ESTIMATION FROM CROSS-SECTIONAL DATA UNDER A SEMIPARAMETRIC TRUNCATION MODEL

Cédric Heuchenne, Jacobo De uña Alvarez, Géraldine Laurent

REPRINT | 2021 / 32

Biometrika (201.), **..**, .., *pp*. 1–19

Printed in Great Britain

Advance Access publication on ..

Estimation from cross-sectional data under a semiparametric truncation model

BY J. DE UÑA-ÁLVAREZ

Department of Statistics and Operations Research & CINBIO, University of Vigo, Campus Lagoas-Marcosende, 36310 Vigo, Spain

jacobo@uvigo.es

C. HEUCHENNE

HEC-Liege, University of Liege and Institute of Statistics, Biostatistics and Actuarial Sciences, Université catholique de Louvain, Rue Louvrex, 14, 4000 Liege, Belgium

C.Heuchenne@ulg.ac.be

G. LAURENT

HEC-Liege, University of Liege, Rue Louvrex, 14, 4000 Liege, Belgium

G.Laurent@ulg.ac.be

Cross-sectional sampling is often used when investigating inter-event times, resulting in left-truncated and right-censored data. In this paper we consider a semiparametric truncation model in which the truncating variable is assumed to belong to a certain parametric family. Two methods to estimate both the truncation and lifetime distributions are considered. Asymptotic representations of the estimators for the lifetime distribution are obtained, and their weak convergence is established. One of the conclusions of our research is that both estimators perform in practice better than Wang's nonparametric maximum likelihood estimator in the sense of the integrated mean squared error when the parametric family for the truncation is sufficiently close to its true distribution. Moreover, the full likelihood approach is preferable to the conditional likelihood approach for the estimation of the lifetime distribution but not necessarily for the estimation of the truncation distribution. A real data application about Alzheimer's disease is carried out together with related bootstrap inference. Hypothesis tests reject the uniform truncation distribution but several other parametric estimations lead to similar behaviors for the distributions of both the truncation and the lifetime after disease onset.

Some key words: Bootstrap; Cross-sectional sampling; Left truncation; Length bias; Right censoring.

1. INTRODUCTION

Cross-sectional survival data are encountered in many applications in which the variable of interest is an inter-event time. Cross-sectional sampling implies that only individuals who did not yet experience at a given time point the event of interest, e.g. death, are recruited. In Section 5, we study the time from the onset of the Alzheimer's disease to death, but the corresponding sample is constructed at a fixed time point with individuals already suffering from this disease. In such a sample, the time from the disease onset to death tends to be larger than in the target population. Such an issue is typically referred to as length-biased sampling. This bias is present in other settings, like reliability studies (Horváth, 1985) or econometrics (de Uña-Álvarez et al., 2009) and whenever the observer arrives at the process at a random time point. Under another viewpoint, cross-sectional sampling introduces some random truncation from the left on the lifetimes of interest. By defining T as the time from onset to the cross-section date and Y as the total survival time, under cross-sectional sampling only pairs (T, Y) satisfying $T \leq Y$ are recruited. This issue has been extensively investigated in the literature. See for example Tsai et al. (1987), Wang (1991), or Gijbels and Wang (1993), who investigated properties of the conditional nonparametric maximum-likelihood estimator of a survival function with left-truncated and right-censored data.

Recently much attention has been paid to the situation in which the distribution of the truncation variable is partially or completely known. For example in the stationary case, when investigating the survival time for a disease which exhibits a constant incidence rate over a certain time interval, a uniform distribution is known to hold for the truncation time T. In the uncensored case, Wang (1989) proposed an estimator of the survival function when the truncation distribution belongs to a parametric family, which improves the variance of the conditional nonparametric maximum likelihood estimator. Gilbert et al. (1999) extended this result to the case of s selection-biased samples where selection bias is characterized by a finite-dimensional parameter. With uniform truncation, Asgharian et al. (2002), see also Asgharian and Wolfson (2005), introduced the nonparametric maximum likelihood estimator of the survival function, and demonstrated its superiority compared to the conditional nonparametric maximum likelihood estimator. Luo and Tsai (2009) proposed two explicit-form approximations to the nonparametric maximum likelihood estimator with censored data and arbitrary, albeit known, truncation distribution; they proved that their estimators have an asymptotic variance smaller than that of the conditional nonparametric maximum likelihood estimator, and they showed by simulations that they are competitive to Asgharian et al. (2002)'s estimator. When the censoring distribution degenerates and the truncation is uniform, one of Luo and Tsai (2009)'s proposals reduces to the estimator previously introduced in de Uña-Álvarez (2004). See also Huang and Qin (2011) for a related estimator. To sum up, estimators which incorporate information about the truncation time are preferred to Wang (1991)'s conditional nonparametric maximum likelihood estimator, and the corresponding theory for known truncation distribution has been fully developed. See Mandel (2007) and Brunel et al. (2008) for other perspectives on this problem.

The choice of a suitable model for the truncation distribution has been considered by Mandel and Betensky (2007), who developed goodness-of-fit tests for a fully specified truncation model. These authors have recognized, however, the lack of testing procedures for a parametric truncation distribution when there is censored information. Although Wang (1989) considered the parametric case, she restricted herself to the uncensored setting. To the best of our knowledge, the problem of fitting a parametric truncation model with left-truncated and right-censored data has been partially addressed by Shen (2007, 2009). However, the proposed techniques suffer from both theoretical and practical problems. These articles only provide sketches of the proofs and partial sets of assumptions at best. In terms of inference, Shen (2007) suggests to develop goodness-of-fit testing procedures for the truncation distribution and to construct confidence intervals for the corresponding parameters based on the bootstrap. However, methods are not fully described and no numerical study is conducted.

In the present paper, we propose a full theoretical framework where relevant inferential techniques are developed and studied. We also compare the finite-sample performance of our proposed approach to that of Shen (2007).

2. Methodology

Let Y be the lifetime of ultimate interest, and let T be the left-truncation time, i.e. $\mathrm{pr}(Y \geq T) > 0$. If $T \leq Y$ then (T, X, Δ) is observed, where $X = \min(Y, C)$, $\Delta = I(Y \leq C)$, and C is a potential right-censoring time. If $T > Y$ nothing is observed. We assume that:

Assumption 1. T and Y are independent.

Assumption 2. $C - T$ is independent of $(T, Y - T)$ conditionally on $T \leq Y$.

Assumption 3. $\mathrm{pr}(C \geq T) = 1$,

Assumption 1 is typical in left-truncated scenarios. Assumption 2 implicitly states that censoring on Y is only possible after the no truncation condition $T \leq Y$ is satisfied; this is the case for most applications with cross-sectional sampling. Assumption 2 implies $\{T \leq Y\} = \{T \leq X\}$ with probability one. Finally, assumption 3 says that the residual censoring time in the observable world, i.e. when $X \geq T$, is independent of the truncation and the survival times. The residual censoring time is then also independent of the birth process given $X \geq T$, i.e., the process that generates the starting point of the studied time and that can be described by the truncation distribution.

Let $(T_1, X_1, \Delta_1), \ldots, (T_n, X_n, \Delta_n)$ be n independent observations with the same distribution as (T, X, Δ) given $T \leq X$. Under assumptions 1-3, the likelihood $\mathcal{L}$ can be decomposed as the product of the conditional likelihood of the truncation times T_i given the (X_i, Δ_i), say $\mathcal{L}_c$, and the marginal likelihood of the (X_i, Δ_i), say $\mathcal{L}_m$: $\mathcal{L} = \mathcal{L}_c \times \mathcal{L}_m$. Straightforward calculations give

$$\begin{aligned}\log \mathcal{L}_c &= \log \left\{ \prod_{i=1}^{n} \frac{dF_T(T_i)}{E_{F_T,\mathcal{G},1}(X_i)^{\Delta_i} E_{F_T,\mathcal{G},0}(X_i)^{1-\Delta_i}} \right\} \\ &\quad + \log \left[\prod_{i=1}^{n} \{1 - \mathcal{G}([X_i - T_i]^-)\}^{\Delta_i} d\mathcal{G}(X_i - T_i)^{1-\Delta_i} \right] \\ &\equiv \mathcal{L}_{c,1} + \mathcal{L}_{c,2},\end{aligned}$$

where F_T is the distribution function of T, $\mathcal{G}$ is the distribution function of $C - T$ given $T \leq X$, and

$$E_{F_T,\mathcal{G},1}(y) = E\left[\left\{1 - \mathcal{G}([y - T]^-)\right\} I(T \leq y)\right] = \int_{t \leq y} \left\{1 - \mathcal{G}([y - t]^-)\right\} dF_T(t),$$

$$E_{F_T,\mathcal{G},0}(y) = E[d\mathcal{G}(y - T) I(T \leq y)] = \int_{t \leq y} d\mathcal{G}(y - t) dF_T(t).$$

On the other hand,

$$\log \mathcal{L}_m = \log \left[\prod_{i=1}^{n} \frac{E_{F_T,\mathcal{G},1}(X_i)^{\Delta_i} E_{F_T,\mathcal{G},0}(X_i)^{1-\Delta_i}}{\int F_T dF} dF(X_i)^{\Delta_i} \{1 - F(X_i)\}^{1-\Delta_i} \right],$$

where F is the df of Y.

If F_T were known, a consistent estimator of F is (cf. Luo and Tsai, 2009)

$$\widehat{F}_{F_T}(y) = \frac{\int_{x \leq y} w^{-1}_{F_T,\mathcal{G}_n}(x) \mathcal{H}^1_n(dx)}{\int w^{-1}_{F_T,\mathcal{G}_n}(x) \mathcal{H}^1_n(dx)},$$

where $\mathcal{H}_n^1(x) = n^{-1}\sum_{i=1}^n I(X_i \le x, \Delta_i = 1)$ is the empirical subdistribution function of the uncensored lifetimes, i.e. $\mathcal{H}^1(z) = \text{pr}(X \le z, \Delta = 1 \mid X \ge T)$, $\mathcal{G}_n(\cdot)$ is the Kaplan–Meier estimator of the residual censoring distribution,

$$w_{F_T,\mathcal{G}_n}(y) = E_{F_T,\mathcal{G}_n,1}(y) = \int_0^y \left\{1 - \mathcal{G}_n((y-t)^- \wedge \mathcal{T})\right\} dF_T(t), \quad \mathcal{T} < \tau_{\mathcal{G}(\cdot)} \wedge \tau_{\mathcal{H}_{Y-T}(\cdot)},$$

where $\tau_W(\cdot) = \inf\{y : W(y) = 1\}$ defines the upper bound of the support of a random variable following a given cumulative distribution function $W(\cdot)$ and $\mathcal{H}_{Y-T}(\cdot)$ is the distribution of $Y - T$ given $Y \ge T$. The role of $\mathcal{T}$ here is to allow for the application of the asymptotic representation of the Kaplan–Meier estimator given in Lo and Singh (1986). In practice, this threshold can be chosen as the largest residual time $X_i - T_i$.

