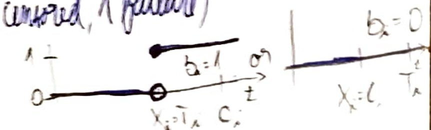
EXERCISE 2 - NON-PARAMETRIC ESTIMATION of CUM HAZARD and SURVIVAL FUNCTION

$T_i \overset{iid}{\sim} F$ with survival function $S(t) = 1-F(t)$, $t>0$ and cumulative hazard Λ

C_i censoring times such that $(T_1,C_1),\dots,(T_n,C_n)$ are independent and independent censoring condition holds

$X_i = T_i \wedge C_i$ censored failure times and $\delta_i = \mathbb{1}(T_i \le C_i)$ is failure indicator (0 censored, 1 failure)



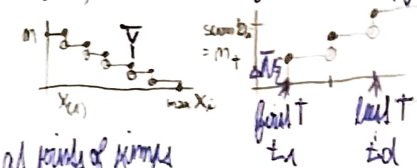
$N_i(t) = \mathbb{1}(X_i \le t, \delta_i = 1)$ counting process which reaches 1 after $X_i = T_i$, otherwise 0

$Y_i(t) = \mathbb{1}(X_i \ge t)$ at-risk process of i-th individual

left-continuous, drops after X_i -> no longer at risk

$\bar N(t) = \sum_{i=1}^{n} N_i(t)$ counting process of total number of observed deaths at t $\Rightarrow \Delta \bar N(t) = \bar N(t) - \bar N(t_-)$ is the number of new deaths at t

$\bar Y(t) = \sum_{i=1}^{n} Y_i(t)$ total number of individuals still at risk at time t - nonincreasing



Nelson-Aalen for Λ

$$\hat\Lambda^{NA}(t) = \int_0^t \frac{1}{\bar Y(s)}\, d\bar N(s)$$ ← counting process ~> integral becomes sum at points of jumps

$$= \sum_{\{j:\, t_j \le t\}} \frac{\Delta \bar N(t_j)}{\bar Y(t_j)}$$

← number of new deaths at time t_j (should be 1 if perfectly continuous)
← number of people still alive (and in the study) at t_j

where $\bar N$ jumps: $0 < t_1 < t_2 < \dots < t_d$ observed times of deaths

right continuous

only jumps preceding time t

$\hat\lambda_j$ ... estimates probability of dying at t_j after surviving till t_j

Fleming-Harrington for S based on $\hat\Lambda^{NA}$

$$\hat S^{FH}(t) = \exp\{-\hat\Lambda^{NA}(t)\}$$ based on well known relationship $S(t) = \exp\{-\Lambda(t)\}$

$$= \prod_{\{j:\, t_j \le t\}} \exp\left\{-\frac{\Delta \bar N(t_j)}{\bar Y(t_j)}\right\}$$

... but sum first and then use exp due to numerical issues

negative close to zero -> exp is close to one -> further multiplication reduces the estimate

Kaplan-Meier for S

$$\hat S^{KM}(t) = \hat P(\text{survive till time } t)\cdot \hat P(\text{survive beyond } t \mid \text{survived till time } t)$$

$$= \hat S^{KM}(t_-)\cdot\left(1 - \frac{\Delta \bar N(t)}{\bar Y(t)}\right)$$ ← estimates prob of surviving beyond t | alive at t)

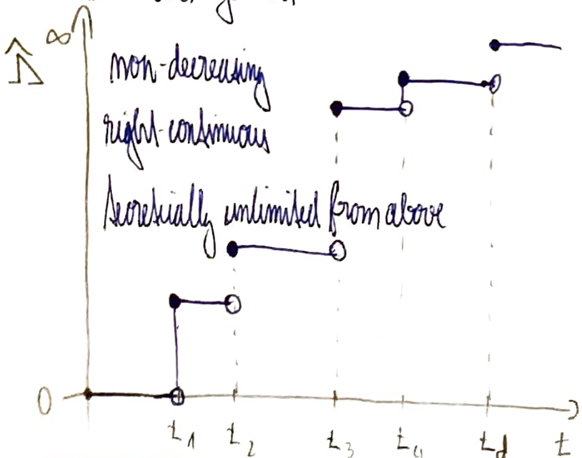
1 - prob of dying

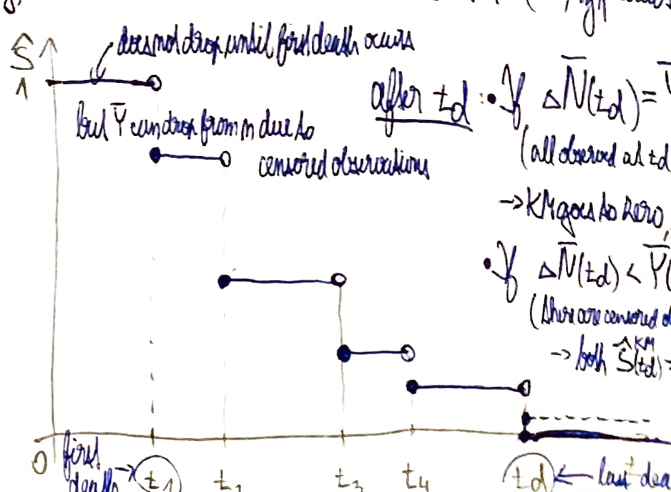
nonzero only at jumps of $\bar N$

$$= \prod_{\{j:\, t_j \le t\}} \left(1 - \frac{\Delta \bar N(t_j)}{\bar Y(t_j)}\right)$$

because the estimate starts at 1 (empty product is zero)

notice for h small is $\exp\{-h\} \approx 1-h$ -> $\hat S^{KM}$ and $\hat S^{FH}$ are very similar





<u>after t_d</u>: • If $\Delta \bar N(t_d) = \bar Y(t_d)$ ——
(all observed at t_d died)
-> KM goes to zero, but not FH

• If $\Delta \bar N(t_d) < \bar Y(t_d)$ - - -
(there are censored obs after t_d)
-> both $\hat S^{KM}(t_d) > 0$, $\hat S^{FH}(t_d) > 0$