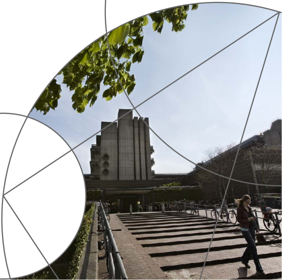
Survival Analysis

Week 4: Nonparametric procedures for survival data

Section of Biostatistics



Recap: hazard and survival

For a survival time T^*, the survival function is $S(t) = pr(T^* > t)$ and the hazard function is

$$\alpha(t) = \lim_{dt \to 0} \frac{pr(T^* < t + dt | T^* \geq t)}{dt}$$

For **absolutely continuous** distributions, survival and the hazard are related as

$$\alpha(t) = -\frac{\frac{\partial}{\partial t} S(t)}{S(t)} = -\frac{\partial}{\partial t} \log S(t)$$

and

$$S(t) = \exp\left(-\int_0^t \alpha(s) ds\right) = \exp(-A(t))$$

where $A(t) = \int_0^t \alpha(s) ds$ is the cumulative hazard.

Recap: hazard and survival

For general (continuous, discrete or mixed) distributions

- Note that

$$\text{pr}(T^* \geq t) = 1 - \text{pr}(T^* < t) = 1 - F(t-) = S(t-)$$

- For infinitesimal dt

$$\begin{aligned} dA(t) &= \text{pr}(T^* < t + dt | T^* \geq t) \\ &= \frac{\text{pr}(t \leq T^* < t + dt)}{pr(T^* \geq t)} \\ &= \frac{\text{pr}(T^* \geq t) - \text{pr}(T^* \geq t + dt)}{\text{pr}(T^* \geq t)} \\ &= \frac{\overbrace{S(t-) - S((t+dt)-)}^{=-dS(t)}}{S(t-)} \\ &= -\frac{dS(t)}{S(t-)} \end{aligned}$$

- We define

$$A(t) = -\int_0^t \frac{dS(s)}{S(s-)}$$

Recap: The Nelson-Aalen estimator

Consider censored survival times ($T_i = T_i^* \wedge C_i, \Delta_i = I(T_i^* \leq C_i)$), where $T_i^* \perp\!\!\!\perp C_i$, $i = 1, \ldots, n$. Define the aggregated processes

- $N_\bullet(t) = \sum_{i=1}^n N_i(t)$, $N_i(t) = \Delta_i I(T_i \leq t)$
- $Y_\bullet(t) = \sum_{i=1}^n Y_i(t)$, $Y_i(t) = I(T_i \geq t)$

The process $N_\bullet(t)$ has intensity $\lambda_\bullet(t) = \alpha(t) Y_\bullet(t)$. The Nelson-Aalen estimator is

$$\hat{A}(t) = \int_0^t \frac{dN_\bullet(s)}{Y_\bullet(s)}$$

Let $J(s) = I(Y_\bullet(s) > 0)$ and define the proxy

$$A^*(t) = \int_0^t J(s)\alpha(s)ds$$

Recall that, with $M_\bullet(t) = N_\bullet(t) - \int_0^t Y_\bullet(s)dA(s)$,

$$\hat{A}(t) = \underbrace{\int_0^t J(s)dA(s)}_{=A^*(t)} + \underbrace{\int_0^t \frac{J(s)}{Y_\bullet(s)} dM_\bullet(s)}_{\text{mean zero martingale}}$$

so that

$$E\left(\hat{A}(t)\right) = E\left(A^*(t)\right)$$

We use the convention that $0/0 = 0$.

Recap: Asymptotics for the Nelson-Aalen estimator

For $t \in [0, \tau]$, where $\tau < \infty$ and such that $\text{pr}(T \geq \tau) = S(\tau-)S^C(\tau-) > 0$, where $S^C(\cdot)$ is the survival function of C, the process $n^{1/2}\left(\hat{A}(t) - A(t)\right)$ converges weakly to a mean zero Gaussian martingale with variance

$$\sigma^2(t) = \int_0^t \frac{dA(s)}{\text{pr}(T \geq s)},$$

which can be estimated by

$$\hat{\sigma}^2(t) = \int_0^t \frac{dN_\bullet(s)}{Y_\bullet^2(s)/n},$$

Product integral

Let $a = t_0 < t_1 < \ldots < t_K = b$ partition $(a, b]$, and let $t_i - t_{i-1} \to 0$ when $K \to \infty$. The product integral of the right-continuous function with left-hand limits $G(u)$ is

$$\prod_{(a,b]} (1 + dG(u)) = \lim_{K \to \infty} \prod_{i=1}^{K} (1 - G(t_i) - G(t_{i-1}))$$

If $G(u)$ is continuous with derivative $\partial/(\partial u)G(u) = g(u)$, then

$$\prod_{(a,b]} (1 + dG(u)) = \exp\left(\int_a^b g(u)du\right)$$

If G is discrete with jumps at points a_j, $j = 1, 2, \ldots$, with jump sizes g_j then

$$\prod_{(a,b]} (1 + dG(u)) = \prod_{j:a<a_j\leq b} (1 + g_j).$$

If G is a mixed distribution with jumps at points a_j, $j = 1, 2, \ldots$, with jump sizes g_j then

$$\prod_{(a,b]} (1 + dG(u)) = \exp\left(\int_a^b g(u)du\right) \prod_{j:a<a_j\leq b} (1 + g_j).$$

Discrete time hazard and survival

Assume that time is measured on a discrete time scale $t_0 = 0 < t_1 < \ldots t_K = t$.

