Journal of Statistical Planning and Inference 139 (2009) 1734–1749

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Interval censored and truncated data: Rate of convergence of NPMLE of the density

Catherine Huber[a,*], Valentin Solev[b,1], Filia Vonta[c]

[a]Université René Descartes, Paris 5, 45 rue des Saints-Péres, 75006 Paris, France
[b]Steklov Institute of Mathematics at St. Petersburg, nab. Fontanki, 27 St. Petersburg 191023, Russia
[c]Department of Mathematics and Statistics, University of Cyprus, P.O. Box 20537, CY-1678 Nicosia, Cyprus

ARTICLE INFO

Available online 24 May 2008

Keywords:
Censored and truncated data
Nonparametric maximum likelihood estimation
Consistency
Bracketing entropy
Hellinger distance
Rate of convergence

ABSTRACT

We consider survival data that are both interval censored and truncated. Under appropriate assumptions on the involved distributions, the censoring, truncation and survival, we prove the consistency of the NPMLE of the density of the survival, and give the rate of convergence. Finally, we give an example where the joint law of the censoring and truncation can be explicitly computed.

1. Introduction

Survival data are often both interval censored and truncated, as observation of the process is not continuous in time and is done through a window of time which could exclude totally some individuals from the sample. For example, the time to onset of a disease in a patient, like AIDS from HIV infection or toxicity of a treatment, is not exactly known, but it is usually known to have taken place between two dates t_1 and t_2; this occurs in particular when the event of interest results in an irreversible change of state of the individual: at time t_1, the individual is in state one, while at time t_2, he is in state two. Moreover, some people can escape the sample if they are observed during a period of time not including some pair of dates t_1, t_2 having the above property. Such data, censored and truncated were at first introduced by Turnbull (1976), and further studied by several authors such as Finkelstein et al. (1993), Frydman (1994), Huber et al. (2005). NPML estimators of the cumulative hazard together with finite dimensional parameters associated to covariates are obtained in Alioum and Commenges (1996) for the Cox model and in Huber and Vonta (2004) for a generalization of the Cox model: a frailty model or transformation model, but without any consistency result. We give here conditions on the involved distributions, the censoring, truncation and survival distributions, all three of them assumed mutually independent, under which the consistency of the nonparametric maximum likelihood estimator of the density of the survival is established together with the rate of convergence.

The proofs use results of Stein (1993), Wong and Shen (1995) and van der Vaart and Wellner (1998, 2000) on nonparametric estimation.

In Section 2, we give a representation of the censoring and truncation mechanisms. As it is due to a noncontinuous observation of the survival process, the censoring mechanism is represented as a denumerable partition of the total interval of observation time (a, b). Then a truncation is added to the censoring, conditioning the observations both of the survival and the censoring

* Corresponding author.
E-mail addresses: catherine.huber@univ-paris5.fr (C. Huber), solev@pdmi.ras.ru (V. Solev), vonta@ucy.ac.cy (F. Vonta).

[1] The research of the second author was supported by Grants RFBR 02-01-00262, RFBR-DFG 04-01-04000.

0378-3758/$ - see front matter © 2008 Published by Elsevier B.V.
doi:10.1016/j.jspi.2008.05.028

processes. Without restriction of the generality, we consider the special case of right truncation. In the next three sections, third, fourth and fifth, three distributions are successively studied, each one being conditional on fixed values which become random in the next section, leading finally to the joint law of censoring and truncation.

More specifically, in Section 3, the distribution associated with a random covering, which is a censoring set conditional on a fixed value x of the survival process, is considered. It is the sum of a denumerable number of elementary probabilities, and is proved to have a density with respect to a baseline probability.

In Section 4, we define the joint distribution of a pair of intervals, a censoring $(L(x), R(x))$ and a truncating one $(L(z), R(z))$, conditional on fixed values x and z, respectively, of the survival X and the right truncation Z.

In Section 5, we consider the distribution of the incomplete observation of X, that is, $(L(X), R(X), L(z), R(z))$, conditional on the truncating variable $Z = z$.

In Section 6, we give assumptions on the set $\mathscr{F}$ of densities f of X which allow us to derive consistency of the NPMLE estimator of f along with the convergence rate.

In the last section we provide an example where the joint law of the censoring and truncation can be explicitly computed, and which satisfies the conditions to get consistency and convergence rate of the NPMLE of the density f of the survival.

2. Partitioning the total observation time

2.1. Random covering

Let τ be a random partition defined on $(a; b)$, where usually a will be equal to 0 and b is either $+\infty$ or a finite strictly positive number:

$$\tau = \left\{ Y_0 = a < Y_1 < \cdots < Y_K < Y_{K+1} = b, \bigcup_{j=0}^{K} (Y_j, Y_{j+1}] = (a, b] \right\}, \tag{1}$$

where K is a random number in $\{2, \dots, K_0\}$ for some given K_0 such that $2 < K_0 < \infty$.

For each $x \in (a; b)$ we define

$$k = k(x) = \inf\{j : \ x \leqslant Y_{j+1}\}, \tag{2}$$

$$\vartheta(x) = (Y_{k(x)}, Y_{k(x)+1}] := (L(x), R(x)], \quad x \in (a, b). \tag{3}$$

where $L(x)$ and $R(x)$ may be thought of as the left and right values in partition τ that “bracket” (the survival) $X = x$.

Then it is clear that

$$\vartheta(x) = \vartheta(y) \quad \text{or} \quad \vartheta(x) \cap \vartheta(y) = \emptyset \tag{4}$$

and we call $\vartheta(x)$ a simple random covering of (a, b).

2.2. Short-cut covering

Let $\vartheta(x) = (L(x), R(x)]$, $x \in \mathbb{R}$, be a simple random covering and τ the partition associated with $\vartheta(x)$. Then, we consider a fixed interval $\triangle = (z_1, z_2]$, and z the associated vector (z_1, z_2), $z_1 \leqslant z_2$, which will play the role of truncating interval.

For a fixed value of $\tau = t$,

$$t = \{y_j, j = 1, 2, \dots, s\},$$

and for each $\triangle = (z_1, z_2]$ such that $z_2 < y_s$, define functions

$$\kappa_1 = \kappa_1(t, z_1) = \inf\{k : y_k \geqslant z_1\}, \quad \mathfrak{z}_1 = \mathfrak{z}_1(t, z_1) = Y_{\kappa_1} := R(z_1),$$

$$\kappa_2 = \kappa_2(t, z_2) = \sup\{k : y_k \leqslant z_2\}, \quad \mathfrak{z}_2 = \mathfrak{z}_2(t, z_2) = Y_{\kappa_2} := L(z_2).$$

The short-cut covering $\vartheta_\triangle(x) = (L_\triangle(x), R_\triangle(x)]$, $x \in \triangle$, is defined as follows:

$$\vartheta_\triangle(x) = (z_1, z_2] \quad \text{if } R(z_1) > L(z_2),$$

else

$$(L_\triangle(x), R_\triangle(x)] = \begin{cases} (L(x), R(x)] & \text{if } x \in (R(z_1), L(z_2)], \\ (z_1, R(z_1)] & \text{if } x \in (z_1, R(z_1)], \\ (L(z_2), z_2] & \text{if } x \in (L(z_2), z_2]. \end{cases}$$

In the special case of right truncation

$$\triangle = (-\infty, z]$$

and we will use the following notations for the corresponding short-cut covering $\vartheta_\triangle(x)$, $x \in \triangle$, and related objects

$$\vartheta_z(x) = \vartheta_\triangle(x),$$
$$\kappa_z = \kappa(t,z) = \sup\{k : y_k \leqslant z\}, \quad \mathfrak{z}_z = \mathfrak{z}(t,z) = Y_{\kappa_z} := L(z),$$
$$L_z(x) = L_\triangle(x), \quad R_z(x) = R_\triangle(x). \tag{5}$$

The short-cut covering $\vartheta_z(x) = (L_z(x), R_z(x)]$, $x \in \triangle$, is defined as follows:

$$\vartheta_z(x) = (L_z(x), R_z(x)] = \begin{cases} (L(x), R(x)] & \text{if } x \in (a, L(z)], \\ (L(z), z] & \text{if } x \in (L(z), z]. \end{cases}$$

2.3. The mechanism of censoring and truncation

The mechanism of censoring and truncating of a random variable X is defined as follows. Let X be a random variable, $\varDelta = (Z_1, Z_2]$ be a random interval, $\vartheta(x) = (L(x), R(x)]$, $x \in \mathbb{R}$, be a random covering, generated by a partition τ defined in (1).

