Survival analysis 2020/2021

Exercises week 4

Exercise 1 : Restricted mean life

Assume we have n (possibly censored) i.i.d. observations on the continuous event time T^* with absolutely continous survival function $S(t) = \text{pr}(T^* > t)$. Assume that $E(T^*) < \infty$. You may assume any additional regularity that you need.

(a) Show that $E(\tilde{T}) = \int_0^\infty S(t)dt$

(b) Typically the mean $E(\tilde{T})$ is difficult to estimate, but one can estimate the restricted mean $E(\tilde{T} \wedge \tau)$, for some $\tau < \infty$, why? Show that $E(\tilde{T} \wedge \tau) = \int_0^\tau S(t)dt$. Another useful quantity is the restricted expected residual life, $E((\tilde{T} \wedge \tau) - s|\tilde{T} > s)$. Find a formula for this quantity in terms of $S(\cdot)$

(c) Load the `TRACE` data:

```
library(timereg)
data(TRACE)
```

and consider the time to death given by `time` and the event status defined from `status!=0`. Estimate $E(\tilde{T} \wedge \tau)$, with $\tau = 7$, for people with (`vf`=1) or without (`vf`=0) based on Kaplan-Meier estimates for the two groups. How do we interpret the numbers and their difference $E(\tilde{T} \wedge \tau | \texttt{vf} = 1) - E(\tilde{T} \wedge \tau | \texttt{vf} = 0)$? The restricted mean (with standard error) can be obtained from `print.survit`.

(d) Fit a model with constant hazard for each group (`vf`=0, `vf`=1) in the `TRACE` data and estimate the restricted mean life (up to $\tau = 7$ years) for the two groups. Compute standard errors based on the delta-theorem. That is, theoretically derive the estimator and the expression for the standard error, and use the estimators to estimate the quantities from the data). Compare your estimates to those based on the Kaplan-Meier estimate (question 3).

Exercise 2 : Kaplan-Meier

Assume we have n i.i.d. observations on the continuous event time $\tilde{T}$ with hazard $\alpha(t)$. Let $A(t) = \int_0^t \alpha(s)ds$ denote the cumulative hazard and $S(t) = \text{pr}(\tilde{T} > t) = \exp(-A(t))$ the survival function. You may assume all additional regularity that you need.

(a) Let $\mathbb{F}_n(t) = n^{-1}\sum_{i=1}^{n} I(\tilde{T}_i \leq t)$ denote the empirical distribution function of $\tilde{T}$. Show that, when there is no censoring, the Kaplan-Meier estimator is equivalent to $1 - \mathbb{F}_n(t)$.
Hint: $1 - 1/n = (n-1)/n$.

(b) Still without censoring, show that Greenwood's formla for the variance of the Kaplan-Meier estimator (cf. MS, p. 83) equals the binomial variance for $\hat{S}(t)$. That is, show that, without censoring, with $N_\bullet$ and $Y_\bullet$ the aggregated observed event and at-risk processes,

$$\hat{S}(t)^2 \int_0^t \frac{dN_\bullet(s)}{Y_\bullet(s)(Y_\bullet(s) - \Delta N_\bullet(s))}$$

(where ΔA denotes the jumps of A) equals

$$\frac{\hat{S}(t)(1-\hat{S}(t))}{n}$$

(c) Now assume random censoring, and let $\hat{S}_C$ denote the the Kaplan-Meier estimator of the censoring distribution and let $\hat{S}(t)$ denote the Kaplan-Meier estimator of $S(t)$. Compute $\hat{S}_C(t)\hat{S}(t)$.

(d) Let $\hat{A}(t)$ denote the Nelson-Aalen estimator of $A(t)$. Argue that $\exp(-\hat{A}(t))$ is asymptotically equivalent to the Kaplan-Meier estimator $\hat{S}(t)$

(e) Load the TRACE data:

```
library(timereg)
data(TRACE)
```

Consider the time to death given by time and the event status defined from `status!=0`. Plot and compare the Kaplan-Meier estimator and the estimator $\exp(-\hat{A}(t))$.

Exercise 3 : Duhamel's equation

Consider a survival time T^* with absolutely continuous hazard α. Let $\hat{S}(t)$ denote the Kaplan-Meier estimator for the survival function $S(t) = \exp\Big(-\int_0^t \alpha(u)du\Big)$

(a) Show that

$$\frac{\hat{S}(t)}{S(t)} - 1 = -\int_0^t \frac{\hat{S}(s-)}{S(s)} d(\hat{A} - A)(s)$$

Hint: It is sufficient to show that both sides of the expression have the same increments. Why?

A more formal argument can be established using integration by parts for Stieltjes integrals: For right-continuous real valued functions F and G of bounded variation on any finite interval,

$$F(t)G(t) - F(0)G(0) = \int_0^t F(u-)dG(u) + \int_0^t G(u)dF(u).$$

(b) Show the identity (MS 4.3)

$$\frac{\hat{S}(t)}{S^*(t)} - 1 = -\int_0^t \frac{\hat{S}(s-)}{S^*(s)} \frac{I(Y_\bullet(s) > 0)}{Y_\bullet(s)} dM_\bullet(s)$$

where $Y_\bullet$ and $M_\bullet$ denote the aggregated at-risk and martingale processes, $\hat{S}(t)$ denotes the Kaplan-Meier estimator for the survival function $S(t)$, and $S^*(t) = \exp(-\int_0^t I(Y_\bullet(s) > 0)\alpha(s)ds)$.