = \prod_{u \le t} (1 - \alpha_u).$$

More generally, for a mixed distribution whose cumulative hazard $A(t)$ decomposes as the sum $A(t) = A_C(t) + A_D(t)$ of an absolutely continuous component $A_C(t)$ and a discrete component $A_D(t)$:

$$A_C(t) = \int_0^t \alpha(u)du, \qquad A_D(t) = \sum_{u \le t} \alpha_u,$$

the product integral becomes:

$$S(t) = \prod_{u \le t} \{1 - \alpha(u)du\} = e^{-A_C(t)} \prod_{u \ge t} \{1 - \alpha_u\}.$$

♠

3.2 Estimator

Definition 3.2.1. The **Kaplan-Meier** estimator of $S(t)$ is the product integral of the Nelson-Aalen estimator:

$$\hat{S}(t) = \prod_{u \le t} \{1 - d\hat{A}(t)\}.$$

The Kaplan-Meier estimator is expressible as:

$$\hat{S}(t) = \prod_{k=1}^{K} \left\{1 - \frac{dN(t_k)}{Y(t_k)}\right\},$$

where the finite product is taken across all distinct event times $0 < t_1 < \cdots < t_K < \tau$. ■

Definition 3.2.2. Greenwood's estimator for the variance of $\hat{S}(t)$ is:

$$\hat{\sigma}^2_{\mathrm{KM}}(t) = \hat{S}^2(t) \int_0^t \frac{dN(s)}{Y(s)\{Y(s) - dN(s)\}}.$$

■

3.3 Asymptotics

Proposition 3.3.3 (Consistency). If $\inf_{s\in[0,\tau]} Y(s) \xrightarrow{p} \infty$ as $n \to \infty$, then:

$$\sup_{t\in[0,\tau]} \big|\hat{S}(t) - S(t)\big| \xrightarrow{p} 0.$$

See [1] (IV.3.1). ♦

Proposition 3.3.4 (Asymptotic Normality). Define the *modified survival function*:

$$S^*(t) = \prod_{u\le t} \big\{1 - dA^*(u)\big\},$$

and let $\hat{S}(t)$ denote the Kaplan-Meier estimator. Suppose there exists a deterministic function $y(s) = \operatorname*{plim}_{n\to\infty} n^{-1}Y(s)$ strictly positive on $[0,\tau]$. The normalized process $\sqrt{n}\big\{\hat{S}(t) - S^*(t)\big\}$ converges weakly to a Gaussian martingale:

$$\sqrt{n}\big\{\hat{S}(t) - S^*(t)\big\} \rightsquigarrow W\big\{\sigma^2_{\text{KM}}(t)\big\}$$

with variance function:

$$\sigma^2_{\text{KM}}(t) = S^2(t)\int_0^t \frac{\alpha(s)}{y(s)}ds = S^2(t)\sigma^2_{\text{NA}}(t).$$

Moreover,

$$\sup_{s\in[0,\tau]} \big|n\cdot\hat{\sigma}^2_{\text{KM}}(s) - \sigma^2_{\text{KM}}(s)\big| \xrightarrow{p} 0,$$

where:

$$\hat{\sigma}^2_{\text{KM}}(t) = \hat{S}^2(t)\int_0^t \frac{J(s)}{Y^2(s)}dN(s),$$

or alternatively Greenwood's estimator. See [1] (IV.3.2) for details. ♦

Proof (*Sketch*)**.** Consider T absolutely continuous. Let $\tilde{S}(t) = \exp\{-\hat{A}(t)\}$ denote the estimator of $S^*(t)$ obtained by exponentiating the Nelson-Aalen estimator $\hat{A}(t)$. From (2.3.10) and the functional delta method (1.1.2):

$$\sqrt{n}\big\{\tilde{S}(t) - S^*(t)\big\} \rightsquigarrow -S^*(t)\cdot W\big\{\sigma^2_{\text{NA}}(t)\big\}.$$