The estimator $\widehat{F}_{F_T}$ is less efficient than the nonparametric maximum likelihood estimator of F based on a known truncation distribution F_T, but the variance increase is typically small and, unlike the nonparametric maximum likelihood estimator, it has an explicit form which facilitates its practical implementation and the theoretical analysis (Luo and Tsai, 2009). However, both $\widehat{F}_{F_T}$ and the nonparametric maximum likelihood estimator rely on having a known F_T.

In order to introduce a more flexible estimator, let $\{F_{T,\theta}\}_{\theta\in\Theta}$ be a parametric family which F_T belongs to, i.e., $F_T = F_{T,\theta_0}$ for some θ_0 and a compact subset $\Theta \in \mathbb{R}^d$, $d \ge 1$, including θ_0. Introduce the hereafter named conditional maximum likelihood estimator:

$$\widehat{\theta} = \arg\max_\theta \mathcal{L}_{c,1}(\theta; \mathcal{G}_n),$$

where

$$\mathcal{L}_{c,1}(\theta; \mathcal{G}_n) = n^{-1} \log\left(\prod_{i=1}^n \frac{dF_{T,\theta}(T_i)}{E_{F_{T,\theta},\mathcal{G}_n,1}(X_i)^{\Delta_i} E_{F_{T,\theta},\mathcal{G}_n,0}(X_i)^{1-\Delta_i}}\right),$$

where $E_{F_{T,\theta},\mathcal{G}_n,1}$ and $E_{F_{T,\theta},\mathcal{G}_n,0}$ are defined in an obvious way. The estimator $\widehat{\theta}$ is a conditional maximum likelihood estimator of θ_0 except for the fact that $\mathcal{G}$ is replaced by $\mathcal{G}_n$ in $\mathcal{L}_{c,1}(\theta; \mathcal{G})$. We have

$$E_{\theta_0}\left\{\frac{\partial}{\partial\theta}\mathcal{L}_{c,1}(\theta; \mathcal{G})|_{\theta=\theta_0} \mid X_i, \Delta_i, i = 1, ..., n\right\} = 0$$

and hence usual properties as consistency and asymptotic normality can be established for

$$\widehat{\theta}(\mathcal{G}) \equiv \arg\max_\theta \mathcal{L}_{c,1}(\theta; \mathcal{G})$$

and therefore for $\widehat{\theta} = \widehat{\theta}(\mathcal{G}_n)$ because of the properties of Kaplan–Meier estimation. See Theorem 1 in Section 3.

On the basis of $\widehat{\theta}$, one may introduce a semiparametric estimator for F as:

$$\widehat{F}_{\widehat{\theta}}(y) \equiv \widehat{F}_{F_{T,\widehat{\theta}}}(y) = \frac{\int_{x\le y} w^{-1}_{\widehat{\theta},\mathcal{G}_n}(x)\mathcal{H}_n^1(dx)}{\int w^{-1}_{\widehat{\theta},\mathcal{G}_n}(x)\mathcal{H}_n^1(dx)},$$

where

$$w_{\widehat{\theta},\mathcal{G}_n}(y) \equiv w_{F_{T,\widehat{\theta}}\mathcal{G}_n}(y) = E_{F_{T,\widehat{\theta}}\mathcal{G}_n,1}(y) = \int_0^y \left\{1 - \mathcal{G}_n((y-t)^- \wedge \mathcal{T})\right\} dF_{T,\widehat{\theta}}(t).$$

In the uncensored case, this is the semiparametric estimator in Wang (1989). Under censoring, however, this is a new estimator which has to be investigated. As in Wang (1989), we expect that

$\widehat{F}_{\widehat{\theta}}(y)$ will be more efficient than the conditional nonparametric maximum likelihood estimator (cfr. e.g. Wang, 1991).

An interesting question is whether the estimation procedure for θ can be improved, in the sense of obtaining a more acurate estimator for θ_0 and also for F. This is irrelevant in the uncensored case, because $(F_{T,\widehat{\theta}}, \widehat{F}_{\widehat{\theta}})$ maximizes the full likelihood in this case. In the censored setting, note however that parts of the likelihood other than $\mathcal{L}_{c,1}$ could contain information on θ. To be more precise, note that $\mathcal{L}_m$ includes the truncation distribution F_T, and hence in principle one may ask for the maximizer of the full likelihood, which is proportional to

$$L(\theta, F) \equiv \prod_{i=1}^{n} \frac{dF_{T,\theta}(T_i)}{\int F_{T,\theta} dF} dF(X_i)^{\Delta_i}(1 - F(X_i))^{1-\Delta_i}.$$

This suggests a two-step procedure to find estimators $\left(\widetilde{\theta}, \widehat{F}_{\widetilde{\theta}}\right)$.

Step 1. Plug $\widehat{F}_\theta$ into $L(\theta, F)$ to obtain

$$L(\theta, \widehat{F}_\theta) \equiv \prod_{i=1}^{n} \frac{dF_{T,\theta}(T_i)}{\int F_{T,\theta} d\widehat{F}_\theta} d\widehat{F}_\theta(X_i)^{\Delta_i}(1 - \widehat{F}_\theta(X_i))^{1-\Delta_i},$$

where

$$\widehat{F}_\theta(y) \equiv \widehat{F}_{F_{T,\theta}}(y) = \frac{\int_{x \le y} w^{-1}_{\theta,\mathcal{G}_n}(x)\mathcal{H}^1_n(dx)}{\int w^{-1}_{\theta,\mathcal{G}_n}(x)\mathcal{H}^1_n(dx)},$$

where

$$w_{\theta,\mathcal{G}_n}(y) \equiv w_{F_{T,\theta},\mathcal{G}_n}(y) = E_{F_{T,\theta},\mathcal{G}_n,1}(y) = \int_{t \le y} \left\{1 - \mathcal{G}_n((y-t)^- \wedge \mathcal{T})\right\} dF_{T,\theta}(t).$$

However, to use the above profile likelihood function in practice, let's rewrite its log-likelihood version for some $\tilde{\mathcal{T}} < \tau_{F_{\theta_0}(\cdot)}$:

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left[\Delta_i\left\{\log f_{T,\theta}(T_i) - \log\int_0^{+\infty} F_{T,\theta}(t)d\widehat{F}_\theta(t) + \log\frac{w^{-1}_{\theta,\mathcal{G}_n}(X_i)}{\int w^{-1}_{\theta,\mathcal{G}_n}(t)\mathcal{H}^1_n(dt)} - \log n\right\}\right.\\
&\qquad\left. + I(C_i \le \tilde{\mathcal{T}})(1-\Delta_i)\left\{\log f_{T,\theta}(T_i) - \log\int_0^{+\infty} F_{T,\theta}(t)d\widehat{F}_\theta(t) + \log(1 - \widehat{F}_\theta(X_i))\right\}\right]\\
&= \mathcal{L}_p(\theta, \widehat{F}_\theta, \mathcal{G}_n, \mathcal{H}^1_n) - \frac{1}{n}\sum_{i=1}^{n}\Delta_i \log n,
\end{aligned}$$

where $f_{T,\theta}$ stands for the density of the truncation parametric model (assumed to exist). In practice, the point $\tilde{\mathcal{T}}$ can be chosen as the largest uncensored X_i.

Step 2. Compute the hereafter named full maximum likelihood estimator

$$\widetilde{\theta} = \arg\max_\theta \mathcal{L}_p(\theta, \widehat{F}_\theta, \mathcal{G}_n, \mathcal{H}^1_n).$$

Shen (2007) proposed a method which maximizes iteratively both marginal likelihood functions of $\{(X_1, \Delta_1), \ldots, (X_n, \Delta_n)\}$ and $\{T_1, \ldots, T_n\}$ given $Y \ge T$. The joint full likelihood of $\{(X_1, T_1, \Delta_1), \ldots, (X_n, T_n, \Delta_n)\}$ is thus not maximized. The first likelihood function is artificially transformed into a function that depends on the weights of a specific distribution function.

For a given value of θ, it can then be maximized by another iterative procedure; this begins with initial not necessarily consistent weights of this specific distribution function. The marginal likelihood of $\{T_1, \ldots, T_n\}$ is maximized with respect to θ for given weights estimating $F(\cdot)$; these are initialized with a product-limit estimator. Shen (2009) replaces the initial product-limit estimator of $F(\cdot)$ by another one based on an estimator of θ; this latter is obtained by maximizing a preliminary conditional likelihood of the T_i's given the X_i's, $\Delta_i = 1$ and $X \geq T$, while the former is computed by using the above iterative procedure that maximizes the marginal likelihood function of $\{(X_1, \Delta_1), \ldots, (X_n, \Delta_n)\}$. In our experience, these two methods are very similar whatever the initial estimator of $F(\cdot)$ and, interestingly, they are outperformed by our approach in simulated settings; see Section 4 for details.

3. MAIN RESULTS

In this section, asymptotic properties of $\widehat{\theta}$, $\tilde{\theta}$, $\widehat{F}_{\widehat{\theta}}$ and $\widehat{F}_{\tilde{\theta}}$ are established. We show the convergence in probability of both estimators of θ_0 and the almost sure consistency of the corresponding conditional and profile log-likelihood functions. For detailed computations, see Lemma 4 in the Supplementary Material of the paper available at *Biometrika* online. As a byproduct, two other basic Lemmas (2 and 3) about $\widehat{F}(\cdot)$ used to prove the results of this section are also displayed in the Supplementary Material. This allows us to develop asymptotic representations for $\widehat{\theta}$ and $\tilde{\theta}$ which in turn give rise to the asymptotic normality of these estimators (Theorems 1 below). Next, asymptotic representations for $\widehat{F}_{\widehat{\theta}}(y)$ and $\widehat{F}_{\tilde{\theta}}(y)$ are developed and, for $F_\theta(y) = \int_{x \leq y} w_{\theta,\mathcal{G}}^{-1}(x) d\mathcal{H}^1(x) / \int w_{\theta,\mathcal{G}}^{-1}(x) d\mathcal{H}^1(x)$, Donsker's Theorem enables to deduce weak convergence of the processes

$$n^{1/2}(\widehat{F}_{\widehat{\theta}}(y) - F_{\theta_{0,c}}(y)), \quad n^{1/2}(\widehat{F}_{\tilde{\theta}}(y) - F_{\theta_{0,p}}(y))$$

to Gaussian processes (see Assumption $(A5)$ below for definitions of $\theta_{0,c}$ and $\theta_{0,p}$).

