Kaplan-Meier estimation of survival function

• Kaplan and Meier (1958)

• One of the most-referenced scientific papers

• Non-parametric estimation of survival function

• Simulation tool at http://www.ms.uky.edu/∼mai/java/stat/KapMei.html

• Kaplan-Meier estimate of $S(t)$ is $\hat{S}(t)$, defined as follows:

$$\hat{S}(t) = \prod_{j=1}^{k} \left(\frac{n_j - d_j}{n_j} \right), t \in \left[t_{(k)}, t_{(k+1)} \right)$$

• $t_{(j)}$ is the j^{th} distinct event (death) time, $j = 0, 1, \ldots, r$

• d_j is the number of events (deaths) at time $t_{(j)}$

• n_j is the number alive (deaths) at time $t_{(j)^-}$

• Note the '$-$' in the definition of n_j, i.e. n_j is the number alive *immediately before* the death(s) at time $t_{(j)}$

• $\hat{S}(0) = 1$

• $\hat{S}(t) = 0, \forall t \geq t_{(r)}$ if $n_r = d_r$

• $\hat{S}(t)$ undefined if the last observation is censored

• Variance of $\hat{S}(t)$ estimated using Greenwood's formula, defined as follows:

$$\text{Var}\left(\hat{S}(t) \right) \approx \left(\hat{S}(t) \right)^2 \sum_{j=1}^{k} \frac{d_j}{n_j(n_j - d_j)}, t \in \left[t_{(k)}, t_{(k+1)} \right)$$

Recipe for Kaplan-Meier

• Ideal exam question!

• Order the data by event times

• If censored times tie with event (death) times, *place them after the deaths*

• Count observations, total deaths and total censored observations

• Write the five column headings: (i) Interval, (ii) n_j, (iii) d_j, (iv) $\dfrac{(n_j - d_j)}{n_j}$ and (v) $\hat{S}(t)$

• Fill in interval column. Begin with time zero, then list the distinct death times only (*not* censored times)

• Fill in the deaths column. d_j is zero at time zero and all other values should be 1 or greater

• Fill in the n_j column. n_j is the total number of observations at the start of the interval who either make it to the end or die during the interval, i.e. excluding observations censored during the interval

• Compute $\dfrac{(n_j - d_j)}{n_j}$

• Compute $\hat{S}(t)$