Recall that the hazard is,

$$dA(s) = \frac{S(s-) - S((s+ds)-)}{S(s-)}$$

Then the discrete time hazard is

$$\begin{aligned} dA(t_k) &= \frac{S(t_{k-1}) - S(t_k)}{S(t_{k-1})} \\ &= \frac{\text{pr}(T^* > t_{k-1}) - \text{pr}(T^* > t_k)}{\text{pr}(T^* > t_{k-1})} \\ &= \text{pr}(T^* = t_k | T^* > t_{k-1}) \end{aligned}$$

and the cumulated hazard is

$$A(t) = \sum_{k: t_k \leq t} dA(t_k)$$

Discrete time hazard and survival

The discrete time survival function is

$$\begin{aligned}
S(t) &= \mathrm{pr}(T^* > t) = \mathrm{pr}(T^* > t_K) = \mathrm{pr}(\{T^* > t_K\} \cap \{T^* > t_{K-1}\}, \ldots \cap \{T^* > t_0\}) \\
&= \mathrm{pr}(T^* > t_0)\mathrm{pr}(T^* > t_1 | T^* > t_0) \cdots \mathrm{pr}(T^* > t_K | T^* > t_{K-1}) \\
&= \prod_{k=1}^{K} \mathrm{pr}(T^* > t_k | T^* > t_{k-1}) \\
&= \prod_{k=1}^{K} \left(1 - \underbrace{\mathrm{pr}(T = t_k | T^* > t_{k-1})}_{\text{discrete time hazard } dA(t_k)} \right) \\
&= \prod_{k=1}^{K} (1 - dA(t_k))
\end{aligned}$$

Survival probability as a product integral

For discrete time,

$$S(t) = \prod_{k=1}^{K} (1 - (A(t_i) - A(t_{i-1}))) \tag{1}$$

Let $K \to \infty$ such that $\max |t_i - t_{i-1}| \to 0$. Then

$$S(t) = \lim_{\max |t_i - t_{i-1}| \to 0} \prod (1 - (A(t_i) - A(t_{i-1}))) = \prod_{0 \le s \le t} (1 - dA(s)).$$

For continuous distributions this yields

$$S(t) = \exp(-A(t)),$$

for mixed distributions we have a mix of this and (1).

Kaplan-Meier

The Kaplan-Meier estimator is achieved by plugging the Nelson-Aalen estimator

$$\hat{A}(t) = \int_0^t \frac{dN_\bullet(u)}{Y_\bullet(u)} = \sum_{k:\tau_k \leq t} \frac{\Delta N(\tau_k)}{Y_\bullet(\tau_k)},$$

where $\tau_1, \ldots, \tau_K$ are the (ordered) jump times of $N_\bullet$ (the unique uncensored event times), into the product-integral expression for the survival function to get the finite product

$$\hat{S}(t) = \prod_{0 \leq s \leq t} (1 - d\hat{A}(s)) = \prod_{k:\tau_k \leq t} (1 - \Delta\hat{A}(\tau_k)) = \prod_{k:\tau_k \leq t} \left(1 - \frac{\Delta N(\tau_k)}{Y_\bullet(\tau_k)}\right)$$

The factor

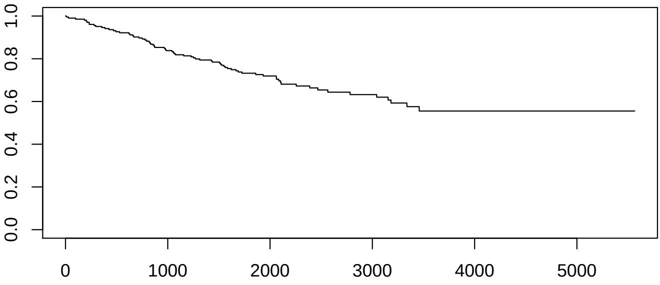
$$\left(1 - \frac{\Delta N(\tau_k)}{Y_\bullet(\tau_k)}\right)$$

estimates the conditional probability of surviving the interval $(\tau_k, \tau_{k+1}]$ given survival up to τ_k.

Survival after melanoma surgery

```
library(timereg)
data(melanoma)
melanoma$dead <- melanoma$status!=2

## Kaplan-Meier curve
kmfit <- survfit(Surv(days, dead)~1, data=melanoma)
plot(kmfit, conf.int=FALSE)
```



Properties of the Kaplan-Meier estimator

Recall that $A^*(t) = \int_0^t I(Y_\bullet(s) > 0) dA(s)$. Let

$$S^*(t) = \prod_{0<s\leq t} (1 - dA^*(s)).$$

Duhamel's equation (see Exercise 3 for this week)

$$\frac{\hat{S}(t)}{S^*(t)} - 1 = -\int_0^t \frac{\hat{S}(s-)}{S^*(s)} d(\hat{A} - A^*)(s).$$

Since $\hat{S}(t-)/S^*(t)$ is predictable and $\hat{A} - A^*$ is a martingale, it follows that

$$\frac{\hat{S}(t)}{S^*(t)} - 1$$

is a martingale.
Hence

$$E\left(\frac{\hat{S}(t)}{S^*(t)}\right) = 1.$$

For large n, we will have $E(\hat{S}(t)) \approx S(t)$ and $\hat{S}(t)/S^*(t) \approx 1$, and

$$\frac{\hat{S}(t)}{S(t)} - 1 \approx -\left(\hat{A}(t) - A(t)\right)$$

so

$$\hat{S}(t) - S(t) \approx -S(t)\left(\hat{A}(t) - A(t)\right)$$

Can we be more precise about $E(\hat{S}(t)) \approx S(t)$?

- Let $\tilde{T} = \inf\{s : J(s) = 0\} = \inf\{s : Y_\bullet(s) = 0\} = \max\{T_1, \dots, T_n\}$.
- Recall that $\hat{S}^*(t) = \prod_{s\in(0,t]}(1 - J(s)\alpha(s)ds)$ and $S(t) = \prod_{s\in(0,t]}(1 - \alpha(s)ds)$
- For $t < \tilde{T}$, $S^*(t) = S(t)$
- For $t \geq \tilde{T}$, $S^*(t) = S(\tilde{T})$ and $\hat{S}(t) = \hat{S}(\tilde{T})$.

