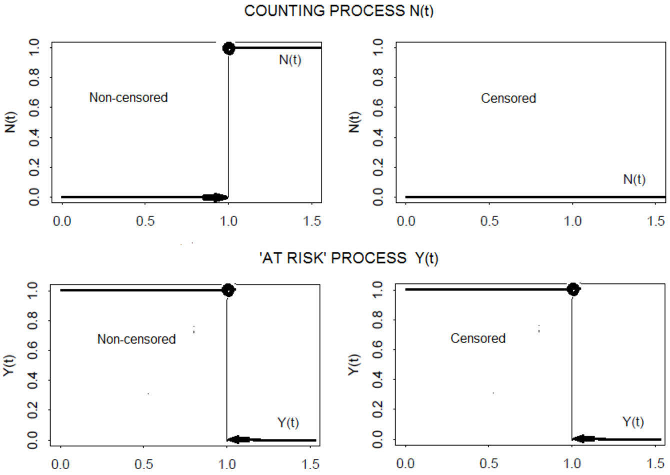
Right-censored data

$$N_i(t) = I(\tilde{T}_i \le t, D_i = 1) = \textit{counting process}$$
$$Y_i(t) = I(\tilde{T}_i \ge t) = \textit{'at risk' indicator}$$
$$\lambda_i(t) = \alpha_i(t)Y_i(t) = \textit{intensity function}$$



The aggregated counting process for right-censored sample with hazard $\alpha(t)$: Aalen's multiplicative model

$$
\begin{aligned}
N(t) &= \sum_{i=1}^{n} N_i(t) \\
Y(t) &= \sum_{i=1}^{n} Y_i(t) \\
\lambda(t) &= \sum_{i=1}^{n} \lambda_i(t) = \sum_{i=1}^{n} \alpha(t) Y_i(t) \\
&= \alpha(t) Y(t)
\end{aligned}
$$

Independent censoring

Let $\mathcal{F}_t$ be the history of all individuals, their censorings and failures, i.e., $\mathcal{F}_t$ contains $\{(N_i(s), Y_i(s)), s \leq t, i = 1, \ldots, n\}$

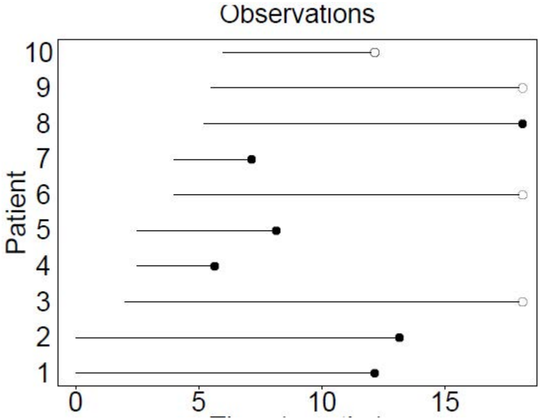
Independent censoring means by definition (p. 30-31 in book):

$$\begin{aligned} P(t \leq \tilde{T}_i < t + dt, D_i = 1 | \tilde{T}_i \geq t, \mathcal{F}_{t-}) &= P(t \leq T_i < t + dt | T_i \geq t) \\ &= \alpha_i(t)dt \end{aligned}$$

NOTE:

$$\textit{independent censoring} \Longleftrightarrow \lambda_i(t) = \alpha_i(t) Y_i(t)$$

Left truncation and right-censoring



Under *left-truncation*, the observation for the ith individual is

$$(V_i, \tilde{T}_i, D_i)$$

where V_i is the *left-truncation* time (i.e., the time of entry) for the individual, and $\tilde{T}_i$ and D_i are as before.

The counting process martingale is

$$M(t) = N(t) - \int_0^t \lambda(s)ds$$

Hence:

$$\begin{aligned} N(t) &= \int_0^t \lambda(s)ds + M(t) \\ &= \textit{predictable} \text{ increasing process} + \text{zero-mean } \textit{martingale} \\ &= \text{the (unique) } \textit{Doob-Meyer decomposition} \text{ for } N(t) \end{aligned}$$

Notation:

$\Lambda(t) \equiv \int_0^t \lambda(s)ds$ is called the **compensator** of $N(t)$.