It can be shown that:

$$\sqrt{n}\big\{\hat{S}(t) - \tilde{S}(t)\big\} = o_p(1),$$

therefore:

$$\begin{aligned}
\sqrt{n}\big\{\hat{S}(t) - S^*(t)\big\} &= \sqrt{n}\big\{\hat{S}(t) - \tilde{S}(t)\big\} + \sqrt{n}\big\{\tilde{S}(t) - S^*(t)\big\} \\
&= \sqrt{n}\big\{\tilde{S}(t) - S^*(t)\big\} + o_p(1) \\
&\rightsquigarrow -S^*(t)\cdot W\big\{\sigma^2_{\text{NA}}(t)\big\}.
\end{aligned}$$

■

Discussion 3.3.1 (**Kaplan-Meier to Nelson-Aalen**). The Kaplan-Meier and Nelson-Aalen estimators are asymptotically related via:

$$\frac{\sqrt{n}\big\{\hat{S}(t) - S(t)\big\}}{-S(t)} = \sqrt{n}\big\{\hat{A}(t) - A(t)\big\} + o_p(1).$$

Using the martingale representation of the Nelson-Aalen estimator:

$$\frac{\sqrt{n}\big\{\hat{S}(t) - S(t)\big\}}{-S(t)} = \int_0^t \frac{\sqrt{n}}{Y(s)} dM(s) + o_p(1).$$

♠

3.4 Confidence Bands

Discussion 3.4.1. Since $\sqrt{n}\big\{\hat{A}(t) - A^*(t)\big\}$ and $\sqrt{n}\big\{\hat{S}(t) - S^*(t)\big\}/S^*(t)$ have the same limiting distributions, confidence bands for the Nelson-Aalen estimator may be adapted to provide confidence bands for the Kaplan-Meier estimator. In particular, the *Hall-Wellner band* takes the form:

$$\hat{S}(t) - \frac{\gamma_{1-\alpha}\{1 + \sigma^2_{\mathrm{NA}}(t)\}\hat{S}(t)}{\sqrt{n}} \leq S^*(t) \leq \hat{S}(t) + \frac{\gamma_{1-\alpha}\{1 + \sigma^2_{\mathrm{NA}}(t)\}\hat{S}(t)}{\sqrt{n}},$$

where $\gamma_{1-\alpha}$ is a critical value such that:

$$\mathbb{P}\left\{\sup_{u \in [0, K(\tau)]} |B(u)| \leq \gamma_{1-\alpha}\right\} = 1 - \alpha,$$

The *equi-precision band* takes the form:

$$\hat{S}(t) - \frac{\gamma_{1-\alpha}\sigma_{\mathrm{NA}}(t)\hat{S}(t)}{\sqrt{n}} \leq S^*(t) \leq \hat{S}(t) + \frac{\gamma_{1-\alpha}\sigma_{\mathrm{NA}}(t)\hat{S}(t)}{\sqrt{n}},$$

where $\gamma_{1-\alpha}$ is a critical value such that:

$$\mathbb{P}\left\{\sup_{u \in [0, K(\tau)]} |B(u)| \cdot \{u(1-u)\}^{-1/2} \leq \gamma_{1-\alpha}\right\} = 1 - \alpha,$$

♠

3.5 Restricted Mean Survival Time

Definition 3.5.1. The **restricted mean survival time** (RMST) $U(\tau)$ is the area under the survival curve up to time τ:

$$U(\tau) = \int_0^\tau S(t)dt.$$

An estimator for $U(\tau)$ is given by:

$$\hat{U}(\tau) = \int_0^\tau \hat{S}(t)dt,$$

where $\hat{S}(t)$ is the Kaplan-Meier (KM) estimator of the survival function. ■