We suppose that the random covering $\vartheta(\cdot)$, the random variable X and the random interval $\varDelta$ are independent, but we do not have complete observations. More precisely, we suppose that the random vector (X, Z_1, Z_2) is partly observable only in the case where $(L(X), R(X)] \subset \varDelta$:

$$Z_1 \leqslant L(X) < R(X) \leqslant Z_2.$$

In that case the available observations are the interval $(L(X), R(X)]$ of the covering $\vartheta(\cdot)$, which contains X, and the random interval $\varDelta^* = (R(Z_1), L(Z_2)]$. When $(L(X), R(X)] \not\subset \varDelta$ we do not have any observation.

Let us define

(1) Conditionally on a fixed value t of τ the random interval $\varDelta$ is taken from the truncated distribution

$$\mathscr{P}_t\{A\} = P\{\varDelta \in A | \text{ the interval } [Z_1, Z_2] \text{ contains at least two points of } t_t\}.$$

In other words, conditionally on fixed values of $\tau = t$ the random vector $Z = (Z_1, Z_2)$ is taken from the truncated distribution

$$P_t\{B\} = P\{Z \in B | \mathfrak{z}_1(t, Z_1) < \mathfrak{z}_2(t, Z_2)\}.$$

(2) Conditionally on a fixed value of $\tau = t$ and $\varDelta = \triangle = (z_1, z_2]$, the random variable X is taken from the truncated distribution

$$P_{t,\triangle}\{C\} = P\{X \in C | X \in (R(z_1), L(z_2)]\}.$$

In other words conditionally on fixed values of $\tau = t$ and $Z_1 = z_1, Z_2 = z_2$ the random variable X is taken from the truncated distribution

$$P\{C | t, z_1, z_2\} = P\{X \in C | X \in (\mathfrak{z}_1(t, z_1), \mathfrak{z}_2(t, z_2)]\}. \tag{6}$$

We consider now the simple case where for a random variable Z the random interval $\varDelta = (-\infty, Z]$, and use the notations that were given in (5). We denote by $\mathfrak{Z}$ the random variable

$$\mathfrak{Z} = \mathfrak{z}(\tau, Z).$$

Recall that the random covering $\vartheta(\cdot)$, the survival X and the truncating random variable Z are independent.

In this case, definition (6) above becomes: Conditionally on fixed values of $\tau = t$ and $Z = z$ the random variable X is taken from the truncated distribution

$$P\{C | t, z\} = P\{X \in C | X \leqslant \mathfrak{z}(t, z)\}.$$

3. The distribution associated with the random covering

Let $\vartheta(x)=(L(x),R(x)]$, $x\in\mathbb{R}$, be a simple random covering. The distribution P_x of random vector $\upsilon(x)=(L(x),R(x))$ will be called the distribution, associated with the random covering $\vartheta(x)$.

We assume that for all x the distribution P_x has density with respect to Lebesgue measure λ^2 on the plane $\mathbb{R}^2$,

$$r_x(u,v)=\frac{\mathrm{d}P_x}{\mathrm{d}\lambda^2},$$

and plan to prove in this case, that there exists a nonnegative function $r(u,v)$ such that for all x

$$r_x(u,v)=r(u,v)\mathbb{1}_{(u,v]}(x)\quad\text{(a.s.)}$$

The function $r(u,v)$ will be called the *baseline density of the simple random covering* $\vartheta(x)$. It is clear that the function $r(u,v)$ is the density of a σ-finite measure, but, for all x, the function $r(u,v)\mathbb{1}_{(u,v]}(x)$ is the density of a probability measure.

It is clear that for all x

$$r_x(u,v)=r_x(u,v)\mathbb{1}_{(u,v]}(x).$$

For positive $x<y$ and nonnegative measurable function $\psi(u,v)$ such that

$$\psi(u,v)=0\quad\text{if } u<x\leqslant v<y \text{ or } x\leqslant u<y\leqslant v. \tag{7}$$

Condition (7) is equivalent to the condition (on function ψ)

$$\psi(u,v)\mathbb{1}_{(u,v]}(x)=\psi(u,v)\mathbb{1}_{(u,v]}(y).$$

Therefore

$$\begin{aligned}\mathbf{E}\psi(L(x),R(x))&=\mathbf{E}\psi(L(x),R(x))\sum_k\mathbb{1}_{(Y_k,Y_{k+1}]}(x)\\&=\sum_k\mathbf{E}\psi(Y_j,Y_{j+1})\mathbb{1}_{(Y_j,Y_{j+1}]}(x)=\sum_k\mathbf{E}\psi(Y_j,Y_{j+1})\mathbb{1}_{(Y_j,Y_{j+1}]}(y)\\&=\mathbf{E}\psi(L(x),R(x))\sum_k\mathbb{1}_{(Y_k,Y_{k+1}]}(y)=\mathbf{E}\psi(L(y),R(y)).\end{aligned}$$

Thus, under condition (7) on function ψ

$$\iint_{u<v}\psi(u,v)r_x(u,v)\,\mathrm{d}u\,\mathrm{d}v=\iint_{u<v}\psi(u,v)r_y(u,v)\,\mathrm{d}u\,\mathrm{d}v,$$

and we obtain for all $u<x\leqslant y\leqslant v$

$$r_x(u,v)=r_y(u,v). \tag{8}$$

From (8) we conclude that there exists a nonnegative function $r(u,v)$, whose support is the set $\{(u,v):\ u<v\}$, and such that for x

$$r_x(u,v)=r(u,v)\mathbb{1}_{(u,v]}(x)\quad\text{(a.s.)}.$$

It is easy to see that the baseline density $r(u,v)$ depends only on the joint distributions of vectors (Y_j,Y_{j+1}).

Lemma 1. *The measure P_x is absolutely continuous with respect to the Lebesgue measure for all x if and only if*

(i) *for all j the distribution of the vector (Y_j,Y_{j+1}) has density $r^j(u,v)$ with respect to the Lebesgue measure,*
(ii) *the series $\sum_j r^j(u,v)$ converges a.s. to a function $r(u,v)$,*
(iii) *the function $r(u,v)$ satisfies the following condition: for all x*

$$r_x(u,v)=r(u,v)\mathbb{1}_{(u,v]}(x).$$

Proof. Suppose that for all x the distribution P_x has density $r_x(u,v)$. Let $\psi(u,v)$ be a nonnegative function, then for all j

$$\begin{aligned}\mathbf{E}\psi(Y_j,Y_{j+1})\mathbb{1}_{(Y_j,Y_{j+1}]}(x)&\leqslant\mathbf{E}\psi(L(x),R(x))\\&=\iint\psi(u,v)r(u,v)\mathbb{1}_{(u,v]}(x)\,\mathrm{d}u\,\mathrm{d}v.\end{aligned}$$

Therefore, for all x the distribution of the vector $(Y_j, Y_{j+1})\mathbb{1}_{(Y_j,Y_{j+1}]}(x)$ has a density. Hence, the distribution of the vector (Y_j, Y_{j+1}) also has a density $r^j(u, v)$.

