STK4080/9080 SURVIVAL AND EVENT HISTORY ANALYSIS

Slides 7: The Nelson-Aalen estimator

Bo Lindqvist
Department of Mathematical Sciences
Norwegian University of Science and Technology
Trondheim

https://www.ntnu.edu/employees/bo.lindqvist
bo.lindqvist@ntnu.no
boli@math.uio.no

University of Oslo, Spring 2021

The multiplicative intensity model (Ch. 3.1.2 in book)

The counting process $N(t)$ has intensity function

$$\lambda(t) = \alpha(t)Y(t)$$

where $\alpha(t) \geq 0$ is a *deterministic* parameter function and $Y(t)$ is a predictable process that does not depend on unknown parameters.

Typically, $Y(t)$ counts the number of units "at risk" at time t.

We have the general expression

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

so we can write

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

The Nelson-Aalen estimator in a minute (see Slides 6)

$$\begin{aligned} dN(s) &= \alpha(s)Y(s)ds + dM(s) \\ \frac{1}{Y(s)}dN(s) &= \alpha(s)ds + \frac{1}{Y(s)}dM(s) \\ \int_0^t \frac{1}{Y(s)}dN(s) &= A(t) + \int_0^t \frac{1}{Y(s)}dM(s) \end{aligned}$$

This suggests the (Nelson-Aalen) estimator

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

which is unbiased (?) and asymptotically normal, with variance estimator

$$\hat{\sigma}^2(t) = \int_0^t \frac{1}{Y(s)^2}dN(s) = \sum_{T_j \le t} \frac{1}{Y(T_j)^2}$$

WHY the (?) above: *A slight modification of the argument is needed since $Y(t)$ may be 0 ... See the last pages of these slides*

Calculation of Nelson-Aalen estimator (from Slides 4)

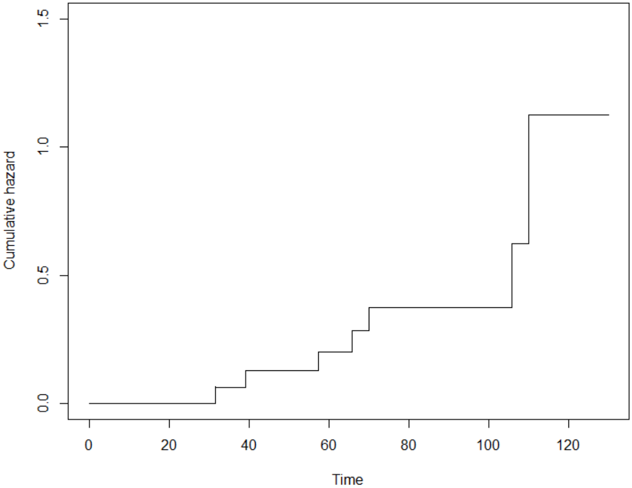
Let the survival data be (+ means right censored)

31.7	39.2	57.5	65.0+	65.8	70.0	75.0+	75.2+
87.5+	88.3+	94.2+	101.7+	105.8	109.2+	110.0	130.0+

Time	at risk	$\frac{d_i}{n_i}$	Nelson-Aalen estimate	
31.7	16	$\frac{1}{16}$	$\frac{1}{16}$	$= 0.06250$
39.2	15	$\frac{1}{15}$	$\frac{1}{16} + \frac{1}{15}$	$= 0.12917$
57.5	14	$\frac{1}{14}$	$\frac{1}{16} + \frac{1}{15} + \frac{1}{14}$	$= 0.20060$
65.8	12	$\frac{1}{12}$	$\frac{1}{16} + \frac{1}{15} + \frac{1}{14} + \frac{1}{12}$	$= 0.28393$
70.0	11	$\frac{1}{11}$	$\frac{1}{16} + \frac{1}{15} + \frac{1}{14} + \frac{1}{12} + \frac{1}{11}$	$= 0.37484$
105.8	4	$\frac{1}{4}$	$\frac{1}{16} + \frac{1}{15} + \frac{1}{14} + \frac{1}{12} + \frac{1}{11} + \frac{1}{4}$	$= 0.62484$
110.0	2	$\frac{1}{2}$	$\frac{1}{16} + \frac{1}{15} + \frac{1}{14} + \frac{1}{12} + \frac{1}{11} + \frac{1}{4} + \frac{1}{2}$	$= 1.12484$

