EXAM C QUESTIONS OF THE WEEK

S. Broverman, 2005

Week of December 26

The survival probability of the random variable T, $S(t)$ is estimated using product-limit estimation. Suppose that you are given the estimates of $S(y_k)$ and $S(y_{k+1})$ at two successive failure times as well as the Greenwood approximations for the variances of those estimates. Show how to find the Greenwood approximation of the variance of the product-limit estimate of the conditional survival probability $P(T > y_{k+1} | T > y_k)$.

Solution can be found below.

Week of December 26 - Solution

We use the usual notation s_i for the number of deaths at the i-th death point, and r_i for the number at risk at the i-th death point. The product-limit estimates are

$S_n(y_k) = (1 - \frac{s_1}{r_1})(1 - \frac{s_2}{r_2}) \cdots (1 - \frac{s_k}{r_k})$

and $S_n(y_{k+1}) = (1 - \frac{s_1}{r_1})(1 - \frac{s_2}{r_2}) \cdots (1 - \frac{s_k}{r_k})(1 - \frac{s_{k+1}}{r_{k+1}})$.

The product-limit estimate of the conditional survival probability $P(T > y_{k+1} | T > y_k)$ is $1 - \frac{s_{k+1}}{r_{k+1}}$, since we are measuring survival from time y_k (we "reset the survival clock" at that point for those who are still at risk).

The Greenwood approximations of the variances of $S_n(y_k)$ and $S_n(y_{k+1})$ are

$\widehat{V}((S_n(y_k)) = [S_n(y_k)]^2 \cdot \sum_{i=1}^{k} \frac{s_i}{(r_i - s_i) r_i} = [S_n(y_k)]^2 \cdot [\frac{s_1}{(r_1 - s_1) r_1} + \cdots + \frac{s_k}{(r_k - s_k) r_k}]$

and $\widehat{V}((S_n(y_{k+1})) = [S_n(y_{k+1})]^2 \cdot \sum_{i=1}^{k+1} \frac{s_i}{(r_i - s_i) r_i}$

$= [S_n(y_{k+1})]^2 \cdot [\frac{s_1}{(r_1 - s_1) r_1} + \cdots + \frac{s_k}{(r_k - s_k) r_k} + \frac{s_{k+1}}{(r_{k+1} - s_{k+1}) r_{k+1}}]$.

The Greenwood approximation of the variance of the product-limit estimate of the conditional survival probability $P(T > y_{k+1} | T > y_k)$ is $(1 - \frac{s_{k+1}}{r_{k+1}})^2 [\frac{s_{k+1}}{(r_{k+1} - s_{k+1}) r_{k+1}}]$.

This can be written as $(1 - \frac{s_{k+1}}{r_{k+1}})^2 [\frac{s_{k+1}/r_{k+1}}{(1 - \frac{s_{k+1}}{r_{k+1}}) r_{k+1}}]$.

Since we are assuming that the product limit estimates of $S(y_k)$ and $S(y_{k+1})$ are known, we have $\frac{S_n(y_{k+1})}{S_n(y_k)} = 1 - \frac{s_{k+1}}{r_{k+1}}$, and $1 - \frac{S_n(y_{k+1})}{S_n(y_k)} = \frac{s_{k+1}}{r_{k+1}}$.

Therefore, the only factor still needed to calculate $(1 - \frac{s_{k+1}}{r_{k+1}})^2 [\frac{s_{k+1}/r_{k+1}}{(1 - \frac{s_{k+1}}{r_{k+1}}) r_{k+1}}]$ is the value of r_{k+1} .

From the known Greenwood approximations of the variances of $S_n(y_k)$ and $S_n(y_{k+1})$, we see that $\frac{\widehat{V}((S_n(y_k))}{[S_n(y_k)]^2} = \frac{s_1}{(r_1 - s_1) r_1} + \cdots + \frac{s_k}{(r_k - s_k) r_k}$

and $\frac{\widehat{V}((S_n(y_{k+1}))}{[S_n(y_{k+1})]^2} = \frac{s_1}{(r_1 - s_1) r_1} + \cdots + \frac{s_k}{(r_k - s_k) r_k} + \frac{s_{k+1}}{(r_{k+1} - s_{k+1}) r_{k+1}}$.

Therefore, $\frac{\widehat{V}((S_n(y_{k+1}))}{[S_n(y_{k+1})]^2} - \frac{\widehat{V}((S_n(y_k))}{[S_n(y_k)]^2} = \frac{s_{k+1}}{(r_{k+1} - s_{k+1}) r_{k+1}}$, which can be written as

$\frac{\widehat{V}((S_n(y_{k+1}))}{[S_n(y_{k+1})]^2} - \frac{\widehat{V}((S_n(y_k))}{[S_n(y_k)]^2} = \frac{s_{k+1}/r_{k+1}}{(1 - \frac{s_{k+1}}{r_{k+1}}) r_{k+1}}$.

We have already seen that $1 - \frac{s_{k+1}}{r_{k+1}}$ and $\frac{s_{k+1}}{r_{k+1}}$ can be found from $S_n(y_k)$ and $S_n(y_{+1})$.

Since we are assuming that $\widehat{V}((S_n(y_k))$ and $\widehat{V}((S_n(y_{k+1}))$ are also known, we can find r_{k+1} to complete the calculation of $(1 - \frac{s_{k+1}}{r_{k+1}})^2 [\frac{s_{k+1}/r_{k+1}}{(1 - \frac{s_{k+1}}{r_{k+1}}) r_{k+1}}]$.