Thus, for all $t \in [0, \tau)$,

$$\frac{\hat{S}(t)}{S^*(t)} = \frac{\hat{S}(t)}{S(t)} + I(t \geq \tilde{T}) \overbrace{\left(\frac{\hat{S}(\tilde{T})}{S(\tilde{T})} - \frac{\hat{S}(\tilde{T})}{S(t)}\right)}^{=\hat{S}(\tilde{T})\frac{S(t)-S(\tilde{T})}{S(\tilde{T})S(t)}}$$

Taking expectations on both sides yields

$$\overbrace{E\left(\frac{\hat{S}(t)}{S^*(t)}\right)}^{=1} = \frac{E(\hat{S}(t))}{S(t)} + E\left(I(t \geq \tilde{T})\hat{S}(\tilde{T})\frac{S(t) - S(\tilde{T})}{S(\tilde{T})}\right)\frac{1}{S(t)}$$

so that

$$E(\hat{S}(t)) - S(t) = E\left(I(t \geq \tilde{T})\overbrace{\hat{S}(\tilde{T})}^{\leq 1}\overbrace{\frac{S(\tilde{T}) - S(t)}{S(\tilde{T})}}^{\leq 1 - S(t)}\right) \leq E\left(I(t \geq \tilde{T})\right)(1 - S(t))$$

Let $S^C(\cdot)$ denote the survival function of C. With (T^*, C) i.i.d. and $T^* \perp\!\!\!\perp C$,

$$\begin{aligned} E\left(I(t \geq \tilde{T})\right) &= \text{pr}(t \geq T_1)\cdots\text{pr}(t \geq T_n) \\ &= (\text{pr}(T_1 \geq t))^n = \left(1 - (S(t)S^C(t)\right)^n \end{aligned}$$

Thus,

$$0 \leq E(\hat{S}(t)) - S(t) \leq (1 - S(t))\left(1 - S(t)S^C(t)\right)^n$$

The Kaplan-Meier estimator is biased upward if there is a positive probability that $\hat{S}(\tilde{T}) > 0$ and $S(t) < S(\tilde{T})$. The bias goes to zero exponentially fast.

Estimating $\mathrm{var}(\hat{S}(t))$

By the same arguments,

$$\mathrm{var}\left(\frac{\hat{S}(t)}{S^*(t)} - 1\right) - \mathrm{var}\left(\frac{\hat{S}(t)}{S(t)} - 1\right) \xrightarrow{P} 0$$

exponentially fast.

The variance of $\hat{S}(t)$ can be approximated by

$$\mathrm{var}\left(\frac{\hat{S}(t)}{S^*(t)} - 1\right) = \left\langle \frac{\hat{S}}{S^*} - 1 \right\rangle (t).$$

When $S(t)$ is continuous, this is

$$\int_0^t \left(\frac{\hat{S}(s-)}{S^*(s)} \frac{J(s)}{Y_\bullet(s)}\right)^2 Y_\bullet(s)\alpha(s)ds = \int_0^t \left(\frac{\hat{S}(s-)}{S^*(s)}\right)^2 \frac{J(s)}{Y_\bullet(s)}\alpha(s)ds.$$

We replace $\alpha(s)ds$ by $d\hat{A}(s)$ and, motivated by the continuity of $S(s)$, $\hat{S}(s-)$ and $S^*(s)$ by $\hat{S}(s)$, yielding

$$\int_0^t \frac{J(s)}{Y_\bullet(s)} d\hat{A}(s) = \int_0^t \frac{dN_\bullet(s)}{(Y_\bullet(s))^2} = \hat{\sigma}^2(t),$$

the large sample variance estimator of $\mathrm{var}(\hat{A} - A)(t)$.

Estimating var($\hat{S}(t)$)

An alternative estimator is found by using the predictable variation for a cumulative hazard that is allowed to have jumps

$$\langle M \rangle(t) = \int_0^t Y_\bullet(s)(1 - \Delta A(s))dA(s),$$

yielding

$$\left\langle \frac{\hat{S}}{S^*} - 1 \right\rangle (t) = \int_0^t \left(\frac{\hat{S}(s-)}{S^*(s)} \right)^2 \frac{J(s)}{Y_\bullet(s)}(1 - \Delta A(s))dA(s)$$

Replacing $S^*(s)$ by $\hat{S}(s)$ and $A(s)$ by $\hat{A}$ and noting that

$$\hat{S}(s) = \hat{S}(s-) \left(1 - \frac{\Delta N_\bullet(s)}{Y_\bullet(s)} \right),$$

we get the estimator for the variance of $\hat{S}(t)/S(t)$

$$\begin{aligned}
&\int_0^t \left(1 - \frac{\Delta N_\bullet(s)}{Y_\bullet(s)} \right)^{-2} \frac{1}{Y_\bullet(s)} \left(1 - \frac{\Delta N_\bullet(s)}{Y_\bullet(s)} \right) \frac{dN_\bullet(s)}{Y_\bullet(s)} \\
&= \int_0^t \frac{Y_\bullet(s)}{Y_\bullet(s) - \Delta N_\bullet(s)} \frac{1}{Y_\bullet(s)} \frac{dN_\bullet(s)}{Y_\bullet(s)} \\
&= \int_0^t \frac{dN_\bullet(s)}{(Y_\bullet(s) - \Delta N_\bullet(s))Y_\bullet(s)}
\end{aligned}$$

This is Greenwood's formula. The literature suggests that this formula would be preferred in practice, although both are consistent.

Recap: Asymptotics for the Nelson-Aalen estimator

When establishing the large sample properties of the Nelson-Aalen estimator, the following results were used. With independent right-censoring, they follow if $S(\tau)S^C(\tau-) > 0$ (see MS Example 2.5.1 and MS Exercise 2.8).

For $t \in [0, \tau]$, as $n \to \infty$,

A1

$$\int_0^t \frac{J(s)}{Y_\bullet(s)} \alpha(s) ds \xrightarrow{P} 0$$

A2

$$\int_0^t \frac{J(s)}{Y_\bullet(s)/n} \alpha(s) ds \xrightarrow{P} \int_0^t \frac{\alpha(s)}{\mathrm{pr}(T \geq s)} ds = \sigma^2(t)$$

A3

$$\int_0^t (1 - J(s)) \alpha(s) ds \xrightarrow{P} 0$$

A4 For all $\epsilon > 0$,

$$\int_0^t \frac{J(s)}{Y_\bullet(s)/n} I\left(\sqrt{n}\frac{J(s)}{Y_\bullet(s)} > \epsilon\right) \alpha(s) ds \xrightarrow{P} 0$$

Lenglart's inequality (MS p. 41)

For a martingale $M(t)$ with non-decreasing predictable variation process $\langle M(t)\rangle$

$$\mathrm{pr}\left(\sup_{s\in[0,t]} |M(s)| > \eta\right) \leq \frac{\delta}{\eta^2} + \mathrm{pr}\left(\langle M\rangle(t) > \delta\right)$$

for any $\eta > 0$ and $\delta > 0$.

Hence $\sup_{s\in[0,t]} |M(s)| \xrightarrow{P} 0$ if $\langle M\rangle(t) \xrightarrow{P} 0$.