Nelson-Aalen plot using R

31.7	39.2	57.5	65.0+	65.8	70.0	75.0+	75.2+
87.5+	88.3+	94.2+	101.7+	105.8	109.2+	110.0	130.0+



Pointwise confidence limits (p. 72 in ABG)

Recall the variance estimator

$$\hat{\sigma}^2(t) = \int_0^t \frac{1}{Y(s)^2} dN(s) = \sum_{T_j \leq t} \frac{1}{Y(T_j)^2}$$

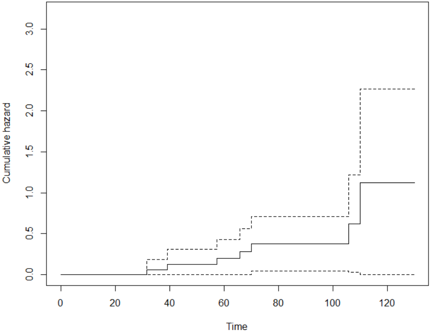
$100(1-\alpha)\%$ pointwise confidence limits are obtained as

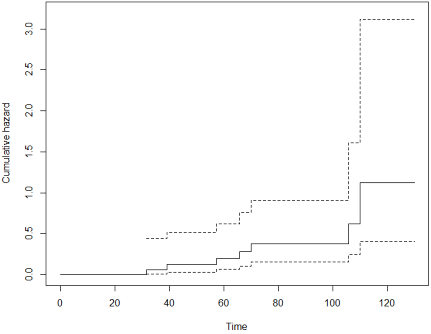
$$\hat{A}(t) \pm z_{1-\alpha/2} \hat{\sigma}(t)$$

Alternatively, by a log-transformation (Exercise 3.3) we can use

$$\hat{A}(t) \exp \left\{ \pm z_{1-\alpha/2} \frac{\hat{\sigma}(t)}{\hat{A}(t)} \right\}$$

Nelson-Aalen plots with confidence intervals (using R)





`conf.type="plain"`

`conf.type="log"`

Competing risks (ABG p. 77)

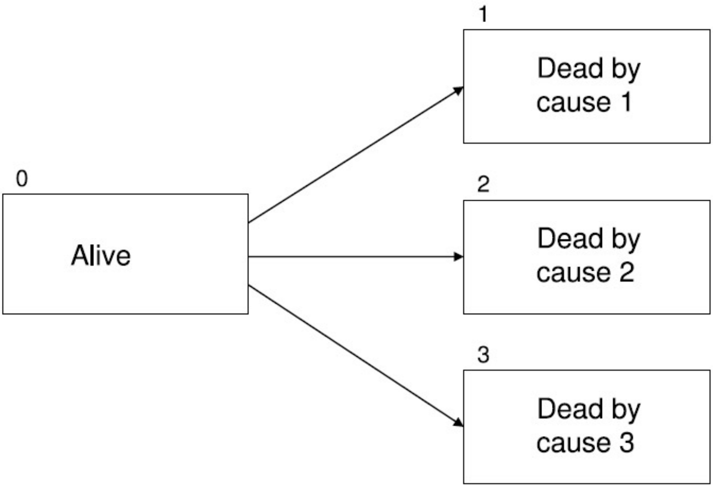


Fig. 3.5 *A model for competing risks with* $k = 3$.

The multiplicative intensity model in competing risks

Consider a **competing risks model** with k causes of death.