Proposition 3.5.5. Define:

$$\mu_\tau(t) = \int_t^\tau S(u)du.$$

The standardized process $\sqrt{n}\big\{\hat{U}(\tau) - U(\tau)\big\}$ converges weakly to a Gaussian martingale:

$$\sqrt{n}\big\{\hat{U}(\tau) - U(\tau)\big\} \rightsquigarrow W\big\{\sigma^2_{\text{RMST}}(\tau)\big\},$$

where:

$$\sigma^2_{\text{RMST}}(\tau) = \int_0^\tau \frac{\mu^2_\tau(t)\alpha(t)}{y(t)}dt$$

and $y(t)$ is the probability limit of $n^{-1}Y(t)$. ♦

Proof. Noting that $d\mu_\tau(t) = -S(t)dt$,

$$\begin{aligned}
\sqrt{n}\big\{\hat{U}(\tau) - U(\tau)\big\} &= \int_0^\tau \sqrt{n}\big\{\hat{S}(u) - S(u)\big\}du \\
&= \int_0^\tau \frac{\sqrt{n}\big\{\hat{S}(u) - S(u)\big\}}{-S(u)} \cdot \{-S(u)du\} \\
&= \int_0^\tau \sqrt{n}\big\{\hat{A}(u) - A(u)\big\}d\mu_\tau(t) + o_p(1).
\end{aligned}$$

Integrating by parts:

$$\begin{aligned}
\int_0^\tau \sqrt{n}\big\{\hat{A}(u) - A(u)\big\}d\mu_\tau(t) =& \Big[\sqrt{n}\big\{\hat{A}(u) - A(u)\big\}\mu_\tau(t)\Big]_{t=0}^{t=\tau} \\
&- \int_0^\tau \mu_\tau(t) \cdot \sqrt{n}d\big\{\hat{A}(u) - A(u)\big\}
\end{aligned}$$

The first term on the RHS vanishes since $\hat{A}(0) = A(0) = 0$ and $\mu_\tau(\tau) = 0$. Using the martingale representation of $\sqrt{n}\big\{\hat{A}(u) - A(u)\big\}$:

$$\sqrt{n}\big\{\hat{U}(\tau) - U(\tau)\big\} = -\int_0^\tau \frac{\mu_\tau(t)\sqrt{n}}{Y(u)}dM(u) + o_p(1).$$

The predictable variation is:

$$\begin{aligned}\left\langle\sqrt{n}\left\{\hat{U}(\tau)-U(\tau)\right\}\right\rangle &= \int_0^{\tau}\frac{n\mu_{\tau}^2(t)}{Y^2(t)}d\langle M(t)\rangle + o_p(1)\\ &= \int_0^{\tau}\frac{n\mu_{\tau}^2(t)}{Y^2(t)}Y(t)\alpha(t)dt + o_p(1)\\ &= \int_0^{\tau}\frac{\mu_{\tau}^2(t)\alpha(t)}{n^{-1}Y(t)}dt + o_p(1)\\ &\xrightarrow{p} \int_0^{\tau}\frac{\mu_{\tau}^2(t)\alpha(t)}{y(t)}dt,\end{aligned}$$

where $\alpha(t)$ is the hazard and $y(t)$ is the limit in probability of $n^{-1}Y(t)$. For additional details, see [4]. ■

Discussion 3.5.1. The optional variation of $\sqrt{n}\left\{\hat{U}(\tau)-U(\tau)\right\}$ is:

$$\begin{aligned}\left[\sqrt{n}\left\{\hat{U}(\tau)-U(\tau)\right\}\right] &= \int_0^{\tau}\frac{n\mu_{\tau}^2(t)}{Y^2(t)}d[M(t)]\\ &= \int_0^{\tau}\frac{\mu_{\tau}^2(t)}{n^{-1}Y^2(t)}dN(t).\end{aligned}$$

The estimated variance:

$$\hat{\sigma}^2_{\text{RMST}}(\tau) = \int_0^{\tau}\frac{\hat{\mu}_{\tau}^2(t)}{n^{-1}Y^2(t)}dN(t),$$

where:

$$\hat{\mu}_{\tau}(t) = \int_t^{\tau}\hat{S}(u)du.$$

The variance estimator is expressible as:

$$\hat{\sigma}^2_{\text{RMST}}(\tau) = \int_0^{\tau}\frac{\hat{\mu}_{\tau}^2(t)}{n^{-1}Y(t)}d\hat{A}(t),$$

where $\hat{A}(t)$ is the Nelson-Aalen estimator.