To present these results, we need to use the asymptotic representation of the Kaplan–Meier estimator given by Lo and Singh (1986):

$$\mathcal{G}_n(t) - \mathcal{G}(t) = \frac{1}{n} \sum_{i=1}^{n} \xi(X_i - T_i, \Delta_i, t) + S_n(t),$$

where $\sup\{|S_n(t)| : -\infty < t < \mathcal{T}\} = o_P(n^{-1/2})$, and for $\mathcal{H}^0(z) = \text{pr}(X \leq z, \Delta = 0 \mid X \geq T)$, and $\mathcal{H}(z) = \text{pr}(X \leq z \mid X \geq T)$,

$$\xi(z, \delta, t) = (1 - \mathcal{G}(t)) \left(- \int_0^{z \wedge t} \frac{d\mathcal{H}^0(s)}{(1 - \mathcal{H}(s))^2} + \frac{I(z < t, \delta = 0)}{1 - \mathcal{H}(z)} \right).$$

We next describe and comment the set of assumptions needed for the main theorems.

Assumption 4. The distributions $\mathcal{G}(\cdot)$ and $\mathcal{H}_{Y-T}(\cdot)$ are continuous distribution functions.

Assumption 5. Let $\rho_{W(\cdot)} = \sup\{y : W(y) = 0\}$ define the lower bound of the support of a random variable following a given cumulative distribution function $W(\cdot)$. We have:
(i) $\tau_{F(\cdot)} = \tau_{F_T(\cdot)}$,
(ii) $\rho_{F(\cdot)} - \rho_{F_T(\cdot)} > 0, \quad \rho_{F_T(\cdot)} = 0 \; (\tau_{F(\cdot)} > \rho_{F(\cdot)})$.

Assumption 6. We have for the truncation distribution:
(i) $\sup_{\theta,t} |f_{T,\theta}(t)| < +\infty$, $\sup_{\theta,t} |\partial f_{T,\theta}(t)/\partial t| < +\infty$,
(ii) $\inf_\theta F_{T,\theta}(\rho_{F(\cdot)}) > 0$,

(iii) $\sup_{\theta\in\Theta} E(|\log f_{T,\theta}(T)|) < +\infty$, $\sup_{\theta\in\Theta} E(|\partial \log f_{T,\theta}(T)/\partial\theta_k|) < +\infty$ $(k = 1,\ldots,d$; $\theta_1,\ldots,\theta_d$ denote the components of $\theta)$,

(iv) $E(|\partial^2 \log f_{T,\theta_0}(T)/\partial\theta_j\partial\theta_k|) < +\infty$, $E(\sup_{\theta\in\Theta} |\partial^3 \log f_{T,\theta}(T)/\partial\theta_j\partial\theta_k\partial\theta_l|) < +\infty$ $(j,k,l = 1,\ldots,d)$.

(v) for all t, all derivatives with respect to components of θ of $F_{T,\theta}(t)$ up to order three are bounded uniformly in t and θ.

(vi) for all t, all derivatives with respect to components of θ of $f_{T,\theta}(t)$ up to order three are bounded uniformly in t and θ.

(vii) for all t, all derivatives with respect to components of θ of $\partial f_{T,\theta}(t)/\partial t$ up to order two are bounded uniformly in t and θ.

Assumption 7. The third moment $E(|X|^3 \mid X \geq T) < +\infty$.

Assumption 8. There exists:

(i) $\theta_{0,c} \in \Theta$ such that for all $\varepsilon > 0$, $\sup_{\theta:\|\theta-\theta_{0,c}\|\geq\varepsilon} E(\mathcal{L}_{c,1}(\theta,\mathcal{G})) < E(\mathcal{L}_{c,1}(\theta_{0,c},\mathcal{G}))$, where $\|\cdot\|$ denotes the classical Euclidean norm.

(ii) $\theta_{0,p} \in \Theta$ such that for all $\varepsilon > 0$, $\sup_{\theta:\|\theta-\theta_{0,p}\|\geq\varepsilon} E(\mathcal{L}_p(\theta, F_\theta, \mathcal{G}, \mathcal{H}^1)) < E(\mathcal{L}_p(\theta_{0,p}, F_{\theta_{0,p}}, \mathcal{G}, \mathcal{H}^1))$.

Assumption 9. We have:

(i) For $\Omega_{1,\theta_{0,c}} = -E(\partial^2\mathcal{L}_{c,1}(\theta_{0,c},\mathcal{G})/\partial\theta\partial\theta^T)$, $\det(\Omega_{1,\theta_{0,c}}) > 0$.

(ii) For $\Omega_{2,\theta_{0,p}} = -E(\partial^2\mathcal{L}_p(\theta_{0,p}, F_{\theta_{0,p}}, \mathcal{G}, \mathcal{H}^1)/\partial\theta\partial\theta^T)$, $\det(\Omega_{2,\theta_{0,p}}) > 0$.

Assumption 10. For $\mathcal{G}(\cdot)$, we have $\rho_{\mathcal{G}(\cdot)} > 0$, $\inf_{\theta\in\Theta} \inf_{\rho_{\mathcal{G}(\cdot)}\leq x\leq\tau_{F(\cdot)}} \int_0^{x\wedge\mathcal{T}} f_{T,\theta}(x - t)d\mathcal{G}(t) > 0$.

The above assumptions are very classical. Assumption 4 is used to ensure asymptotic properties of the Kaplan–Meier estimator (Lo and Singh, 1986). Assumptions 5, 6, 7, 9 and 10 are purely technical conditions mainly enabling to bound terms that appear in the development of the proofs. Assumption 8 allows identifying a unique maximum for the mean of each likelihood function (given $X \geq T$). In this assumption, we observe that $\theta_{0,c}$ and $\theta_{0,p}$ are different since they are actually solutions of different maximization problems with not necessarily the same maximum due to $\mathcal{T}$ and $\tilde{\mathcal{T}}$. However, $\mathcal{T}$ and $\tilde{\mathcal{T}}$ can be made arbitrarily close to $\tau_{\mathcal{G}(\cdot)} \wedge \tau_{\mathcal{H}_{Y-T}(\cdot)}$ and $\tau_{F(\cdot)}$ so that $\theta_{0,c}$ and $\theta_{0,p}$ become arbitrarily close to the same value θ_0. In addition, under this remark, in assumption $8(ii)$ the likelihood function should actually be

$$\mathcal{L}_p^*(\theta, F_\theta) = \frac{1}{n}\sum_{i=1}^n \left[\Delta_i \left\{\log f_{T,\theta}(T_i) - \log\int_0^{+\infty} F_{T,\theta}(t)dF_\theta(t) + \log f_\theta(X_i)\right\}\right.$$
$$\left. +(1-\Delta_i)I(C_i \leq \tilde{\mathcal{T}})\left\{\log f_{T,\theta}(T_i) - \log\int_0^{+\infty} F_{T,\theta}(t)dF_\theta(t) + \log(1 - F_\theta(X_i))\right\}\right].$$

However, with $h^1(t) = \frac{\partial\mathcal{H}^1(t)}{\partial t}$,

$$E(\mathcal{L}_p^*(\theta, F_\theta)) = E(\mathcal{L}_p(\theta, F_\theta, \mathcal{G}, \mathcal{H}^1)) + E(\Delta_1 \log h^1(X_1)),$$

where the last term does not depend on θ. As a consequence, assumption $8(ii)$ being true for $E(\mathcal{L}_p(\theta, F_\theta, \mathcal{G}, \mathcal{H}^1))$ is equivalent to assumption $8(ii)$ being true for $E(\mathcal{L}_p^*(\theta, F_\theta))$.

Assumption 9 is impossible to check theoretically. However, it is possible to bring a plausibility argument using Theorem 4 in Asgharian (2016): the set of zeros of the Hessian of the log-likelihood functions used without any estimator is a nowhere dense set of Lebesgue measure zero. $\det(\Omega_{1,\theta_{0,c}}) > 0$ $\det(\Omega_{2,\theta_{0,p}}) > 0$.

$\int_{t\leq y}\{1-\mathcal{G}([y-t]^-)\}\,dF_T(t)$

Finally, we observe that, in practice, due to the unknown forms of $\mathcal{H}(\cdot)$ and $\mathcal{G}(\cdot)$, it is not possible to theoretically check several assumptions above. However, this problem is classical when some quantities are estimated in a nonparametric way. In the present case, assumptions 8 and 9 cannot be checked theoretically under simple conditions; empirical/numerical check should be achieved. If there are several possible maxima in assumptions $8(i)$ or $8(ii)$, the domain of Θ should be restricted so that the resulting part only contains one maximum.

We can next display the first theorem that mainly provides tools and arguments to develop further statistical inference for $\theta_{0,c}$, $\theta_{0,p}$ and the resulting truncation distributions.

