Estimation of Survival Function $S(t)$

- Life table
- product-limit / Kaplan-Meier

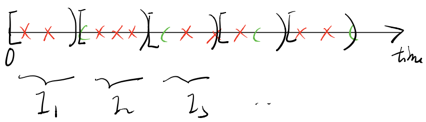
Assume: mixed of right-censored + random censored
non informative censoring

data: (T_i, δ_i) $i=1 \dots n$
$T_i = \min(X_i, C_i)$

① Life table

Using intervals:

Divide observation period into a series of time intervals
intervals depend on # samples in study, but usually 5 - 15



$I_j = [t_j, t_{j+1})$

let d_j : # death
c_j : # censored
n_j : $\underbrace{\text{\# at risk}}_{\text{alive}}$ at the start of I_j

Assume: b/c censoring, number of effective # at risk is not n_j

censoring time occurs uniformly over intervals I_j
so on average, withdrawn halfway through the interval
so that ave. # at risk during I_j is

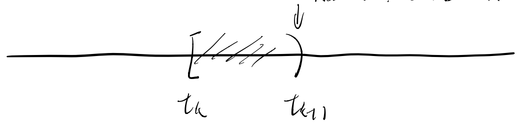
$$n_j' = n_j - \frac{c_j}{2}$$

derive estimate $S(t)$ during I_k:

$t_k \le t < t_{k+1}$:

$S(t) = P(X \geq t)$ prob of surviving beyond time t

Assume: events all happen at the end of intervals

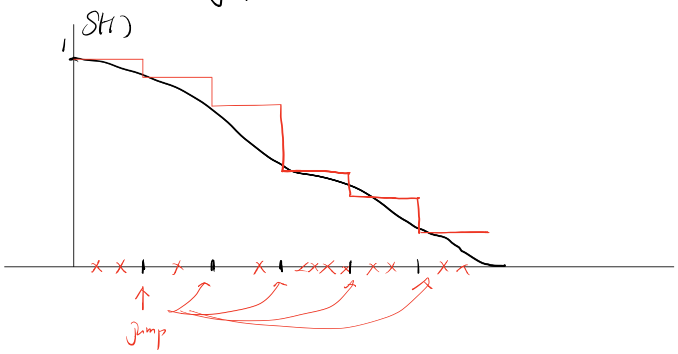


$= P(X \geq t_k)$

$= \underbrace{P(X \geq t_k \mid X \geq t_{k-1})}_{1 - \frac{d_{k-1}}{n'_{k-1}}} P(X \geq t_{k-1} \mid X \geq t_{k-2}) \cdots \underbrace{P(T \geq t_2 \mid T \geq t_1)}_{1 - \frac{d_1}{n_1}} \underbrace{P(T \geq t_1)}_{=1}$

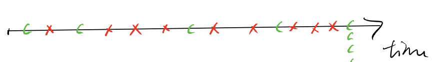
prob. of surviving interval I_{k-1}
given surviving through I_{k-2} = at risk during I_{k-1}

$$= \prod_{j=1}^{k-1} \left(1 - \frac{d_j}{n'_j}\right)$$



Sensitive to choice of intervals
well suited to grouped survival data

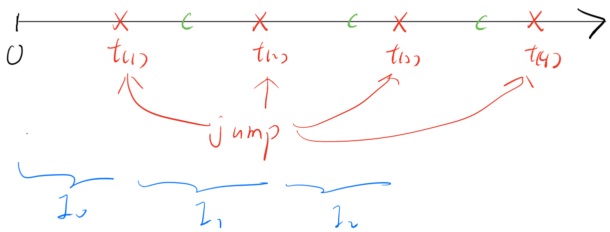
② Product-Limit / Kaplan Meier (KM)



if time:
- continuous: all T_i's will be distinct
- discrete: some T_i's will be same, in practice