Summing up key results on counting processes

- $N(t)$ is a counting process with history $\mathcal{F}_t$
- $P(dN(t) = 1|\mathcal{F}_{t-}) = \lambda(t)dt$
- $\Lambda(t) = \int_0^t \lambda(s)ds$ is predictable *(compensator)*
- $M(t) = N(t) - \Lambda(t)$ is a mean-zero martingale (Doob-Meyer)

Variation processes and stochastic integrals of counting process martingales

Predicitve: $\langle M \rangle (t) = \int_0^t \lambda(s) ds$

Optional: $[M](t) = N(t)$

$$\left\langle \int H dM \right\rangle (t) = \int_0^t H^2(s)\, \lambda(s) ds$$
$$\left[\int H dM \right] (t) = \int_0^t H^2(s)\, dN(s)$$

Multiplicative intensity and the Nelson-Aalen estimator

Multiplicative intensity model: $\lambda(t) = \alpha(t)Y(t)$

Thus

$$N(t) \overset{general}{=} \int_0^t \lambda(s)ds + M(t) \overset{mult.intens.}{=} \int_0^t \alpha(s)Y(s)ds + M(t)$$

so

$$dN(s) = \alpha(s)Y(s) + dM(s)$$

Dividing through by $Y(t)$ we get

$$\frac{1}{Y(s)}dN(s) = \alpha(s)ds + \frac{1}{Y(s)}dM(s)$$

and integrating we get

$$\int_0^t \frac{1}{Y(s)}dN(s) = A(t) + \int_0^t \frac{1}{Y(s)}dM(s)$$

The Nelson-Aalen estimator is then

$$\hat{A}(t) = \int_0^t \frac{1}{Y(s)}dN(s) = \sum_{T_j \leq t} \frac{1}{Y(T_j)}$$

Properties of the Nelson-Aalen estimator

We have

$$\hat{A}(t) - A(t) = \int_0^t \frac{1}{Y(s)} dM(s)$$

and hence $\hat{A}(t)$ is an unbiased estimator (if $Y(t) > 0$ for all t). Further, by formulas on previous slides,

$$\left\langle \hat{A} - A \right\rangle (t) = \int_0^t \frac{1}{Y^2(s)} \alpha(s) Y(s) ds = \int_0^t \frac{1}{Y(s)} \alpha(s) ds$$

$$\left[\hat{A} - A \right] (t) = \int_0^t \frac{1}{Y^2(s)} dN(s) = \sum_{T_j \leq t} \frac{1}{Y^2(T_j)} \tag{1}$$

We also know from our earlier formulas that

$$Var(\hat{A}(t)) = \text{Var}(\hat{A}(t) - A(t)) = E\left(\left[\hat{A} - A \right] (t) \right)$$

so that (1) is an unbiased estimator of $Var(\hat{A}(t))$ (if. $Y(t) > 0$ for all t)

Asymptotic distribution of Nelson-Aalen estimator

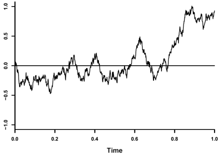
Suppose that our data were based on observation of a large number n units. We shall let n tend to infinity in the relation

$$\sqrt{n}(\hat{A}(t) - A(t)) = \int_0^t \sqrt{n}\frac{1}{Y(s)}dM(s)$$

and hence we need to look at the limiting behaviour of the mean zero martingale on the right hand side.

This brings us to the need for an asymptotic theory for martingales, see Slides 6.

Wiener process



A *Wiener process with variance parameter 1* is a stochastic process $W(t)$ with values in the real numbers satisfying

1. $W(0) = 0$
2. $W(t)$ has independent increments
3. For $s < t$, $W(t) - W(s)$ is normally distributed with expected value 0 and variance $(t - s)$
4. The paths are continuous

Gaussian martingales

$W(t)$ is itself a

- zero-mean martingale
- with predictable variation process $\langle W \rangle (t) = t$

Let $V(t)$ be a strictly increasing continuous function with $V(0) = 0$. Then the process

$$U(t) = W(V(t))$$

is also a

- a zero-mean martingale
- and has predictable variation process $\langle U \rangle (t) = V(t)$

(exercise 2.12 in book).

