Mortality Estimation

Nelson-Åalen Estimator

1. This estimator estimates the cumulative hazard function.
2. Suppose:
   The cumulative hazard rate before time y_1 is known to be b.
   If at the time s_1 lives out of a risk set of r_1 die, the hazard at that time y_1 is $\frac{s_1}{r_1}$.
   Therefore, the cumulative hazard function is increased by that amount $\left(\frac{s_j}{r_j}\right)$, and becomes $b + \frac{s_1}{r_1}$.
   The Nelson-Åalen estimator sets $\widehat{H}(0) = 0$ and then at each time y_j at which an event occurs, $\widehat{H}(y_j) = \widehat{H}(y_{j-1}) + \frac{s_j}{r_j}$
3. Formula:

$$\widehat{H}(t) = \sum_{i=1}^{j-1} \frac{s_i}{r_i}, \qquad y_{j-1} \leq t \leq y_j$$

Variance of Kaplan-Meier and Nelson-Åalen Estimator

1. Formulas:
   The Kaplan-Meier estimator is an unbiased estimator of the survival function. Greenwood's approximation of the variance is:

$$\widehat{Var}\left(\hat{S}(t)\right) = \hat{S}(t)^2 \sum_{y_j \leq t} \frac{s_j}{r_j(r_j - s_j)}$$

   Klein's estimate of the variance of the Nelson-Åalen estimator is:

$$\widehat{Var}\left(\widehat{H}(t)\right) = \sum_{j|y_j \leq t} \frac{s_j(r_j - s_j)}{r_j^3}$$

$$\widehat{Var}\left(\widehat{H}(t)\right) = \sum_{j|y_j \leq t} \frac{s_j}{r_j^2}$$

   Both formulas are recursive formulas. So,

$$\widehat{Var}\left(\widehat{H}(y_j)\right) = \widehat{Var}\left(\widehat{H}(y_{j-1})\right) + \frac{s_j}{r_j^2}$$

   Using the delta method, the estimated variance of $\widehat{H}(y)$ is multiplied by $\hat{S}(y)^2$.
   Using Åalen's formula, $\widehat{Var}\left(\hat{S}(y)\right) = \hat{S}(y)^2 \sum_{y_j \leq y} \frac{s_i}{r_i^2}$
   Using Klein's formula, $\widehat{Var}\left(\hat{S}(y)\right) = \hat{S}(y)^2 \sum_{y_j \leq y} \frac{s_i(r_i - s_i)}{r_i^3}$
   The variance for estimates of S(y) with $y \geq y_{max}$ depend on the tail correction used. For Efron's tail correction the variance is 0, while for Klein-Moeschberger's it is the same as the variance of $S_n(y_k)$ when $y < \gamma$, 0 otherwise. For the exponential tail correction, the following formula follows from the delta method:

$$\widehat{Var}(S_n(y)) = \left(\frac{y}{y_{max}}\right)^2 \left(S_n(y_k)^{\frac{y}{y_{max}}-1}\right)^2 \widehat{Var}(S_n(y_k)) = \left(\frac{y}{y_{max}}\right)^2 \left(\frac{S_n(y)}{S_n(y_k)}\right)^2 \widehat{Var}(S_n(y_k))$$

2. Confidence intervals
   The usual symmetric normal confidence intervals for $S_n(t)$ and $\widehat{H}(t)$ can be constructed by adding and subtracting to the estimator the standard deviation times a coefficient from the standard normal distribution based on the confidence level. If z_p is the $100p^{th}$ quantile of a standard normal distribution, then the linear confidence interval for S(t) is,

$$\left(S_n(t) - z_{\frac{(1+p)}{2}}\sqrt{\widehat{Var}(S_n(t))}, S_n(t) + z_{\frac{(1+p)}{2}}\sqrt{\widehat{Var}(S_n(t))}\right)$$

   The resulting interval for $S_n(t)$ is, $\left(S_n(t)^{\frac{1}{U}}, S_n(t)^U\right)$, where $U = \exp\left(\frac{z_{\frac{(1+p)}{2}}\sqrt{\widehat{Var}(S_n(t))}}{S_n(t) ln S_n(t)}\right)$

   And the resulting interval for $\widehat{H}(t)$ is $\left(\frac{\widehat{H}(t)}{U}, \widehat{H}(t)U\right)$, where $U = \exp\left(\frac{z_{\frac{(1+p)}{2}}\sqrt{\widehat{Var}(\widehat{H}(t))}}{\widehat{H}(t)}\right)$