The effect of ignoring censoring in survival analysis: theoretical and practical considerations

Olivier Bouaziz

Laboratory MAP5, University Paris Descartes and CNRS, Sorbonne Paris Cité, Paris, France

Abstract

The paper considers the use of naive estimation techniques that do not take into account censoring in survival analysis. These naive estimators are often proposed by students who are not familiar with the topic of survival analysis. Using simple mathematical arguments and simulation experiments these estimators are shown to behave poorly in presence of censored data. We stress that censored data should be analyzed using only adapted methods such as the Kaplan-Meier estimator. The article is addressed to both researchers who are not familiar with survival analysis and to teachers who want to emphasize the needs for new estimation techniques in an introductory lesson in survival analysis.

Keywords: Biased estimators, Censoring, Follow-up studies, Kaplan-Meier estimator, Statistics education, Survival analysis.

1 Introduction

Teaching survival analysis has become standard in many statistical courses and several textbooks on the subject are nowadays available. Usually these books either focus on complex martingale theory or directly introduce standard estimation techniques in survival analysis without discussing the need for specific tools in this context. Yet estimation strategies can be easily discussed with students: when asked they usually propose three naive estimators that are constructed by neglecting the censoring effect or treating it as a missing at random variable. As shown in the paper, bias of these estimators can be easily exhibited, using only basic probability calculations. While many textbooks present dataset to introduce the concept of censoring (which is indeed of major interest) we emphasize the implementation of basic simulations which illustrate how all naive estimators give biased results. Simulating data has been shown to be an effective way to teach statistics (see Sigal and Chalmers (2016)): in the context of survival analysis it is especially useful as it helps students to understand the concept of censoring and how follow-up data are collected. Finally, modifying these naive estimators can lead to a more efficient estimation strategy. We explain that a natural quantity of interest in survival analysis is the hazard rate as it can be estimated unbiasedly and a direct extension is the famous Kaplan-Meier estimator (Kaplan and Meier (1958)) of the survival function.

Simulation studies are presented in the paper which illustrate the strong performance of the Kaplan-Meier estimator and how all naive methods will lead to biased results. In particular, increasing the sample size will not improve the performance of naive methods and even in a case of low censoring rate they will perform poorly compared to the Kaplan-Meier estimator.

2 Naive estimation methods lead to biased results when dealing with censored data

When introducing survival analysis it is important to start with some standard examples of follow-up data such as the remission data of leukemia patients from Freireich et al. (1963). This is a double blind clinical trial where some of the patients were administrated a treatment, others a placebo and the variable of interest is the remission time, that is the time in weeks between two relapses of a patient (see Table 1). The times with a plus sign next to them are censored times while other times represent the relapse times of interest. A censored time is not directly of interest but it provides the information that the relapse time has not yet occurred at the censoring time and will occur at some time after it (for instance, the fourth observed time in the treatment group of Table 1 is a censored time which informs us that this patient was still in remission after 6 weeks but its exact relapse time is unknown). This dataset can be discussed to explain the concept of clinical trial and then to focus on censored data. An easy conclusion reached by students is that taking the drug seems to result in longer remission times and therefore into more censoring. Consequently, removing censoring will provide an under estimation of the remission times of patients who took the drug. Discussion of the dataset can be found for instance in Kleinbaum and Klein (2006). Other interesting dataset for introductory lessons in survival analysis can also be found in Andersen et al. (1993) or in Collett (2003).

Table 1: Remission times (weeks) for two groups of leukemia patients from the Freireich dataset (1963). + denotes censoring due to study end before relapse, lost to follow-up or withdraws.

Treatment group ($n = 21$)							Placebo group ($n = 21$)						
6	6	6	6^+	7	9^+	10	1	1	2	2	3	4	4
10^+	11^+	13	16	17^+			5	5	8	8	8		
19^+	20^+	22	23	25^+			8	11	11	12	12		
32^+	32^+	34^+	35^+				15	17	22	23			