We have

$$\begin{aligned}\mathbf{E}\psi(L(x), R(x)) &= \sum_j \mathbf{E}\psi(Y_j, Y_{j+1})\mathbb{1}_{(Y_j,Y_{j+1}]}(x)\\ &= \sum_j \iint \psi(u, v)r^j(u, v)\mathbb{1}_{(u,v]}(x)\,\mathrm{d}u\,\mathrm{d}v\\ &= \iint \psi(u, v)\left\{\sum_j r^j(u, v)\mathbb{1}_{(u,v]}(x)\right\}\mathrm{d}u\,\mathrm{d}v.\end{aligned}$$

So, we obtain

$$r(u, v) = \sum_j r^j(u, v) \quad \text{(a.s.)}.$$

Now suppose that (i), (ii) are fulfilled. Then we obtain for a nonnegative measurable function $\psi(u, v)$ (by the same way as above)

$$\mathbf{E}\psi(L(x), R(x)) = \iint \psi(u, v)\left\{\sum_j r^j(u, v)\right\}\mathbb{1}_{(u,v]}(x)\,\mathrm{d}u\,\mathrm{d}v.$$

From this equality we conclude that the series

$$r(u, v) = \sum_j r^j(u, v) < \infty \quad \text{(a.s.)}$$

and

$$r_x(u, v) = r(u, v)\mathbb{1}_{(u,v]}(x). \qquad \square$$

4. The distribution of the random vector $(L(x), R(x), L(z), R(z))$

From now on we concentrate on the case of right truncation. Due to censoring by the partition t, z is not observed. Instead we have $]L(z); R(z)] \ni z$, and only $L(z)$ is observed. Now for $x < z$ we denote by $P_{x,z}$ the distribution of the random vector $(L(x), R(x), L(z), R(z))$.

Denote by λ^n the Lebesgue measure on $\mathbb{R}^n$. The distribution $P_{x,z}$ is not absolutely continuous with respect to the suppress on λ^4. Denote by v the measure, which is defined for continuous nonnegative functions $\psi(s) = \psi(s_1, s_2, s_3, s_4)$ by the relation

$$\begin{aligned}\iiiint \psi(s)\,\mathrm{d}v = &\iint \psi(s_1, s_2, s_1, s_2)\,\mathrm{d}s_1\,\mathrm{d}s_2\\ &+ \iiint \psi(s_1, s_2, s_2, s_4)\,\mathrm{d}s_1\,\mathrm{d}s_2\,\mathrm{d}s_4 + \iiiint \psi(s_1, s_2, s_3, s_4)\,\mathrm{d}s_1\,\mathrm{d}s_2\,\mathrm{d}s_3\,\mathrm{d}s_4.\end{aligned}$$

We suppose that the distribution $P_{x,z}$ is absolutely continuous with respect to the measure v and denote its density by $q_{x,z}(s)$:

$$q_{x,z}(s) = q_{x,z}(s_1, s_2, s_3, s_4) = \frac{\mathrm{d}P_{x,z}}{\mathrm{d}v}.$$

We suppose that for all $n, m > 0$ the random vector $(Y_{-m}, \ldots, Y_n)$ has a density with respect to the corresponding Lebesgue measure. For $i + 1 < j$, let function

$r_{i,j}(y_1, y_2, y_3, y_4)$ be the density of random vector $(Y_i, Y_{i+1}, Y_j, Y_{j+1})$,

$r_j(y_1, y_2, y_3)$ be the density of the random vector (Y_{j-1}, Y_j, Y_{j+1}),

and

$r^j(y_1, y_2)$ be the density of the random vector (Y_j, Y_{j+1}).

We assume that

$$\mathfrak{d}_4(y_1, y_2, y_3, y_4) = \sum_{\substack{i,j:\\ i+1<j}} r_{ij}(y_1, y_2, y_3, y_4) < \infty \quad (\lambda^4\text{-a.s.}),$$

$$\mathfrak{d}_3(s_1,s_2,s_3)=\sum_j r_j(s)<\infty \quad (\lambda^3\text{-a.s.}),$$

and

$$\mathfrak{d}_2(y_1,y_2)=\sum_j r^j(y_1,y_2)<\infty \quad (\lambda^2\text{-a.s.}).$$

For a nonnegative function $\psi(x), x=(x_1,x_2,x_3,x_4)$ and $x<z$ we have

$$\begin{aligned}\mathbf{E}\psi(L(x),R(x),L(z),R(z)) &= \mathbf{E}\sum_{i,j}\psi(Y_i,Y_{i+1},Y_j,Y_{j+1})\mathbb{1}_{(Y_i,Y_{i+1}]}(x)\mathbb{1}_{(Y_j,Y_{j+1}]}(z)\\ &= \sum_j \mathbf{E}\psi(Y_j,Y_{j+1},Y_j,Y_{j+1})\mathbb{1}_{(Y_j,Y_{j+1}]}(x)\mathbb{1}_{(Y_j,Y_{j+1}]}(z)\\ &\quad+\sum_j \mathbf{E}\,\psi(Y_{j-1},Y_j,Y_j,Y_{j+1})\mathbb{1}_{(Y_{j-1},Y_j]}(x)\mathbb{1}_{(Y_j,Y_{j+1}]}(z)\\ &\quad+\sum_{\substack{i,j:\\ i+1<j}} \mathbf{E}\psi(Y_i,Y_{i+1},Y_j,Y_{j+1})\mathbb{1}_{(Y_i,Y_{i+1}]}(x)\mathbb{1}_{(Y_j,Y_{j+1}]}(z).\end{aligned}$$

Thus,

$$\begin{aligned}&\mathbf{E}\,\psi(L(x),R(x),L(z),R(z))\\ &\quad=+\iint \psi(s_1,s_2,s_1,s_2)\mathfrak{d}_2(s_1,s_2)\mathbb{1}_{(s_1,s_2]}(x)\mathbb{1}_{(s_1,s_2]}(z)\,\mathrm{d}s_1\,\mathrm{d}s_2\\ &\qquad+\iiint \psi(s_1,s_2,s_2,s_3)\mathfrak{d}_3(s_1,s_2,s_3)\mathbb{1}_{(s_1,s_2]}(x)\mathbb{1}_{(s_2,s_3]}(z)\,\mathrm{d}s_1\,\mathrm{d}s_2\,\mathrm{d}s_3\\ &\qquad\iiint\int \psi(s_1,s_2,s_3,s_4)\mathfrak{d}_4(s_1,s_2,s_3,s_4)\mathbb{1}_{(s_1,s_2]}(x)\mathbb{1}_{(s_3,s_4]}(z)\,\mathrm{d}s_1\,\mathrm{d}s_2\,\mathrm{d}s_3\,\mathrm{d}s_4.\end{aligned}$$