Uniform consistency of the Kaplan-Meier estimator

Using Lenglart's inequality and that

$$\left\langle \frac{\hat{S}}{S^*} - 1 \right\rangle (t) = \int_0^t \left(\frac{\hat{S}(s-)}{S^*(s)} \right)^2 \frac{J(s)}{Y_\bullet(s)} \alpha(s) ds,$$

we get, for any $\delta, \eta > 0$,

$$\text{pr}\left(\sup_{s \in [0,\tau]} \left| \frac{\hat{S}(s)}{S^*(s)} - 1 \right| > \eta \right) \leq \frac{\delta}{\eta^2} + \text{pr}\left(\int_0^\tau \overbrace{\left(\frac{\hat{S}(s-)}{S^*(s)} \right)^2}^{\leq \frac{1}{(S^*(s))^2} \leq \frac{1}{(S(\tau))^2}} \frac{J(s)}{Y_\bullet(s)} \alpha(s) ds > \delta \right)$$

$$\leq \frac{\delta}{\eta^2} + \text{pr}\left(\frac{1}{(S(\tau))^2} \underbrace{\int_0^\tau \frac{J(s)}{Y_\bullet(s)} \alpha(s) ds}_{\xrightarrow{P} 0 \text{ by } A1} > \delta \right)$$

Hence,

$$\sup_{s \in [0,\tau]} \left| \frac{\hat{S}(s)}{S^*(s)} - 1 \right| \xrightarrow{P} 0 \qquad (2)$$

Uniform consistency of the Kaplan-Meier estimator

Using Duhamel's equation again (see Exercise 3 for this week)

$$\left|\frac{S(s)}{S^*(s)} - 1\right| = \int_0^s \frac{S(u-)}{S^*(u)} \overbrace{d(A - A^*)(u)}^{\alpha(u)du - J(u)\alpha(u)du} = \int_0^s \overbrace{\frac{S(u-)}{S^*(u)}}^{\leq \frac{1}{S(s)}} (1 - J(u))\alpha(u)du$$

$$\leq \frac{1}{S(s)} \underbrace{\int_0^s (1 - J(u))\alpha(u)du}_{\xrightarrow{P} 0 \text{ by } A3}$$

Hence,

$$\sup_{s \in [0,\tau]} \left|\frac{S(s)}{S^*(s)} - 1\right| \xrightarrow{P} 0. \tag{3}$$

Combining (2) and (3),

$$\begin{aligned}\sup_{s \in [0,\tau]} \left|\hat{S}(s) - S(s)\right| &\leq \sup_{s \in [0,\tau]} \frac{\left|\hat{S}(s) - S(s)\right|}{S^*(s)} \\ &\leq \sup_{s \in [0,\tau]} \left|\frac{\hat{S}(s)}{S^*(s)} - 1\right| + \sup_{s \in [0,\tau]} \left|\frac{S(s)}{S^*(s)} - 1\right| \xrightarrow{P} 0\end{aligned}$$

Weak convergence of $\sqrt{n}(\hat{S} - S)$

Let

$$\tilde{U}(t) = -\sqrt{n}\left(\frac{\hat{S}(s)}{S^*(s)} - 1\right) = \sqrt{n}\int_0^t \frac{\hat{S}(s-)}{S^*(s)} \frac{J(s)}{Y_\bullet(s)} dM_\bullet(s)$$

$$\tilde{U}_\epsilon(t) = \sqrt{n}\int_0^t \frac{\hat{S}(s-)}{S^*(s)} \frac{J(s)}{Y_\bullet(s)} I\left(\sqrt{n}\frac{\hat{S}(s-)}{S^*(s)} \frac{J(s)}{Y_\bullet(s)} > \epsilon\right) dM_\bullet(s)$$

By Rebolledo's martingale central limit theorem, to show weak convergence of $\tilde{U}$, it is sufficient to establish that, as $n \to \infty$

1. For $t \in [0, \tau]$

$$\langle \tilde{U} \rangle(t) \xrightarrow{P} \sigma^2(t)$$

2. For $t \in [0, \tau]$ and all $\epsilon > 0$

$$\langle \tilde{U}_\epsilon \rangle(t) \xrightarrow{P} 0$$

- First consider the first condition $\langle U\rangle(t) \xrightarrow{P} \sigma^2(t)$

$$\langle \tilde{U}\rangle(t) = n\int_0^t \left(\frac{\hat{S}(s-)}{S^*(s)}\frac{J(s)}{Y_\bullet(s)}\right)^2 Y_\bullet(s)\alpha(s)ds = \int_0^t \overbrace{\left(\frac{\hat{S}(s-)}{S^*(s)}\right)^2}^{\xrightarrow{P} 1 \text{ uniformly}} \frac{J(s)}{Y_\bullet(s)/n}\alpha(s)ds$$
$$\xrightarrow{P} \int_0^t \frac{\alpha(s)}{\mathrm{pr}(T \geq s)}ds = \sigma^2(t)$$

- Now consider the second condition $\langle \tilde{U}_\epsilon\rangle(t) \xrightarrow{P} 0$

$$\begin{aligned}\langle \tilde{U}_\epsilon\rangle(t) &= n\int_0^t \left(\frac{\hat{S}(s-)}{S^*(s)}\frac{J(s)}{Y_\bullet(s)}\right)^2 I\left(\frac{\sqrt{n}J(s)}{Y_\bullet(s)} > \epsilon\right) Y_\bullet(s)\alpha(s)ds \\ &= n\int_0^t \left(\frac{\hat{S}(s-)}{S^*(s)}\right)^2 \frac{J(s)}{Y_\bullet(s)} I\left(\sqrt{n}\frac{J(s)}{Y_\bullet(s)} > \epsilon\right)\alpha(s)ds \\ &\leq \frac{1}{S(t)^2}\underbrace{\int_0^t \frac{J(s)}{Y_\bullet(s)/n} I\left(\sqrt{n}\frac{J(s)}{Y_\bullet(s)} > \epsilon\right)\alpha(s)ds}_{\xrightarrow{P} 0 \text{ by } A4} \xrightarrow{P} 0\end{aligned}$$

Weak convergence

We have established that

$$\sqrt{n}\left(\frac{\hat{S}}{S^*} - 1\right) \tag{4}$$

converges weakly to a mean zero Gaussian martingale $-U$ with covariance function $\mathrm{cov}(U(s), U(t)) = \sigma^2(s \wedge t)$.

