About an adaptively weighted Kaplan-Meier estimate

Jean-François Plante, *HEC Montréal*

Abstract The minimum averaged mean squared error nonparametric adaptive weights use data from m possibly different populations to infer about one population of interest. The definition of these weights is based on the properties of the empirical distribution function. We use the Kaplan-Meier estimate to let the weights accommodate right-censored data and use them to define the weighted Kaplan-Meier estimate. The proposed estimate is smoother than the usual Kaplan-Meier estimate and converges uniformly in probability to the target distribution. Simulations show that the performances of the weighted Kaplan-Meier estimate on finite samples exceed that of the usual Kaplan-Meier estimate. A case study is also presented.

Keywords: Adaptive weights · Borrowing strength · Kaplan-Meier estimate · Nonparametrics · Survival analysis · Weighted inference

1 Introduction

The relevance weighted empirical distribution function introduced by Hu & Zidek (1993), further developed in Hu (1994), then published in Hu & Zidek (2002), is designed to estimate a target distribution by using data from possibly different distributions whose similarity is encoded through "relevance" weights.

The main work of Hu (1994) concerns the weighted likelihood which is linked to the relevance weighted empirical distribution function. Wang (2001) as well as Wang et al. (2004) and Wang & Zidek (2005) investigate the weighted likelihood under a specific paradigm that we adopt in this paper: we suppose that data comes from m populations and that for each fixed $i = 1, \dots, m$, we observe $X_{i1}, \dots, X_{in_i} \overset{iid}{\sim} F_i$.

Plante (2008) shows that the maximum weighted likelihood can be seen as a special case of the Maximization Entropy Principle of Akaike (1977) where a form of weighted empirical distribution function is used to estimate the unknown and unknowable distribution F_1. In the same article, the author proposes a nonparametric adaptive method to determine likelihood weights: the minimum averaged mean squared error weights (the MAMSE weights).

Let $\hat{F}_i(x) = (1/n_i) \sum_{j=1}^{n_i} \mathbb{1}(X_{ij} \leq x)$ be the empirical distribution function based on data from Population i. Then

$$\hat{F}_\lambda(x) = \sum_{i=1}^{m} \lambda_i \hat{F}_i(x)$$

is a weighted empirical distribution function when $\lambda_i \geq 0$ and $\sum_{i=1}^{m} \lambda_i = 1$.

For complete data, Plante (2008) suggests a criterion to determine weights that make $\hat{F}_\lambda(x)$ close to $\hat{F}_1$ but less variable. A few pre-processing steps that may set some weights to zero are first applied, then the weights are determined by minimizing the objective function

$$P(\lambda) = \int \left[\left\{ \hat{F}_1(x) - \hat{F}_\lambda(x) \right\}^2 + \widehat{\mathrm{var}} \left\{ \hat{F}_\lambda(x) \right\} \right] \mathrm{d}\hat{F}_1(x), \tag{1}$$

under the constraints $\lambda_i \geq 0$ and $\sum_{i=1}^{m} \lambda_i = 1$, with

$$\widehat{\mathrm{var}} \left\{ \hat{F}_\lambda(x) \right\} = \sum_{i=1}^{m} \lambda_i^2 \, \widehat{\mathrm{var}} \left\{ \hat{F}_i(x) \right\} \quad \text{and} \quad \widehat{\mathrm{var}} \left\{ \hat{F}_i(x) \right\} = \frac{1}{n_i} \hat{F}_i(x) \{1 - \hat{F}_i(x)\}.$$

Plante (2009) shows that these weights make $\hat{F}_\lambda(x)$ converge uniformly to F_1.

In situations where measurements such as the time of death, or the time to failure, are of interest, the exact value of the outcome may not be observed for all individuals, yielding for instance right-censored data. In such situations, Kaplan & Meier (1958) propose a nonparametric estimate of the survival function that takes into consideration both the observed and the censored data.

In this paper, we use the Kaplan-Meier estimate to build a modified version of the minimum averaged mean squared error weights that accommodates right-censored data. These weights are used to define a weighted Kaplan-Meier estimate, proved to be uniformly weakly consistent.

When data are available from populations that are similar, the weighted Kaplan-Meier is a nonparametric estimate of the distribution of survival times that borrows strength from the additional populations. Bayesian hierarchical models are designed for situations akin to our paradigm where data comes from m different sources, but such models require parametric assumptions, including the specification of an hyper-parameter to link the populations together. The weighted Kaplan-Meier requires no such assumptions.

Plante (2008) discusses situations where the minimum averaged mean squared error weights are useful. This includes cases where a mixture of the additional populations is similar to the target population. Consider for instance situations where the additional populations are demographic subgroups of the population of interest, or when data from similar studies are available.

The proposed weighted Kaplan-Meier estimate is not based on a specific definition of similarities between the populations. In particular, it does not require testing against discrepancies in the data. The weights adjust automatically and discard data that are too different. Better performances will occur when the distributions of the data are similar, but the method remains consistent even if they are very different.

The Kaplan-Meier estimate has jumps only at times of death. The weighted Kaplan-Meier estimate will typically be smoother since steps can occur at times of deaths from all the populations. The simulations in Section 5 and the case study in Section 6 illustrate this.

Section 2 introduces the notation used in this document by reviewing properties of the Kaplan-Meier estimate. The minimum averaged mean squared error weights for censored data are defined in Section 3. In Section 4, the uniform convergence of the weighted Kaplan-Meier estimate is proved – technical details of the proofs can be found in the Appendix. Section 5 presents simulation results that explore the performance of the weighted Kaplan-Meier estimate on finite samples and illustrate the use of the bootstrap to determine confidence intervals. Finally, Section 6 presents a case-study.

2 The Kaplan-Meier Estimate

We use the pretext of a review of the well-known Kaplan-Meier estimate to introduce the notation that prevails throughout this document. For simplicity, we also adopt a survival analysis terminology where the measurements of interest are the survival of individuals.

Consider a probability space $(\Omega, \mathcal{B}(\Omega), P)$ and for Population i, let X_{ij} be the time of death of individual j and V_{ij} its censoring time. The positive random variables X_{ij} and V_{ij} are independent. We assume that the distributions of X_{ij} is continuous and denote it by F_i. For any fixed $k \in \mathbb{N}$, we observe (Z_{ij}, δ_{ij}) for $i = 1, \ldots, m$ and $j = 1, \ldots, n_{ik}$, where $Z_{ij} = \min(X_{ij}, V_{ij})$ and $\delta_{ij} = \mathbb{1}(V_{ij} \geq X_{ij})$. The index k is used to express the asymptotic results of Section 4. It allows keeping track of the m sample sizes at once: we assume that the sample sizes are non-decreasing with k and that $n_{1k} \to \infty$ as $k \to \infty$.

Let $H_i(t) = \mathrm{pr}(Z_{i1} \leq t)$ and let $\tau_{H_i} = \sup\{t : H_i(t) < 1\}$ be the largest value that Z_{ij}

can attain. The possibility that $\tau_{H_i} = \infty$ is not ruled out although it is unlikely to occur in practice. In addition, let $H_i^*(t) = \text{pr}(Z_{i1} \leq t, \delta_{i1} = 1)$ be the distribution of observed death times for Population i. We adopt the usual notation

$$N_{ik}(s) = \sum_{j=1}^{n_{ik}} \mathbb{1}(Z_{ij} \leq s, \delta_{ij} = 1) \quad , \quad \mathrm{d}N_{ik}(s) = N_{ik}(s) - N_{ik}(s^-),$$

$$Y_{ik}(s) = \sum_{j=1}^{n_{ik}} \mathbb{1}(Z_{ij} \geq s) \quad , \quad \mathrm{d}Y_{ik}(s) = Y_{ik}(s) - Y_{ik}(s^+).$$

For Population i, the Kaplan-Meier estimate of the probability of dying at time t or earlier is written (see e.g. Kaplan & Meier 1958)

$$\hat{F}_{ik}(t) = 1 - \prod_{0 \leq s \leq t} \left\{1 - \frac{\mathrm{d}N_{ik}(s)}{Y_{ik}(s)}\right\}.$$

The Kaplan-Meier estimate is an increasing step function with jumps at each observed times of death. The number of deaths observed in Population 1 is $\mathcal{N}_k = N_{1k}(\tau_{H_1})$, hence $t_{k\mathcal{N}_k}$ represents the largest observed time of death. For $k \in \mathbb{N}$, let $t_{k1} < \cdots < t_{k\mathcal{N}_k}$ be the ordered times of these deaths, distinct by the continuity of F_1.

Using the convention that $t_{k0} = 0$, we have $J_{kj} = \hat{F}_{1k}(t_{kj}) - \hat{F}_{1k}(t_{k(j-1)})$ for $j \in \{1, \ldots, \mathcal{N}_k\}$. Then $\sum_{j=1}^{\mathcal{N}_k} J_{kj} \leq 1$. and $J_{k1} \leq J_{k2} \leq \cdots \leq J_{k\mathcal{N}_k}$.

