

Producing multistage survival curves

DJS 8th October 2020

- Run a Cox proportional hazards model for each individual cause of leaving the list, removing all other losses as censoring.
- For cause i and patient with covariates x , we can then estimate a cumulative hazard model $H_i(t, x) = H_{i0}(t) \exp[a'_i x]$, where
 - t is week post-listing (or this could be done in days if week is not smooth enough)
 - a_i are the estimated log-hazard ratios for the i th cause of loss from the waiting list,
 - $H_{i0}(t)$ is a smooth model for the cumulative hazard, say through using fractional polynomials.
 - $h_i(t, x)$ is the instantaneous hazard at time t , ie the risk of leaving with cause i , given still on list at time t
- Assuming independent causes of leaving the list, the probability of still being on the list at time t is

$$S(t, x) = \exp\left[-\sum_i H_i(t, x)\right].$$

- We want to estimate the probability $F_i(t, x)$ of an individual having left the list because of cause i by time t , which is what we want to plot. A loss of type i in time period $t, t + dt$ occurs with probability $h_i(t, x)S(t, x)dt$, so that

$$F_i(t, x) = \int h_i(t, x)S(t, x)dt.$$

- This is not available in closed-form, but can be obtained in discrete time. Set $S(1, x) = 1, F_i(1, x) = 0$. Now run through the following steps for $t = 1, 2, \dots$
 - Assuming a patient is still on the list at the start of week t , estimate the probability of a loss of type i as $h_i(t, x) = H_i(t + 1, x) - H_i(t, x)$.
 - Calculate $p_i(t, x) = h_i(t, x)S(t, x)$ as the probability of leaving the list with cause i
 - Add these to form $F_i(t + 1, x) = F_i(t, x) + p_i(t, x)$.
 - Set $S(t + 1, x) = S(t, x) - \sum_i p_i(t, x)$
- Then $S(t, x), F_i(t, x)$ will add to 1 and form the bands to be plotted.
- Alternatively this could be carried out with simulation, although not sure there is much point
 - Assuming a patient is still on the list at the start of week j , estimate the probability of a loss of type i as $h_i(t, x) = H_i(t + 1, x) - H_i(t, x)$. Form a vector $p = (h_1(t, x), h_1(t, x) + h_2(t, x), \dots, 1 - \sum_i h_i(t, x))$.

- Start with 10,000 (or maybe 1000 is enough) individuals at time 0. Cycle through each week, picking a random number between 0 and 1, and if it lies in one of appropriate categories of p , then consider that event to have occurred, and stop the cycle. If it lands in the final category, no event has occurred and the cycle continues.
- Continue until you have 10,000 trajectories, and then plot accordingly.
- If the predictors are in categories, then this process could be carried out for all combinations and the resulting curves stored. Otherwise could be done on the fly, maybe with 1000 trajectories.