By $A4$ and arguments from the consistency proof,

$$\sup_{s \in [0,t]} \left| \sqrt{n}\left(\frac{S(s)}{S^*(s)} - 1\right) \right| \xrightarrow{P} 0 \tag{5}$$

Combining (4) and (5), we have that

$$\frac{\sqrt{n}(\hat{S} - S)}{S^*} \xrightarrow{\mathcal{L}} -U$$

Pointwise convergence in distribution

Let $\hat{\sigma}(t)^2$ be either of the two estimators of the variance of the Nelson-Aalen estimator from before. Then

$$\frac{\sqrt{n}(\hat{S}(t) - S(t))}{\hat{S}(t)\hat{\sigma}(t)} \overset{\text{approx.}}{\sim} N(0,1)$$

Pointwise convergence in distribution

Let $\hat{\sigma}(t)^2$ be either of the two estimators of the variance of the Nelson-Aalen estimator from before. Then

$$\frac{\sqrt{n}(\hat{S}(t) - S(t))}{\hat{S}(t)\hat{\sigma}(t)} \overset{\text{approx.}}{\sim} N(0, 1)$$

A $100(1-a)\%$ confidence interval for $S(t)$ is given by

$$\hat{S}(t) \pm c_{a/2} \frac{\hat{S}(t)\hat{\sigma}(t)}{\sqrt{n}}$$

where $c_{a/2}$ is the $1 - a/2$ quantile of the standard normal distribution.

This confidence interval may include probabilities larger than 1 and smaller than 0.

log-minus-log transformation

Consider the log-minus-log transformation. By the delta theorem,

$$\begin{aligned}\operatorname{var}\left(\log(-\log \hat{S}(t)\right) &\approx \left(\frac{\partial}{\partial x}\log(-\log x)\Big|_{x=S(t)}\right)^2 \operatorname{var}(\hat{S}(t)) \\ &= \frac{1}{(S(t)\log S(t))^2}\operatorname{var}(\hat{S}(t))\end{aligned}$$

which can be estimated by

$$\frac{\hat{S}(t)^2\hat{\sigma}(t)^2}{n\left(\hat{S}(t)\log \hat{S}(t)\right)^2} = \frac{\hat{\sigma}(t)^2}{n\left(\log \hat{S}(t)\right)^2}.$$

For large n,

$$\frac{\log(-\log \hat{S}(t)) - \log(-\log S(t))}{\hat{\sigma}(t)/(\sqrt{n}\log \hat{S}(t))} \overset{\text{approx.}}{\sim} N(0,1)$$

log-minus-log transformation

From

$$\frac{\log(-\log \hat{S}(t)) - \log(-\log S(t))}{\hat{\sigma}(t)/(\sqrt{n}\log \hat{S}(t))} \overset{\text{approx.}}{\sim} N(0,1)$$

A $100(1-a)\%$ confidence interval for $S(t)$ is given by

$$\begin{aligned}
&\exp\left(-\exp\left(\log\left(-\log \hat{S}(t)\right) \pm c_{a/2}\frac{\hat{\sigma}(t)}{\sqrt{n}\log \hat{S}(t)}\right)\right) \\
&= \exp\left(\overbrace{-\exp\left(\log\left(-\log \hat{S}(t)\right)\right)}^{=\log \hat{S}(t)} \exp\left(\pm c_{a/2}\frac{\hat{\sigma}(t)}{\sqrt{n}\log \hat{S}(t)}\right)\right) \\
&= \exp\left(\log\left(\hat{S}(t)^{\exp\left(\pm c_{a/2}\frac{\hat{\sigma}(t)}{\sqrt{n}\log \hat{S}(t)}\right)}\right)\right) \\
&= \hat{S}(t)^{\exp\left(\pm c_{a/2}\frac{\hat{\sigma}(t)}{\sqrt{n}\log \hat{S}(t)}\right)}
\end{aligned}$$

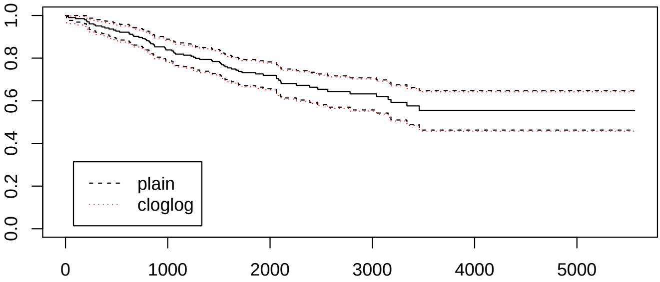
The log-minus-log transformation not only forces the interval to stay within $[0,1]$, it also has superior small sample properties.

Survival after melanoma surgery

```
## Kaplan-Meier curve
kmfit1 <- survfit(Surv(days, dead)~1, conf.type="plain", data=melanoma)
kmfit2 <- survfit(Surv(days, dead)~1, conf.type="log-log", data=melanoma)

plot(kmfit1)
lines(kmfit2$time, kmfit2$lower, lty=3, type="s", col="red")
lines(kmfit2$time, kmfit2$upper, lty=3, type="s", col="red")

legend("bottomleft",inset=.05,lty=2:3,col=1:2,legend=c("plain","cloglog"))
```



Quantiles

The pth quantile of the survival distribution, is the value t_p such that

$$p = \mathrm{pr}(T^* \leq t_p) = 1 - S(t_p).$$

The median corresponds to $p = 0.5$. t_p can be estimated by

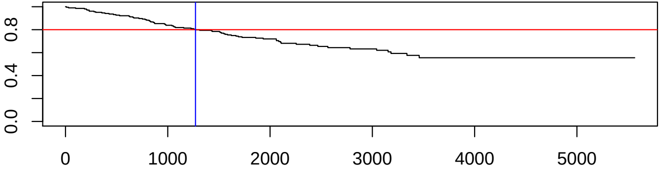
$$\hat{t}_p = \inf_t \left\{ t \geq 0 : \hat{S}(t) \leq 1 - p \right\}$$

That is, we draw a horizontal line in the Kaplan Meier plot at height $1 - p$ until it crosses the Kaplan Meier curve.

```
kmfit <- survfit(Surv(days, dead)~1, data=melanoma)
quantile(kmfit1, p=.2)$quantile

##   20
## 1271

plot(kmfit1, conf.int = FALSE)
abline(h=.8,col="red")
abline(v=quantile(kmfit1, p=.2)$quantile,col="blue")
```