If we define a ν-measurable function $\mathfrak{d}(s|x,z), s=(s_1,s_2,s_3,s_4)$, by

$$\mathfrak{d}(s|x,z)=\mathbb{1}_{(s_1,s_2]}(x)\mathfrak{d}_*(s|z),$$

where

$$\mathfrak{d}_*(s|z)=\begin{cases}\mathfrak{d}_2(s_1,s_2)\mathbb{1}_{(s_1,s_2]}(z) & \text{if } s_1=s_3<s_2=s_4,\\ \mathfrak{d}_3(s_1,s_2,s_4)\mathbb{1}_{(s_2,s_4]}(z) & \text{if } s_1<s_2=s_3<s_4,\\ \mathfrak{d}_4(s_1,s_2,s_3,s_4)\mathbb{1}_{(s_3,s_4]}(z) & \text{if } s_1<s_2<s_3<s_4,\\ 0 & \text{else}\end{cases} \tag{9}$$

then we obtain for $x<z$

$$\mathbf{E}\psi(L(x),R(x),L(z),R(z))=\iiint\int \psi(s)\mathfrak{d}(s|x,z)\,\mathrm{d}\nu,$$

and therefore

$$q_{x,z}(s_1,s_2,s_3,s_4)=\mathbb{1}_{(s_1,s_2]}(x)\mathfrak{d}_*(s_1,s_2,s_3,s_4|z). \tag{10}$$

5. The distribution of the random vector $(L(X),R(X),L(Z),R(Z))$

For the right truncated density function $f(x)$ we shall use the following notation:

$$f_a(x)=\frac{f(x)}{\int_{u\leqslant a}f(u)\,\mathrm{d}u}\mathbb{1}_{(-\infty,a]}(x).$$

Now we suppose that for fixed z and fixed value of $\tau=t$, the random variable X is taken from the truncated distribution with density $f_{\mathfrak{z}}(x)$. Here $\mathfrak{z}=\mathfrak{z}(t,z)=L(z)$. It follows from (10) that in that case the distribution P_z of random vector $(L(X),R(X),L(z),R(z))$ has density (with respect to the measure ν) $q(s_1,s_2,s_3,s_4|z)$,

$$q(s_1,s_2,u,v|z)=\int q_{x,z}(s_1,s_2,u,v)f_u(x)\,\mathrm{d}x,$$

and (see (9))

$$q(s_1,s_2,u,v|z)=\int_{s_1}^{s_2} f_u(x)\,\mathrm{d}x\times \mathfrak{d}_*(s_1,s_2,s_3,s_4|z),$$

where for $s=(s_1,s_2,s_3,s_4)$

$$\mathfrak{d}_*(s|z)=\begin{cases}\mathfrak{d}_3(s_1,s_2,s_4)\mathbb{1}_{(s_2,s_4]}(z) & \text{if } s_1<s_2=s_3<s_4,\\ \mathfrak{d}_4(s_1,s_2,s_3,s_4)\mathbb{1}_{(s_3,s_4]}(z) & \text{if } s_1<s_2<s_3<s_4,\\ 0 & \text{else.}\end{cases}$$

Therefore the distribution P_z is absolutely continuous with respect to the measure v_*, which is defined for continuous nonnegative functions $\psi(s)$ by the relation

$$\iiint\int \psi(s)\,\mathrm{d}v_*=\iiint \psi(s_1,s_2,s_2,s_4)\,\mathrm{d}s_1\,\mathrm{d}s_2\,\mathrm{d}_4+\iiint\int \psi(s_1,s_2,s_3,s_4)\,\mathrm{d}s_1\,\mathrm{d}s_2\,\mathrm{d}s_3\,\mathrm{d}s_4,$$

and

$$\frac{\mathrm{d}P_z}{\mathrm{d}v_*}=q(s|z).$$

Now suppose that Z is a random variable with density g, which is independent from the random covering $\vartheta(\cdot)$. For fixed values $Z=z$ and $\tau=t$, random variable X is taken from the truncated distribution with density $f_{\mathfrak{z}}(x)$, $\mathfrak{z}=\mathfrak{z}(t,z)=L(z)$. Denote by P_* the distribution of the random vector $(L(X),R(X),L(Z),R(Z))$. It is clear that the distribution P_* has density $q(s)$ with respect to the measure v_*,

$$\begin{aligned}q(s_1,s_2,u,s_4)&=\int_{s_1}^{s_2} f_u(x)\,\mathrm{d}x\times\int \mathfrak{d}_*(s_1,s_2,u,s_4|z)g(z)\,\mathrm{d}z\\&=\int_{s_1}^{s_2} f_u(x)\,\mathrm{d}x\times \mathfrak{d}(s_1,s_2,u,s_4).\end{aligned}$$

Now consider the random vector $W=(L(X),R(X),L(Z))$. Let v^{**} be the measure on $\mathbb{R}^3$, defined for continuous nonnegative functions ψ by

$$\iiint \psi(s_1,s_2,s_3)\,\mathrm{d}v=\iint \psi(s_1,s_2,s_2)\,\mathrm{d}s_1\,\mathrm{d}s_2+\iiint \psi(s_1,s_2,s_3)\,\mathrm{d}s_1\,\mathrm{d}s_2\,\mathrm{d}s_3,$$

It is clear that the distribution P_W of random vector W is absolutely continuous with respect to the measure v^{**} and

$$p(y)=p(y_1,y_2,y_3)=\frac{\mathrm{d}P_z}{\mathrm{d}v^{**}}=\int q(y_1,y_2,y_3,u)\,\mathrm{d}u.$$

Therefore,

$$p(u,v,z)=\int_u^v f_z(x)\,\mathrm{d}x\times r(u,v,z),$$

where

$$r(u,v,z)=\int \mathfrak{d}(u,v,z,x)\,\mathrm{d}x.$$

6. Estimation of the density of survival

The problem that we are faced with could be formulated as follows. Let $W,W_1,\ldots,W_n$ be i.i.d. random vectors, $W=(L(X),R(X),L(Z))$, with unknown density

$$p(u,v,w)=r(u,v,w)\times\frac{\int_u^v f(x)\,\mathrm{d}x}{\int_{x\leqslant w} f(x)\,\mathrm{d}x}. \tag{11}$$

We assume that the baseline density r and density f belong to given sets $\mathscr{G}$ and $\mathscr{F}$ correspondingly, and specify these sets later. We set

$$\varphi(f;u,v,w)=\frac{\int_u^v f(x)\,\mathrm{d}x}{\int_{x\leqslant w} f(x)\,\mathrm{d}x},$$

$$\mathscr{P}=\{p:p=r\varphi(f;\cdot),(r,f)\in\mathscr{G}\times\mathscr{F}\}. \tag{12}$$

Denote by P_n the empirical measure,

$$P_n\{A\}=\frac{1}{n}\sum_{j=1}^{n}\mathbb{1}_A(W_j).$$

Consider the maximum likelihood estimator $\widehat{p}_n$ for unknown $p\in\mathscr{P}$,

$$\int \ln\widehat{p}_n\,\mathrm{d}P_n=\max_{q\in\mathscr{P}}\int \ln q\,\mathrm{d}P_n. \tag{13}$$

It is clear that $\widehat{p}_n=\widehat{r}_n\times\varphi(\widehat{f}_n;\cdot)$, where $\widehat{r}_n$ and $\widehat{f}_n$ are maximum likelihood estimators for r and f,

$$\int \ln\varphi(\widehat{f}_n;\cdot)\,\mathrm{d}P_n=\max_{q\in\mathscr{F}}\int \ln\varphi(q;\cdot)\,\mathrm{d}P_n,$$

$$\int \ln\widehat{r}_n\,\mathrm{d}P_n=\max_{q\in\mathscr{G}}\int \ln q\,\mathrm{d}P_n.$$

We are interested in the estimation of f in the presence of the nuisance parameter r. First, we need to define the Hellinger distance between two densities f and g.