In the following we show how the teacher can illustrate, first using some mathematical arguments and then using simulation experiments, that naive estimators are not adapted to deal with such types of data and how they will result in biased estimates. We introduce the classical notations for right-censored data: $T_i = \min(\tilde{T}_i, C_i)$, $\Delta_i = I(\tilde{T}_i \leq C_i)$ for $i = 1, \ldots, n$, where $\tilde{T}_i$ is the variable of interest (time to relapse in the Freireich dataset), C_i the censoring variable (time to end of study, lost to follow-up or withdraw) and I represents the indicator function. For instance, in the treatment group of Table 1, $T_1 = T_2 = T_3 = T_4 = 6$ and $\Delta_1 = \Delta_2 = \Delta_3 = 1$, $\Delta_4 = 0$. We also note F, G the cumulative distribution functions of $\tilde{T}$, C and we denote by S the survival function of $\tilde{T}$, that is $S = 1 - F$. The focus will be on the estimation of this survival function which represents, in the Freireich dataset, the probability of being still in remission at a given time point. When students are asked to estimate the survival function using only the observed data (T_i, Δ_i), they usually propose one of the following three estimators:

1. A first estimator is obtained by computing the empirical cumulative survival function directly from the observations T_i:

$$\hat{S}_1(t) = \frac{1}{n} \sum_{i=1}^{n} I(T_i > t) \cdot$$

2. A second estimator is obtained by computing the empirical cumulative survival function, but from the uncensored observations $\{T_i, \Delta_i = 1\}$ only:

$$\hat{S}_2(t) = \frac{1}{n} \sum_{i=1}^{n} I(T_i > t, \Delta_i = 1) \cdot$$

3. A third estimator is built by rescaling the previous estimator by the proportion of uncensored observations:

$$\hat{S}_3(t) = \frac{1}{\sum_j \Delta_j} \sum_{i=1}^{n} I(T_i > t, \Delta_i = 1)\cdot$$

Table 2: Estimation of the survival function in the treatment group of the Freireich dataset using the three naive estimators and the Kaplan-Meier estimator (defined at page 6).

t	$\hat{S}_1(t)$	$\hat{S}_2(t)$	$\hat{S}_3(t)$	$\hat{S}_{KM}(t)$
0	1	$9/21 \simeq 0.43$	1	1
6	$17/21 \simeq 0.81$	$6/21 \simeq 0.29$	$6/9 \simeq 0.67$	$1 - 3/21 \simeq 0.86$
7	$16/21 \simeq 0.76$	$5/21 \simeq 0.24$	$5/9 \simeq 0.56$	$(1 - 3/21) \times (1 - 1/17) \simeq 0.81$
9	0.71	0.24	0.56	0.81
10	0.62	0.19	0.44	0.75
11	0.57	0.19	0.44	0.75
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Computation of these three estimators on the treatment group of the Freireich dataset are presented in Table 2, in the first three columns.

It is very common for students to use the first estimator, often because they cannot understand the concept of censoring and they think the observed time variable is the variable of interest. The estimator $\hat{S}_1(t)$ is an unbiased and consistent estimator of $H(t) := \mathbb{P}[T > t]$. Under the independent censoring assumption, which assumes that $\tilde{T}$ is independent of C, it is straightforward to write

$$H(t) = \mathbb{P}[\tilde{T} > t, C > t] = S(t)(1 - G(t)),$$

which implies that $H(t) \leq S(t)$ for all $t \geq 0$ and the equality holds true for $t = 0$, for $t = \infty$ and for all t such that there is no censoring in the dataset occurring before time t. One sees that for very small t, since there is a lower risk to get censored for lower times, the bias will be small. But as t gets larger the bias gets bigger (in absolute value) until it reaches a maximum and then decreases and equals 0 at infinity.

The second estimator focuses on observed events of interest only and treats the censored events as missing at random variables. As a matter of fact, $\hat{S}_2(t)$ is an unbiased and consistent estimator of $\mathbb{P}[T > t, \Delta = 1]$ so, if censoring could be treated as missing variables, then focusing on the observations $\{T_i, \Delta_i = 1\}$ should only deteriorate the sample size of the estimator and one should have $\mathbb{P}[T > t, \Delta = 1] = \mathbb{P}[\tilde{T} > t]$. This is a common mistake made by students and the teacher can easily show that this is not the case, no matter what is the distribution of C. Using Fubini's theorem, the mathematical arguments are as follows:

$$\begin{aligned}
\mathbb{P}[T > t, \Delta = 1] &= \iint I(u > t, u \leq v) dF(u) dG(v) \\
&= \int I(u > t) \left(\int I(u \leq v) dG(v) \right) dF(u) \\
&= \int_t^\infty (1 - G(u)) dF(u) = S(t) - \int_t^\infty G(u) dF(u). \qquad (1)
\end{aligned}$$