THEOREM 1. *a. Under assumptions* $1-5$, 6*(i)-(iv),(vi),(vii),* 7, 8*(i),* 9*(i) and* 10, $n^{1/2}(\widehat{\theta}-\theta_{0,c})$ *converges in law to a zero mean normal random vector with covariance matrix*

$$\Sigma_1 = \Omega^{-1}_{1,\theta_{0,c}} E(\eta_{1,\theta_{0,c}}(T,X,\Delta)\eta^T_{1,\theta_{0,c}}(T,X,\Delta) \mid X \geq T)\Omega^{-1}_{1,\theta_{0,c}},$$

where $\eta^T_{1,\theta_{0,c}}(T,X,\Delta)$ *denotes the transpose of the vector* $\eta_{1,\theta_{0,c}}(T,X,\Delta)$, $\eta_{1,\theta_{0,c}}(t,x,\delta) = \partial\mathcal{L}_{c,1}(\theta_{0,c},\mathcal{G})/\partial\theta + \eta^0_{1,\theta_{0,c}}(t,x,\delta)$ *for a function* $\eta^0_{1,\theta_{0,c}}(\mathcal{G}_n-\mathcal{G})$ *that satisfies the development* $E(\partial\mathcal{L}_{c,1}(\theta_{0,c},\mathcal{G}_n(\cdot))/\partial\theta - \partial\mathcal{L}_{c,1}(\theta_{0,c},\mathcal{G}(\cdot))/\partial\theta \mid \mathcal{G}_n) = \eta^0_{1,\theta_{0,c}}(\mathcal{G}_n-\mathcal{G}) + o_P(n^{-1/2})$; $E(\cdot \mid \mathcal{G}_n)$ *is the mean given the data used to construct* $\mathcal{G}_n(\cdot)$ *but not the argument of* $\mathcal{G}_n(\cdot)$, *and* $\eta^0_{1,\theta_{0,c}}(t,x,\delta)$ *is* $\eta^0_{1,\theta_{0,c}}(\mathcal{G}_n-\mathcal{G})$ *with* $\mathcal{G}_n-\mathcal{G}$ *replaced by the main term of its asymptotic representation.*

b. Under assumptions $1-5$, 6*(ii)-(vi),* 7, 8*(ii) and* 9*(ii),* $n^{1/2}(\tilde{\theta}-\theta_{0,p})$ *converges in law to a zero mean normal random vector with covariance matrix*

$$\Sigma_2 = \Omega^{-1}_{2,\theta_{0,p}} E(\eta_{2,\theta_{0,p}}(T,X,\Delta)\eta^T_{2,\theta_{0,p}}(T,X,\Delta) \mid X \geq T)\Omega^{-1}_{2,\theta_{0,p}},$$

where $\eta_{2,\theta_{0,p}}(t,x,\delta) = \partial\mathcal{L}_p(\theta_{0,p},F_{\theta_{0,p}},\mathcal{G},\mathcal{H}^1)/\partial\theta + \eta^0_{2,\theta_{0,p}}(t,x,\delta)$ *for a function* $\eta^0_{2,\theta_{0,p}}(\mathcal{G}_n-\mathcal{G},\mathcal{H}^1_n-\mathcal{H}^1)$ *that satisfies*

$$E\left(\frac{\partial\mathcal{L}_p(\theta_{0,p},\widehat{F}_{\theta_{0,p}}(\cdot),\mathcal{G}_n(\cdot),\mathcal{H}^1_n(\cdot))}{\partial\theta} - \frac{\partial\mathcal{L}_p(\theta_{0,p},F_{\theta_{0,p}}(\cdot),\mathcal{G}(\cdot),\mathcal{H}^1(\cdot))}{\partial\theta} \mid \mathcal{H}^1_n,\mathcal{G}_n\right)$$
$$= \eta^0_{2,\theta_{0,p}}(\mathcal{G}_n-\mathcal{G},\mathcal{H}^1_n-\mathcal{H}^1) + o_P(n^{-1/2});$$

$E(\cdot \mid \mathcal{H}^1_n,\mathcal{G}_n)$ *is the mean given the data used to construct* $\mathcal{H}^1_n(\cdot)$ *and* $\mathcal{G}_n(\cdot)$ *but not their argument, and* $\eta^0_{2,\theta_{0,p}}(t,x,\delta)$ *is* $\eta^0_{2,\theta_{0,p}}(\mathcal{G}_n-\mathcal{G},\mathcal{H}^1_n-\mathcal{H}^1)$ *with* $\mathcal{G}_n-\mathcal{G}$ *and* $\mathcal{H}^1_n-\mathcal{H}^1$ *replaced by the main term of their asymptotic representation.*

For the next result, we need the following functions

$$\xi_{2,\theta}(y,z,\delta,t) = \int_0^t w^{-2}_{\theta,\mathcal{G}}(x)\int_0^x \xi(z,\delta,(x-s)\wedge\mathcal{T})dF_{T,\theta}(s)d\mathcal{H}^1(x) + I(y\leq t)\delta w^{-1}_{\theta,\mathcal{G}}(y)$$
$$-\int_0^t w^{-1}_{\theta,\mathcal{G}}(x)d\mathcal{H}^1(x)$$

and

$$\xi_{3,\theta}(y,z,\delta,t) = \left(\int_0^{+\infty} w_{\theta,\mathcal{G}}^{-1}(x)d\mathcal{H}^1(x)\right)^{-1} \xi_{2,\theta}(y,z,\delta,t)$$
$$- \frac{\int_0^t w_{\theta,\mathcal{G}}^{-1}(x)d\mathcal{H}^1(x)}{\left(\int_0^{+\infty} w_{\theta,\mathcal{G}}^{-1}(x)d\mathcal{H}^1(x)\right)^2}\xi_{2,\theta}(y,z,\delta,+\infty).$$

The above results focus on inference about the truncation distribution and its parameters. Since our aim is also to make inference about the lifetime distribution, the next theorem states main useful properties to achieve this goal.

THEOREM 2. *Under assumptions* $1-5$, 6*(ii),(iii),(iv),(vi) and* 7,

1. *If in addition assumptions* 6*(i),(vii),* 8*(i),* 9*(i) and* 10 *are met,*

$$\widehat{F}_{\widehat{\theta}}(t) - F_{\theta_{0,c}}(t) = \frac{1}{n}\sum_{i=1}^n \nu_1(T_i,X_i,\Delta_i,t) + R_n(t),$$

where $\sup\{|R_n(t)| : -\infty < t < +\infty\} = o_P(n^{-1/2})$ *and*

$$\nu_1(T_i,X_i,\Delta_i,t) = \frac{\partial F_{\theta_{0,c}}(t)}{\partial\theta^T}\Omega_{1,\theta_{0,c}}^{-1}\eta_{1,\theta_{0,c}}(T_i,X_i,\Delta_i) + \xi_{3,\theta_{0,c}}(X_i,X_i-T_i,\Delta_i,t).$$

Under the same assumptions, the process $Z_{1n}(t) = n^{1/2}(\widehat{F}_{\widehat{\theta}}(t) - F_{\theta_{0,c}}(t))$ *converges weakly to a zero mean Gaussian process* $Z_1(t)$ *with covariance function*

$$cov(Z_1(t),Z_1(t')) = E(\nu_1(T,X,\Delta,t)\nu_1(T,X,\Delta,t') \mid X \geq T).$$

2. *Alternatively, if in addition assumptions* 6*(v),* 8*(ii) and* 9*(ii) are met, then*

$$\widehat{F}_{\tilde{\theta}}(t) - F_{\theta_{0,p}}(t) = \frac{1}{n}\sum_{i=1}^n \nu_2(T_i,X_i,\Delta_i,t) + R_n(t),$$

where $\sup\{|R_n(t)| : -\infty < t < +\infty\} = o_P(n^{-1/2})$ *and*

$$\nu_2(T_i,X_i,\Delta_i,t) = \frac{\partial F_{\theta_{0,p}}(t)}{\partial\theta^T}\Omega_{2,\theta_{0,p}}^{-1}\eta_{2,\theta_{0,p}}(T_i,X_i,\Delta_i) + \xi_{3,\theta_{0,p}}(X_i,X_i-T_i,\Delta_i,t).$$

Under the same assumptions, the process $Z_{2n}(t) = n^{1/2}(\tilde{F}_{\tilde{\theta}}(t) - F_{\theta_{0,p}}(t))$ *converges weakly to a zero mean Gaussian process* $Z_2(t)$ *with covariance function*

$$cov(Z_2(t),Z_2(t')) = E(\nu_2(T,X,\Delta,t)\nu_2(T,X,\Delta,t') \mid X \geq T).$$

These theorems show that although we use data in which T_i is always smaller or equal to X_i, $i = 1,\ldots,n$, consistency rates are the same as if data were generated from the true truncation and lifetime distributions. Beside the above formulas, the explicit expressions of the $\eta_{1,\theta_{0,c}}(t,x,\delta)$ and $\eta_{2,\theta_{0,p}}(t,x,\delta)$ are provided in the Appendix while the proofs of the two above theorems are postponed to the Supplementary Material.

In practice, we can compute standard errors and confidence intervals by using the generalized obvious bootstrap procedure of Wang (1991) with our estimators for the truncation and lifetime distributions. We can achieve this as follows.

1. for $i = 1,\ldots,n$,
   Step 1. Generation of $Y_{i,b}^*$ from $\widehat{F}_\theta(\cdot)$ with θ equal to $\widehat{\theta}$ or $\tilde{\theta}$

Step 2. Generation of $T^*_{i,b,\theta}$ from the distribution $F_{T,\theta}(\cdot)$ with θ equal to $\widehat{\theta}$ or $\tilde{\theta}$. If the value of $T^*_{i,b,\theta}$ is larger than the value of $Y^*_{i,b}$, then the pair $(Y^*_{i,b}, T^*_{i,b,\theta})$ is rejected and the algorithm goes back to Step 1. Otherwise, the pair is kept and the algorithm carries on.

Step 3. Generation of $V^*_{i,b}$ from the Kaplan–Meier estimator $\mathcal{G}_n(\cdot)$.

Definition of $X^*_{i,b} = \min(Y^*_{i,b}, T^*_{i,b,\theta} + V^*_{i,b})$ and of $\Delta^*_{i,b} = I(Y^*_{i,b} \leq T^*_{i,b,\theta} + V^*_{i,b})$.

2. Computation of parameters estimator $\widehat{\theta}^*_b$, respectively $\tilde{\theta}^*_b$, obtained with the bootstrap sample

$$\{(T^*_{i,b,\theta}, X^*_{i,b}, \Delta^*_{i,b}) : i = 1, ..., n\}$$

and the conditional, respectively full, log-likelihood method, where the vector of parameters θ is $\widehat{\theta}$, respectively $\tilde{\theta}$.