Quantiles

We have that

$$\frac{\hat{S}(t) - S(t)}{\hat{S}(t)\hat{\sigma}(t)/\sqrt{n}} \overset{\text{approx.}}{\sim} N(0,1)$$

In order to test

$$H_0 : t_p = t_p^0 \text{ versus } H_1 : t_p \neq t_p^0$$

we use the test statistic

$$\frac{\hat{S}(t_p^0) - S(t_p^0)}{\hat{S}(t_p^0)\hat{\sigma}(t_p^0)/\sqrt{n}} = \frac{\hat{S}(t_p^0) - (1-p)}{\hat{S}(t_p^0)\hat{\sigma}(t_p^0)/\sqrt{n}} \overset{\text{approx. under } H_0}{\sim} N(0,1)$$

For a test with (approximate) significance level a, we reject H_0 when

$$\left| \frac{\hat{S}(t_p^0) - (1-p)}{\hat{S}(t_p^0)\hat{\sigma}(t_p^0)/\sqrt{n}} \right| > c_{a/2}.$$

Quantiles

Confidence intervals are obtained by inverting the confidence limits for the survival function. We get a $100(1-a)\%$ confidence interval for t_p as all t_p^0 values that are not rejected, i.e. all values t such that

$$\left|\frac{\hat{S}(t)-(1-p)}{\hat{S}(t)\hat{\sigma}(t)/\sqrt{n}}\right| \leq c_{a/2}$$

i.e. all t such that

$$\left|\hat{S}(t)-(1-p)\right| \leq c_{a/2}\hat{S}(t)\hat{\sigma}(t)/\sqrt{n}.$$

The standard $100(1-a)\%$ confidence interval for $S(t)$ is

$$\hat{S}(t) \pm c_{a/2}\hat{S}(t)\hat{\sigma}(t)/\sqrt{n} = [\hat{S}_L(t), \hat{S}_U(t)]$$

The confidence interval for t_p consists of all t where $\hat{S}(t)$ is closer to $1-p$ than the confidence limits $\hat{S}_L(t)$ and $\hat{S}_U(t)$. That is,

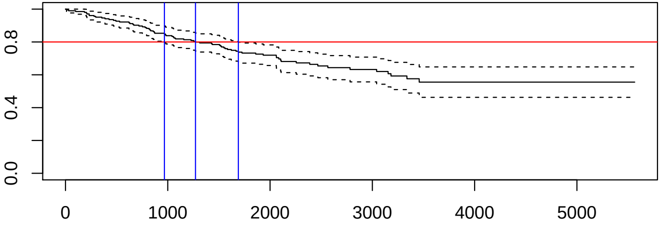
$$\left[\inf_t \left\{t \geq 0 : \hat{S}_L(t) \leq 1-p\right\}, \inf_t \left\{t \geq 0 : \hat{S}_U(t) \leq 1-p\right\}\right]$$

Quantiles with confidence intervals

```
kmfit <- survfit(Surv(days, dead)~1, data=melanoma)
plot(kmfit1, conf.int = TRUE)
abline(h=.8,col="red")
km20pct <- quantile(kmfit1, p=.2)
km20pct

## $quantile
##   20
## 1271
##
## $lower
##  20
## 967
##
## $upper
##   20
## 1690

abline(v=c(km20pct$lower,km20pct$quantile,km20pct$upper),col="blue")
```



The log-rank test

Consider $n = n_1 + n_2$ (censored) event times T_{ik}, $i = 1, \ldots, n_k$, from groups $k = 1, 2$. The processes $N_{\bullet k}(t)$ and $Y_{\bullet k}(t)$, count the number of events and individuals at risk in group k.

Assume that $N_{\bullet,k}(t)$ has the multiplicative intensity

$$\lambda_k(t) = \alpha_k(t) Y_{\bullet k}(t).$$

We will test the nullhypothesis

$$H_0 : \alpha_1(t) = \alpha_2(t) (= \alpha(t)) \text{ for all } t \in [0, \tau].$$

If H_0 is true, then $N_\bullet(t) = N_{\bullet 1}(t) + N_{\bullet 2}(t)$ is a univariate counting process with intensity $\alpha(t) Y_\bullet(t)$, where $Y_\bullet(t) = Y_{\bullet 1}(t) + Y_{\bullet 2}(t)$.

Idea: compare the group-specific Nelson-Aalen estimators

$$\hat{A}_k(t) = \int_0^t \frac{dN_{\bullet k}(s)}{Y_{\bullet k}(s)}$$

to those under H_0

$$\hat{A}(t) = \int_0^t \frac{dN_\bullet(s)}{Y_\bullet(s)}$$

using both groups.

The log-rank statistic

Consider the processes

$$R_k(t) = \int_0^t Y_{\bullet k}(s)\left(d\hat{A}_k(s) - d\hat{A}(s)\right) = \int_0^t dN_{\bullet k}(s) - \int_0^t \frac{Y_{\bullet k}(s)}{Y_{\bullet}(s)} dN_{\bullet}(s)$$

Under $H_0 : \alpha_1 = \alpha_2 = \alpha$,

$$dM_{\bullet k}(t) = dN_{\bullet k}(t) - Y_{\bullet k}(t)\alpha_k(t)dt = dN_{\bullet k}(t) - Y_{\bullet k}(t)\alpha(t)dt,$$

$$\begin{aligned} R_1(t) &= \overbrace{\int_0^t Y_{\bullet 1}(s)\alpha(s)ds - \int_0^t \frac{Y_{\bullet 1}(s)}{Y_{\bullet}(s)} Y_{\bullet}(s)\alpha(s)ds}^{=0} \\ &\quad + \int_0^t dM_{\bullet 1}(s) - \int_0^t \frac{Y_{\bullet 1}(s)}{Y_{\bullet}(s)} dM_{\bullet}(s) \\ &= \int_0^t \left(1 - \frac{Y_{\bullet 1}(s)}{Y_{\bullet}(s)}\right) dM_{\bullet 1}(s) - \int_0^t \frac{Y_{\bullet 1}(s)}{Y_{\bullet}(s)} dM_{\bullet 2}(s) \\ &= \int_0^t \frac{Y_{\bullet 2}(s)}{Y_{\bullet}(s)} dM_{\bullet 1}(s) - \int_0^t \frac{Y_{\bullet 1}(s)}{Y_{\bullet}(s)} dM_{\bullet 2}(s) \end{aligned}$$

Thus, $R_1(t)$ is a mean zero martingale under H_0.