Let $(\mathscr{Y},\mathscr{B},\mu)$ be a measurable space and $Y_1,\dots,Y_n$ be i.i.d. random elements of $\mathscr{Y}$ with common distribution $P\in\mathscr{P}$ and density f,

$$f(y)=\frac{\mathrm{d}P}{\mathrm{d}\mu}(y),\quad f\in\mathscr{F}=\left\{f:f=\frac{\mathrm{d}P}{\mathrm{d}\mu},\ \text{for some } P\in\mathscr{P}\right\}.$$

For nonnegative f,g let $h(f,g)$ be the Hellinger distance,

$$h^2(f,g)=\frac{1}{2}\int_{\mathscr{Y}}(\sqrt{f}-\sqrt{g})^2\,\mathrm{d}\mu.$$

For a pair of nonnegative functions $g^L\leqslant g^R$ denote by $V(g^L,g^R)$ the set

$$V(g^L,g^R)=\{g:g^L\leqslant g\leqslant g^R\}.$$

Denote by $N_{[\ ]}(\varepsilon,\mathscr{F},h(\mu))$ the smallest value of m such that

$$\mathscr{F}\subset\bigcup_{j=1}^{m}V(g_j^L,g_j^R)\quad\text{where } h(g_j^L,g_j^R)\leqslant\varepsilon,\ \ j=1,\dots,m.$$

The bracketing Hellinger ε-entropy $H(\varepsilon,\mathscr{F},h(\mu))$ is defined as

$$H(\varepsilon,\mathscr{F},h(\mu))=\ln N_{[\ ]}(\varepsilon,\mathscr{F},h(\mu)).$$

The bracketing Hellinger ε-entropy $H(\varepsilon,\mathscr{F},L_2(\mu))$ in relation to the usual L_2 norm with respect to the measure μ is defined as

$$H(\varepsilon,\mathscr{F},L_2(\mu))=\ln N_{[\ ]}(\varepsilon,\mathscr{F},L_2(\mu)).$$

Theorem 1 (*Wong and Shen*). *Let $\mathscr{F}$ be a set of densities on a measurable space $(\mathscr{Y},\mathscr{B},\mu)$ and $Y^n=(Y_1,\dots,Y_n)$ a random sample with density $f\in\mathscr{F}$. The bracketing Hellinger s-entropy of $\mathscr{F}$ is denoted by $H(s,\mathscr{F},h(\mu))$ and ε is a positive number assumed to verify, for some constant c*:

$$\int_{\varepsilon^2}^{\varepsilon}H^{1/2}(s,\mathscr{F},h(\mu))\,\mathrm{d}s\leqslant c\varepsilon^2\sqrt{n}. \tag{14}$$

Then there exist positive constants c_1, c_2, C such that

$$P\left\{\sup_{\substack{h(g,f)\geqslant\varepsilon,\\ g\in\mathscr{F}}}\prod_{j=1}^{n}\frac{g(Y_j)}{f(Y_j)}\geqslant\exp\{-c_1n\varepsilon^2\}\right\}\leqslant C\exp\{-c_2n\varepsilon^2\}. \tag{15}$$

Corollary 1. *Under the assumptions of Theorem* 1, *the maximum likelihood estimator $\hat{f}_n$ of f verifies*:

$$P\{h(\hat{f}_n,f)\geqslant\varepsilon\}\leqslant C\exp\{-c_2n\varepsilon^2\}. \tag{16}$$

Proof. Let $A=\{Y^n : h(\hat{f}_n,f)\geqslant\varepsilon\}$. $\tilde{f}_n$ is defined as the "maximum likelihood estimator chosen in $\mathscr{F}$ outside the Hellinger ball with center f and radius ε":

$$\prod_{j=1}^{n}\frac{\tilde{f}_n(Y_j)}{f(Y_j)}\geqslant\prod_{j=1}^{n}\frac{g(Y_j)}{f(Y_j)}\quad\forall g\in\mathscr{F}\cap\{h(f,g)\geqslant\varepsilon\}.$$

On A, $\tilde{f}_n\equiv\hat{f}_n$ so that, on A we have

$$\prod_{j=1}^{n}\frac{\tilde{f}_n(Y_j)}{f(Y_j)}=\prod_{j=1}^{n}\frac{\hat{f}_n(Y_j)}{f(Y_j)}\geqslant 1\geqslant\exp(-c_1n\varepsilon^2).$$

As a consequence

$$A\subset B=\left\{Y^n:\prod_{j=1}^{n}\frac{\tilde{f}_n(Y_j)}{f(Y_j)}\geqslant\exp(-c_1n\varepsilon^2)\right\}.$$

Thus $P(A)\leqslant P(B)$ and finally, due to Theorem 1

$$P\{h(\hat{f}_n,f)\geqslant\varepsilon\}\leqslant C\exp\{-c_2n\varepsilon^2\}. \tag{17}$$

which proves the corollary. □

Remark. In order to be able to prove consistency, we need that the smallest ε verifying (14), as a function of n, tends to 0 when n tends to ∞ and that $n\varepsilon^2$ tends to ∞, which is automatically verified as soon as $H(\varepsilon,\mathscr{F},h(\mu))\to\infty$ when $\varepsilon\to 0$. Then we have consistency together with the rate of convergence.

Consider the following definition and theorem which can be found in van der Vaart and Wellner (2000, pp. 154–155, 157), respectively.

Definition 1. Let χ be a bounded set in $\mathbb{R}^d$, $0<\alpha\leqslant 1$, $r\in\mathbb{N}$, and $\beta=r+\alpha$.

Then $\mathscr{C}^{\beta}_{M_0}$ is the set of all functions from χ onto $\mathbb{R}$ that possess uniformly bounded partial derivatives up to order r and whose highest partial derivatives are Lipschitz functions of order α. More precisely, for any $k=(k_1,\ldots,k_d)$

$$\|g\|_\beta=\max_{\sum k_i\leqslant r}\sup_x|D^kg(x)|+\max_{\sum k_i=r}\sup_{x,y}\frac{|D^kg(x)-D^kg(y)|}{\|x-y\|^\alpha}\leqslant M_0, \tag{18}$$

where the supremum is taken over all x,y in the interior of χ with $x\neq y$.

Theorem 2. *Let χ be a bounded convex subset of $\mathbb{R}^d$ with nonempty interior. There exists a constant K_0 depending on β,* diam(χ), *p, M_0 and d such that*

$$\log(N_{[\ ]}(\varepsilon,\mathscr{C}^{\beta}_{M_0},L_p(Q))\leqslant K_0\left(\frac{1}{\varepsilon}\right)^{d/\beta} \tag{19}$$

for every $p\geqslant 1$, $\varepsilon>0$ and probability measure Q.

We will use the above definition to impose a smoothness condition on our set of functions f or $\phi(f)$ which along with the above theorem and the following lemma will allow us to give bounds on the covering numbers of these sets.

Lemma 2. *We assume that for the space $\mathscr{P}$ defined in (12) the function r is bounded by a constant $r_0 > 0$ and that the space $\mathscr{F}$ of densities with respect to Lebesgue measure λ is uniformly bounded above and below*:

$$\mathscr{F} = \{f : f \text{ has compact support } \chi \text{ and } 0 < c_l \leqslant f \leqslant c_u < \infty\}.$$

Then if

$$H(\varepsilon, \mathscr{F}, L_2(\lambda)) = h_0 < \infty \tag{20}$$

there exist constants c' and c'' such that

$$H(c'\varepsilon, \mathscr{F}, h(\lambda)) \leqslant h_0, \tag{21}$$

$$H(c''\varepsilon, \mathscr{P}, h(v^{**})) \leqslant h_0. \tag{22}$$

Proof. In order to prove (21) and (22) we shall show that to each bracket of size ε in the space $\mathscr{F}$ with respect to the Hellinger distance corresponds a bracket of size $c'\varepsilon$ for L_2 distance in $\mathscr{F}$ and a bracket of size $c''\varepsilon$ in the space $\mathscr{P}$ with respect to the Hellinger distance. The first result is a consequence of the equivalence of the L^2 and Hellinger distances in the space $\mathscr{F}$, due to the property of uniform bi-boundedness of $\mathscr{F}$. And as we show below that this property holds also for $\mathscr{P}$ we get the second result.