4. SIMULATIONS

In the following settings, the variables T and Y have Weibull densities $f(t) = \alpha\lambda(\lambda t)^{\alpha-1}\exp(-(\lambda t)^{\alpha})$ with respectively $(\lambda;\alpha) = (\lambda_T;\alpha_T)$ and $(\lambda;\alpha) = (0.75; 1.25)$. The variable $C - T$ follows the same distribution as $5 \times W$, where $W \sim$ Beta$(0{\cdot}75;\gamma)$ for which the density is $f(t) = \Gamma(0{\cdot}75+\gamma)/(\Gamma(0{\cdot}75)\Gamma(\gamma))t^{-0{\cdot}25}(1-t)^{\gamma-1}I(t \in [0;1])$ with $\gamma = 1, 3, 5,$ 7 or 9. Each simulation result in Tables 1 and 2 below is obtained for samples of size $n = 50, 100$ or 200. For the purpose of the analysis, theoretical truncation and censoring percentages, respectively pr$(T > Y)$ and pr$(C - T < Y - T \mid X \geq T)$ and denoted Trunc. % and Cens. %, are indicated while estimated integrated mean squared errors of the type $\frac{1}{R}\sum_{k=1}^{R}\int(\widehat{F}(y) - F(y))^2 dy$ along $R = 1000$ Monte Carlo trials are computed for $\widehat{F}(\cdot) = F_{T,\widehat{\theta}}(\cdot)$, $F_{T,\tilde{\theta}}(\cdot)$ corresponding to $F(\cdot) = F_{T,\theta_0}(\cdot)$ and $\theta_0 = (\lambda_T;\alpha_T)$, $F_{\widehat{\theta}}(\cdot)$, $F_{\tilde{\theta}}(\cdot)$ and $F_n(\cdot)$ corresponding to $F(\cdot) = F_{\theta_0}(\cdot)$.

Table 1 presents the results for the estimators of F_{T,θ_0}. In general, the integrated mean squared error decreases when n increases and the censoring percentage decreases. Whereas the relation with n is intuitively obvious, an increase of censoring has a positive impact on the quality of $\mathcal{G}_n(\cdot)$. In other simulations not reported here, censoring does not lead to such a clear effect, especially for the conditional log-likelihood case which does not depend on $\widehat{F}_{\theta}(\cdot)$. Indeed, this log-likelihood function provides a solution even though censoring percentage is 100%. Another observation is a slight decrease of the ratio of integrated mean squared error of $F_{T,\widehat{\theta}}$ and $F_{T,\tilde{\theta}}$ between $n = 50$ and $n = 200$, which suggests a possible reduction of the conditional effect of the log-likelihood with respect to the nonparametric effect in the full log-likelihood. Improvement of $F_{T,\widehat{\theta}}(\cdot)$ with respect to $F_{T,\tilde{\theta}}(\cdot)$ seems to occur for smaller truncation percentages as well: in the simulations presented here, increase of information about T improves globally more $F_{T,\widehat{\theta}}(\cdot)$ than $F_{T,\tilde{\theta}}(\cdot)$.

Both semiparametric estimators $\widehat{F}_{\widehat{\theta}}(\cdot)$ and $\widehat{F}_{\tilde{\theta}}(\cdot)$ outperform $F_n(\cdot)$ for all the cases considered in Table 2. Since the distribution $F_{\theta_0}(\cdot)$ to estimate is the same for both couples of truncation parameters, the integrated mean squared errors can be compared for different truncation percentages. As it can be expected in this case, the smaller pr$(T > Y)$, the larger the available information about Y and therefore the better the estimation is if the censoring percentage does not vary too much. Now, the main observation is that $\widehat{F}_{\tilde{\theta}}(\cdot)$ often outperforms $\widehat{F}_{\widehat{\theta}}(\cdot)$, suggesting a positive impact of the full likelihood method on the estimation of $F_{\theta_0}(\cdot)$. Finally, note that an increase of the censoring percentage does not necessarily lead to an increase of the integrated

Table 1. *Practical performance of $F_{T,\hat{\theta}}$ and $F_{T,\tilde{\theta}}$*

n	$(\lambda_T;\alpha_T)$	Trunc. %	γ	Cens. %	IMSE($F_{T,\hat{\theta}}$)	IMSE($F_{T,\tilde{\theta}}$)	RIMSE$_{T,\hat{\theta},\tilde{\theta}}$
50	$(0.5,3)$	72.13	1	26.69	1.68	1.81	0.92
			3	49.53	2.44	2.42	1.01
			5	61.08	3.04	3.36	0.91
			7	68.23	4.57	3.90	1.17
			9	73.13	5.84	4.90	1.19
	$(0.75,3)$	55.66	1	27.61	6.13	6.00	1.02
			3	50.81	6.44	6.47	0.99
			5	62.37	7.25	8.53	0.85
			7	69.46	7.46	8.95	0.83
			9	74.28	9.76	11.07	0.88
100	$(0.5,3)$	72.13	1	26.69	0.75	0.79	0.95
			3	49.53	0.93	0.98	0.95
			5	61.08	1.11	1.17	0.95
			7	68.23	1.66	1.69	0.98
			9	73.13	1.75	2.20	0.80
	$(0.75,3)$	55.66	1	27.61	2.99	2.88	1.04
			3	50.81	3.12	3.11	1.00
			5	62.37	3.15	3.52	0.89
			7	69.46	3.60	4.44	0.81
			9	74.28	4.12	4.61	0.89
200	$(0.5,3)$	72.13	1	26.69	0.39	0.41	0.94
			3	49.53	0.42	0.43	0.97
			5	61.08	0.49	0.55	0.90
			7	68.23	0.59	0.61	0.96
			9	73.13	0.77	0.85	0.91
	$(0.75,3)$	55.66	1	27.61	1.46	1.54	0.95
			3	50.81	1.52	1.67	0.91
			5	62.37	1.61	1.84	0.87
			7	69.46	1.75	2.06	0.85
			9	74.28	1.91	2.38	0.80

Distributions, $Y \sim$Weibull$(0.75;1.25)$, $T \sim$Weibull$(\lambda_T;\alpha_T)$, $C-T \sim 5\times$Beta$(0.75;\gamma)$; Trunc. $\%$ and Cens. $\%$, the truncation and censoring percentages; IMSE, the estimated integrated mean squared error $(\times 10^{-2})$; $F_{T,\hat{\theta}}$ and $F_{T,\tilde{\theta}}$, the truncation distributions based on the conditional and the full maximum likelihood estimators respectively; RIMSE$_{T,\hat{\theta},\tilde{\theta}}$, the ratio of the estimated integrated mean squared errors of $F_{T,\hat{\theta}}$ and $F_{T,\tilde{\theta}}$ based on 1000 replications. The standard errors for the integrated squared errors in columns 6 and 7 are bounded by 4.4×10^{-3} for $= 50$, 2.2×10^{-3} for $n = 100$ and 8.1×10^{-4} for $n = 200$ for a truncation percentage of 72.13%, and for a truncation percentage equal to 55.66%, they are bounded by 6.1×10^{-4} for $n = 50$, 2.5×10^{-4} for $n = 100$ and 1.3×10^{-4} for $n = 200$.

mean squared error for given $F_{T,\theta_0}(\cdot)$ and $F_{\theta_0}(\cdot)$, due to the informative character of C. This phenomenon has been observed on other simulations not reported here.

In order to obtain a global study of the proposed semiparametric estimators, the case where the truncation distribution is misspecified is next considered. We simulated 5 times 1000 samples of size $n = 100$ with $Y \sim$Weibull$(0{\cdot}75;1{\cdot}25)$, $T \sim$Weibull$(0{\cdot}5,3)$ and $C-T \sim 5\times$Beta$(0{\cdot}75;1)$. On each set of 1000 samples, we applied our methods with a fixed parameter of the truncation distribution, namely $\alpha_T^* = 1,2,2{\cdot}5,3{\cdot}5,4$. In practice, cases where a usual well-known wrong parametric family of distributions is chosen to fit the truncation distribution, for example a Weibull distribution fitted on gamma distributed truncation values, are more often encountered; these are also illustrated in the Supplementary Material. Table 3 hereunder describes the obtained results.

Table 2. *Practical performance of $\widehat{F}_{\hat{\theta}}$, $\widehat{F}_{\tilde{\theta}}$ and F_n*

γ	Cens. %	IMSE($\widehat{F}_{\hat{\theta}}$)	IMSE($\widehat{F}_{\tilde{\theta}}$)	IMSE(F_n)	RIMSE$_{Y,\hat{\theta},n}$	RIMSE$_{Y,\tilde{\theta},n}$
		$n=50$ - $(\lambda_T;\alpha_T)=(0.5;3)$ - Trunc. %$=72.13$				
1	26.69	9.9	9.8	11.8	0.84	0.83
3	49.53	11.2	10.8	12.4	0.90	0.87
5	61.08	12.1	11.6	13.5	0.90	0.86
7	68.23	12.9	12.6	14.3	0.90	0.88
9	73.13	14.7	13.4	15.5	0.95	0.86
		$n=50$ - $(\lambda_T;\alpha_T)=(0.75;3)$ - Trunc. %$=55.66$				
1	27.61	7.3	7.1	8.7	0.83	0.82
3	50.81	8.2	7.8	9.2	0.89	0.84
5	62.37	9.3	8.3	9.9	0.93	0.83
7	69.46	9.4	9.0	10.4	0.91	0.86
9	74.28	10.1	9.7	11.4	0.88	0.85
		$n=100$ - $(\lambda_T;\alpha_T)=(0.5;3)$ - Trunc. %$=72.13$				
1	26.69	7.2	7.0	9.0	0.80	0.78
3	49.53	7.6	7.2	9.2	0.82	0.78
5	61.08	7.8	7.8	9.8	0.80	0.80
7	68.23	8.7	8.4	10.2	0.84	0.82
9	73.13	9.2	8.7	10.7	0.86	0.82
		$n=100$ - $(\lambda_T;\alpha_T)=(0.75;3)$ - Trunc. %$=55.66$				
1	27.61	5.2	5.1	6.6	0.78	0.77
3	50.81	5.8	5.4	7.2	0.81	0.75
5	62.37	6.2	5.7	7.0	0.88	0.81
7	69.46	6.6	6.3	7.4	0.89	0.85
9	74.28	7.4	7.0	8.3	0.89	0.84
		$n=200$ - $(\lambda_T;\alpha_T)=(0.5;3)$ - Trunc. %$=72.13$				
1	26.69	5.0	4.9	6.8	0.74	0.72
3	49.53	5.4	5.2	7.1	0.75	0.73
5	61.08	5.6	5.6	7.4	0.76	0.75
7	68.23	5.9	5.8	7.7	0.76	0.75
9	73.13	7.2	6.0	7.7	0.94	0.78
		$n=200$ - $(\lambda_T;\alpha_T)=(0.75;3)$ - Trunc. %$=55.66$				
1	27.61	3.7	3.6	4.8	0.77	0.75
3	50.81	3.8	3.6	4.9	0.79	0.74
5	62.37	4.6	3.9	4.6	0.99	0.83
7	69.46	4.5	4.3	5.0	0.89	0.85
9	74.28	5.5	4.8	5.8	0.94	0.83

Distributions, $Y\sim$Weibull$(0.75;1.25)$, $T\sim$Weibull$(\lambda_T;\alpha_T)$, $C-T\sim 5\times$Beta$(0.75;\gamma)$; Trunc. % and Cens. %, the truncation and censoring percentages; IMSE, the estimated integrated mean squared error ($\times 10^{-2}$); $\widehat{F}_{\hat{\theta}}$ and $\widehat{F}_{\tilde{\theta}}$, the semi-parametric estimators of F_{θ_0} based on the conditional and the full maximum likelihood estimators respectively; F_n, the product-limit estimator of F_{θ_0}; RIMSE$_{Y,\widehat{\theta},n}$ (respectively RIMSE$_{Y,\tilde{\theta},n}$) the ratio of the estimated integrated mean squared errors of $\widehat{F}_{\hat{\theta}}$ (respectively $\widehat{F}_{\tilde{\theta}}$) and F_n. The standard errors for the integrated squared errors in columns 3, 4 and 5 are bounded by 6.2×10^{-3}.