Observed and expected number of events

The martingale property suggests interpreting the terms of R_1 as the observed and expected number of events

$$R_1(\tau) = \underbrace{\int_0^\tau dN_{\bullet 1}(s)}_{\text{observed}} - \underbrace{\int_0^\tau \frac{Y_{\bullet 1}(s)}{Y_\bullet(s)} dN_\bullet(s)}_{\text{expected}}$$

$$= O_1 - E_1$$

- O_1 is the observed number of events in group one
- E_1 can, for many purposes, be thought of as the "expected" number of events in group1 under H_0
- E_1 is stochastic, so it is not really an expectation, but when H_0 is true, $R_1(t)$ is a martingale, and thus $E(O_1) = E(E_1)$.

Variance of R_1

When M_1, M_2 are orthogonal martingales,

$$\langle M_1 + M_2\rangle(t) = \langle M_1\rangle(t) + \langle M_2\rangle(t).$$

Hence, the predictable variation of R_1 is

$$\begin{aligned}
\langle R_1\rangle(t) &= \int_0^t \left(\frac{Y_{\bullet 2}(s)}{Y_{\bullet}(s)}\right)^2 d\langle M_{\bullet 1}(s)\rangle + \int_0^t \left(\frac{Y_{\bullet 1}(s)}{Y_{\bullet}(s)}\right)^2 d\langle M_{\bullet 2}(s)\rangle \\
&= \int_0^t \frac{Y_{\bullet 2}(s)^2}{Y_{\bullet}(s)^2} Y_{\bullet 1}(s)\alpha(s)ds + \int_0^t \frac{Y_{\bullet 1}(s)^2}{Y_{\bullet}(s)^2} Y_{\bullet 2}(s)\alpha(s)ds \\
&= \int_0^t \left(\frac{Y_{\bullet 1}(s)Y_{\bullet 2}(s)^2}{Y_{\bullet}(s)^2} + \frac{Y_{\bullet 1}(s)^2 Y_{\bullet 2}(s)}{Y_{\bullet}(s)^2}\right)\alpha(s)ds \\
&= \int_0^t \frac{Y_{\bullet 1}(s)Y_{\bullet 2}(s)}{Y_{\bullet}(s)^2} \underbrace{(Y_{\bullet 2}(s) + Y_{\bullet 1}(s))}_{Y_{\bullet}(s)} \alpha(s)ds \\
&= \int_0^t \frac{Y_{\bullet 1}(s)Y_{\bullet 2}(s)}{Y_{\bullet}(s)}\alpha(s)ds,
\end{aligned}$$

which may be estimated by replacing $\alpha(s)ds$ by $d\hat{A}(t) = dN_{\bullet}(s)/Y_{\bullet}(s)$, as

$$\int_0^t \frac{Y_{\bullet 1}(s)Y_{\bullet 2}(s)}{Y_{\bullet}(s)} \frac{dN_{\bullet}(s)}{Y_{\bullet}(s)}.$$

Asymptotics

We verify the conditions for the marginale central limit theorem.
Assume that

- $\lim_{n\to\infty} n_k/n \to a_k > 0$
- $S_k(\tau)S_k^C(\tau-) > 0$, where $S_k(\cdot)$ and $S_k^C(\cdot)$ denote the survial functions of the event and censoring in group k.

Let

$$\tilde{R}_1(t) = \frac{1}{\sqrt{n}}\int_0^t \frac{Y_{\bullet 2}(s)}{Y_\bullet(s)}dM_{\bullet 1}(s) - \frac{1}{\sqrt{n}}\int_0^t \frac{Y_{\bullet 1}(s)}{Y_\bullet(s)}dM_{\bullet 2}(s)$$

The predictable variation of $\tilde{R}_1$ is

$$\begin{aligned}
\langle \tilde{R}_1\rangle(t) &= \frac{1}{n}\int_0^t \frac{Y_1(s)Y_2(s)}{Y_\bullet(s)}\alpha(s)ds \\
&= \frac{n_1 n_2}{n^2}\int_0^t \frac{Y_1(s)/n_1\, Y_2(s)/n_2}{Y_\bullet(s)/n}\alpha(s)ds \\
&\xrightarrow{P} a_1 a_2 \int_0^t \frac{\overbrace{S_1(s-)S_1^C(s-)}^{\text{pr}(T_{\mathbf{11}}\geq s)}\,\overbrace{S_2(s-)S_2^C(s-)}^{\text{pr}(T_{\mathbf{12}}\geq s)}}{a_1 S_1(s-)S_1^C(s-) + a_2 S_2(s-)S_2^C(s-)}\alpha(s)ds \\
&= a_1 a_2 \int_0^t \frac{S(s-)S_1^C(s-)S_2^C(s-)}{a_1 S_1^C(s-) + a_2 S_2^C(s-)}\alpha(s)ds = \sigma_R^2(t)
\end{aligned}$$

where $S(t) = \prod_{s\in(0,t]}(1-\alpha(s)ds) = \exp(-\int_0^t \alpha(s)ds)$.

Asymptotics

Let

$$\tilde{R}_{1,\epsilon}(t) = \frac{1}{\sqrt{n}} \int_0^t \frac{Y_{\bullet 2}(s)}{Y_\bullet(s)} I\left(\frac{1}{\sqrt{n}} \frac{Y_{\bullet 2}(s)}{Y_\bullet(s)} > \epsilon\right) dM_{\bullet 1}(s)$$

$$- \frac{1}{\sqrt{n}} \int_0^t \frac{Y_{\bullet 1}(s)}{Y_\bullet(s)} I\left(\frac{1}{\sqrt{n}} \frac{Y_{\bullet 1}(s)}{Y_\bullet(s)} > \epsilon\right) dM_{\bullet 2}(s)$$