We will consider the Kaplan-Meier estimate on a bounded interval $[0, U]$ with $U < \tau_{H_1}$. It is well known that $\sup_{t \leq U} |\hat{F}_{ik}(t) - F_i(t)| \to 0$ almost surely as $n_{ik} \to \infty$, see e.g. Winter et al. (1978) or Földes & Rejtö (1981).

Efron (1967) and Breslow & Crowley (1974) assume that the distribution of censoring time is continuous and show that $\hat{F}_{ik}(t)$ is approximately normal with mean $F_i(t)$ and a variance that can be estimated using Greenwood's formula

$$\widehat{\text{var}}\{\hat{F}_{ik}(t)\} \approx \widetilde{\text{var}}\{\hat{F}_{ik}(t)\} = \{1 - \hat{F}_{ik}(t)\}^2 \sum_{0 \leq s \leq t} \frac{\mathrm{d}N_{ik}(s)}{Y_{ik}(s) Y_{ik}(s^+)}, \tag{2}$$

an expression that becomes less reliable as t approaches τ_{H_i}.

Defining minimum averaged mean squared error weights based on the Kaplan-Meier estimate involves using an estimate of its variance. Equation (2) will be used for that purpose even though the continuity of the distribution of censoring time is not assumed.

3 Minimum Averaged Mean Squared Error Weights for Right-Censored Data

We extend the idea of Plante (2008) to censored data by replacing the empirical distribution functions in (1) with the corresponding Kaplan-Meier estimates.

We assume that we can specify an upper bound $U < \tau_{H_1}$ and limit our study of the lifetime distribution to the interval $[0, T]$ where $T < U$ is such that $H_1^*(T) < H_1^*(U)$, meaning that there is a non-null probability that a death is observed in the interval $(T, U]$. This will be the case whenever the probability of death, observed or not, is non-null in that interval.

A few pre-processing steps are first applied. For a fixed k and $i = 2, \ldots, m$, let $m_{ik} = \min_{\{j \leq n_{ik}:\delta_{ij}=1\}} Z_{ij}$ and $M_{ik} = \max_{\{j \leq n_{ik}:\delta_{ij}=1\}} Z_{ij}$ be the smallest and largest times of death observed in Population i. The weight allocated to the sample from Population i is set to 0 if one of the following two conditions fails:

1. $U \in [m_{ik}, M_{ik}]$, i.e. at least one observed death from Population i is in the interval $[0, U]$ and at least one observed death occurs after U;

2. $\sum_{\{j \leq n_{1k}:\delta_{1j}=1\}} \mathbb{1}\{X_{1j} \in [m_{ik}, \min(M_{ik}, U)]\} \geq 1$, i.e. at least one observed death from Population 1 which occurred in $[0, U]$ falls within the range of the observed times of death in Population i.

Condition 1 ensures that Formula (2) is well defined on $[0, U]$ and not null everywhere on that interval. Condition 2 means that the same formula will be strictly positive for at least one of the times of death from Population 1 in $[0, U]$, ensuring the unicity of the minimum averaged mean squared error weights and the convergence of the algorithm used to calculate them. See Section 4.3 of Plante (2007) for more detailed explanations. The pre-processing requirements appear technical, but they avoid using the Kaplan-Meier estimate in situations where it becomes unreliable.

Define the objective function

$$P_k(\lambda) = \int_0^U \left[\left\{ \hat{F}_{1k}(t) - \sum_{i=1}^m \lambda_i \hat{F}_{ik}(t) \right\}^2 + \sum_{i=1}^m \lambda_i^2 \widetilde{\mathrm{var}} \left\{ \hat{F}_{ik}(t) \right\} \right] \mathrm{d}\hat{F}_{1k}(t), \tag{3}$$

a special case of Equation (1) where $\widehat{\mathrm{var}}\{\hat{F}_{ik}(t)\}$ is estimated by $\widetilde{\mathrm{var}}\{\hat{F}_{ik}(t)\}$ from Equation (2) and $\hat{F}_{ik}(t)$ now represent Kaplan-Meier estimates.

Note that none of the pre-processing steps can remove Population 1 since it is the population of interest. In cases where $M_{1k} = t_{k\mathcal{N}_k} < U$, the expression for $\widetilde{\mathrm{var}}\{\hat{F}_{1k}(t_{k\mathcal{N}_k})\}$ involves a division by 0 since $Y_{ik}(t^+_{k\mathcal{N}_k}) = 0$. In that case, we substitute the ill-defined term by its value just before $t_{k\mathcal{N}_k}$, the largest observed time of death. This adjustment will affect at most one term of the integral $P_k(\lambda)$.

The weights are chosen to minimize $P_k(\lambda)$ subject to $\lambda \geq 0$ and $\sum_{i=1}^m \lambda_i = 1$. We call the solution of this optimization problem the survival minimum averaged mean squared error weights. These optimal weights, denoted $\mu_k = [\mu_{1k}, \ldots, \mu_{mk}]^\mathsf{T}$, are random variables defined on Ω since they depend on the data. For values of t in the interval $[0, T]$, the weighted Kaplan-Meier estimate of the lifetime's cumulative distribution function is given by

$$\hat{G}_k(t) = \sum_{i=1}^{m} \mu_{ik} \hat{F}_{ik}(t). \tag{4}$$

Whether a sample is rejected in the pre-processing or not may vary with k, but it does not affect the distribution of probabilities calculated in expressions such as (6) since Population 1 is never excluded from the optimization problem. Moreover, pre-processing does not change the fact that $\lambda = [1, 0, \ldots, 0]^\mathsf{T}$ is a suboptimal choice of weights.

For fixed k, let us write

$$V(x) = \begin{bmatrix} \widetilde{\mathrm{var}}\{\hat{F}_{2k}(x)\} & & 0 \\ & \ddots & \\ 0 & & \widetilde{\mathrm{var}}\{\hat{F}_{mk}(x)\} \end{bmatrix}, \ \mathcal{F}(x) = \begin{bmatrix} \hat{F}_{1k}(x) - \hat{F}_{2k}(x) \\ \vdots \\ \hat{F}_{1k}(x) - \hat{F}_{mk}(x) \end{bmatrix}$$

and $\tilde{\lambda} = [\lambda_2, \ldots, \lambda_m]^\mathsf{T}$. Then, following Plante (2008),

$$P_k(\lambda) = \tilde{\lambda}^\mathsf{T} \bar{\mathsf{A}} \tilde{\lambda} - 2\tilde{\lambda}^\mathsf{T} \vec{1} \bar{b} + \bar{b} \tag{5}$$

where $\vec{1}$ is a vector of ones,

$$\bar{\mathsf{A}} = \int \left[\mathcal{F}(x)\mathcal{F}(x)^\mathsf{T} + V(x) + \vec{1}\vec{1}^\mathsf{T} \widetilde{\mathrm{var}} \left\{ \hat{F}_{1k}(x) \right\} \right] \mathrm{d}\hat{F}_{1k}(x) \quad , \quad \bar{b} = \int \widetilde{\mathrm{var}} \left\{ \hat{F}_{1k}(x) \right\} \mathrm{d}\hat{F}_{1k}(x).$$

The minimum of $P_k(\lambda)$ without the constraints on λ would be the solution to the equation $\bar{\mathsf{A}}\tilde{\lambda} = \bar{b}\vec{1}$. To enforce the constraints, the following algorithm is applied

1. Solve the equation $\bar{\mathsf{A}}\tilde{\lambda} = \bar{b}\vec{1}$;

2. if all the weights obtained are nonnegative, stop. Otherwise set the negative weights to 0, ignore the corresponding samples and repeat from Step 1 with the reduced

system. The weight allocated to Population 1 from Step 1 cannot be negative (see Lemma 4 of Plante (2008)). If no other samples are left, then $\tilde{\lambda} = 0$ and $\lambda_1 = 1$.

The algorithm above works because $P_k(\lambda)$ is quadratic and positive definite. These facts are proved rigorously in Section 2.4 and 4.3 of Plante (2007) who uses the sufficient Kuhn-Tucker conditions. The proof is similar to that found in Plante (2008) for the non-censored case. The pre-processing steps are important to insure the positive definiteness of $\bar{\mathsf{A}}$ which also guarantees that the algorithm above converges and that for any given samples, Equation 5 has a unique solution.

4 Asymptotic Properties of the Weighted Kaplan-Meier Estimate

We prove that $\hat{G}_k(t)$ converges uniformly in probability to $F_1(t)$. The details of the proofs are deferred to the Appendix. Remember that we assumed that the distribution of the times of death is continuous, but the distribution of the times of censoring need not be. Recall also that $n_{1k} \to \infty$ as $k \to \infty$, but that the other n_{ik} may either be bounded or go to ∞ at any rate.

