muddies the picture. Maybe I'll make $\beta = X_n$

I. Review

Kaplan-Meier analysis is based on the alternative data layout.

1/16/2023

II. An Example of Kaplan-Meier Curves

Two groups of lukemia patients, 21 patients per group.
Group 1: treatment, Group 2: placebo

Group 1: 9 failed, 12 censored, ~~meaning all 21 went out of remission during the study period~~
Group 2: 21 failed, 0 censored

Group 1 (n=21) treatment	Group 2 (n=21) placebo
6 6 6 7 10	1 1 2 2 3
13 16 22 23	4 4 5 5
6+ 9+ 10+ 11+	8 8 8 8
17+ 19+ 20+	11 11 12 12
25+ 32+ 32+	15 17 22 23
34+ 35+	

Note + denotes censored

Group	# failed	# censored	Total
1	9	12	21
2	21	0	21

Descriptive statistics:

$\bar{T}_1$ (ignoring multiples) $= 17.1$ $\quad \bar{T}_2 = 8.6$

$\bar{h}_1 = 0.025$, $\bar{h}_2 = 0.115$, $\bar{h}_2 / \bar{h}_1 = 4.6$

Group 1 appears to have better overall survival prognosis than group 2 based on the descriptive statistics, but this does not compare the two groups at different follow up times.

- If no censored data, then can calculate survivorship as # surviving / # @ start

- KM formula involves the product of conditional probability terms; each term in the product is the probability of exceeding a specific ordered failure time $t_{(f)}$ given that a subject survives up to that time.

(Need to write out example).

Group 2

size of risk set → n_f; number of failures → m_f; # censored → q_f; naive survival function → $\hat{S}(t_{(f)})$

$t_{(f)}$	n_f	m_f	q_f	$\hat{S}(t_{(f)})$
0	21	0	0	1
1	21	2	0	19/21
2	19	2	0	17/21
3	17	1	0	16/21
4	16	2	0	14/21
5	14	2	0	12/21
8	12	4	0	8/21
11	8	2	0	6/21
12	6	2	0	4/21
15	4	1	0	3/21
17	3	1	0	2/21
22	2	1	0	1/21
23	1	1	0	0/21

Alternate formula for survival probability if censored

$\hat{S}(0) = 1$ always

Probability of exceeding 4 weeks can also be calculated by product limit

$$\hat{S}(4) = 1 \times \frac{19}{21} \times \frac{17}{19} \times \frac{16}{17} \times \frac{14}{16} = \frac{14}{21} = 0.67$$

$$\mathbb{P}(T>0 \mid T\geq 0)\,\mathbb{P}(T>1 \mid T\geq 1)\,\mathbb{P}(T>2 \mid T\geq 2)\,\mathbb{P}(T>3 \mid T\geq 3)\,\mathbb{P}(T>4 \mid T\geq 4)$$

Notice how the denominator of each term is the size of the remaining risk set before the next failure, and the numerator is that size minus the number of failures. Makes sense.

Now we account for censoring

Group 1

$t_{(f)}$	n_f	m_f	q_f	$\hat{S}(t_{(f)})$
0	21	0	0	1
6	21	3	1	$1 \times 18/21 = 0.8571$
7	17	1	1	$0.8571 \times 16/17 = 0.8067$
10	15	1	2	$0.8067 \times 14/15 = 0.7529$
13	12	1	0	$0.7529 \times 11/12 = 0.6902$
16	11	1	3	$0.6902 \times 10/11 = 0.6275$
22	7	1	0	$0.6275 \times 6/7 = 0.5379$
23	6	1	5	$0.5379 \times 5/6 = 0.4482$

- This estimate does not hit 0
- Note that the risk set shrinks to account for censoring.

III. General features of KM curves

$$\hat{S}(t_{(f)}) = \hat{S}(t_{(f-1)}) \times \hat{\mathbb{P}}(T>t_{(f)} \mid T\geq t_{(f)}) = \prod_{i=1}^{f} \hat{\mathbb{P}}(T>t_{(i)} \mid T\geq t_{(i)})$$

$\mathbb{P}(A \,\&\, B) = \mathbb{P}(A) \times \mathbb{P}(B \mid A)$

$A = T \geq t_{(f)}$, $B = T > t_{(f)}$

$\mathbb{P}(A \,\&\, B) = \mathbb{P}(B) = S(t_{(f)})$

There are no failures during $(t_{(f-1)}, t_{(f)})$ so

$\mathbb{P}(A) = \mathbb{P}(T > t_{(f-1)}) = S(t_{f-1})$

Putting it all together

$S(t_{(f)}) = S(t_{(f-1)})\,\mathbb{P}(T > t_{(f)} \mid T \geq t_{(f)})$