The predictable variation of $\tilde{R}_{1,\epsilon}$ is

$$\langle \tilde{R}_{1,\epsilon} \rangle (t) = \int_0^t \left(\frac{1}{\sqrt{n}} \frac{Y_{\bullet 2}(s)}{Y_\bullet(s)}\right)^2 I\left(\frac{1}{\sqrt{n}} \frac{Y_{\bullet 2}(s)}{Y_\bullet(s)} > \epsilon\right) Y_{\bullet 1}(s)\alpha(s) ds$$

$$+ \int_0^t \left(\frac{1}{\sqrt{n}} \frac{Y_{\bullet 1}(s)}{Y_\bullet(s)}\right)^2 I\left(\frac{1}{\sqrt{n}} \frac{Y_{\bullet 1}(s)}{Y_\bullet(s)} > \epsilon\right) Y_{\bullet 2}(s)\alpha(s) ds$$

and converges to zero in probability as

$$\frac{1}{\sqrt{n}} \frac{Y_{\bullet 1}(s)}{Y_\bullet(s)} = \frac{1}{\sqrt{n}} \frac{Y_{\bullet 1}(s)/n}{Y_\bullet(s)/n} \xrightarrow{P} 0$$

We have verified the two conditions for the martingale central limit theorem. under H_0, $\tilde{R}_1$ converges weakly to a mean zero Gaussian martingale and

$$\frac{R_1(t)}{\sqrt{\int_0^t \frac{Y_{\bullet 1}(s) Y_{\bullet 2}(s)}{Y_\bullet(s)^2} dN_\bullet(s).}} \overset{\text{approx.}}{\sim} N(0,1)$$

Weights

Consider the weighted processes

$$R_k(t) = \int_0^t W(s) Y_{\bullet k}(s) \left(d\hat{A}_k(s) - d\hat{A}(s) \right)$$

where W is a nonnegative predictable weight.

- Choosing $W(t) = 1$, gives the log-rank test that is optimal (efficient) for proportional hazard alternatives, i.e., when, for all t,

$$\frac{\alpha_1(t)}{\alpha_2(t)} = c$$

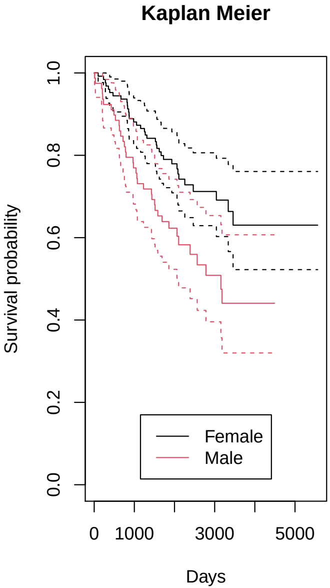
  for a constant c not depending on t. The log-rank test is the score test based on a Cox model.
- Many other suggestions exist, e.g. the Harrington & Fleming weights

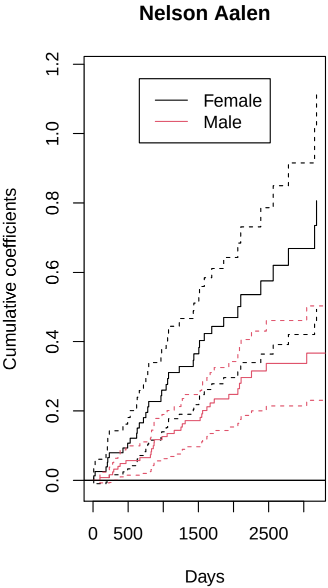
$$W(t) = \hat{S}(t-)^{\rho},$$

  where $\hat{S}(t)$ is the Kaplan-Meier estimator under H_0 and $\rho \in [0, 1]$. Choosing $\rho = 0$ gives the log-rank test, $\rho = 1$ gives a Wilcoxon type test which puts more weight on differences for small t.

Extension to k-sample tests is straightforward. Stratified tests also exist (see MS section 4.2.2).

Survival after melanoma surgery





Melanoma: sex

```
## Log-rank
survdiff(Surv(days, dead)~sex, data=melanoma)

## Call:
## survdiff(formula = Surv(days, dead) ~ sex, data = melanoma)
##
##         N Observed Expected (O-E)^2/E (O-E)^2/V
## sex=0 126       35     46.3      2.75       7.9
## sex=1  79       36     24.7      5.14       7.9
##
##  Chisq= 7.9  on 1 degrees of freedom, p= 0.005

## Gehan-Wilcoxon
survdiff(Surv(days, dead)~sex, rho=1, data=melanoma)

## Call:
## survdiff(formula = Surv(days, dead) ~ sex, data = melanoma, rho = 1)
##
##         N Observed Expected (O-E)^2/E (O-E)^2/V
## sex=0 126     28.3     37.9      2.43      8.28
## sex=1  79     30.0     20.4      4.50      8.28
##
##  Chisq= 8.3  on 1 degrees of freedom, p= 0.004
```

Melanoma: tumour thickness

```
## Three cathegories
melanoma$thickgrp <- cut(melanoma$thick, c(-Inf,120,310,Inf),
                          labels=c("thin","medium","thick"))
table(melanoma$thickgrp)

##
##   thin medium  thick
##     61     74     70

## Log-rank
survdiff(Surv(days, dead)~thickgrp, data=melanoma)

## Call:
## survdiff(formula = Surv(days, dead) ~ thickgrp, data = melanoma)
##
##                   N Observed Expected (O-E)^2/E (O-E)^2/V
## thickgrp=thin   61        9     23.7     9.114    13.711
## thickgrp=medium 74       23     27.1     0.612     0.989
## thickgrp=thick  70       39     20.2    17.399    24.437
##
##  Chisq= 27.2  on 2 degrees of freedom, p= 1e-06
```

Melanoma: tumour thickness stratified for sex

```
table(melanoma$sex, melanoma$thickgrp)

##
##     thin medium thick
##   0   40     55    31
##   1   21     19    39

prop.table(table(melanoma$sex, melanoma$thickgrp),1)

##
##          thin    medium     thick
##   0 0.3174603 0.4365079 0.2460317
##   1 0.2658228 0.2405063 0.4936709

## Stratified log-rank
survdiff(Surv(days, dead)~sex+strata(thickgrp), data=melanoma)

## Call:
## survdiff(formula = Surv(days, dead) ~ sex + strata(thickgrp),
##     data = melanoma)
##
##         N Observed Expected (O-E)^2/E (O-E)^2/V
## sex=0 126       35     42.8      1.43      3.88
## sex=1  79       36     28.2      2.17      3.88
##
##  Chisq= 3.9  on 1 degrees of freedom, p= 0.05
```