It is then directly seen that $\mathbb{P}[T > t, \Delta = 1] \leq S(t)$ for all $t \geq 0$. One should note that for $t = 0$, the formula gives $\mathbb{P}[\Delta = 1]$, the probability to be uncensored, and

$$\mathbb{P}[\Delta = 1] = 1 - \int_0^\infty G(u) dF(u).$$

It means that the estimator is already biased for $t = 0$, with a bias equal to $\int_0^\infty G(u)dF(u)$, which is the probability of being censored. Since $\int_t^\infty G(u)dF(u)$ in (1) is a decreasing function of t, the bias will decrease (in absolute value) with respect to t. To conclude, the bias equals the proportion of censoring for $t = 0$ and as t grows it decreases and reaches 0 when t tends to infinity.

The last estimator tries to improve the performances of $\hat{S}_2$ by appropriately weighting the estimator by the number of observed events of interest. For teaching purposes, simulation experiments are sufficient to assess its performance. However, we provide the mathematical arguments to derive the asymptotic bias of that estimator in the Appendix section where it is shown that this estimator is biased downwards. The weights $\sum_j \Delta_j$ used in $\hat{S}_3$ are not efficient as they do not take into account the time dependent structure of the censoring process (mainly the fact that the longer a patient is followed, the greater chances he has to be censored). We will show that, using the independent censoring assumption, a quantity that can be estimated without bias is the hazard rate. From this quantity the survival function is easily estimated. Before introducing the usual estimator used in survival analysis it is of interest to illustrate the biased performances of these naive estimators through some simulations conducted by the students.

3 Performance of the naive estimators on simulated data

After discussion of the theoretical properties of these naive estimators, the students are asked to simulate some survival data. This implies to first simulate the times of interest (which are therefore all known by the user) and some censoring times. The observed times are then derived by taking the minimum between these two times for each individual. Simulations are beneficials to students for two reasons. First of all, they realize that the observed censored times are not equal to the times of interest but are always lower. Secondly, they can experiment different estimation strategies and directly observe that naive estimators always fail to provide accurate estimations.

In a first scenario the variable of interest is simulated as a uniform variable with parameters equal to 0 and 10 and the right-censoring variable as a uniform variable with parameters equal to 2 and 8. This example is interesting because many students believe that if censoring occurs "at random", then ignoring it will not affect substantially the estimation results. A second scenario considers more realistic random variables: the variable of interest is simulated as a Weibull distribution with shape parameter equal to 2 and scale parameter equal to 5 (using the terminology of the R software) and the right-censoring variable as a Weibull distribution with shape parameter equal to 1 and scale parameter equal to 5.8 (which is an exponential distribution). Both scenarios result in 50% of censoring.

For a single sample of size $n = 60$, estimated survival curves are shown in Figure 1 for both scenarios. The true survival curves are also plotted along with the optimal estimator defined as:

$$\hat{S}_0(t) = \frac{1}{n}\sum_{i=1}^{n} I(\tilde{T}_i > t).$$

This estimator is not allowed in practice as one does not observe all the $\tilde{T}_i$s in the presence of censoring but it is used as a reference for the optimal estimation performance. In both scenarios, all estimators behave poorly compare to $\hat{S}_0$ and they are all biased downwards as previously studied. Interestingly, the uniform scenario clearly shows that censoring cannot be treated as a missing at random variable. In this scenario the less performant estimator is $\hat{S}_2$, then comes $\hat{S}_3$ and finally $\hat{S}_1$. This last estimator has a good performance for $t \leq 2$ as no censoring can occur before time 2 but then quickly deteriorates. In the Weibull scenario, the performance of the two estimators $\hat{S}_1$ and $\hat{S}_3$ is switched with the latter being more accurate than the former. As

expected, in both scenarios, the estimator $\hat{S}_2$ is already biased for $t = 0$, with a value of $\hat{S}_2(0)$ equal to the proportion of censoring in the dataset.