Let us rewrite $p(u, v, w)$ as

$$p(u, v, w) = r^*(u, v, w) \times \varphi^*(f; u, v, w),$$

$$r^*(u, v, w) = \frac{v-u}{w} r(u, v, w),$$

$$\varphi^*(f; u, v, w) = \frac{\dfrac{\int_u^v f(x)\,\mathrm{d}x}{v-u}}{\dfrac{\int_0^w f(x)\,\mathrm{d}x}{w}}.$$

The fact that $\mathscr{F}$ is bi-bounded, $0 < c_l < f < c_u < \infty$ for any $f \in \mathscr{F}$, implies the same property for the set $\mathscr{F}^* = \{\varphi^*(f);\ f \in \mathscr{F}\}$:

$$\frac{c_l}{c_u} < \varphi^*(f) < \frac{c_u}{c_l}, \quad f \in \mathscr{F}. \tag{23}$$

Thus, the Hellinger distance is equivalent to the L^2 distance both for $\mathscr{F}$ and for $\mathscr{F}^*$, as

$$4c_l h^2(f, g) \leqslant L^2(f, g) = \int (f-g)^2 \,\mathrm{d}\lambda = \int (\sqrt{f} - \sqrt{g})^2 (\sqrt{f} + \sqrt{g})^2 \,\mathrm{d}\lambda \leqslant 4c_u h^2(f, g).$$

Let $D = \varphi^*(f) - \varphi^*(g)$. Using $a/c - b/d = (a-b)/c + b(1/c - 1/d)$, we get

$$D = \frac{w}{v-u} \frac{\int_u^v (f-g)}{\int_0^w f} + \frac{w}{v-u} \left(\int_u^v g\right) \frac{\int_0^w (g-f)}{\int_0^w f \int_0^w g}$$

so that

$$|D| \leqslant \frac{\int_u^v \dfrac{|f-g|}{v-u}}{\int_0^w \dfrac{f}{w}} + \int_u^v \frac{g}{(v-u)} \frac{\int_0^w \dfrac{|f-g|}{w}}{\int_0^w \dfrac{f}{w} \int_0^w \dfrac{g}{w}}.$$

We now need to show that $\psi(f, g) = \int D^2(u, v, w)\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w$ tends to 0 if $h^2(f, g)$ tends to 0.

Let $Mf(x)$ be defined for any real function f as $Mf(x) = \sup_{I \ni x} \int_I |f(u)|\,\mathrm{d}u/|I|$, then we know Stein (1993) that

$$\int |Mf(u)|^p \,\mathrm{d}u \leqslant C(p)^p \int |f(u)|^p \,\mathrm{d}u, \quad \forall p > 1. \tag{24}$$

Applying (24) to $p=2$ with function f replaced by $f-g$, f and g in $\mathscr{F}$, we get that the L_2^2 norm of D, denoted $\psi(f,g)$ is such that:

$$\begin{aligned}\psi(f,g) &= \int_{0<u<v\leqslant w<1} D^2(u,v,w)\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w\\ &\leqslant 2\left[\frac{1}{c_l^2}\int_0^1 (M(f-g)(u))^2\,\mathrm{d}u + \frac{c_u^2}{c_l^4}\int_0^1 (M(f-g)(w))^2\,\mathrm{d}w\right]\\ &\leqslant \frac{2c_l^2+2c_u^2}{c_l^4}C(2)^2\times\int_0^1((f-g)(x))^2\,\mathrm{d}x\\ &\leqslant \frac{2c_l^2+2c_u^2}{c_l^4}C(2)^2\times 4c_u h^2(f,g).\end{aligned}$$

Since the function r is assumed to be bounded by $r_0>0$, we have that

$$\begin{aligned}h^2(p^{r,f},p^{r,g}) &\leqslant r_0 h^2(\varphi^*(f),\varphi^*(g))\\ &\leqslant r_0 4\frac{c_u}{c_l}\psi(f,g)\\ &\leqslant r_0\frac{16c_u^2(2c_l^2+2c_u^2)\;C(2)^2}{c_l^5}h^2(f,g).\end{aligned}$$

In order to complete the proof we have to follow the above steps for $f=f_R$ and $g=f_L$ for $f_L\leqslant f\leqslant f_R$ and $h^2(f_L,f_R)\leqslant\varepsilon$. Actually, $(f_L*=\max(f_L,c_l), f_R*=\min(f_R,c_u))$ is also an ε-bracketing interval for f. Then we obtain that $h^2(p^{r,f_L*},p^{r,f_R*})\leqslant c''\varepsilon$ as desired. □

Lemma 3. *Let us assume that $f\in\mathscr{F}\subset\mathscr{C}^{\beta}_{M_0}(\chi)$ where $\chi=[0;b]$, $b>0$ and $\mathscr{F}$ is a set of densities on χ with respect to Lebesgue measure λ. $\mathscr{F}$ is also assumed to be uniformly bounded below by some $c_l>0$. Moreover, the censoring and truncating law, with density r with respect to v^{**} lies in $\mathscr{G}$ such that*

$$\forall\varepsilon>0 \text{ and } \forall x\in[\varepsilon;b],\quad \inf_{r\in\mathscr{G}}\int_{0\leqslant u<v<w, v-u\leqslant\varepsilon} r(u,v,w)\,\mathrm{d}v^{**}(u,v,w)>0. \tag{25}$$

Then if $h^2(f_1,f_2)>0$, we have

$$h^2(p^{r_1,f_1},p^{r_2,f_2})>0. \tag{26}$$

Proof. Let us assume that $2h^2(f_1,f_2)\geqslant\varepsilon>0$. Then there exists at least one $x_0\in[0,b]$ such that $|\sqrt{f_1(x_0)}-\sqrt{f_2(x_0)}|>\sqrt{\varepsilon}/(2\sqrt{b})$. As f_1 and f_2 are in $\mathscr{C}^{\beta}_{M_0}$ with $\beta>\frac{1}{2}$, there exist a constant $c_m>0$, such that

$$|f_1(x)-f_1(x+t)|\leqslant c_m t^{\alpha}$$

uniformly on $\mathscr{F}$, where $\beta=r+\alpha$, $r\in\mathbb{N}$, $0<\alpha\leqslant 1$. So,

$$|f_1-f_2| = |\sqrt{f_1}-\sqrt{f_2}|(\sqrt{f_1}+\sqrt{f_2})>|\sqrt{f_1}-\sqrt{f_2}|2\sqrt{c_l}>\sqrt{c_l/b}\sqrt{\varepsilon}.$$

Moreover, there exists $t_0>0$ derived below, such that

$$\forall t\in[-t_0\;t_0],\quad |f_1(x_0+t)-f_1(x_0)|\leqslant c_m t^{\alpha}<\tfrac{1}{4}\{\sqrt{c_l/b}\sqrt{\varepsilon}\}.$$

This last inequality holds as soon as

$$t_0=\left[\frac{\sqrt{c_l\varepsilon}}{4\sqrt{b}c_m}\right]^{1/\alpha}.$$

Then, uniformly on $[x_0-t_0;x_0+t_0]$, $|f_1(x)-f_2(x)|>\sqrt{c_l\varepsilon/2b}$.

