The mortality table context

- left truncation $\sim$ age at entry to the study
  right censoring $\sim$ end of study or time the subject leaves the study (time of "surrender")
- **Notation:**
  For the j^{th} observation,
- d_j ... truncation point ($d_j = 0$ if no truncation);
- x_j ... value of observation if **not** censored;
- u_j ... value of observation if censored.

The mortality table context (cont'd)

- **Notation of data summary:**
- $y_1 < y_2 < \cdots < y_k$ ... k **unique values of** $x_j's$;
- s_j ... the number of times y_j appears in the sample ($j = 1, \ldots, k$);
- r_j ... the risk set at the j^{th} ordered observation, i.e.,

$$\begin{aligned} r_j &= \sum_{i=j}^{k} s_i + \text{"number of } u_i's \geq y_j\text{"} \\ &\quad - \text{"number of } d_i's \geq y_j\text{"} \\ &= \text{"number of } d_i's < y_j\text{"} - \sum_{i=1}^{k-1} s_i \\ &\quad - \text{"number of } u_i's < y_j\text{"} \end{aligned}$$

- Recursion:

$$\begin{aligned} r_j = r_{j-1} &+ \text{"number of } d_i's \text{ between } y_{j-1} \text{ and } y_j\text{"} - s_{j-1} \\ &- \text{"number of } u_i's \text{ between } y_{j-1} \text{ and } y_j\text{"} \end{aligned}$$

with "between" a and b meaning in $[a, b)$

The Kaplan-Meier product-limit estimator

- This is an estimator of the survival function based on the mortality table described above.

$$S_n(t) = \begin{cases} 1, & 0 \le t < y_1, \\ \prod_{i=1}^{j-1} \left(\frac{r_i - s_i}{r_i} \right), & y_{j-1} \le t < y_j,\ j = 2, \ldots, k \\ \prod_{i=1}^{k} \left(\frac{r_i - s_i}{r_i} \right) \quad \text{OR} \quad 0 \quad \text{OR} \ldots, & t \ge y_k \end{cases}$$

- If $s_k = r_k$, then $S_n(t) = 0$ for $t \ge y_k$
- Otherwise, we can keep the survival function flat (the first option above) or set it at zero or do something else
- One possibility is to use exponential continuation with the Kaplan-Meier product limit estimator:

$$S_n(t) = e^{(t/w)\ln(s^*)}$$

with $w = \max\{x_1, x_2, \ldots, x_n, u_1, \ldots, u_n\}$ and

$$s^* = \prod_{i=1}^{k} \left(\frac{r_i - s_i}{r_i} \right)$$

A modification of the Nelson-Åalen estimate

- We can also recycle the Nelson-Åalen estimate interpreting the same notation in the context of the present section and set

$$\hat{S}(t) = e^{-\hat{H}(t)}, t < w$$

and

$$\hat{S}(t) = 0 \quad \text{OR} \quad \hat{S}(t) = (\hat{S}(y_k))^{t/w}$$