These are worse and worse when the departure from the true truncation model increases. The estimators for $F(\cdot)$ seem to deteriorate slowly since they stay close to the product-limit estimator, except for the most distant case $\alpha_T^*=1$. Next, the following characteristic already slightly observed in the well-specified case seems here more important: the full likelihood method provides better results for the lifetime distribution and the conditional likelihood method does for the trun-

Table 3. *Misspecified shape α_T^* parameter for a fitted Weibull distribution ($n = 100$)*

α_T^*	IMSE($F_{T,\hat\theta}$) RIMSE$_{T,\hat\theta,\tilde\theta}$	IMSE($F_{T,\tilde\theta}$)	IMSE($\widehat{F}_{\hat\theta}$) RIMSE$_{Y,\hat\theta,n}$	IMSE($\widehat{F}_{\tilde\theta}$) RIMSE$_{Y,\tilde\theta,n}$	dist
3.5	11.2	8.3	4.8	4.8	0.19
	1.35		0.54	0.53	
2.5	13.6	15.1	8.3	8.2	0.3
	0.90		0.92	0.91	
4	12.3	12.7	9.5	8.9	0.61
	0.97		1.06	0.99	
2	84.9	102.9	10.9	10.7	1.6
	0.83		1.21	1.19	
1	380.6	420.6	21.1	19.8	13.90
	0.90		2.34	2.2	

Distributions, Y ∼Weibull$(0.75; 1.25)$, T ∼Weibull$(0.50; 3)$, $C - T \sim 5\times$Beta$(0.75; 1)$; IMSE, the estimated integrated mean squared error ($\times 10^{-3}$ for $F_{T,\hat\theta}$ and $F_{T,\tilde\theta}$, and $\times 10^{-2}$ for $\widehat{F}_{\hat\theta}$, $\widehat{F}_{\tilde\theta}$ and F_n); $F_{T,\hat\theta}$ and $F_{T,\tilde\theta}$, the truncation distributions based on the conditional and full maximum likelihood estimators; $\widehat{F}_{\hat\theta}$ and $\widehat{F}_{\tilde\theta}$, the misspecified semiparametric estimators of F based on the conditional and full maximum likelihood estimators; F_n, the product-limit estimator of F; RIMSE$_{T,\hat\theta,\tilde\theta}$, the ratio of the estimated integrated mean squared errors of $F_{T,\hat\theta}$ and $F_{T,\tilde\theta}$; RIMSE$_{Y,\widehat{\theta},n}$ (respectively RIMSE$_{Y,\tilde\theta,n}$) the ratio of the estimated integrated mean squared errors of $\widehat{F}_{\hat\theta}$ (respectively $\widehat{F}_{\tilde\theta}$) and F_n; dist, the L^2 distance between the true F_{T,θ_0} and the closest distribution constrained by α_T^* ($\times 10^{-2}$).

cation distribution. See also Table 1 in the Supplementary Material. On one side, the conditional technique searches for optimal values of parameters of a misspecified truncation distribution to make it as close as possible to the true $F_T(\cdot)$; on the other side, the full likelihood searches for optimal values of these parameters so as to fit as far as possible both a misspecified truncation distribution and misspecified nonparametrically estimated lifetime distribution.

Next, we compare our approach to the method of Shen (2007); the alternative procedure in Shen (2009) reported similar results and it will not be considered in the sequel. Table 4 shows that both our conditional and full likelihood methods outperform the Shen (2007) method in the simulated settings. Most of the time, the conditional likelihood method provides the best results for the truncation distribution while the full likelihood method does for the lifetime distribution. In particular, the Shen (2007) method seems to get worse with respect to the other methods when the censoring percentage increases. Here as well, it is worth mentioning that the conditional likelihood method delivers results up to 100% censoring; none of the other methods does it, whether it be the full likelihood method or any of the methods of Shen (2007, 2009). Other simulations not reported here provided similar results.

In order to investigate the performance of the bootstrap algorithm introduced in Section 3, simulations are carried out for $n = 50$, 100 and 200 and $B = 250$ bootstrap replications; the results are obtained for 1000 simulations again. The simulations are restricted to the case $(\lambda_T; \alpha_T) = (0{\cdot}5; 3)$. The results corresponding to symmetric two-sided confidence intervals with confidence level of 95% for the basic bootstrap and the percentile method are provided in the Supplementary Material. The full log-likelihood method seems to be the best one in most of the situations; in particular, there is only one case where the average length of its confidence interval is larger. The conditional log-likelihood method is indeed submitted to two penalizing weights:

Table 4. *Comparison with the Shen (2007) method ($n = 100$)*

γ	Cens. %	IMSE($F_{T,\hat{\theta}}$)	IMSE($\widehat{F}_{\hat{\theta}}$)
		IMSE($F_{T,\tilde{\theta}}$)	IMSE($\widehat{F}_{\tilde{\theta}}$)
			IMSE(F_n)
		IMSE(F_{T,θ^s})	IMSE($\widehat{F}_{\theta^s}$)
1	26.69	0.75	7.2
		0.79	7.0
			9.0
		0.77	8.0
3	49.53	0.93	7.6
		0.98	7.2
			9.2
		0.94	8.6
5	61.08	1.11	7.8
		1.17	7.8
			9.8
		1.43	9.4
7	68.23	1.66	8.7
		1.69	8.4
			10.2
		2.12	9.8
9	73.13	1.75	9.2
		2.20	8.7
			10.7
		2.89	10.3

$Y \sim$Weibull$(0.75; 1.25)$, $T \sim$Weibull$(0.5; 3)$, $C - T \sim 5\times$Beta$(0.75; \gamma)$; Cens. %, the censoring percentage; IMSE, estimated integrated mean squared error ($\times 10^{-2}$); $F_{T,\hat{\theta}}$, $F_{T,\tilde{\theta}}$ and F_{T,θ^s}, the truncation distributions based on the conditional, the full and the Shen (2007) maximum likelihood estimators; $\widehat{F}_{\hat{\theta}}$, $\widehat{F}_{\tilde{\theta}}$ and $\widehat{F}_{\theta^s}$, the semiparametric estimators of F_{θ_0} based on the conditional, the full and the Shen (2007) maximum likelihood estimators; F_n, the product-limit estimator of F_{θ_0}.

$w_{\theta,\mathcal{G}_n}(\cdot)$ for uncensored data points and $\int_0^{\cdot\wedge\mathcal{T}} f_{T,\theta}(\cdot - t)d\mathcal{G}_n(t)$ for censored data points. These weights can obviously be very small and deteriorate estimation. However, the full log-likelihood method can only suffer from weights $w_{\theta,\mathcal{G}_n}(\cdot)$ close to zero. When constructing bootstrap confidence intervals, this weights effect is present in both estimation and resampling steps. A solution to partially avoid this problem would be for example to truncate the distribution of T and therefore delete some large values of X: in practice, that could be achieved by simply skipping data points from a given value of T.

Remark 1. As the associate editor of the journal mentioned, giving two-sided intervals conceals the possibility that the one-sided tail errors could be 0% and 5% rather than 2.5% and 2.5%. We therefore checked the behavior of the bootstrap approximations of the distributions of the statistics ($\widehat{\theta} - \theta_{0,c}$ and $\tilde{\theta} - \theta_{0,p}$) through Q-Q plots. These can be obtained in Section 2 of the Supplementary Material.

5. Data analysis

The methods discussed in the previous sections are applied to Alzheimer's disease data. Stern et al. (1997) reported a cross-sectional prospective cohort study of 236 patients followed up for up to 7 years to investigate disease progress from onset to death. In 2006, the available data of this study have up to 16 years of follow-up after the cross section. The lifetime of interest is the time from the onset of Alzheimer's disease to death and the truncation time is the time from disease onset to study entry. We removed one patient with zero truncation time, so finally there were 235 patients and 213 observed deaths among them.

To check the goodness-of-fit of the tested truncation parametric distributions, we used the nonparametric estimator of Wang (1991). Wang (1991) developed an estimator for the distribution of T using the same weighting idea as here but inverting the roles of Y and T: this paper therefore estimates the distribution of T with weights using the distribution of Y. We conducted a hypothesis testing procedure by computing both the Kolmogorov-Smirnov and Cramer-von Mises statistics based on the difference between Wang's estimator and our $\widehat{F}_{\widehat{\theta}}(\cdot)$ or $\widehat{F}_{\tilde{\theta}}(\cdot)$. The distribution of these statistics is simply obtained by using the bootstrap samples constructed with the method proposed in Section 3. At the 5% level, none of the proposed models is rejected: for the Kolmogorov-Smirnov statistic and the conditional likelihood method, $p-$values are $0{\cdot}773$, $0{\cdot}327$ and $0{\cdot}356$ respectively for the gamma, the Weibull and the log-normal distributions. For the Cramer-von Mises, the $p-$values are very similar and, for both Cramer-von Mises and Kolmogorov-Smirnov procedures, they are slightly larger in the conditional likelihood case than in the full likelihood case. We also tested the uniform distribution with a support equal to the largest observed data. The results clearly reject this distribution, with $p-$values of 0 for both statistics.