The results of this section hold for any adaptive criterion ensuring that the following assumption is respected.

Assumption 1 $\int_0^U \left\{\hat{F}_{1k}(t) - \hat{G}_k(t)\right\}^2 \mathrm{d}\hat{F}_{1k}(t) \xrightarrow{P} 0$ *as* $k \to \infty$.

Note that this assumption is indeed respected by the proposed adaptive weights.

Theorem 1 *The proposed survival minimum averages mean squared error weights respect Assumption 1.*

Theorem 2 *Let* $0 < T < U < \tau_{H_1}$ *with* $H_1^*(T) < H_1^*(U)$, *then* $\sup_{t\leq T} \left|\hat{G}_k(t) - F_1(t)\right| \xrightarrow{P} 0$ *as* $k \to \infty$.

The weighted Kaplan-Meier estimate converges uniformly in probability to the lifetime distribution of Population 1 in the interval $[0, T]$.

Knowing the asymptotic distribution of the weighted Kaplan-Meier estimate would be handy for the determination of confidence bands, but such a result is not even known in the simpler case of uncensored data. Plante (2009) describes how the minimum averaged mean squared error weights may remain random even for infinitely large sample sizes, complicating asymptotic derivations. A similar behavior is expected for the survival case. As an alternative, resampling methods such as the bootstrap may be used to determine confidence intervals for the weighted Kaplan-Meier estimate. An example is provided in Section 5.4.

5 Simulations

This section presents the results of simulations performed to evaluate the finite-sample performance of the weighted Kaplan-Meier estimate relative to that of the usual Kaplan-Meier estimate.

Software to calculate the minimum mean squared error weights and the weighted Kaplan-Meier estimate were developed in R and are available as a library (called MAMSE) on the Comprehensive R Archive Network.

After calculating both $\hat{F}_\mu(t)$ and the usual Kaplan-Meier estimate, $\hat{F}_1(t)$, we evaluate their relative performance by comparing $A_\mu = \int_0^T |\hat{F}_\mu(t) - F_1(t)|\,\mathrm{d}t$ and $A_1 = \int_0^T |\hat{F}_1(t) - F_1(t)|\,\mathrm{d}t$.

At the end of this section, we also explore the reliability and effectiveness of bootstrap pointwise confidence intervals for the weighted Kaplan-Meier estimate. Since the asymptotic distribution of $\hat{F}_\mu(t)$ is not known yet, practitioners will have to rely on such resampling methods to take advantage of the weighted Kaplan-Meier estimate.

Simulations use either 10000 or 20000 repetitions. Unless otherwise stated, these numbers are large enough to make the standard deviation of the simulation error smaller than the last digit shown in the tables. Symbols with a bar, e.g. $\bar{A}_1$ or $\bar{\mu}_1$, correspond to an average of the corresponding statistic over the simulated samples.

5.1 Gamma Model

Let us first consider the gamma distribution. Equal samples of size $n \in \{10, 25, 100, 1000\}$ are drawn from four populations with common scale parameter ($\beta = 0.5$), but different shape parameters (α_i), yielding expectations α_i/β. In a first scenario, the shape parameters are 0.5, 0.7, 0.9 and 0.3 respectively, in a second scenario, they are 0.75, 0.5, 1 and 1.25.

The Kaplan-Meier estimate based on Population 1 alone and its weighted counterpart are computed on 20000 repetitions. The results are summarized in Table 2.

Note that $U = 3/2$, $T = 1$ and that the times of censoring are simulated as independent Uniforms on $[0, 3]$, yielding censoring rates of 0.29, 0.39, 0.48 and 0.18 respectively under Scenario 1 and of 0.41, 0.29, 0.52 and 0.61 under Scenario 2.

Table 1: Relative performance of the weighted Kaplan-Meier estimate as measured by $100\bar{A}_1/\bar{A}_\mu$ as well as the average weights. Equal samples of size n are drawn from four Gamma populations under two different scenarios. Note that $U = 3/2$, $T = 1$ and that each entry is based on 20000 repetitions.

	Scenario 1					Scenario 2				
$100\times$	$\bar{A}_1/\bar{A}_\mu$	$\bar{\mu}_1$	$\bar{\mu}_2$	$\bar{\mu}_3$	$\bar{\mu}_4$	$\bar{A}_1/\bar{A}_\mu$	$\bar{\mu}_1$	$\bar{\mu}_2$	$\bar{\mu}_3$	$\bar{\mu}_4$
n =10	110	79	8	7	6	115	72	9	10	8
25	111	67	13	8	12	116	59	17	14	10
100	112	53	18	8	21	113	47	25	17	10
1000	104	70	18	0	12	102	62	16	21	1

The weighted Kaplan-Meier estimate is better for all cases tried since $\bar{A}_1/\bar{A}_\mu \geq 1$. Under Scenario 1, the weights allocated to Populations 2 and 4 remain relatively high even for large sample sizes, meaning that a mixture of Populations 2 and 4 must be quite similar to Population 1. The weak consistency of the weighted Kaplan-Meier estimate however implies that the weight will eventually shift entirely to Population 1 in this case.

A similar behavior is observed under Scenario 2 where Population 2 and 3 share a fairly large proportion of the weight even for large sample sizes.

5.2 Gamma and Weibull Distributions

We now consider distributions of different shapes, but similar locations. Two scenarios are simulated, each with four populations: two Gammas and two Weibulls. The scale parameters (β) of the distributions are chosen as functions of the shape parameters (α) to give them an expected value of 1. For the Gamma, this means $\beta = \alpha$, but for the Weibull, $\beta = \{\Gamma(1+\alpha)\}^{-1}$. In the first scenario, the shape parameters of the Gamma are 0.9 and 1.1 respectively, those of the two Weibulls are 2 and 1.1. A second scenario is also considered where the shape parameters are 0.5, 2, 2 and 5 respectively for the two Gammas and the two Weibulls. Population 1 is of interest: the Gamma distribution with $\alpha = 0.9$ under Scenario 1 and that with $\alpha = 0.5$ under Scenario 2. Each scenario is repeated 10000 times.

Censoring times are simulated as follow. Let X be a random variable drawn from the simulated distribution and $X_1, ..., X_r$ be r additional random variables with the same distribution. The censoring time is defined as $V = \max(X_1, \ldots, X_r)$, yielding a censoring rate of $P(V < X) = 1/(r+1)$. We set $r = 3$ for a censoring rate of 25%.

Table 2: Relative performance of the weighted Kaplan-Meier estimate as measured by $100\bar{A}_1/\bar{A}_\mu$ as well as the average weights. Equal samples of size n are drawn from two Gamma populations and two Weibull populations under two different scenarios. Note that $U = 5/4$, $T = 1$ and that each entry is based on 10000 repetitions.

	Scenario 1					Scenario 2				
$100\times$	$\bar{A}_1/\bar{A}_\mu$	$\bar{\mu}_1$	$\bar{\mu}_2$	$\bar{\mu}_3$	$\bar{\mu}_4$	$\bar{A}_1/\bar{A}_\mu$	$\bar{\mu}_1$	$\bar{\mu}_2$	$\bar{\mu}_3$	$\bar{\mu}_4$
10	122	65	13	8	14	102	82	10	7	1
25	120	59	18	6	17	98	86	10	4	0
100	114	63	19	2	16	98	96	4	0	0
1000	100	85	11	0	4	100	100	0	0	0

Under both scenarios, Population 2 whose shape is closer to Population 1 tends to get more weight than either of the Weibull populations. Under Scenario 1, other populations contribute to the inference with an average weight of up to 40%. Under Scenario 2 however, the discrepancies in shapes are larger and Populations 2, 3 and 4 quickly get dismissed as the sample sizes increase. The performance is not improved by the weighted method under that scenario and small losses are sometimes observed.

Estimating the weights from the data has some cost that can be recovered when the distributions are sufficiently similar. When the distributions are too different, that cost may not be recovered and a small loss of performance is observed. The discrepancies in the shapes did not however cause the weighted method to fail under the simulated scenarios.

5.3 Survival in the United States of America

We use the decennial life tables published by the National Center for Health Statistics in 1997. From this publication, we obtain the distributions of lifetime for four subgroups of the population in the United States of America: white males, white females, males other than white, females other than white. The life tables have a resolution of one year from birth to the age of 109 years. We draw the day and time of death uniformly during the year.

We use the same strategy as in Section 5.2 to simulate censoring times with $r = 4$ yielding a censoring rate of 20%.

Inferential interest concerns the distribution of white males survival based on equal samples drawn from the four demographic groups mentioned above. For different values of the upper bound $U \in \{60, 70, 80, 90, 100\}$, we generate samples of equal size $n \in \{10, 25, 100, 1000\}$ from each of the four populations. Each scenario is repeated 20000 times.