U is called a *Gaussian martingale*.

Rebolledo's martingale convergence theorem

Let $\tilde{M}^{(n)}(t)$ be a sequence of mean zero martingales for $t \in [0, \tau]$.

What is needed is that as $n \to \infty$,

(i) $\left\langle \tilde{M}^{(n)}(t) \right\rangle \to V(t)$ in probability for all $t \in [0, \tau]$ as $n \to \infty$

(ii) The sizes of the jumps of $\tilde{M}^{(n)}(t)$ go to zero in a certain sense

Then (Rebolledo, 1980):

$$\tilde{M}^{(n)}(t) \to U(t)$$

in distribution, as stochastic processes for $t \in [0, \tau]$, where $U(t)$ is a Gaussian martingale with predictable variation $V(t)$.

Special case of counting process martingales

$$\textit{Suppose} \qquad \tilde{M}^{(n)}(t) = \int_0^t H^{(n)}(s)dM^{(n)}(s)$$

$$\textit{where} \qquad M^{(n)}(t) = N^{(n)}(t) - \int_0^t \lambda^{(n)}(s)ds$$

Then

$$\left\langle \tilde{M}^{(n)}(t) \right\rangle (t) = \int_0^t (H^{(n)}(s))^2 \lambda^{(n)}(s)ds \stackrel{?}{\to} V(t)$$

The following are sufficient conditions for convergence in Rebolledo's theorem. If

(i) $(H^{(n)}(s))^2 \lambda^{(n)}(s) \to v(s) > 0$ for all $s \in [0, \tau]$

(ii) $H^{(n)}(s) \to 0$ for all $s \in [0, \tau]$

then $\tilde{M}^{(n)}(t) \to U(t)$ with $\langle U \rangle (t) = V(t) \equiv \int_0^t v(s)ds$.

Nelson-Aalen example

Recall that

$$\sqrt{n}(\hat{A}(t) - A(t)) = \int_0^t \frac{\sqrt{n}}{Y(s)} dM(s)$$

so that we have

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

Assume that there is a deterministic positive function $y(t)$ such that $Y(t)/n \to y(t) > 0$ in probability. Then the two sufficient conditions for Rebolledo's theorem are satisfied:

$$(H^{(n)}(s))^2 \lambda^{(n)}(s) = \frac{n}{Y^2(s)} \cdot \alpha(s) Y(s) = \frac{\alpha(s)}{Y(s)/n} \to \frac{\alpha(s)}{y(s)} \equiv v(s)$$

$$H^{(n)}(s) = \frac{\sqrt{n}}{Y(s)} = \frac{1/\sqrt{n}}{Y(s)/n} \to 0$$

Nelson-Aalen example (cont.)

Conclusion:

$$\sqrt{n}(\hat{A}(t) - A(t)) = \int_0^t \frac{\sqrt{n}}{Y(s)} dM(s)$$

converges in distribution to the mean zero Gaussian martingale $U(t) = W(V(t))$ with predictable variation process

$$V(t) = \int_0^t v(s) ds = \int_0^t \frac{\alpha(s)}{y(s)} ds$$

Thus, for a fixed t,

$$\sqrt{n}(\hat{A}(t) - A(t)) \xrightarrow{d} N\left(0, \int_0^t \frac{\alpha(s)}{y(s)} ds\right)$$

or, informally, $\hat{A}(t)$ is approximately normal with

$$\mathrm{Var}(\hat{A}(t)) \approx \frac{1}{n} \cdot \int_0^t \frac{\alpha(s)}{Y(s)/n} ds = \int_0^t \frac{\alpha(s)}{Y(s)} ds \approx \int_0^t \frac{dN(s)}{Y^2(s)} ds = \sum_{T_j \leq t} \frac{1}{Y^2(T_j)}$$