Figure 1 displays the truncation distribution estimators $F_{T,n}(\cdot)$, i.e. the Wang (1991) estimator, $F_{T,\hat{\theta}}(\cdot)$ and $F_{T,\tilde{\theta}}(\cdot)$. The parameters as well as the confidence intervals for each distribution, log-normal, Weibull and gamma, are given in Table 5 for each method. On Fig. 1, we also represent confidence bands for the truncation distributions. These are also obtained by our bootstrap procedure; we add and subtract to $F_{T,\widehat{\theta}}(\cdot)$ (respectively $F_{T,\tilde{\theta}}(\cdot)$) the 95% percentile of the empirical distribution of $\sup_x |F_{T,\widehat{\theta}^*_b}(x) - F_{T,\widehat{\theta}}(x)|$ (respectively $\sup_x |F_{T,\tilde{\theta}^*_b}(x) - F_{T,\tilde{\theta}}(x)|$; $b = 1,\ldots,B$). The full likelihood method clearly exhibits more variability than the conditional one.

The values of the estimators $\widehat{F}_{\tilde{\theta}}(\cdot)$ and $\widehat{F}_{\widehat{\theta}}(\cdot)$ for the log-normal, Weibull and gamma truncation distributions as well as the product-limit estimator $F_n(\cdot)$ are represented in Fig. 2. The estimators seem to roughly report the same behavior for the cumulative lifetime distribution, even though the fitted truncation models are different from each other. However, the curves obtained with the full likelihood method seem closer to the completely nonparametric estimator than with the conditional likelihood method. In addition, the variability obtained with the former is smaller than with the latter.

Supplementary material

Supplementary material available at *Biometrika* online includes additional analyses of misspecified distributions and bootstrap confidence intervals together with the proofs of the theorems described in Section 3 .

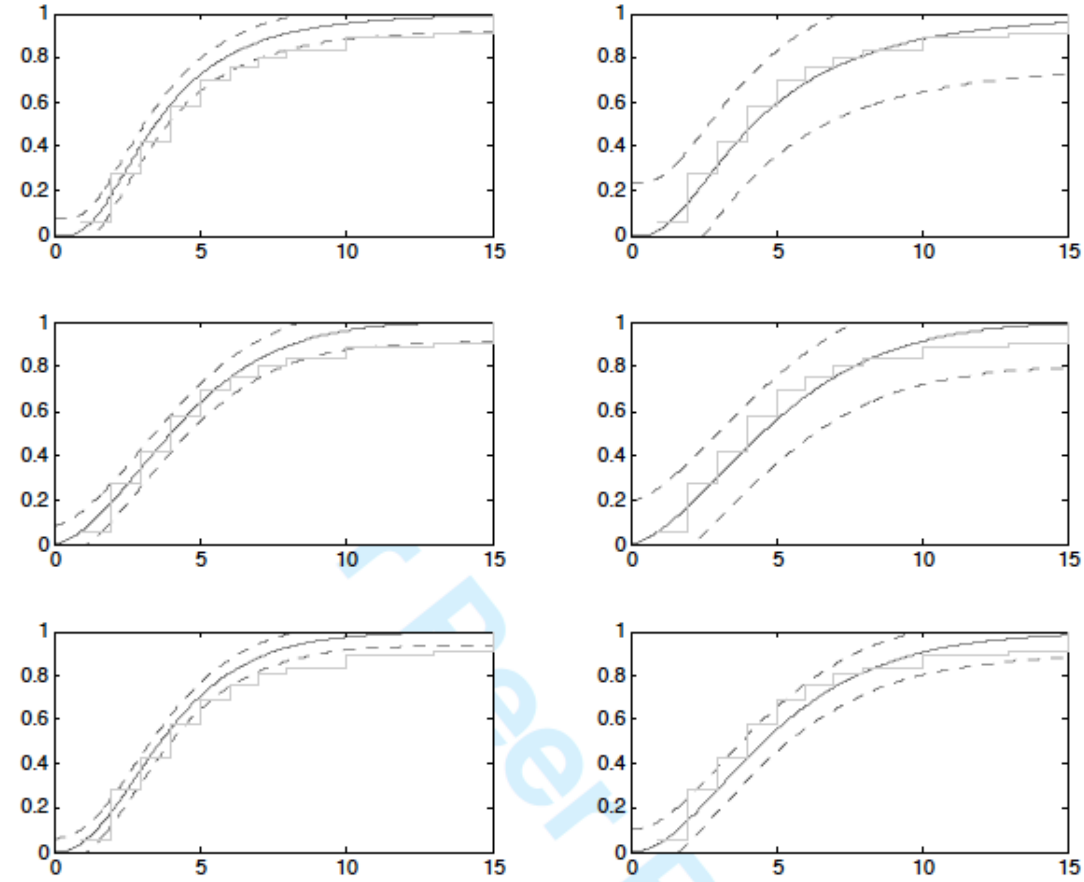

Fig. 1. Alzheimer's disease data: representation of estimated truncation time (in years) distributions: the left (respectively right) column provides estimations obtained by the conditional (respectively full) likelihood method; rows 1, 2 and 3 correspond to fitted log-normal, Weibull and gamma distributions respectively. On each subgraph, the light-grey stairs function corresponds to the Wang (1991) truncation distribution estimator; the solid curve corresponds to the parametric truncation distribution while the dashed curves represent its confidence bands.

APPENDIX

Explicit expressions of $\eta_{1,\theta}(t,x,\delta)$ *and* $\eta_{2,\theta}(t,x,\delta)$

We provide here the explicit expressions of $\eta_{1,\theta}(t,x,\delta)$ and $\eta_{2,\theta}(t,x,\delta)$ appearing in Theorem 1. The notation $\dot{m}(\theta,x)$ is used for the column vector of derivatives with respect to each component of θ of any function m differentiable with respect to each component of θ, depending on θ and other variables. The notation $\dot{m}(\theta_0,x)$ corresponds to this same function computed at the point $\theta=\theta_0$. We have

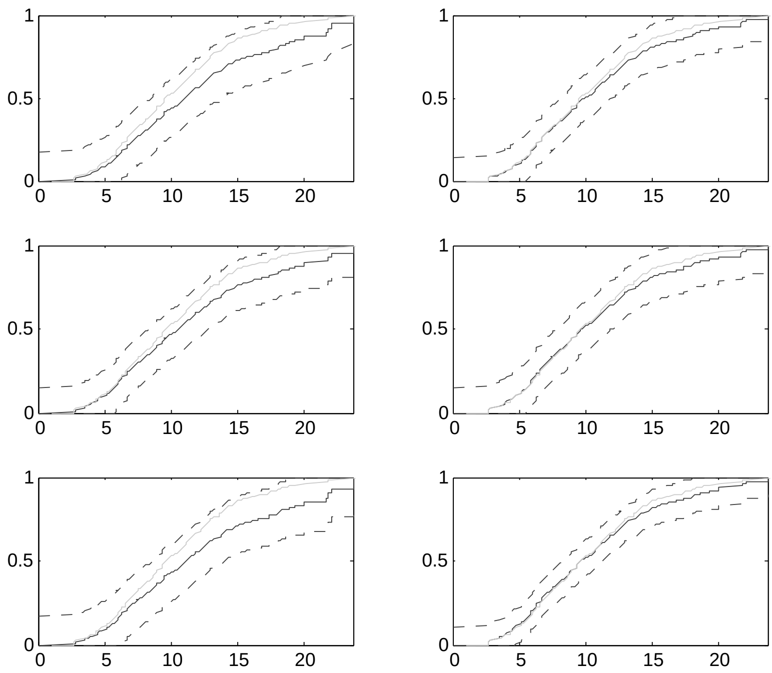

Fig. 2. Alzheimer's disease data: representation of estimated lifetime (in years) distributions: the left (respectively right) column provides estimations obtained by the conditional (respectively full) likelihood method; rows 1, 2 and 3 correspond to fitted log-normal, Weibull and gamma distributions respectively. In each subgraph, the light-grey stairs function corresponds to the completely nonparametric estimator; the solid curve corresponds to the cumulative distribution function while the dashed curves represent its confidence bands.

$$
\begin{aligned}
\eta_{1,\theta}(t,x,\delta) = \frac{\dot{f}_{T,\theta}(t)}{f_{T,\theta}(t)} + \int \Bigg(& \frac{\int_0^z \xi(x-t,\delta,(z-s)\wedge\mathcal{T})\dot{f}_{T,\theta}(s)ds}{w_{\theta,\mathcal{G}}(z)} \\
& - \frac{\dot{w}_{\theta,\mathcal{G}}(z)\int_0^z \xi(x-t,\delta,(z-s)\wedge\mathcal{T})f_{T,\theta}(s)ds}{w_{\theta,\mathcal{G}}^2(z)}\Bigg) d\mathcal{H}^1(z) \\
- \int \Bigg(& \frac{\int_0^{z\wedge\mathcal{T}} \dot{f}_{T,\theta}(z-s)d\xi(x-t,\delta,s)}{\int_0^{z\wedge\mathcal{T}} f_{T,\theta}(z-s)d\mathcal{G}(s)} \\
& - \frac{\int_0^{z\wedge\mathcal{T}} \dot{f}_{T,\theta}(z-s)d\mathcal{G}(s)\int_0^{z\wedge\mathcal{T}} f_{T,\theta}(z-s)d\xi(x-t,\delta,s)}{\left(\int_0^{z\wedge\mathcal{T}} f_{T,\theta}(z-s)d\mathcal{G}(s)\right)^2}\Bigg) d\mathcal{H}^0(z) \\
- \delta\frac{\dot{w}_{\theta,\mathcal{G}}(x)}{w_{\theta,\mathcal{G}}(x)} & - (1-\delta)\frac{\int_0^{x\wedge\mathcal{T}} \dot{f}_{T,\theta}(x-s)d\mathcal{G}(s)}{\int_0^{x\wedge\mathcal{T}} f_{T,\theta}(x-s)d\mathcal{G}(s)}
\end{aligned}
$$

Table 5. *Alzheimer's disease data : estimations of the truncation distribution parameters*