Let us denote for convenience

$$N_i=\frac{1}{v-u}\int_u^v f_i(x)\,\mathrm{d}x,\quad D_i=\frac{1}{w}\int_0^w f_i(x)\,\mathrm{d}x,\;\; i=1,2$$

so that

$$|(\varphi^*(f_1)-\varphi^*(f_2))| = \left|\frac{N_1}{D_1}-\frac{N_2}{D_2}\right|. \tag{27}$$

Taking care first of the numerators N_i, depending on u and v only, we have that on $0<x_0-t_0<u<v<x_0+t_0$, where the difference f_1-f_2 has always the same sign, the following inequality holds:

$$|N_1-N_2| = \frac{1}{v-u}\int_u^v |f_2-f_1| > \frac{1}{2}\sqrt{\frac{c_l\varepsilon}{b}}. \tag{28}$$

Now, taking care of the denominators D_i, depending on w alone, there exists a $b_0<b$ such that

$$\forall w \geqslant b_0,\ \ D_1 \geqslant 1-\eta_1, \quad D_2 \geqslant 1-\eta_2,$$

where

$$\eta_1 < \varepsilon^\gamma, \quad \eta_2 < \varepsilon^\gamma$$

for some $\gamma>0$.

Then, on the set

$$B_\varepsilon = \{0<x_0-t_0 \leqslant u<v \leqslant x_0+t_0<b_0<b\} \tag{29}$$

we have the following inequality, due to (28):

$$|(\varphi^*(f_1)-\varphi^*(f_2))| = \left|\frac{N_1}{D_1}-\frac{N_2}{D_2}\right| > |N_1-N_2| - \max(N_1\eta_2, N_2\eta_1) > \frac{1}{2}\sqrt{\frac{c_l\varepsilon}{b}} - c_u\varepsilon^\gamma. \tag{30}$$

It suffices then to take $\gamma>\frac{1}{2}$ in order that the right term of the inequality is strictly positive for sufficiently small ε. Then we can say that

$$\begin{aligned} 2h^2(\varphi^*(f_1),\varphi^*(f_2)) &\geqslant \int_{B_\varepsilon}\left(\sqrt{\varphi^*(f_1)}-\sqrt{\varphi^*(f_2)}\right)^2 \mathrm{d}\lambda^3 \\ &\geqslant (1/16)(c_l^2/bc_u)(1-2\sqrt{b/c_l}c_u\varepsilon^{\gamma-1/2})^2 \times \lambda^3(B_\varepsilon)\varepsilon \\ &> 0. \end{aligned}$$

We thus proved that $h^2(\varphi^*(f_1),\varphi^*(f_2)$ cannot be equal to 0, as soon as $h^2(f_1,f_2)>0$. Now, due to assumption (25) the same result holds for $h^2(r^*\varphi^*(f_1), r^*\varphi^*(f_2))$. □

Remark. If we do not have a compact support for $\mathscr{F}$, we need that $\mathscr{F}$ be uniformly integrable on $\mathbb{R}$, that is

$$\forall\delta>0,\ \exists 0<b_0<\infty \text{ such that } \sup_{f\in\mathscr{F}} P(X>b_0)<\delta$$

in order to get the same result.

Remark. Note that a sufficient assumption for the condition (25) is that r^* is a.s. positive with respect to the measure ν^{**}.

Theorem 3. *Suppose that the parameter of interest, that is, the true density f, with respect to Lebesgue measure, of the survival time X, belongs to the space*

$$\mathscr{F} = \{f:\ f\in\mathscr{C}^\beta_{M_0} \text{ with compact support } \chi \text{ and } 0<c_l\leqslant f\leqslant c_u<\infty\} \tag{31}$$

with $\beta>\frac{1}{2}$. Also, suppose that the function r which describes the censoring and truncation mechanism is known and bounded by a constant $r_0>0$. Then the nonparametric maximum likelihood estimator $\hat{f}_n$ is consistent in the Hellinger distance for the density f, namely, for any $\varepsilon>0$

$$\sup_{p=p^{rf}\in\mathscr{P}} P_p\{h(\widehat{f}_n,f)>\varepsilon\}\to 0 \quad \text{as } n\to\infty.$$

More specifically, the rate of convergence is given by

$$\sup_{p=p^{rf}\in\mathscr{P}} P_p\{h(\widehat{f}_n,f)>n^{-\beta/(2\beta+1)}\}\leqslant C\exp\{-c_2 n^{1/(2\beta+1)}\}.$$

Proof. By Theorem 2 and Lemma 2 we have that

$$\log(N_{[\,]}(\varepsilon, \mathscr{F}, L_2(\lambda)) \leqslant K_0 \left(\frac{1}{\varepsilon}\right)^{1/\beta}$$

or equivalently

$$\log(N_{[\,]}(c'\varepsilon, \mathscr{F}, h(\lambda)) \leqslant K_0 \left(\frac{1}{\varepsilon}\right)^{1/\beta}$$

or

$$\log(N_{[\,]}(c''\varepsilon, \mathscr{P}, h(v^{**})) \leqslant K_0 \left(\frac{1}{\varepsilon}\right)^{1/\beta}. \tag{32}$$

By Theorem 1, Corollary 1 and Lemma 3 we have the result of consistency along with the rate of the convergence as long as we verify that the smallest ε that satisfies (14), as a function of n, tends to 0 when n tends to ∞ and that $n\varepsilon^2$ tends to ∞.

More specifically, in our case we have

$$\int_{\varepsilon^2}^{\varepsilon} H^{1/2}(s, \mathscr{P}, h(v^{**}))\,\mathrm{d}s \leqslant \int_{\varepsilon^2}^{\varepsilon} \sqrt{K_0}\left(\frac{1}{s}\right)^{1/2\beta} \mathrm{d}s \tag{33}$$

$$= \sqrt{K_0}\frac{2\beta}{2\beta-1}(\varepsilon^{1-1/2\beta} - \varepsilon^{2-1/\beta}) \leqslant c\varepsilon^2\sqrt{n} \tag{34}$$

which provides for the n that

$$\sqrt{n} \geqslant C_0(\varepsilon^{-1-1/2\beta} - \varepsilon^{-1/\beta})$$

with $\beta > \frac{1}{2}$. This implies that the smallest ε for which (14) holds is as a function of n given as

$$\varepsilon = C_0 n^{-\beta/(2\beta+1)}$$

which converges to 0 as $n \to \infty$. $\square$

Also, $n\varepsilon^2 = C_0 n^{1-2\beta/(2\beta+1)} \to \infty$ and that proves the desired consistency and provides the rate of convergence.

7. Example

In this section we will provide an example in order to illustrate the theoretical results provided in the previous sections. We consider as the total interval of observation time the interval $[0,1]$. We also consider the case where the censoring mechanism $\{Y_j, j=1,\dots,K\}$, where K is considered at first fixed, is represented in the form of an ordered sample of size K from the Uniform(0,1) distribution. At the same time, the right-truncating variable Z follows an independent Uniform$(0,M)$ distribution where $M>1$. The survival time X follows a distribution with unknown density $f(x)$, $x \in [0,1]$ with respect to the Lebesgue measure on $\mathbb{R}^+$, which we assume to belong to the space $\mathscr{F}$ defined in (31). What we observe is $W_1,\dots,W_n$, a sample of i.i.d. random vectors, where $W=(L(X),R(X),L(Z))$, with density

$$p(u,v,w) = r(u,v,w) \times \frac{\int_u^v f(x)\,\mathrm{d}x}{\int_{x\leqslant w} f(x)\,\mathrm{d}x}.$$

We would like to estimate f while $r(u,v,w)$ is known and describes the censoring and truncating mechanism, which we chose to fix at a first stage. We could also have a family of densities r if K is random and follows a Poisson(λ) where $2<\lambda<K_0$. We now have to compute $r(u,v,w)$ for K fixed.

