SensSpecConcCmpRsks052017

Weighted Concordance Estimates in the Presence of Competing Risks

Reference: Blanche et al Statist Med 2013, 32: 5381-5397

For unweighted estimates, the authors provide the R package "timeROC".

<u>Assumptions</u>: we wish to obtain a weighted estimate of the sensitivity, specificity and concordance of assigned risks as applied to observed outcomes for the first of K competing events. Here a "risk" is an assigned probability of failing with event 1 within t* time units of cohort entry. Subjects who fail of events 2,…,K before failing of event 1 are considered outcome-negative.

<u>Notation</u>: Let T^k denote a person's unobserved time to outcome k, $k = 1,\ldots,K$, and let $T = min(T^1,\ldots,T^K)$, with $S_T(t) = Pr\ (T>t)$ denoting the overall survival function. Also let Z denote independent time to censoring, with $S_Z(t) = Pr\ (Z>t)$. Let $\hat{S}_Z(t)$ denote the corresponding weighted Kaplan-Meier survival function estimate.

For each of a random sample of N subjects, we observe data (X,δ,R) where R is his/her assigned risk, $X = \min(T,Z)$, $\delta \in \{0,1,\ldots,K\}$, with $\delta=0$ indicating $X = Z$. We wish to use these data to obtain weighted estimates of the model sensitivity at cutpoint c

$$Se(c) = \Pr(R > c \mid T \le t^*, \delta = 1), \tag{1}$$

model specificity at cutpoint c

$$Sp(c) = \Pr(R \le c \mid \{T > t^*\} \cup \{T \le t^*, \delta > 1\}) \tag{2}$$

and model concordance

$$C = \Pr(R_i > R_j \mid T_i \le t^*, \delta_i = 1, \{T_j > t^*\} \cup \{T_j \le t^*, \delta_i > 1\}) \ . \tag{3}$$

In addition we assign each subject a nonnegative weight w satisfying $\sum_{i=1}^N w_i = N$. The weighted versions of the estimates proposed by Blanche et al (2013) for $Se(c), Sp(c)$ & C are

$$\begin{aligned} \widehat{Se}(c) &= \sum_{i=1}^N w_i A_i 1(R_i > c) \Big/ \sum_{i=1}^N w_i A_i \\ \widehat{Sp}(c) &= \sum_{i=1}^N w_i B_i 1(R_i \le c) \Big/ \sum_{i=1}^N w_i B_i \end{aligned} \tag{4}$$

and

$$\hat{C} = \frac{\sum_{i=1}^N \sum_{j=1}^N w_i w_j A_i B_j 1(R_i > R_j)}{\left(\sum_{i=1}^N w_i A_i\right)\left(\sum_{j=1}^N w_j B_j\right)}, \tag{5}$$

where

$$A_i = 1\left(X_i \le t^*, \delta_i = 1\right) / \hat{S}_Z\left(X_i\right)$$
$$B_j = \left[1\left(X_j > t^*\right) / \hat{S}_Z\left(t^*\right)\right] + \left[1\left(X_j \le t^*, \delta_j > 1\right) / \hat{S}_Z\left(X_j\right)\right] \quad (6)$$

The empirical ROC curve is the plot of points $\left(1-\widehat{Sp}\left(c\right), \widehat{Se}\left(c\right)\right)$ as c ranges from 1 to 0.

Special cases:

a) Unweighted case: $w_1 = \cdots = w_N = 1$. Then the estimates (4,5) agree with formula (6,7,9) of Blanche et al.
b) Unweighted case, K = 1 (no competing risks). Then (3) becomes

$$C = \Pr\left(R_i > R_j \mid T_i \le t^*, T_j > t^*\right). \quad (7)$$

and (5) becomes

$$\hat{C} = \left(\sum_{i=1}^{N} \sum_{j=1}^{N} A_i B_j 1\left(R_i > R_j\right)\right) / \left\{\left(\sum_{i=1}^{N} A_i\right)\left(\sum_{j=1}^{N} B_j\right)\right\}. \quad (8)$$

Note that N^{-2} times the numerator of (8) is a consistent estimate of $\Pr\left(T_i \le t^*, T_j > t^*, R_i > R_j\right)$. Also $N^{-1}\sum_{i=1}^{N} A_i = N^{-1}\sum_{i=1}^{N} 1\left(X_i = T_i \le t^*\right) / \hat{S}\left(X_i\right)$ is a consistent estimator of $\Pr\left(T_i \le t^* \mid T_i < Z_i\right) = \Pr\left(T_i \le t^*\right)$ and $N^{-1}\sum_{j=1}^{N} B_j = N^{-1}\sum_{j=1}^{N} 1\left(X_j > t^*\right) / \hat{S}_Z\left(t^*\right)$ is a consistent estimator of $\Pr\left(T_j > t^* \mid Z_j > t^*\right) = \Pr\left(T_j > t^*\right)$. Thus $\hat{C}$ of (8) is consistent for the concordance measure C of (3).

Blanche et also give an estimate for the variance of their concordance estimate, which they show is consistent.

Questions/tasks:

1. Prepare code to calculate the weighted estimates (4,5) and the ROC curve resulting from varying c from 1 to 0 and plotting the points (x,y) = [[]].
2. Generate a data set with times to K = 2 competing outcomes (disease & death) and censored survival data. Check that the estimate obtained using Gail's code agrees with that of Blanche et al when all subjects' weights are one.
3. Send the data set to Scott & ask him to use his code to calculate the Hung and Chiang estimate.