♠

Simulation

4.1 Cumulative Hazard

Discussion 4.1.1 (Generating Sample Paths). Recall from (2.2.7) that:

$$\sqrt{n}\left\{\hat{A}(t)-A^*(t)\right\} = \int_0^t\sqrt{n}\frac{J(s)}{Y(s)}dM(s).$$

Let $H(s) = \sqrt{n}J(s)/Y(s)$ denote the integrand and define the process:

$$\begin{aligned}\Delta(t) &\equiv \sqrt{n}\{\hat{A}(t) - A^*(t)\} \\ &= \int_0^t H(s)\sum_{i=1}^n dM_i(s) \\ &= \sum_{i=1}^n \left\{\int_0^t H(s)dM_i\right\}.\end{aligned}$$

To generate approximate sample paths from $\Delta(t)$, consider the process:

$$\Delta^{(b)}(t) \equiv \sum_{i=1}^n Z_i^{(b)} \left\{\int_0^t H(s)dN_i(s)\right\},$$

where $Z_i^{(b)}$ are IID perturbation weights, with mean 0 and variance 1; for example,

$$Z_i^{(b)} \overset{\text{IID}}{\sim} N(0,1).$$

Practically, the sample path $\Delta^{(b)}(\cdot)$ is a step function, and may be summarized in tabular form. Let $t_1 < \cdots < t_K$ denote the distinct observed event times. Then, sample path is completely characterized by $\Delta^{(b)}(t_k)$ for $k \in \{1, \cdots, K\}$, where:

$$\Delta^{(b)}(t_k) = \sqrt{n}\sum_{k=1}^K \frac{Z_k^{(b)}\delta_k}{Y(t_k)},$$

where δ_k is the number of events observed at time t_k, and $Z_k^{(b)} \sim N(0, \delta_k)$.

The collection of sample paths:

$$\{\Delta^{(1)}, \cdots, \Delta^{(B)}\},$$

may be used to approximate the percentiles for functions of $\Delta(t)$. For example, suppose interest lies in identifying a critical value $\gamma_{1-\alpha}$ such that:

$$\mathbb{P}\left\{\sup_{t\in[0,\tau]} \left|\Delta(t)\right| \le \gamma_{1-\alpha}\right\} = 1 - \alpha.$$

For each of B iterations,

i. Generate the perturbation weights $Z_i^{(b)}$.

ii. Compute and store $M^{(b)} = \sup_{t\in[0,\tau]} |\Delta^{(b)}(t)|$.

Finally, select $\gamma_{1-\alpha}$ as the upper $(1-\alpha)$th percentile of the $\{M^{(b)}\}$, then:

$$\hat{A}(t) \pm \frac{\gamma_{1-\alpha}}{\sqrt{n}}$$

is an asymptotic confidence band for $A^*(t)$. ♠

4.2 Survival Function

Discussion 4.2.1 (**Difference of Survival Curves**). Let $\hat{S}_1(t)$ and $\hat{S}_2(t)$ denote estimates of the survival functions for two independent groups, and suppose $g(x, y)$ is a continuously differentiable measure of the between-group difference, such as $g(x, y) = y - x$, such that:

$$g(\hat{S}_1, \hat{S}_2) = \hat{S}_2(t) - \hat{S}_1(t).$$

Consider generating confidence bands for the following process:

$$\Delta(t) = \sqrt{n}\omega(t)\big\{g(\hat{S}_1, \hat{S}_2) - g(S_1, S_2)\big\}, \tag{4.2.14}$$

where $n = n_1 + n_2$ is the overall sample size and $\omega(t)$ is a weight function. Let $t_{L,j}$ denote the the first time that $\hat{S}_j(t)$ jumps, and let $t_{U,j}$ denote the last time that $\hat{S}_j(t)$ jumps. Define $t_L = \max(t_{L,1}, t_{L,2})$ and $t_U = \min(t_{U,1}, t_{U,2})$. The confidence band is sought over the interval $[t_L, t_U]$.