Table 3 shows the ratio $100\bar{A}_1/\bar{A}_\mu$ for different choices of n, U and T. The weighted Kaplan-Meier estimate performs better under all scenarios considered although this advantage seems more modest for the largest sample size ($n = 1000$).

Table 3: Relative performance of the weighted Kaplan-Meier estimate as measured by $100\bar{A}_1/\bar{A}_\mu$ for different values of U and T. Samples of equal size n are drawn from each of four subpopulations, then used to estimate the distribution of the lifetimje of a white male living in the United States of America. Each scenario is repeated 20000 times.

	$T = 55$					$T = U - 5$				
	$U = 60$	70	80	90	100	$U = 60$	70	80	90	100
$n = 10$	114	135	142	118	100	114	132	137	116	100
25	137	148	149	128	101	137	141	137	122	101
100	135	143	140	128	102	135	134	128	118	101
1000	121	120	108	105	103	121	115	103	101	101

Dissimilarities between populations are considered on the interval $[0, U]$, hence the lack of a clear trend as U varies. For $U = 100$, the samples from other populations are frequently dismissed at pre-processing, thus ignored, especially for small sample sizes. Indeed, 25% of the white females reach the age of 90, but less than 3% survive long enough to celebrate their 100^{th} birthday. An abrupt change in the average weights is observed in Figure 1 and probably explains the drop in performance from $U = 90$ to $U = 100$.

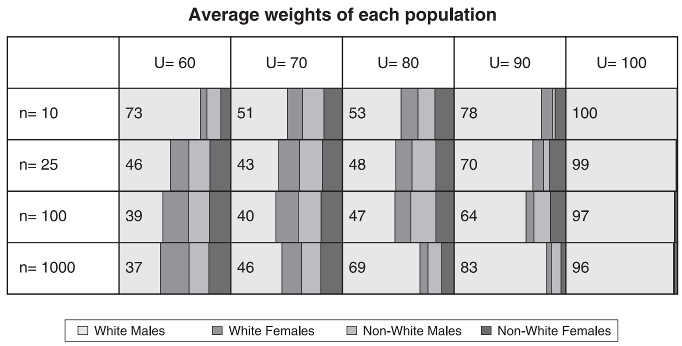

Figure 1: Average weights for four samples of size n drawn from demographic sub-populations of the United States of America. The area of the cells are proportional to the average weight allocated to each population. The numbers correspond to $100\bar{\mu}_1$ and are averaged over 20000 repetitions. Different values of U are considered.

Unless a mixture of the distributions of Populations 2, 3 and 4 is identical to that of the distribution of Population 1, Theorem 2 implies that the weight allocated to Population 1 converges to 1 as the sample sizes increase. This tendency is not observed for $U \in \{60, 70\}$, meaning that a mixture of the other 3 distributions must be very similar to that of Population 1 on $[0, U]$.

Figure 2 displays estimates of the distribution functions for a given simulated sample. The smooth gray line shows the true distribution of the lifetime of a white male in the United States of America, the plain black line shows the Kaplan-Meier estimate based on a sample of size n and the dashed line corresponds to the weighted Kaplan-Meier estimate. The numbers on each panel correspond to $\bar{A}_\mu$ and $\bar{A}_1$ respectively with $T = 75$. As we may expect from the previous tables, $\bar{A}_\mu$ is typically smaller than $\bar{A}_1$, although exceptions arise such as for $n = 1000$ on Figure 2.

A close look at Figure 2 shows an important advantage of the weighted Kaplan-Meier estimate over the Kaplan-Meier estimate: the increased number of points where jumps may occur yields a smoother step function.

Table 4 shows the performances of the weighted Kaplan-Meier in estimating $F_1(55) = 0.11976$ or $F_1^{-1}(0.10) = 52.081$ as measured by the ratio of their mean squared errors. Note

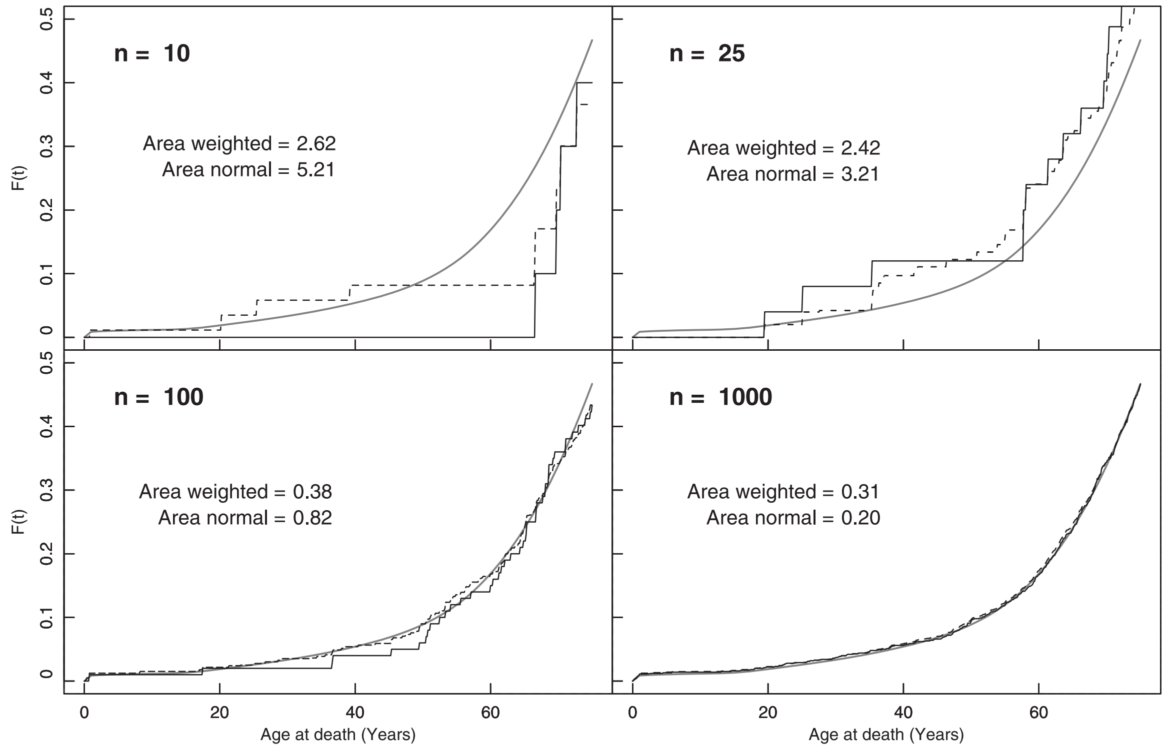

Figure 2: Typical examples of the weighted Kaplan-Meier estimate (dashed line) and of the usual Kaplan-Meier estimate (plain black line) for different sample sizes. Note that $U = 80$ and $T = 75$. The true distribution is depicted by a smooth gray line.

that we write $\hat{q}_1 = \hat{F}_1^{-1}(0.10)$ and $\hat{q}_\mu = \hat{F}_\mu^{-1}(0.10)$.

The estimates obtained from the weighted method feature a smaller mean squared error in almost all cases. Moreover, the magnitude of the gains seems to outweigh that of the occasional losses, especially when we consider that such losses occur when n is large, not the cases where our method would be most useful.

Table 5 explores the effect of different censoring probabilities $p \in \{1/3, 1/4, 1/5, 1/6\}$ while we set $U = 80$ and $T = 75$ (we use diffenrent values of r to simulate the censoring times). The proportion of censored data has little or no effect on the relative performance of the Kaplan-Meier estimate compared to its weighted equivalent. A closer look at the raw data shows that the precision of both estimates are affected by a larger p, but the magnitude of this effect appears to be the same.

Overall, the weighted Kaplan-Meier estimate seems to outperform the usual Kaplan-Meier estimate in almost all the cases explored.

Table 4: Relative performance of the weighted Kaplan-Meier estimate compared to the usual Kaplan-Meier estimate for estimating $F_1(55)$ and $F_1^{-1}(0.10)$ as measured by 100 MSE$\{\hat{F}_1(55)\}$/MSE$\{\hat{F}_\mu(55)\}$ and 100 MSE$(\hat{q}_1)$/MSE$(\hat{q}_\mu)$ respectively. Samples of equal size n are drawn from four subgroups of the American population. Different choices of U are considered; each scenario is repeated 20000 times.