Suppose events occur at D distinct times $t_{(1)} < t_{(2)} < \cdots < t_{(D)}$

KM: using exact event time $t_{(1)} < t_{(2)} < \cdots < t_{(D)}$



Let $I_j = [t_{(j)}, t_{(j+1)})$

n_j : # at risk just before $t_{(j)}$

d_j = # death at $t_{(j)}$

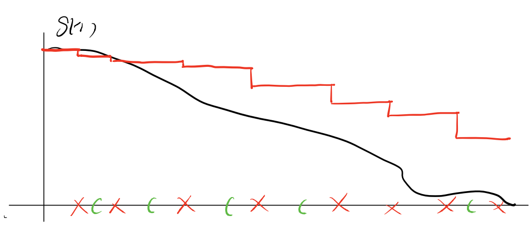
$t \in I_k$, $t_{(k)} \le t < t_{(k+1)}$

$$S(t) = P(X \geq t)$$

$$= P(X \geq t_{(k)})$$

$$= \underbrace{P(X \geq t_{(k)} \mid X \geq t_{(k-1)})}_{\substack{\text{prob. of surv. after } j\text{th death} \\ \text{given survival } (j-1)\text{th death} \\ 1 - \frac{d_k}{n_k}}} \cdots\cdots \underbrace{P(X \geq t_{(1)} \mid X \geq t_{(0)})}_{1 - \frac{d_1}{n_1}} \underbrace{P(X \geq t_{(0)})}_{\substack{\text{origin} \\ 1}}$$

$$= \prod_{j=1}^{k} \left(1 - \frac{d_j}{n_j}\right)$$



Think KM as the limit of life-table estimate if # interval ↗ , hence product-limit

③ Standard error of Estimated Survival Function (Uncertainty for point estimate)

Greenwood's formula: variance of KM estimator at time t:

for $t \in I_k$ i.e. $t_{(k)} \le t < t_{(k+1)}$)

$$\hat{V}(\hat{S}_{KM}(t)) = \hat{S}_{KM}^2(t) \sum_{j=1} \frac{d_j}{n_j(n_j - d_j)}$$

log-transf.

Estimation of Cumulative Hazard Function $H(t)$

$$H(t) = -\log S(t)$$

for $t_{(k)} \le t < t_{(k+1)}$

$$\Rightarrow \hat{H}_{KM}(t) = -\log \hat{S}_{KM}(t)$$

$$= -\log \prod_{j=1}^{k} \left(1 - \frac{d_j}{n_j}\right)$$

$$= -\sum_{j=1}^{k} \log\left(1 - \frac{d_j}{n_j}\right)$$

since $1 - \frac{d_j}{n_j} \approx e^{-\frac{d_j}{n_j}}$

$$\Rightarrow \hat{H}_{NA}(t) = -\sum_{j=1}^{k} \log\left[\exp\left(-\frac{d_j}{n_j}\right)\right]$$

$$= \sum_{j=1}^{k} \frac{d_j}{n_j}$$

$$\widehat{Var}(\hat{H}_{NA}(t)) = \sum_{j=1}^{k} \frac{d_j}{n_j^2}$$

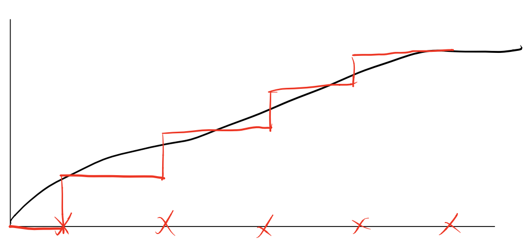


TABLE 1.1

Remission duration of 6-MP versus placebo in children with acute leukemia

Pair	*Remission Status at Randomization*	*Time to Relapse for Placebo Patients*	*Time to Relapse for 6-MP Patients*
1	Partial Remission	1	10
2	Complete Remission	22	7
3	Complete Remission	3	32^+
4	Complete Remission	12	23
5	Complete Remission	8	22
6	Partial Remission	17	6
7	Complete Remission	2	16
8	Complete Remission	11	34^+
9	Complete Remission	8	32^+
10	Complete Remission	12	25^+
11	Complete Remission	2	11^+
12	Partial Remission	5	20^+
13	Complete Remission	4	19^+
14	Complete Remission	15	6
15	Complete Remission	8	17^+
16	Partial Remission	23	35^+
17	Partial Remission	5	6
18	Complete Remission	11	13
19	Complete Remission	4	9^+
20	Complete Remission	1	6^+
21	Complete Remission	8	10^+

$^+$Censored observation

	Time	n_j	d_j	$\hat{S}_{KM}$	$\sqrt{\hat{V}(\hat{S}_{KM})}$	$\hat{H}_{NA}$	$\sqrt{\hat{V}(\hat{H}_{NA})}$
[6, 7)	6	24	4	0.833	0.0761	0.1667	0.083
[7, 10)	7	17	1				
[10, 13)	10	18	1				
[13, 16)	13	14	1				
[16, 22)	16	13	1				
[22, 23)	22	8	1				
[23, 35)	23	7	1				