Efficient Estimation of Survival Prognosis Under Immortal Time Bias

Nima Hejazi, Kevin Benac, Nicholas P. Jewell

Group in Biostatistics & Department of Statistics, University of California, Berkeley

School of Public Health
UNIVERSITY OF CALIFORNIA, BERKELEY

Overview & Motivations

1. We consider the problem of efficiently estimating survival prognosis under a data structure complicated by the presence of immortal time bias.
2. The matter of efficient estimation under a bias induced by time-dependent risks presents a novel challenge that received surprisingly meager attention in the literature.
3. We compare parametric and nonparametric estimators of survival, including variations of the Cox proportional hazards model and the Kaplan-Meier estimator, evaluating the efficiency of each in the estimation of the multiple survival processes that occur under this data-generating process.
4. We are given survival times for patients with a single primary melanoma, and some of the patients develop a second primary melanoma before dying.

Introduction & Data

- Question of interest: **How does the second melanoma change the survival prognosis of the patients?**
- In order to prepare for a real data analysis, we simulate a data structure that matches what we expect — that is, the data-generating process is the the Cox proportional hazards model.
- Survival time T: time before the actual death of the patient,
- Time until second melanoma appears: U,
- Baseline hazard in absence of second melanoma: $\lambda_0(t)$,
- Time-varying covariate: $Z(t) = I(t > U)$.
- Constant baseline hazard $\lambda_0 = \lambda$.
- A second melanoma doubles the hazard.

Methodology I

- Time origin for all sujects: date of their first or index primary melanoma (PM).
- Two hazard functions of essential interest:
- $\lambda_1(t)$ — hazard of an individual, alive at time t who has only experienced one PM.
- $\lambda_2(t)$ — hazard of an individual, alive at time t who has experienced more than one PM.
- Data: of $n = n_1 + n_2$ subjects.
- The first n_1 only experience one PM before death at the observed time t_i. The other n_2 experience a second PM at the observed time u_j and then die at observed time t_j. It is possible to consider censored observations for both sets of subjects but we do not discuss this here for the sake of notation.
- We compare three approaches of this problem, namely, the Cox proportional hazards model, the method presented in Youlden et al. [1] and its correction by Jewell.

The basic proportional hazards model is a semi-parametric model for the hazard function defined by

$$\lambda\left(t; Z = z\right) = \lambda_0(t) \exp\left(\beta^T z\right), \quad t \geq 0. \tag{1}$$

where $\lambda_0(\cdot)$ is the baseline hazard function is estimated non-parametrically, while β is the vector of regression coefficients and is estimated parametrically using Cox's partial likelihood.

Methodology II

- The second approach is non-parametric and uses Kaplan-Meier's estimator defined as

$$\widehat{S}(t) = \prod_{i:t(i)<t} \left(1 - \frac{d_i}{n_i}\right), \quad t \geq 0,$$

where d_i and n_i are the respective numbers of death and individual at risks at the ordered time $t^{(i)}$, $i = 1, \ldots, n$.

- Youlden et al. [1] only uses patients for whom no occurrence of a second melanoma is observed, in the estimation of S_1 and ignores the other patients, which causes a bias.
- Jewell corrects their estimator by including all the patients in the study.
- The ones that were excluded by Youlden et al. [1] still contain information about λ_1: those are censored observations at time U.