	100 MSE$\{\hat{F}_1(55)\}$/MSE$\{\hat{F}_\mu(55)\}$					100 MSE$(\hat{q}_1)$/MSE$(\hat{q}_\mu)$				
	$U = 60$	70	80	90	100	60	70	80	90	100
$n = 10$	117	151	172	134	101	120	140	161	141	100
25	137	159	170	149	102	153	173	172	133	101
100	125	142	142	134	104	124	145	139	126	102
1000	110	107	84	86	103	119	113	86	86	106

5.4 Bootstrap Confidence Intervals

In this section, we use the bootstrap to determine pointwise confidence intervals for the weighted Kaplan-Meier estimate. Davison & Hinkley (1997) discuss techniques to perform bootstrap in the presence of right-censorship, but we use one of the approaches described by Efron (1981).

For a given sample, we calculate the Kaplan-Meier estimate of the survival times as well as the Kaplan-Meier estimate of the censored time (time of censoring becomes the event of interest and we see death as the event that censors it). We simulate the bootstrap sample by drawing from these two distributions. Exactly one of the two Kaplan-Meier curves will not reach 1, e.g. the estimate of the lifetime's cumulative distribution function when the last datum is censored. The remaining probability mass can be allocated to ∞ without complications: taking the minimum between the times of death and censoring will not allow ∞ into the data.

The procedure above is repeated separately in each of the m populations to produce bootstrap samples from which the weighted Kaplan-Meier estimate can be determined.

We generate samples of size 100 from the four subgroups of the American population described in Section 5.3, fix $U = 80$ and $r = 4$ for a censoring rate of 20%. The bootstrap is used to produce 90% pointwise confidence intervals for the weighted Kaplan-Meier estimate

Table 5: Average weights for different rates of censoring p and different sample size n as well as the relative performance of the weighted Kaplan-Meier estimate as measured by $100\bar{A}_1/\bar{A}_\mu$. Samples of equal size n are drawn from four subgroups of the American population. Figures are averaged over 20000 repetitions and the values $U = 80$ and $T = 75$ are used.

	$100\bar{\mu}_1$				$100\,\bar{A}_1/\bar{A}_\mu$			
	$p = \frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{1}{6}$
$n = 10$	57	54	53	52	133	135	137	136
25	50	48	48	47	136	137	137	138
100	47	47	47	47	127	127	128	126
1000	68	69	69	69	102	103	103	102

and likewise for the Kaplan-Meier estimate based on white males alone. For completeness, confidence intervals based on Greenwood's formula are also calculated.

We simulate 10000 samples on which we calculate both estimates and their pointwise confidence intervals. The average length of the confidence interval with respect to time and the estimated coverage probabilities are displayed in Figure 3.

Using more samples decreases the variance of the weighted Kaplan-Meier estimate compared to the Kaplan-Meier estimate as shown by the uniformly shorter confidence intervals of the former.

All three versions of confidence intervals flirt with the 90% confidence level, except for values close to zero. The estimated coverage of the weighted Kaplan-Meier estimate is much smoother than that of the Kaplan-Meier estimate, whether we use Greenwood's formula or the bootstrap. The big jumps in coverage tend to occur around time points that solve $F(t) = k/100$ where k is an integer. The smoothness of the weighted Kaplan-Meier estimate comes from the larger number of data points that it uses, but also from the fact that the corresponding jumps are weighted, allowing a greater variety of sizes. The result is striking in Figure 3.

Although the convergence of resampling methods for the weighted Kaplan-Meier estimate is not proved formally, the example above indicates that such methods may behave

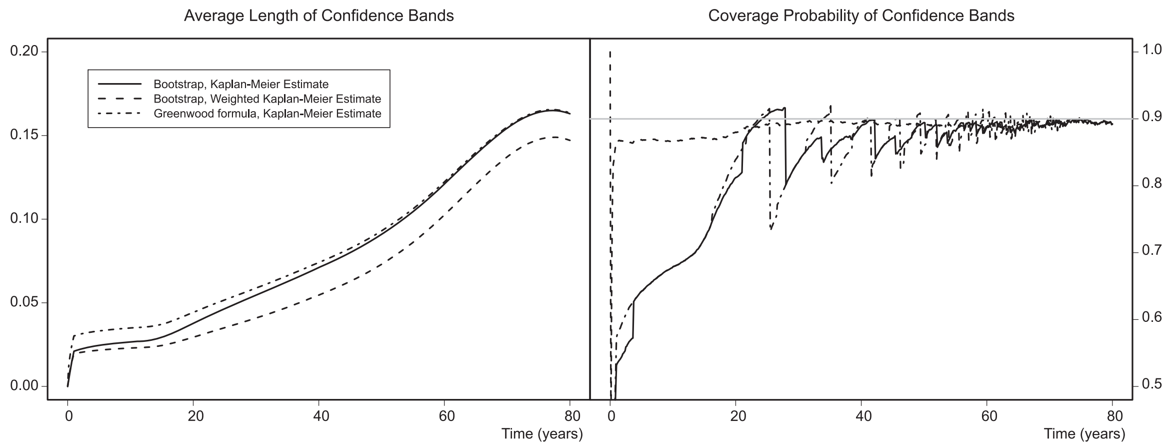

Figure 3: Average length of the confidence intervals (left panel) and coverage probability (right panel) as a function of time. All lines are obtained from averages over 10000 simulated samples.

well. Practitioners who have hesitations with using the bootstrap could test it in small simulations where the pseudo-data is akin to that expected from their study.

When data are available from different sources that are likely to feature a distribution similar to that of the population of interest, it seems preferable to incorporate that information rather than dismissing it.

6 Case Study: Survival After Kidney Transplant

We illustrate the use of the weighted Kaplan-Meier estimate on a real dataset taken from Klein & Moeschberger (1997). The time of death (in days) of 863 kidney transplant patients are given for each of four demographic groups: 432 white males, 280 white females, 92 black males and 59 black females. The number of observed values are 73, 39, 14 and 14 respectively for the four groups, the rest are right-censored. The Kaplan-Meier estimates of the distribution of survival for each of the groups appear in Figure 4.

We address the problem of estimating the distribution of survival in the two smallest groups: black males and black females. For both of these problems, we will compare the weighted Kaplan-Meier to the usual Kaplan-Meier estimate.

Figure 4 shows that the distribution of the survival of black males is somewhere in the middle of that of the white patients, but the distribution of black female seems different,

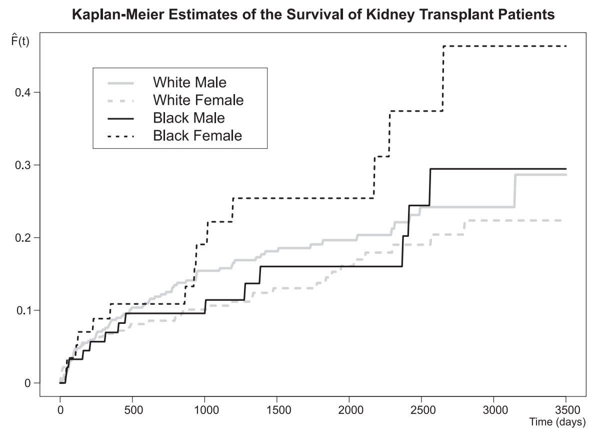

Figure 4: Kaplan-Meier estimate of the distribution of survival for kidney transplant patients in four demographic groups.

featuring the highest risk of all groups. Using the weighted Kaplan-Meier in that case will give an example of what happens when the distributions are not as similar as we wished.

The longest follow-up time for this study is 3434 days (over 9.4 years). Values of T and U must thus be smaller than that figure. Let us suppose that we are interested in the distribution of survival during the five first years. We can set $T = 1825$ (5 years) and choose U just slightly greater, say $U = 2000$.

Keeping a gap between U and T is necessary for the asymptotic convergence of the method. It is however not desirable to choose a very large U because the weights compare the Kaplan-Meier functions on the interval $[0, U]$. Therefore, discrepancies between the populations in the interval $[T, U]$ could play a role in discarding samples containing useful information on $[0, T]$, the interval of interest.

Another way to look at this is to keep in mind that the convergence of the weighted Kaplan-Meier estimate is not guaranteed at U. Choosing T very close to U will thus mean a slower convergence at T.

Let $\hat{F}_{BM}$ be the Kaplan-Meier estimate based on the data from the black males alone and $\hat{F}_{BF}$ that of the black females. We define $\hat{F}_{WM}$ and $\hat{F}_{WF}$ similarly.

The weighted Kaplan-Meier estimate of the distribution of of a black male survival after a kidney transplant corresponds to $0.151\hat{F}_{BM} + 0.375\hat{F}_{WM} + 0.440\hat{F}_{WF} + 0.034\hat{F}_{BF}$. For the black females, we get $0.476\hat{F}_{BF} + 0.524\hat{F}_{WM}$. These functions are displayed in Figure 5 with 90% pointwise confidence intervals determined using 10000 bootstrap samples.

For reference, the corresponding Kaplan-Meier estimate ($\hat{F}_{BM}$ and $\hat{F}_{BF}$) are drawn as a thick gray lines along with 90% pointwise confidence intervals. All confidence intervals are obtained using the bootstrap approach presented in Section 5.4.