Conditional MLE	Boot. method	Confidence interval	Full MLE	Boot. method	Confidence interval
		T ∼LogNormal(μ_T,σ_T)			
$\hat{\mu}_T = 1.25$	Basic	$[1.15; 1.35]$	$\tilde{\mu}_T = 1.45$	Basic	$[0.95; 1.40]$
	Perc.	$[1.15; 1.35]$		Perc.	$[1.50; 1.95]$
$\hat{\sigma}_T = 0.625$	Basic	$[0.550; 0.700]$	$\tilde{\sigma}_T = 0.725$	Basic	$[0.475; 0.750]$
	Perc.	$[0.550; 0.700]$		Perc.	$[0.700; 0.975]$
		T ∼Weibull(λ_T,α_T)			
$\hat{\lambda}_T = 0.200$	Basic	$[0.175; 0.225]$	$\tilde{\lambda}_T = 0.175$	Basic	$[0.175; 0.225]$
	Perc.	$[0.175; 0.225]$		Perc.	$[0.125; 0.175]$
$\hat{\alpha}_T = 1.7$	Basic	$[1.45; 1.85]$	$\tilde{\alpha}_T = 1.6$	Basic	$[1.40; 1.65]$
	Perc.	$[1.55; 1.95]$		Perc.	$[1.55; 1.80]$
		T ∼Gamma(α_T,λ_T)			
$\hat{\alpha}_T = 2.90$	Basic	$[2.3; 3.4]$	$\tilde{\alpha}_T = 2.40$	Basic	$[1.6; 2.7]$
	Perc.	$[2.4; 3.5]$		Perc.	$[2.1; 3.2]$
$\hat{\lambda}_T = 0.70$	Basic	$[0.55; 0.85]$	$\tilde{\lambda}_T = 0.45$	Basic	$[0.20; 0.55]$
	Perc.	$[0.55; 0.85]$		Perc.	$[0.35; 0.70]$

The three left (respectively right) columns correspond to the conditional (respectively full) likelihood method; the first (respectively fourth) column provides the estimated parameters according to the assumed truncation distribution; Boot. method, the bootstrap method; Basic, the basic bootstrap method; Perc., the percentile bootstrap method.

and

$$\begin{aligned}
\eta_{2,\theta}(t,x,\delta) = {} & \delta\left(\frac{\dot{f}_{T,\theta}(t)}{f_{T,\theta}(t)} - \frac{\dot{w}_{\theta,\mathcal{G}}(x)}{w_{\theta,\mathcal{G}}(x)}\right) + (1-\delta)I(x\le\tilde{\mathcal{T}})\left(\frac{\dot{f}_{T,\theta}(t)}{f_{T,\theta}(t)} - \frac{\int_x^\infty \frac{\dot{w}_{\theta,\mathcal{G}}(z)}{w^2_{\theta,\mathcal{G}}(z)}d\mathcal{H}^1(z)}{\int_x^\infty w^{-1}_{\theta,\mathcal{G}}(z)d\mathcal{H}^1(z)}\right) \\
& + \int\left(\frac{\int_0^z \xi(x-t,\delta,(z-s)\wedge\mathcal{T})\dot{f}_{T,\theta}(s)ds}{w_{\theta,\mathcal{G}}(z)}\right. \\
& \qquad \left. - \frac{\dot{w}_{\theta,\mathcal{G}}(z)\int_0^z \xi(x-t,\delta,(z-s)\wedge\mathcal{T})f_{T,\theta}(s)ds}{w^2_{\theta,\mathcal{G}}(z)}\right)d\mathcal{H}^1(z) \\
& + \int_0^{\tilde{\mathcal{T}}}\left(\frac{\dot{\tilde{\xi}}_{2,\theta}(x,x-t,\delta,z)}{\int_z^\infty w^{-1}_{\theta,\mathcal{G}}(s)d\mathcal{H}^1(s)} + \frac{\int_z^\infty \frac{\dot{w}_{\theta,\mathcal{G}}(s)}{w^2_{\theta,\mathcal{G}}(s)}d\mathcal{H}^1(s)\tilde{\xi}_{2,\theta}(x,x-t,\delta,z)}{\left(\int_z^\infty w^{-1}_{\theta,\mathcal{G}}(s)d\mathcal{H}^1(s)\right)^2}\right)d\mathcal{H}^0(z) \\
& - \left(\mathrm{pr}(\Delta=1) + \int_0^{\tilde{\mathcal{T}}}d\mathcal{H}^0(z)\right)\left(\frac{\dot{\xi}^{F_{T,\theta}}_{2,\theta}(x,x-t,\delta,\infty)}{\int_0^{+\infty}F_{T,\theta}(z)w^{-1}_{\theta,\mathcal{G}}(z)d\mathcal{H}^1(z)}\right. \\
& \left. - \frac{\xi^{F_{T,\theta}}_{2,\theta}(x,x-t,\delta,\infty)\int_0^{+\infty}\left(\dot{F}_{T,\theta}(z)w^{-1}_{\theta,\mathcal{G}}(z) - F_{T,\theta}(z)\frac{\dot{w}_{\theta,\mathcal{G}}(z)}{w^2_{\theta,\mathcal{G}}(z)}\right)d\mathcal{H}^1(z)}{\left(\int_0^{+\infty}F_{T,\theta}(z)w^{-1}_{\theta,\mathcal{G}}(z)d\mathcal{H}^1(z)\right)^2}\right) \\
& - \frac{\int_0^{+\infty}\left(\dot{F}_{T,\theta}(z)w^{-1}_{\theta,\mathcal{G}}(z) - F_{T,\theta}(z)\frac{\dot{w}_{\theta,\mathcal{G}}(z)}{w^2_{\theta,\mathcal{G}}(z)}\right)d\mathcal{H}^1(z)}{\int_0^{+\infty}F_{T,\theta}(z)w^{-1}_{\theta,\mathcal{G}}(z)d\mathcal{H}^1(z)}(\delta + (1-\delta)I(x\le\tilde{\mathcal{T}})),
\end{aligned}$$

where

$$\tilde{\xi}_{2,\theta}(y,z,\delta,t) = \int_t^{+\infty} w_{\theta,\mathcal{G}}^{-2}(x) \int_0^x \xi(z,\delta,(x-s)\wedge \mathcal{T}) dF_{T,\theta}(s) d\mathcal{H}^1(x)$$
$$+ I(t \le y)\delta w_{\theta,\mathcal{G}}^{-1}(y) - \int_t^{\infty} w_{\theta,\mathcal{G}}^{-1}(x) d\mathcal{H}^1(x)$$

and

$$\xi_{2,\theta}^{f}(y,z,\delta,t) = \int_0^t f(x) w_{\theta,\mathcal{G}}^{-2}(x) \int_0^x \xi(z,\delta,(x-s)\wedge \mathcal{T}) dF_{T,\theta}(s) d\mathcal{H}^1(x)$$
$$+ I(y \le t) f(y)\delta w_{\theta,\mathcal{G}}^{-1}(y) - \int_0^t f(x) w_{\theta,\mathcal{G}}^{-1}(x) d\mathcal{H}^1(x).$$

References

Asgharian, M., M´Lan, C. E. & Wolfson, D. B. (2002). Length-biased sampling with right-censoring: an unconditional approach. *J. Am. Statist. Assoc.* **97**, 201–209.

Asgharian, M. & Wolfson, D. B. (2005). Asymptotic behavior of the unconditional NPMLE of the length-biased survivor function from right censored prevalent cohort data. *Ann. Statist.* **33**, 2109–2131.

Brunel, E., Comte, F. & Guilloux, A. (2008). Estimation strategies for censored lifetimes with a Lexis-diagram type model. *Scand. J. Stat.* **35**, 557–576.

de Uña-Álvarez, J. (2004). Nonparametric estimation under length-biased sampling and Type I censoring: a moment based approach. *Ann. Inst. Statist. Math.* **56**, 667–681.

de Uña-Álvarez, J., Arévalo-Tomé, R. & Otero-Giráldez, M. S. (2009). Nonparametric estimation of households' duration of residence from panel data. *J. Real Estate Financ. Economics* **39**, 58–73.

Gilbert, P. B., Lele, S. R. & Vardi, Y. (1999). Maximum likelihood estimation in semiparametric selection bias models with application to AIDS vaccine trials. *Biometrika* **86**, 27–43.

Gijbels, I. & Wang, J. L. (1993). Strong representation of the survival function estimator for truncated and censored data with applications. *J. Multivar. Anal.* **47**, 210–229.

Gross, S. T. & Lai, T. L. (1996). Bootstrap methods for truncated and censored data. *Stat. Sinica* **6**, 509–530.

Horváth, L. (1985). Estimation from a length-biased distribution. *Statistics and Decisions* **3**, 91–113.

Huang, C.-Y. & Qin, J. (2011). Nonparametric estimation for length-biased and right-censored data. *Biometrika* **98**, 177–186.

Hyde, J. (1980). Survival analysis with incomplete observations. In *Biostatistics Casebook,* R. G. Miller, B. Efron, B. W. Brown & L. E. Moses, eds., 31–46. New York: John Wiley and Sons.

Lo, S. H. & Singh, K. (1986). The product-limit estimator and the bootstrap: some asymptotic representations. *Probab. Theory Rel.* **71**, 455–465.

Luo, X. & Tsai, W. Y. (2009). Nonparametric estimation for right-censored length-biased data: a pseudo-partial likelihood approach. *Biometrika* **96**, 873–886.

Mandel, M. (2007). Nonparametric estimation of a distribution function under biased sampling and censoring. *IMS Lecture Notes-Monograph Series, Complex Datasets and Inverse Problems: Tomography, Networks and Beyond* **54**, 224–238.

Mandel, M. & Betensky, R. A. (2007). Testing goodness of fit of a uniform truncation model. *Biometrics* **63**, 405–412.

Shen, P.-S. (2007). A general semiparametric model for left-truncated and right-censored data. *J. Nonparametr. Stat.* **19**, 113–129.

Shen, P.-S. (2009). Semiparametric analysis of survival data with left truncation and right censoring. *Comput. Stat. Data An.* **53**, 4417–4432.

Stern, Y., Tang, M. X., Albert, M. S., Brandt, J., Jacobs, D. M., Bell, K., Marder, K., Sano, M., Devanand, D., Albert, S. M., Bylsma, F., Tsai, W. Y. (1997). Predicting time to nursing home care and death in individuals with Alzheimer disease. *J. Am. Med. Assoc.* **277**, 806–812.

Tsai, W. Y., Jewell, N. P. & Wang, M. C. (1987). A note on the product-limit estimator under right censoring and left truncation. *Biometrika* **74**, 883–886.

Wang, M. C. (1989). A semiparametric model for randomly truncated data. *J. Am. Statist. Assoc.* **84**, 742–748.

Wang, M. C. (1991). Nonparametric estimation from cross-sectional survival data. *J. Am. Statist. Assoc.* **86**, 130–143.

[*Received. Revised*]