Results & Discussion

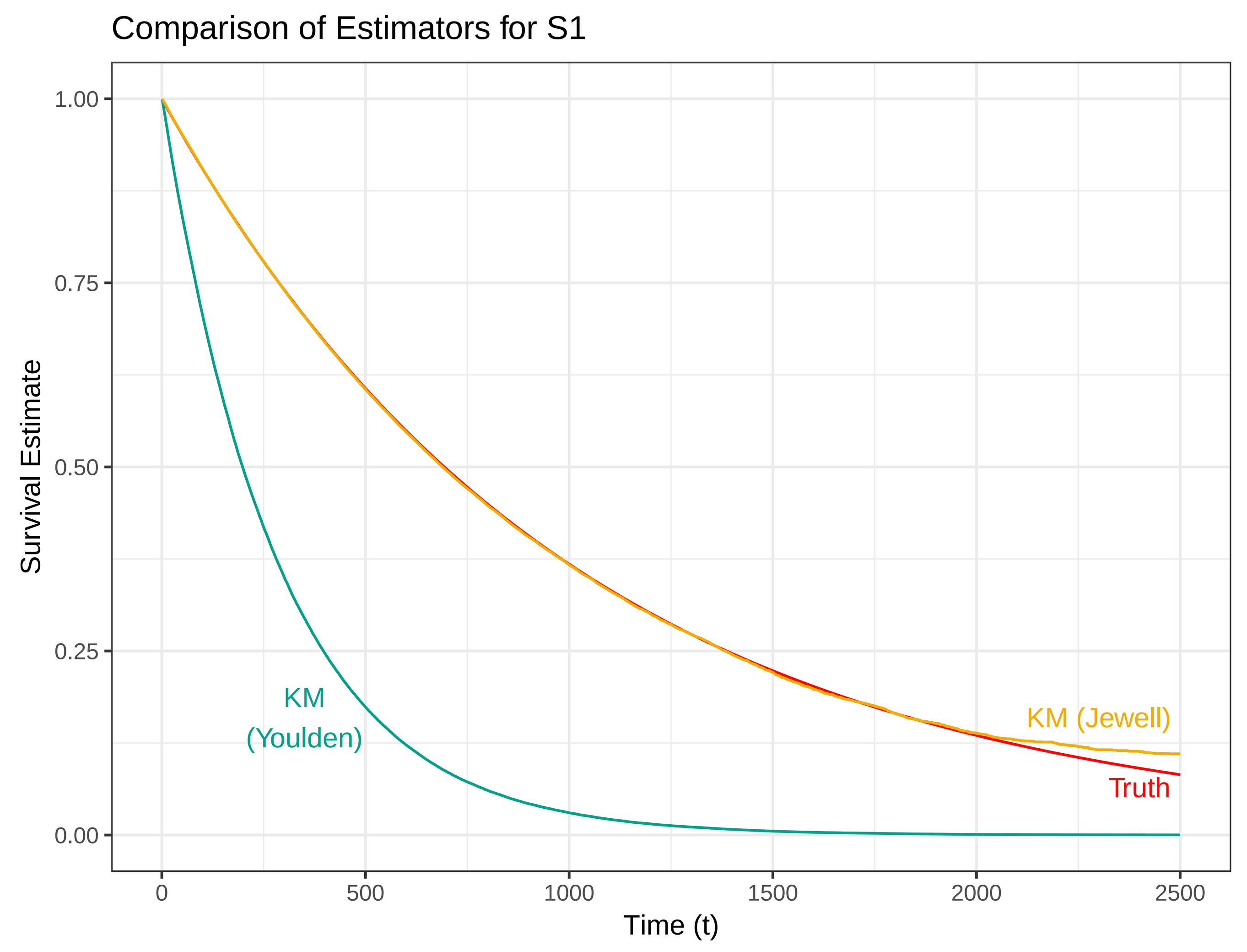

Figure 1: Average performance of estimators for S_1 for a sample of size $n = 1000$, over about 300 simulations.

- The Kaplan-Meier estimator proposed by Youlden displays obvious bias.
- The estimates of the survival curve produced by Cox regression and the Kaplan-Meier estimator with the Jewell correction show no such bias.
- Under the Cox model, Cox regression will outperform other estimators — it draws upon information across both subject groups over all time points.
- The Kaplan-Meier estimator exhibits a slight divergence from the truth in the right tail due to a well-studied finite-sample bias caused by censored observations.
- We display results for $n = 1000$ since this sample size is closest to that from the observational medical study we analyze.

Principal References

[1] Danny R Youlden, Peter D Baade, H Peter Soyer, Philippa H Youl, Michael G Kimlin, Joanne F Aitken, Adele C Green, and Kiarash Khosrotehrani. Ten-year survival after multiple invasive melanomas is worse than after a single melanoma: a population-based study. *Journal of Investigative Dermatology*, 136(11):2270–2276, 2016.

[2] Wei-Yann Tsai, Nicholas P Jewell, and Mei-Cheng Wang. A note on the product-limit estimator under right censoring and left truncation. *Biometrika*, 74(4):883–886, 1987.

[3] Steven M Snapinn, QI Jiang, and Boris Iglewicz. Illustrating the impact of a time-varying covariate with an extended kaplan-meier estimator. *The American Statistician*, 59(4):301–307, 2005.

Contact Information

N. Hejazi: NHEJAZI@BERKELEY.EDU
K. Benac: BENAC@BERKELEY.EDU
N. P. Jewell: JEWELL@BERKELEY.EDU