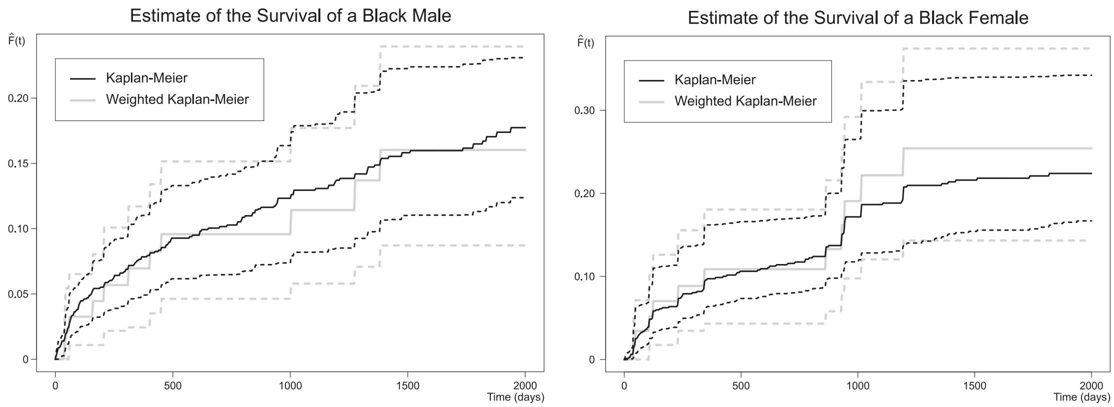

Figure 5: Weighted Kaplan-Meier estimate of the distribution of survival after a kidney transplant for black males (left panel) and black females (right panel) based on data from four demographic groups. The thick gray lines correspond to the usual Kaplan-Meier estimate. The pointwise confidence intervals are obtained with bootstrap.

The minimum averaged mean squared error weights attempt to reduce the variance of the estimate, while limiting the bias it could incur. Therefore, large samples tend to be given important weights, especially when their distributions are similar to the target distribution. This explains why there is so much weight given to the white populations in the estimation of the weighted Kaplan-Meier for the black males. By comparison, the small sample of black females that features a seemingly different distribution provides a rather small contribution.

The left panel of Figure 5 shows that the weighted Kaplan-Meier estimate of the distribution of black males survival is very close to $\hat{F}_{BM}$, but it is much smoother and has narrower confidence intervals. We cannot easily verify the coverage probability of the bootstrap intervals on a single data set, but we can suppose that the behavior observed in the simulations of Section 5.4 would occur here too.

The Kaplan-Meier estimate of the distribution of the survival of black female patients is the furthest away from the 3 other estimates. It is therefore not very surprising that two populations get completely dismissed in the weighted Kaplan-Meier estimate. The importance of the weight allocated to the population of white males – more than half of the total – may however seem unexpected. This can be explained by the proximity of $\hat{F}_{BF}$ and $\hat{F}_{WM}$ in the first 3 years, but also by the large size of the white males sample which means more potential for reduced variance. The right panel of Figure 5 shows that the weighted estimate is close to the usual Kaplan-Meier despite the large weight allocated to the sample of white males.

To better understand the behavior of the method, note that the weight allocated to the white male sample falls to 0.349 when $U = 2500$ and to 0.275 when $U = 3000$. This reflects the fact that when compared on these longer intervals, there is more discrepancies between the Kaplan-Meier estimates of the white males and that of the black females.

When data is available from different similar populations, using the minimum averaged mean squared error weights to build a weighted Kaplan-Meier estimate let us exploit that data to obtain a smoother and typically less variable estimate. This is all done without making parametric assumptions. The simulations of Section 5.4 showed the advantages of the method; the case-study of this section illustrates that it can work well with real data.

7 Conclusions

Plante (2008) defines a data-based criterion to determine mixing probabilities that make $\hat{F}_\lambda$ close to $\hat{F}_1$, and less variable. Extending the minimum averaged mean squared error weights to right-censored data allows to define the weighted Kaplan-Meier estimate, a nonparametric estimate of the distribution of lifetimes that borrows strength from similar populations.

The weighted Kaplan-Meier estimate converges weakly and uniformly to the target distribution. Simulations confirmed that the addition of other samples allows to outperform the usual Kaplan-Meier estimate for different scenarios involving finite samples. This addition also means improved smoothness and better coverage.

Determining the asymptotic distribution of the weighted Kaplan-Meier estimate would be useful, but the data-dependence of the weights prevents us from building a proof using the usual strategies. Meanwhile, resampling methods may be used to determine confidence intervals or variance estimates as illustrated in the simulations of this paper.

Although almost all the simulated scenarios yield results showing improved inference, we could not determine a specific criterion which would guarantee the superiority of the weighted Kaplan-Meier estimate over the usual estimate. Finding such a criterion would constitute a neat contribution. For now, if one hesitates, simulation studies with distributions that are akin to those expected for the study could justify their choice.

Other avenues for future research include the determination of optimal weights for the weighted Kaplan-Meier estimate and the development of the weighted partial likelihood.

With the promising results shown in this paper, we are looking forward to seeing the weighted Kaplan-Meier estimate used successfully in different case studies or applications.

Acknowledgements

This paper constitutes part of the author's doctoral thesis supervised by Professor James V. Zidek at the University of British Columbia. I am thankful to

Professor Zidek for his guidance and support in discovering the weighted likelihood and in developing the results that led to this publication. For partial support of this work through graduate scholarships and research grants, thanks are due to the Natural Sciences and Engineering Research Council of Canada and to the Fonds québécois de la recherche sur la nature et les technologies.

Appendix

The mathematical proofs of the theoretical results are presented below.

Proof of Theorem 1. By the definition of the minimum averaged mean squared error weights, $[1, 0, \ldots, 0]^\mathsf{T}$ is a suboptimal choice of weights. Following Equations (2) and (3), we thus have

$$\begin{aligned}\int_0^U \left\{\hat{F}_{1k}(t) - \hat{G}_k(t)\right\}^2 \mathrm{d}\hat{F}_{1k} \leq P_k\{\mu_k\} \;\leq\; P_k\{[1, 0, \ldots, 0]^\mathsf{T}\} &= \int_0^U \widetilde{\mathrm{var}}\left\{\hat{F}_{1k}(t)\right\} \mathrm{d}\hat{F}_{1k}(t) \\ &\leq \left\{\sum_{0 \leq s \leq U} \frac{\mathrm{d}N_{1k}(s)}{Y_{1k}(s)Y_{1k}(s^+)}\right\} \int_0^U \{1 - \hat{F}_{1k}(t)\}^2 \,\mathrm{d}\hat{F}_{1k}(t)\end{aligned}$$

Now consider the term

$$\sum_{0 \leq s \leq U} \frac{\mathrm{d}N_{1k}(s)}{Y_{1k}(s)Y_{1k}(s^+)} \leq \sum_{0 \leq s \leq U} \frac{Y_{1k}(s) - Y_{1k}(s^+)}{Y_{1k}(s)Y_{1k}(s^+)} \leq \frac{1}{Y_{1k}(U)}$$

since $\mathrm{d}N_{1k}(s) \leq \mathrm{d}Y_{1k}(s)$ for all s and the sum in the middle term telescopes, even in the case of concurrent censoring times. Since the Z_{1j}'s are independent, $Y_{1k}(U)$ has a Binomial distribution with parameters n_{1k} and $p_{H_1} = 1 - H_1(U)$. Let $\epsilon > 0$, then

$$\text{pr}\left\{\sum_{0\le s\le U} \frac{\mathrm{d}N_{1k}(s)}{Y_{1k}(s)Y_{1k}(s^+)} > \epsilon\right\} \le \text{pr}\left\{\frac{1}{Y_{1k}(U)} > \epsilon\right\} \approx \Phi\left[\frac{1/\epsilon - n_{1k}p_{H_1}}{\sqrt{n_{1k}p_{H_1}\{1-p_{H_1}\}}}\right] \to 0 \qquad (6)$$

as $k \to \infty$ since the argument inside the standard normal cumulative distribution function Φ tends to $-\infty$. ◇

Let $\mathcal{J}_k = \max_{t\le U} |\hat{F}_{1k}(t) - \hat{F}_{1k}(t^-)| = J_{k\nu_k}$, be the biggest jump of $\hat{F}_{1k}(t)$ on the interval $[0, U]$. By the uniform convergence of the Kaplan-Meier estimate, $\mathcal{J}_k \to 0$ almost surely as $k \to \infty$.

Recalling that $T < U$ is such that $H_1^*(T) < H_1^*(U)$, we let $D_k = N_{1k}(U) - N_{1k}(T)$ be the number of deaths observed in the interval $(T, U]$ among individuals sampled from Population 1. Since the Z_{1j} are independent, D_k follows a Binomial distribution with parameters n_{1k} and $H_1^*(U) - H_1^*(T)$.