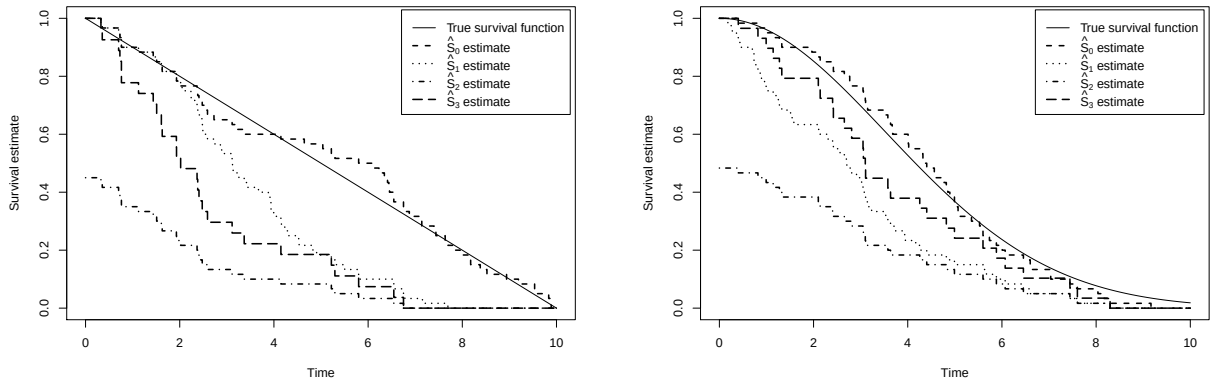

Figure 1: Naive survival estimates in the uniform scenario (left panel) and the Weibull scenario (right panel) with 50% of censoring.

Finally, computations of mean estimates were also performed using the estimation techniques corresponding to the different naive estimators. The optimal mean estimator is defined as $\hat{E}_0 = \sum_i \tilde{T}_i/n$ and the other mean estimators as:

$$\hat{E}_1 = \frac{1}{n}\sum_{i=1}^{n} T_i,$$

$$\hat{E}_2 = \frac{1}{n}\sum_{i=1}^{n} T_i\Delta_i,$$

$$\hat{E}_3 = \frac{1}{\sum_j \Delta_j}\sum_{i=1}^{n} T_i\Delta_i.$$

As an illustration, on the samples used to obtain Figure 1, we found for the uniform scenario: $\hat{e}_0 = 5.10$, $\hat{e}_1 = 3.37$, $\hat{e}_2 = 1.15$ and $\hat{e}_3 = 2.57$ while the true mean is equal to 5. For the Weibull scenario the results are: $\hat{e}_0 = 4.42$, $\hat{e}_1 = 2.88$, $\hat{e}_2 = 1.76$ and $\hat{e}_3 = 3.65$ while the true mean is approximately equal to 4.43. These results give the same tendency as for the survival estimates, the mean estimates having the advantage to summarize the properties of the estimators over the whole period of time.

We also considered a low censoring rate case, which is of interest to illustrate the need for appropriate estimation methods to deal with survival data even when only few observations are censored. We took, for the censoring variable, a uniform distribution of parameters 2 and 34 in the uniform scenario and a Weibull distribution with scale parameter equal to 42 in the Weibull scenario which lead to 10% of censoring for both scenarios. Monte-Carlo simulations were then performed on $50,000$ replications. Mean estimates of the different estimators are shown in Table 3 for the two scenarios and the two different censoring rates (10% and 50%). It is seen that a substantial bias still remains even in the low rate censoring case when using naive estimators. Other experiments were conducted using different sample sizes. Interestingly, increasing the sample size does not improve at all the performance of the naive estimators and each of them has the same bias in each case (results not shown here).

Table 3: Bias of the naive mean estimators for the uniform and Weibull scenarios with 10% or 50% of censoring. True mean equals 5 in the uniform scenario and equals 4.43 in the Weibull scenario.

	10% of censoring				50% of censoring			
	$\hat{e}_0$	$\hat{e}_1$	$\hat{e}_2$	$\hat{e}_3$	$\hat{e}_0$	$\hat{e}_1$	$\hat{e}_2$	$\hat{e}_3$
Uniform scenario	0.001	-0.267	-0.736	-0.260	0.000	-1.400	-3.600	-2.199
Weibull scenario	0.000	-0.282	-0.551	-0.125	0.000	-1.540	-2.621	-0.817

4 Unbiased estimation of the hazard rate

The next step is to show that the crucial quantity of interest in survival analysis is the hazard rate defined as:

$$\lambda(t) = \lim_{\triangle t \to 0} \frac{\mathbb{P}[t \leq \tilde{T} < t + \triangle t | \tilde{T} \geq t]}{\triangle t} = \frac{dF(t)}{S(t)}.$$