For each cause h we define the **cause-specific hazard** given by

$$\alpha_{oh}(t) = P(\text{die from cause } h \text{ in } [t, t + dt) \mid \text{alive at } t-)$$

Based on a sample from a population, we let N_{0h} count the number of observed $0 \rightarrow h$-transitions in $[0, t]$, and let $Y_0(t)$ be the number at risk (i.e. in state 0) just prior to time t.

The intensity process of N_{0h} takes the multiplicative form

$$\lambda_{0h}(t) = \alpha_{0h}(t) Y_0(t)$$

so the Nelson-Aalen estimator can be applied.

Competing risks and causes of death in Norway

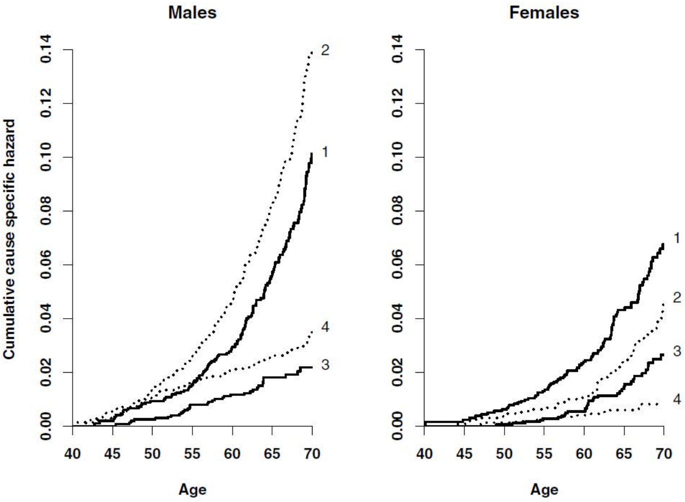


Fig. 3.6 *Nelson-Aalen estimates of the cumulative cause-specific hazard rates for four causes of death among middle-aged Norwegian males (left) and females (right). 1) Cancer; 2) cardiovascular disease including sudden death; 3) other medical causes; 4) alcohol abuse, chronic liver disease, and accidents and violence.*

Mulitiplicative model in *relative mortality*

Consider survival data where $N_i(t)$ as usual counts the observed number of deaths (0 or 1) for individual i. Assume that the intensity process takes the form

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

Here $Y_i(t)$ is the usual 'at risk' indicator for individual i, while $\mu_i(t)$ is the (*assumed known*) mortality rate of an individual of the same gender, age, etc. as individual i.

Assume that the individuals $i = 1, \dots n$ under study are from a specific population where one wants to study the relative mortality, denoted $\alpha(t)$, of this population as compared to the general population.

The aggregated counting process $N(t) = \sum_{i=1}^{n} N_i(t)$ has intensity process of the multiplicative form $\lambda(t) = \sum_{i=1}^{n} \lambda_i(t) = Y(t)\alpha(t)$, with

$$Y(t) = \sum_{i=1}^{n} Y_i(t)\mu_i(t)$$

which is a predictable process. **Nelson-Aalen estimator** is hence at hand.

Relative mortality after hip replacements (ABG p. 79)

Let t measure time since hip replacement.

Let $\mu_f(a), \mu_m(a)$ be (known) mortality rates for females and males, respectively, of age a.

Let g_i be gender and a_i the age of the ith patient.

Then the intensity function of the ith patient can be modeled by

$$\lambda_i(t) = Y_i(t)\alpha(t)\mu_{g_i}(a_i + t)$$

The aggregated process $N(t)$ thus has intensity of the multiplicative form $\lambda(t) = Y(t)\alpha(t)$ where

$$Y(t) = \sum_{i=1}^{n} Y_i(t)\mu_{g_i}(a_i + t)$$

(and is not integer valued as in earlier applications...)