Let $\ell_k = N_{1k}(T)$ be the number of deaths observed in the interval $[0, T]$, and their corresponding times of death $t_{k1} < \ldots < t_{k\ell_k} \le T$. By convention, we set $t_{k(\ell_k+1)} = \tau_{H_1}$ if no death is observed after $t_{k\ell_k}$ and we define $\mathcal{T}_k = \{t_{k1}, \ldots, t_{k(\ell_k+1)}\}$.

Lemma 1 *Let* $\mathcal{A}_k = \left\{\omega \in \Omega : \max_{0\le t\le T}\left|\hat{F}_{1k}(t) - \hat{G}_k(t)\right| \le \mathcal{J}_k + \max_{t\in\mathcal{T}}\left|\hat{F}_{1k}(t) - \hat{G}_k(t)\right|\right\}$, *then* $\text{pr}(\mathcal{A}_k) \to 1$ *as* $k \to \infty$.

Proof of Lemma 1. Fix $k \in \mathbb{N}$, $\omega \in \Omega$, and let $x_0 \in [0, T]$ be the value maximizing $|\hat{F}_{1k}(t) - \hat{G}_k(t)|$. That maximum exists since $|\hat{F}_{1k}(t) - \hat{G}_k(t)|$ is a bounded function being optimized on a compact set. Three disjoint cases are considered:

<u>Case 1:</u> $\hat{G}_k(x_0) \le \hat{F}_{1k}(x_0)$ and $D_k \ge 1$.

Let $j_1 = \max\{j \le \ell_k : t_{kj} \le x_0\}$ be the index of the largest time of death from Population 1 inferior to x_0. By the choice of j_1, t_{kj_1} belongs to the same step as x_0 and hence $\hat{F}_{1k}(t_{kj_1}) = \hat{F}_{1k}(x_0)$. Recalling that x_0 maximizes the difference between $\hat{F}_{1k}(t)$ and $\hat{G}_k(t)$ and that $\hat{G}_k(t)$ is nondecreasing, we can write that $\max_{0\le t\le T} |\hat{F}_{1k}(t) - \hat{G}_k(t)|$ equals

$$\hat{F}_{1k}(x_0) - \hat{G}_k(x_0) \le \hat{F}_{1k}(t_{kj_1}) - \hat{G}_k(t_{kj_1}) \le \mathcal{J}_k + \max_{t\in\mathcal{T}_k}\left|\hat{F}_{1k}(t) - \hat{G}_k(t)\right|,$$

meaning that the maximum will always occur at a time of death from Population 1.

<u>Case 2:</u> $\hat{G}_k(x_0) > \hat{F}_{1k}(x_0)$ and $D_k \ge 1$.

Let $j_2 = \min\{j \le \ell_k + 1 : t_{kj} \ge x_0\}$ be the index of the smallest time of death greater than x_0. The condition $D_k \ge 1$ ensures that $t_{k(\ell_k+1)}$ exists, hence j_2 is well defined. The choice of j_2 ensures that it belongs to the step of $\hat{F}_{1k}(t)$ that immediately follows x_0, hence $\hat{F}_{1k}(t_{k(j_2-1)}) = \hat{F}_{1k}(x_0)$. For the same reasons as in Case 1, we write

$$\max_{0\le t\le T} |\hat{F}_{1k}(t) - \hat{G}_k(t)| = \hat{G}_k(x_0) - \hat{F}_{1k}(x_0) \le \hat{G}_k(t_{kj_2}) - \hat{F}_{1k}(t_{k(j_2-1)})$$
$$= \{\hat{F}_{1k}(t_{kj_2}) - \hat{F}_{1k}(t_{k(j_2-1)})\} + \hat{G}_k(t_{kj_2}) - \hat{F}_{1k}(t_{kj_2}) \le \mathcal{J}_k + \max_{t\in\mathcal{T}_k} \left|\hat{F}_{1k}(t) - \hat{G}_k(t)\right|,$$

meaning that under Case 2, the maximum will occur immediately before a jump of $\hat{F}_{1k}(t)$.

<u>Case 3:</u> $D_k = 0$.

This event has probability $[1 - \{H_1^*(U) - H_1^*(T)\}]^{n_{1k}} \to 0$ as $k \to \infty$.

Combining all three cases, implies that $\text{pr}(\mathcal{A}_k) \ge \text{pr}(D_k \ge 1) \to 1$ as $k \to \infty$. $\diamond$

Proof of Theorem 2. Let $\epsilon > 0$ be such that $\epsilon < H_1^*(U) - H_1^*(T)$. We first show that

$$\text{pr}\left\{\max_{t\in\mathcal{T}_k} \left|\hat{F}_{1k}(t) - \hat{G}_k(t)\right| > \epsilon\right\} \to 0$$

as $k \to \infty$.

For a large k, let $x_k \in \{1, \ldots, \ell_k+1\}$ be the index of a time of death where the difference $\left|\hat{F}_{1k}(t) - \hat{G}_k(t)\right|$ is maximized. We define the following three events:

$$A_k = \left\{\omega \in \Omega : \hat{F}_{1k}(t_{kx_k}) - \hat{G}_k(t_{kx_k}) > \epsilon\right\},$$

$$B_k = \left\{\omega \in \Omega : \hat{G}_k(t_{kx_k}) - \hat{F}_{1k}(t_{kx_k}) > \epsilon\right\} \quad \text{and} \quad C_k = \{\omega \in \Omega : D_k \ge \epsilon n_{1k} + 1\}.$$

Then,

$$\text{pr}\left\{\max_{t\in\mathcal{T}_k} \left|\hat{F}_{1k}(t) - \hat{G}_k(t)\right| > \epsilon\right\} \le \text{pr}(C_k^C) + \text{pr}(A_k \cap C_k) + \text{pr}(B_k \cap C_k)$$

and we show that each of the three probabilities on the right hand side go to zero as $k \to \infty$.

<u>Case 1:</u> $\text{pr}(C_k^C) \to 0$.

Recalling that D_k follows a Binomial distribution with n_{1k} trials and probability of success $\{H_1^*(U) - H_1^*(T)\}$, we have

$$\text{pr}(C_k^C) \approx \Phi\left(\frac{1 + n_{1k}\left[\epsilon - \{H_1^*(U) - H_1^*(T)\}\right]}{\sqrt{n_{1k}\{H_1^*(U) - H_1^*(T)\}\{1 - H_1^*(U) + H_1^*(T)\}}}\right) \to 0$$

as $k \to \infty$ by the choice of a small enough ϵ.

Case 2: $\text{pr}(A_k \cap C_k) \to 0$.

Let $u_k = \min\{u : \sum_{i=u+1}^{x_k} J_{ki} \le \epsilon\}$. This index exists when k is large enough since $J_{kx_k} \le \mathcal{J}_k \to 0$ and $\sum_{i=1}^{x_k} J_{ki} = \hat{F}_{1k}(t_{kx_k}) > \hat{G}_k(t_{kx_k}) + \epsilon \ge \epsilon$.

For a large enough k, $J_{kx_k} < \epsilon$ and hence $u_k \le x_k - 1$. For $j \in \{u_k, \ldots, x_k - 1\}$, we have

$$\hat{F}_{1k}(t_{kj}) - \hat{G}_k(t_{kj}) \ge \hat{F}_{1k}(t_{kx_k}) - \hat{G}_k(t_{kx_k}) - \sum_{i=j+1}^{x_k} J_{ki} \ge \epsilon - \sum_{i=j+1}^{x_k} J_{ki} \ge 0.$$

The last inequality holds because of the choice of u_k. The function $\hat{F}_{1k}(t)$ gives a mass of J_{kj} to the point t_{kj}, and hence

$$\begin{aligned}\int_0^U |\hat{G}_k(t) - \hat{F}_{1k}(t)|^2 \,\mathrm{d}\hat{F}_{1k}(t) &\ge \sum_{j=u_k}^{x_k} J_{kj} |\hat{G}_k(t_{kj}) - \hat{F}_{1k}(t_{kj})|^2 \\ &\ge \; J_{kx_k}\epsilon^2 + \sum_{j=u_k}^{x_k-1} J_{kj} \left(\epsilon - \sum_{i=j+1}^{x_k} J_{ki}\right)^2 \to \int_0^\epsilon (\epsilon - x)^2 \,\mathrm{d}x = \frac{\epsilon^3}{3} \qquad (7)\end{aligned}$$

since the summation corresponds to the Riemann sum for the integral $\int_0^\epsilon (\epsilon - x)^2 \,\mathrm{d}x$ depicted in Figure 6. The sum converges as $k \to \infty$ because the width of the columns J_{kj} tend to zero.