The reason for that comes from our ability to estimate it without bias (except on the tail of the distribution of $\tilde{T}$) using only the observations (T_i, Δ_i). Recall that $H(t) = \mathbb{P}[T > t] = S(t)(1 - G(t))$ under independent censoring and note $H_1(t) := \mathbb{P}[T \leq t, \Delta = 1]$. This is the cumulative distribution function of the observed events of interest, and assuming the variable T to be continuous, it has a density which we will represent by $dH_1(t)$. Similarly to formula (1), we have:

$$\begin{aligned} H_1(t) &= \int_0^t (1 - G(u)) dF(u), \text{ and} \\ dH_1(t) &= (1 - G(t)) dF(t). \end{aligned}$$

Dividing both sides of the equation by $H(t)$ gives $dH_1(t)/H(t) = dF(t)/S(t) = \lambda(t)$. This is an important result because both functions dH_1 and H can be easily estimated from the observations (T_i, Δ_i). Note that the only hypothesis used here to prove the relation $dH_1(t)/H(t) = \lambda(t)$ is the independent censoring assumption. This equality can be rewritten as:

$$\lim_{\triangle t \to 0} \frac{\mathbb{P}[t \leq T < t + \triangle t, \Delta = 1 | T \geq t]}{\triangle t} = \lim_{\triangle t \to 0} \frac{\mathbb{P}[t \leq \tilde{T} < t + \triangle t | \tilde{T} \geq t]}{\triangle t}, \tag{2}$$

which means that conditionally on not having experienced yet the event of interest at time t ($\tilde{T} \geq t$), the probability that it occurs before time $t + \triangle t$ is the same as the probability of observing the event of interest before time $t + \triangle t$ knowing the individual is at risk at time t ($T \geq t$). The notion of individual at risk ($T \geq t$) is crucial in survival analysis, due to this relationship, and it means that the individual has not yet experienced the event of interest nor been censored.

The left hand side of Equation (2) suggests a direct estimator of the hazard rate at any observed time T_i as: $\hat{\lambda}(T_i) = \Delta_i / R_i$ where $R_i = \sum_j I(T_j \geq T_i)$ represents the number of individuals at risk at time T_i. In case of ties, the estimator is easily adapted by replacing Δ_i by $\sum_j I(T_j = T_i, \Delta_i = 1)$. This estimator can be directly used to construct the Kaplan-Meier estimator. The basic idea is that on the ordered time events $T_{(1)} < T_{(2)} < \cdots < T_{(k)}$, where $k \leq n$ represents the unique different time values ($k = n$ in case of no ties), one can easily write:

$$S(T_{(i)}) = \prod_{j=1}^{i} \left(1 - p(T_{(j)} | T_{(j-1)})\right),$$

with $p(t_j | t_{j-1}) := \mathbb{P}[\tilde{T} \leq t_j | \tilde{T} > t_{j-1}]$ and the Kaplan-Meier estimator is then defined as:

$$\hat{S}_{KM}(t) := \prod_{j=1}^{i} \left(1 - \hat{\lambda}(T_{(j)})\right), \text{ for } t \text{ such that } T_{(i)} \leq t < T_{(i+1)}.$$

See Table 2 (last column) for the computation of the estimator on the Freireich dataset for some time values. For more details about the Kaplan-Meier estimator, see among many authors Andersen et al. (1993) or Collett (2003).

5 Comparison with the Kaplan-Meier estimator

Finally, the lesson ends by comparing the performance of the Kaplan-Meier estimator to the naive estimators under the different scenarios presented in the previous simulation setting. Plots of the survival estimates are represented in Figure 2 for the same samples of size $n = 60$ as previously, with 50% of censored data. Only the best of the three naive estimators are included in these plots ($\hat{S}_1$ in the uniform scenario and $\hat{S}_3$ in the Weibull scenario). Clearly the Kaplan-Meier is a very accurate estimator with similar performance as the optimal estimator $\hat{S}_0$.