Relative mortality after hip replacements (ABG p. 79)

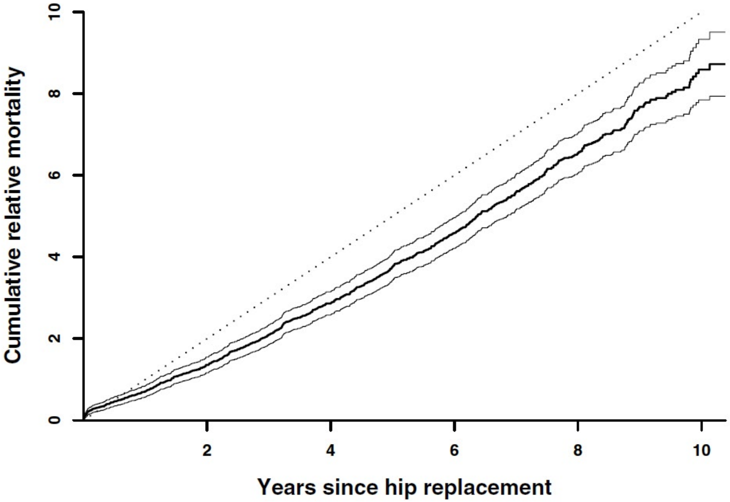


Fig. 3.7 *Nelson-Aalen estimate of the relative cumulative relative mortality with 95% standard confidence intervals for patients who have had a hip replacement in Norway in the period 1987–97. A dotted line with unit slope is included for easy reference.*

Small sample properties of the Nelson-Aalen estimator (3.1.5 in ABG)

Since there is a possibility that some $Y(t) = 0$, we introduce the indicator $J(t) = I\{Y(t) > 0\}$ and use the convention $0/0 = 0$.

Then, since $dN(s) = Y(s)\alpha(s)ds + dM(s)$,

$$\begin{aligned}
\hat{A}(t) &= \int_0^t \frac{dN(s)}{Y(s)} = \int_0^t \frac{J(s)}{Y(s)} dN(s) \\
&= \int_0^t \frac{J(s)}{Y(s)} \{Y(s)\alpha(s)ds + dM(s)\} \\
&= \int_0^t J(s)\alpha(s)ds + \int_0^t \frac{J(s)}{Y(s)} dM(s) \\
&\equiv A^*(t) + I(t) \\
(&\approx A(t) + \text{ mean-zero martingale })
\end{aligned}$$

Expectation and variance of the Nelson-Aalen estimator

$$\begin{aligned} E\{\hat{A}(t)\} &= E\{A^*(t) + I(t)\} = E\{A^*(t)\} \\ &= E\{\int_0^t J(s)\alpha(s)ds\} = \int_0^t P(Y(s) > 0)\alpha(s)ds \\ &\approx A(t) \end{aligned}$$

Furthermore, since $\hat{A}(t) - A^*(t) = \int_0^t \frac{J(s)}{Y(s)} dM(s)$, it follows that

$$\left[\hat{A} - A^*\right](t) = \int_0^t \frac{J(s)}{Y^2(s)} dN(s) = \sum_{T_j \le t} \frac{1}{Y^2(T_j)} \equiv \hat{\sigma}^2(t)$$

which is hence an unbiased estimator for the variance of $\hat{A}(t) - A^*(t)$ and hence *an approximately unbiased estimator of the variance of* $\hat{A}(t)$.

Here we use the general result that $Var(M(t)) = E[M](t)$

Asymptotics of Nelson-Aalen estimator (3.1.6 in ABG)

(See Slides 6 for details)

Asymptotically there is no difference between $A(t)$ and $A^*(t)$, and it hence follows from Rebolledo's theorem that both

$$\sqrt{n}(\hat{A}(t) - A^*(t)) \text{ and } \sqrt{n}(\hat{A}(t) - A(t))$$

converge 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 value t, the Nelson-Aalen estimator $\hat{A}(t)$ is approximately normally distributed with variance that can be estimated by

$$\hat{\sigma}^2(t) = \sum_{T_j \leq t} \frac{1}{Y^2(T_j)}$$