The difference process (4.2.14) is asymptotically equivalent to:

$$\dot{\Delta}(t) = \sqrt{n}\omega(t)\big\{g_2(\hat{S}_1, \hat{S}_2) \cdot (\hat{S}_2 - S_2) + g_1(\hat{S}_1, \hat{S}_2) \cdot (\hat{S}_1 - S_1)\big\},$$

which is expressible as:

$$\dot{\Delta}(t) = g_2(\hat{S}_1, \hat{S}_2)U_2(t) + g_1(\hat{S}_1, \hat{S}_2)U_1(t),$$

where:

$$U_j(t) = \sqrt{n}\omega(t)\big\{\hat{S}_j(t) - S_j(t)\big\}$$

Now, $U_j(t)$ is asymptotically equivalent to the process:

$$\begin{aligned} U_j(t) &\doteq -\omega(t)\hat{S}_j(t) \cdot \sqrt{n}\big\{\hat{A}_j(t) - A_j(t)\big\} \\ &= -\omega(t)\hat{S}_j(t)\sum_{i=1}^{n_j}\int_0^t H(s)dM_{ij}(s), \end{aligned}$$

with $H(s) = \sqrt{n}J(s)/Y(s)$ and $M_{ij}(s) = N_{ij}(s) - \Lambda_{ij}(s)$. Let $Z_{ij}^{(b)}$ denote IID $(0, 1)$ perturbation weights, then sample paths from $U_j(t)$ may be simulated via:

$$U_j^{(b)}(t) = -\omega(t)\hat{S}_j(t)\sum_{i=1}^{n_j} Z_{ij}^{(b)} \int_0^t H(s)dN_{ij}(s).$$

The confidence band for $\Delta(t)$ may be generated as follows. For each of B iterations,

i. Generate the perturbation weights $Z_{ij}^{(b)}$.

ii. Approximate a sample path of $\Delta(t)$ via:

$$\dot{\Delta}^{(b)}(t) = g_2(\hat{S}_1, \hat{S}_2)U_2^{(b)}(t) + g_1(\hat{S}_1, \hat{S}_2)U_1^{(b)}(t).$$

iii. Compute and store $M^{(b)} = \sup_{t\in[t_L,t_U]} \big|\dot{\Delta}^{(b)}(t)\big|$

Let $\gamma_{1-\alpha}$ denote the upper $(1-\alpha)$th percentile of the $\{M^{(b)}\}$, then:

$$g(\hat{S}_1, \hat{S}_2) \pm \frac{\gamma_{1-\alpha}}{\sqrt{n}\omega(t)}$$

is an asymptotic confidence band for $g(S_1, S_2)$.

Finally, let:

$$T_{\text{KS}} = \sqrt{n} \cdot \sup_{t\in[t_L,t_U]} \omega(t)\big|g(\hat{S}_1, \hat{S}_2)\big|$$

denote a KS statistics of $H_0 : S_1 = S_2$. An approximate p-value is given by:

$$\hat{p} = \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\big\{M^{(b)} \geq T_{\text{KS}}\big\}.$$

See [3]. ♠

References

[1] PK Andersen et al. *Statistical Models Based on Counting Processes*. 2nd. Springer-Verlag, 1997.

[2] V Nair. "Confidence Bands for Survival Functions with Censored Data: A Comparative Study". In: *Technometrics* 26.3 (1984), pp. 265–275.

[3] MI Parzen, LJ Wei, and Z Ying. "Simultaneous Confidence Intervals for the Difference of Two Survival Functions". In: *Scandinavian Journal of Statistics* 24 (1997), pp. 309–314.

[4] L Zhou et al. "Utilizing the integrated difference of two survival functions to quantify the treatment contrast for designing, monitoring, and analyzing a comparative clinical stud". In: *Clinical Trials* 9.5 (2012), pp. 570–577.