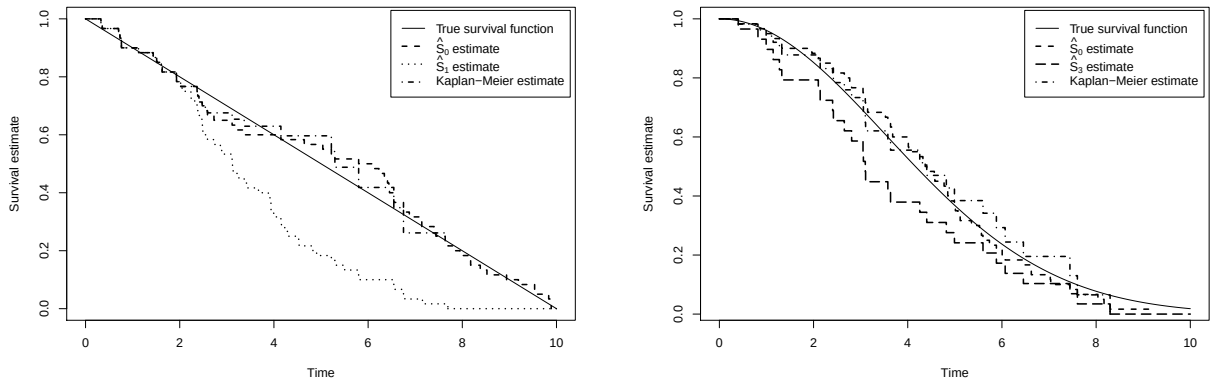

Figure 2: Estimation of the survival function with the best of the naive estimators and the Kaplan-Meier estimator in the uniform scenario (left panel) and the Weibull scenario (right panel) with 50% of censoring.

It is well know that the Kaplan-Meier estimator behaves poorly in the tail of the distribution due to censoring. As a consequence, usual indicators are quantiles of the distribution such as the median. However, another useful indicator in terms of estimation performance is the restricted mean estimator (see for instance Andersen et al. (2004)) which has the advantage to summarize the Kaplan-Meier estimator performance on a time interval. For a given $\tau > 0$, the restricted mean survival time is defined as

$$\mu(\tau) := \mathbb{E}[\min(\tilde{T}, \tau)] = \int_0^\tau S(t)dt,$$

and therefore its estimator for right-censored data is defined as:

$$\hat{\mu}_{KM}(\tau) = \int_0^\tau \hat{S}_{KM}(t)dt.$$

We will compare this estimator to the restricted mean survival estimators corresponding to each

of the three naive estimation methods:

$$\hat{\mu}_1(\tau) = \frac{1}{n}\sum_{i=1}^{n} T_i \wedge \tau,$$

$$\hat{\mu}_2(\tau) = \frac{1}{n}\sum_{i=1}^{n} \Delta_i (T_i \wedge \tau),$$

$$\hat{\mu}_3(\tau) = \frac{1}{\sum_j \Delta_j}\sum_{i=1}^{n} \Delta_i (T_i \wedge \tau),$$

and to the optimal (unobserved) restricted mean survival estimator,

$$\hat{\mu}_0(\tau) = \frac{1}{n}\sum_{i=1}^{n} \tilde{T}_i \wedge \tau\cdot$$

As before Monte-Carlo simulations were performed with the choice of $\tau = 6$ and are presented in Table 4. The results show that the Kaplan-Meier estimator has similar performance as the optimal estimator (almost no bias at all) while naive estimators have a bias related to the amount of censoring in the dataset. Note that even in the low censoring case, minimal bias are respectively equal to -0.133 and -0.085 in the uniform and Weibull cases which emphasizes the importance of using the Kaplan-Meier estimator even when only a few proportion of observations are censored.

Table 4: Bias of the restricted mean estimators for the uniform and Weibull scenarios with 10% or 50% of censoring. True restricted mean equals 4.2 in the Uniform scenario and equals 4 in the Weibull scenario.

	10% of censoring					50% of censoring				
	$\hat{e}_0$	$\hat{e}_1$	$\hat{e}_2$	$\hat{e}_3$	$\hat{e}_{KM}$	$\hat{e}_0$	$\hat{e}_1$	$\hat{e}_2$	$\hat{e}_3$	$\hat{e}_{KM}$
Uniform	0.001	-0.133	-0.569	-0.163	0.001	0.000	-0.710	-2.822	-1.443	0.001
Weibull	0.000	-0.218	-0.475	-0.085	0.000	0.000	-1.255	-2.312	-0.597	-0.001

6 Concluding remarks

The aim of the paper was to emphasize the need for specific statistical tools to perform estimation in survival analysis. When introducing the topic, estimation techniques can be discussed with students and through a didactic lesson the teacher can conduct the course in a gradual way until the derivation of the Kaplan-Meier estimator. Classical statistical methods for censored data can then be taught in the usual way, and they will be much easily understood after having discussed the basic concepts.