To clarify the link between the Riemann sum and the integral, consider the change of variable $p = x_k - j$ and let

$$c_{kp} = \begin{cases} 0 & p = 0 \\ \sum_{i=1}^{p} J_{k(x_k - i + 1)} & p = 1, \ldots, x_k - u_k \end{cases}.$$

Note that $c_{k(p+1)} - c_{kp} = J_{k(x_k - p)} = J_{kj}$ and with respect to the variable j, c_{kp} equals $\sum_{i=j+1}^{x_k} J_{ki}$ when $p > 0$. We can thus write the expression in (7) as

$$\sum_{p=0}^{x_k - u_k - 1} (c_{k(p+1)} - c_{kp})(\epsilon - c_{kp})^2 \to \int_0^\epsilon (\epsilon - x)^2 \,\mathrm{d}x = \int_0^\epsilon x^2 \,\mathrm{d}x = \frac{\epsilon^3}{3} \qquad (8)$$

Consequently, there exists a k_0 such that $\int_0^U |\hat{G}_k(t) - \hat{F}_{1k}(t)|^2 \,\mathrm{d}\hat{F}_{1k}(t) > \epsilon^3/6$ for all $k \ge k_0$, an event of probability 0 according to Theorem 1. We conclude that $\text{pr}(A_k \cap C_k) \to 0$ as $k \to \infty$.

Case 3: $\text{pr}(B_k \cap C_k) \to 0$.

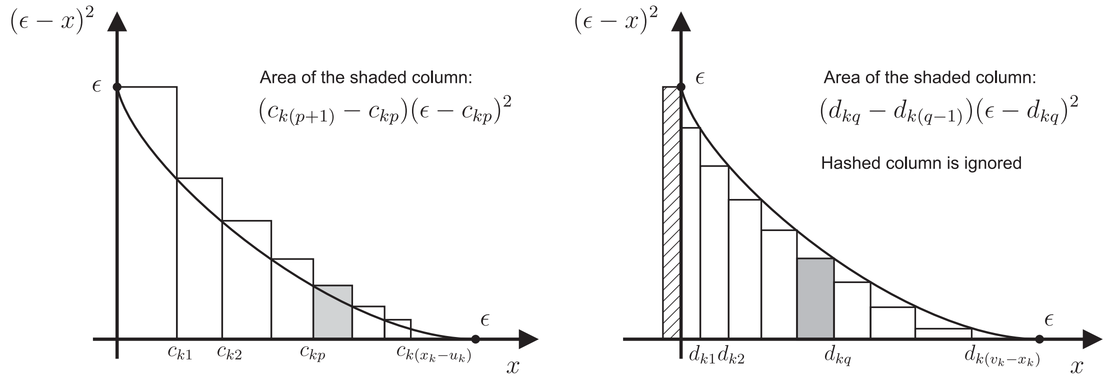

Figure 6: Graphics representing the Riemann sums used in the proof of Case 2 (left panel) and Case 3 (right panel).

Recall that the smallest possible size of a jump in $\hat{F}_{1j}(t)$ is $1/n_{1k}$. Therefore, $J_{kj} \geq 1/n_{1k}$ and $D_k \geq \epsilon n_{1k} + 1$ implies that

$$\sum_{\{j: t_{kj} \in (T,U], j > \ell_k + 1\}} J_{kj} \geq \frac{\epsilon n_{1k} + 1}{n_{1k}} > \epsilon.$$

Let $v_k = \max\{v : \sum_{j=x_k+1}^{v} J_{kj} \leq \epsilon\}$. For a large enough k, $J_{k(x_k+1)} \leq \mathcal{J}_k < \epsilon$ and thus $v_k \geq x_k + 1$. For $j \in \{x_k + 1, \ldots, v_k\}$,

$$\hat{G}_k(t_{kj}) - \hat{F}_{1k}(t_{kj}) \geq \hat{G}_k(t_{kx_k}) - \hat{F}_{1k}(t_{kx_k}) - \sum_{i=x_k+1}^{j} J_{ki} \geq \epsilon - \sum_{i=x_k+1}^{j} J_{ki} \geq 0,$$

the last inequality holding because of the choice of v_k. Using again the fact that $\mathrm{d}\hat{F}_{1k}(t)$ allocates a mass of J_{kj} to t_{kj}, we find that

$$\begin{aligned} \int_0^U |\hat{G}_k(t) - \hat{F}_{1k}(t)|^2 \, \mathrm{d}\hat{F}_{1k}(t) &\geq \sum_{j=x_k}^{v_k} J_{kj} |\hat{G}_k(t_{kj}) - \hat{F}_{1k}(t_{kj})|^2 \\ &\geq \sum_{j=x_k+1}^{v_k} J_{kj} \left(\epsilon - \sum_{i=x_k+1}^{j} J_{ki} \right)^2 \to \int_0^\epsilon (\epsilon - x)^2 \, \mathrm{d}x = \frac{\epsilon^3}{3} \end{aligned} \quad (9)$$

since the summation corresponds to the Riemann sum for the integral $\int_0^\epsilon (\epsilon - x)^2 \, \mathrm{d}x$ depicted on the right panel of Figure 6. The term $J_{kx_k}\epsilon^2$ ignored in Equation (9) corresponds to the hashed column. The sum converges as $k \to \infty$ because the width of the columns J_{kj} tend to zero. Figure 6 uses the change of variables $q = j - x_k$ and

$$d_{kq} = \begin{cases} 0 & q = 0 \\ \sum_{i=1}^{q} J_{k(x_k+i)} & q = 1, \ldots, v_k - x_k \end{cases}.$$

Combining the three cases implies that $\max_{t\in\mathcal{T}_k} |\hat{F}_{1k}(t) - \hat{G}_k(t)|$ converges weakly to 0. The addition of Lemma 1 and the fact that $\mathcal{J}_k \to 0$ imply that $\sup_{t\leq T} |\hat{G}_k(t) - \hat{F}_{1k}(t)|$ converges weakly to 0. Finally, the triangular inequality yields

$$\sup_{0\leq t\leq T} \left|\hat{G}_k(t) - F_1(t)\right| \leq \sup_{0\leq t\leq T} \left|\hat{G}_k(t) - \hat{F}_{1k}(t)\right| + \sup_{0\leq t\leq T} \left|\hat{F}_{1k}(t) - F_1(t)\right|.$$

The uniform convergence of the Kaplan-Meier estimate completes the proof. ◇

References

Akaike H (1977) On entropy maximization principle. Applications of statistics 27–42

Breslow NE & Crowley J (1974) A large sample study of the life table and product limit estimates under random censorship. The Annals of Statistics, 2:437–453

Davison AC & Hinkley DV (1997) Bootstrap Methods and their Application. Cambridge University Press, New York

Efron B (1967) The two sample problem with censored data Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. University of California, Berkeley 4:831–853

Efron B (1981) Censored data and the bootstrap. Journal of the American Statistical Association. 76:312–319

Földes A & Rejtö L (1981) Strong uniform consistency for nonparametric survival curve estimators from randomly censored data. The Annals of Statistics 9:122–129

Hu F (1994) Relevance weighted smoothing and a new bootstrap method, unpublished doctoral dissertation. Department of Statistics, The University of British Columbia, 177 pp

Hu F & Zidek JV (1993) A relevance weighted nonparametric quantile estimator. Technical report no 134, Department of Statistics, The University of British Columbia, Vancouver

Hu F & Zidek JV (2002) The weighted likelihood. The Canadian Journal of Statistics, 30:347–371

Kaplan EL & Meier P (1958) Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 53:457–481

Klein JP & Moeschberger ML (1997) Survival Analysis: techniques for censored and truncated data. Springer-Verlag, New York

National Center for Health Statistics (1997) US decennial life tables for 1989–91 vol 1 no 1. Hyattsville, Maryland, 44 pp

Plante J-F (2008) Nonparametric adaptive likelihood weights. The Canadian Journal of Statistics, 36:443–461

Plante J-F (2009) On the Asymptotic Properties of the MAMSE Adaptive Likelihood Weights. Journal of Statistical Planning and Inference, in press.

Plante J-F (2007) Adaptive likelihood weights and mixtures of empirical distributions. Unpublished doctoral dissertation, Department of Statistics, University of British Columbia, 171 pp

Wang X (2001) Maximum weighted likelihood estimation. unpublished doctoral dissertation, Department of Statistics, The University of British Columbia, 151 pp

Wang X, van Eeden, C, Zidek JV (2004) Asymptotic properties of maximum weighted likelihood estimators. Journal of Statistical Planning and Inference 119:37–54

Wang X, Zidek JV (2005) Selecting likelihood weights by cross-validation. The Annals of Statistics 33:463–501

Winter BB, Földes A, Rejtö L (1978) Glivenko-Cantelli theorems for the PL estimate. Problems of Control and Information Theory, 7:213–225

---

10 February 2009