In some studies it can be difficult to realise that the observations suffer from censoring. Famous examples include the two studies on survival of left-handed people (see Altman and Bland (2005)) where the researchers only looked at dead individuals. By ignoring censoring effect, the original studies concluded that right-handed people live longer than left handed (9 years longer on average!), a result that has been refuted since then.

The present paper only considered nonparametric aspects of estimation. Confusion on censored data is also frequent when trying to implement regression models. A classical situation is when the outcome is binary but can only be observed for individuals that had a complete follow-up. As an example consider the recent study in Clausen et al. (2016) where researchers aimed to assess the effect of mode of delivery in babies on the onset of type 1 diabetes before age 15. If all children had been followed for 15 years the outcome would have been binary (children

can either have diabetes or not). However all children did not have the same follow-up duration (mainly due to study end) and using regression methods such as logistic regression would result in a large bias in the estimations. The correct method, which is the one conducted in the article, is to implement a Cox model (see Cox (1972)) which directly models the hazard rate for time to onset of diabetes and take into account right-censoring in the estimation method.

Appendix: study of the asymptotic bias of $\hat{S}_3$

The estimator $\hat{S}_3(t)$ is a consistent estimator of $\mathbb{P}[T > t|\Delta = 1] = \mathbb{P}[T > t, \Delta = 1]/\mathbb{P}[\Delta = 1]$. The numerator was calculated in (1) and the denominator was obtained from (1) by taking the limit as t tends to infinity. We have:

$$\mathbb{P}[T > t|\Delta = 1] = \frac{S(t) - \int_t^\infty G(u)dF(u)}{1 - \int_0^\infty G(u)dF(u)}.$$

The asymptotic bias can be studied by subtracting $S(t)$ to the previous quantity, and direct calculations give

$$\mathbb{P}[T > t|\Delta = 1] - S(t) = \frac{\int_0^t G(u)dF(u) - F(t)\int_0^\infty G(u)dF(u)}{1 - \int_0^\infty G(u)dF(u)}.$$

Notice then that the function on the right hand side of the equation equals 0 for $t = 0$ or $t = \infty$. Simple analysis of the derivative of that function shows that $\mathbb{P}[T > t|\Delta = 1] - S(t)$ is decreasing for $t \in [0, t^*)$, reaches a minimum at t^* and is increasing for $t \in [t^*, \infty)$, where t^* is defined such that $G(t^*) = \int_0^\infty G(u)dF(u) = \mathbb{P}[\Delta = 0]$. For continuous distribution functions, such a value of t^* always exists with the two degenerated cases corresponding to $\mathbb{P}[\Delta = 0] = 0$ (meaning there are no censored observations) which implies that $t^* = 0$ and $\mathbb{P}[\Delta = 0] = 1$ (all observations are censored) which implies that $t^* = \infty$. This concludes that

$$\mathbb{P}[T > t|\Delta = 1] - S(t) \leq 0,$$

and the equality holds true only for $t = 0$, $t = \infty$, or for all $t \geq 0$ in the case of no censoring.

References

Altman, D. and Bland, M. (2005). Do the left-handed die young? *Significance*, 2(4):166–170.

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical models based on counting processes*. Springer Series in Statistics. Springer-Verlag, New York.

Andersen, P. K., Hansen, M. G., and Klein, J. P. (2004). Regression analysis of restricted mean survival time based on pseudo-observations. *Lifetime data analysis*, 10(4):335–350.

Clausen, T. D., Bergholt, T., Eriksson, F., Rasmussen, S., Keiding, N., and Løkkegaard, E. C. (2016). Prelabor cesarean section and risk of childhood type 1 diabetes: A nationwide register-based cohort study. *Epidemiology*, 27(4):547–555.

Collett, D. (2003). *Modelling survival data in medical research*. CRC press.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, 34:187–220.

Freireich, E. J., Gehan, E., Frei, E., Schroeder, L. R., Wolman, I. J., Anbari, R., Burgert, E. O., Mills, S. D., Pinkel, D., Selawry, O. S., et al. (1963). The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukemia: A model for evaluation of other potentially useful therapy. *Blood*, 21(6):699–716.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481.

Kleinbaum, D. G. and Klein, M. (2006). *Survival analysis: a self-learning text*. Springer Science & Business Media.

Sigal, M. J. and Chalmers, R. P. (2016). Play it again: Teaching statistics with monte carlo simulation. *Journal of Statistics Education*, 24(3):136–156.