Lemma 4. *For the observational scheme described above, the density $r(u,v,w)$ of the censoring and truncating mechanism, has two parts, one which is absolutely continuous with respect to λ^3, denoted by r_3, and one that is absolutely continuous with respect to λ^2,* *denoted by* r_2 *which are given as*

$$r_3(u,v,w) = K(K-1)\{(u+1-v)^{K-2} - (u+w-v)^{K-2} + (K-2)(M-1)(u+w-v)^{K-3}\}/M,$$
$$0 \leqslant u < v < w \leqslant 1,$$

$$r_2(u,v) = K\{(u+1-v)^{K-1} - u^{K-1}) + (K-1)(M-1)u^{K-2}\}/M, \quad 0 \leqslant u < v \leqslant 1.$$

Proof. For any measurable function $\psi(u,v,w)$, v^{**} denotes the following measure:

$$\int \psi(u,v,w)\,\mathrm{d}v^{**} = \int \psi(u,v,w)\,\mathrm{d}\lambda^3(u,v,w) + \int \psi(u,v,v)\,\mathrm{d}\lambda^2(u,v).$$

The measure Q is absolutely continuous with respect to v^{**} and one has

$$\int \psi(u,v,w)\,\mathrm{d}Q = \int \psi(u,v,w) r_3(u,v,w)\,\mathrm{d}\lambda^3(u,v,w) + \int \psi(u,v,v) r_2(u,v)\,\mathrm{d}\lambda^2(u,v). \tag{35}$$

Let us first compute r_3. The fact that v is strictly smaller than w means that we have to consider the density $r_{Y_j,Y_{j+1},Y_k,Y_{k+1},Z}(u,v,w,w',z)$ with $k > j+1$:

$$r_{Y_j,Y_{j+1},Y_k,Y_{k+1},Z}(u,v,w,w',z) = \frac{1}{M}\frac{K!}{(j-1)!(k-j-2)!(K-k-1)!}u^{j-1}(w-v)^{k-j-2}(1-w')^{K-k-1^*}1\{0<u<v<w<z<w'<M\}$$

which has to be integrated with respect to w' and z, and summed over all $K-1 \geqslant k > j+1$ and j from 1 to $K-3$. This is done hereafter and given in Eq. (36). Doing so, however, we miss the borderline term where Z lies between Y_K, the last time we look at a patient, and M. Adding also this part (37) to (36), leads to the final value for r_3 given in (38).

Integration with respect to z and w' gives

$$r_{Y_j,Y_{j+1},Y_k}(u,v,w) = \frac{1}{M}\frac{K!}{(j-1)!(k-j-2)!(K-k-1)!}u^{j-1}(w-v)^{k-j-2}\int_w^1 (1-w')^{K-k-1}(w'-w)\,\mathrm{d}w'$$

and consequently,

$$r_{Y_j,Y_{j+1},Y_k}(u,v,w) = \frac{1}{M}\frac{K!}{(j-1)!(k-j-2)!(K-k+1)!}u^{j-1}(w-v)^{k-j-2}(1-w)^{K-k+1}.$$

We need now to sum r_{Y_j,Y_{j+1},Y_k} first over k and then over j. For the first sum we have (apart from M in the denominator)

$$\frac{u^{j-1}}{(j-1)!}\sum_{k=j+2}^{K-1}\frac{K!}{(k-j-2)!(K-k+1)!}(w-v)^{k-j-2}(1-w)^{K-k+1}$$

which with the changes $K' = K-1$ and $k' = k-j-2$ becomes

$$\frac{u^{j-1}}{(j-1)!}\sum_{k'=0}^{K'-j-2}\frac{K'+1!}{k'!(K'-j-k')!}(w-v)^{k'}(1-w)^{K'-j-k'}$$
$$= \frac{u^{j-1}(K'+1)!}{(j-1)!(K'-j)!}\{(1-v)^{K'-j} - (K'-j)(w-v)^{K'-j-1}(1-w) - (w-v)^{K'-j}\}.$$

Then, changing again to $K = K'+1$ and summing for j from 1 to $K-3$ we obtain the sum of three terms which we will denote by S_1, S_2 and S_3. First,

$$S_1 = \sum_{j=1}^{K-3}\frac{K!}{(j-1)!(K-1-j)!}u^{j-1}(1-v)^{K-1-j}$$

which with the changes $K' = K+3$ and $j' = j-1$ becomes

$$\sum_{j'=0}^{K'-1}\frac{(K'+3)!}{j'!(K'+1-j')!}u^{j'}(1-v)^{K'+1-j'}$$
$$= (K'+3)(K'+2)\{(u+1-v)^{K'+1} - u^{K'+1} - (K'+1)u^{K'}(1-v)\}$$
$$= K(K-1)\{(u+1-v)^{K-2} - u^{K-2} - (K-2)u^{K-3}(1-v)\}.$$

The second term

$$S_2 = \sum_{j=1}^{K-3} \frac{K!}{(j-1)!(K-1-j)!}(K-1-j)u^{j-1}(1-w)(w-v)^{K-j-2}$$

with the changes $K' = K+3$ and $j' = j-1$ becomes

$$(1-w)\sum_{j'=0}^{K'-1} \frac{(K'+3)!}{j'!(K'-j')!}u^{j'}(w-v)^{(K'-j')} = (K'+3)(K'+2)(K'+1)(1-w)\{(u+w-v)^{K'} - u^{K'}\}$$
$$= K(K-1)(K-2)(1-w)\{(u+w-v)^{K-3} - u^{K-3}\}.$$

The third term

$$S_3 = \sum_{j=1}^{K-3} \frac{K!}{(j-1)!(K-1-j)!}u^{j-1}(w-v)^{K-j-1}$$

with the changes $K' = K+3$ and $j' = j-1$ becomes

$$\sum_{j'=0}^{K'-1} \frac{(K'+3)!}{j'!(K'+1-j')!}u^{j'}(w-v)^{(K'+1-j')}$$
$$= (K'+3)(K'+2)\{(u+w-v)^{K'+1} - u^{K'+1} - u^{K'}(K'+1)(w-v)\}$$
$$= K(K-1)\{(u+w-v)^{K-2} - u^{K-2} - (K-2)u^{K-3}(w-v)\}.$$

After some simplifications, $(S_1+S_2+S_3)/M$ becomes

$$K(K-1)\{(u+1-v)^{K-2} - (u+w-v)^{K-2} - (K-2)(u+w-v)^{K-3}(1-w)\}/M \tag{36}$$

for $0 \leqslant u < v < w \leqslant 1$.

We now add the borderline case where Y_k coincides with Y_K and the truncation variable Z falls within the interval (Y_K, M). The density of (Y_j, Y_{j+1}, Y_K, Z) is given as

$$r_{Y_j,Y_{j+1},Y_K,Z}(u,v,w,z) = \frac{1}{M}\frac{K!}{(j-1)!(K-j-2)!}u^{j-1}(w-v)^{K-j-2}.$$

Then, the density $r_{Y_j,Y_{j+1},Y_K}(u,v,w)$ is equal to

$$\frac{1}{M}\frac{K!}{(j-1)!(K-j-2)!}u^{j-1}(w-v)^{K-j-2}(M-w).$$

Now we need to sum r_{Y_j,Y_{j+1},Y_K} over j from 1 to $K-2$. Then, by making also the changes $K' = K-2$ and $j' = j-1$ we get

$$\frac{1}{M}(M-w)\sum_{j'=0}^{K'-1} \frac{(K'+2)!}{j'!(K'-j'-1)!}u^{j'}(w-v)^{K'-j'-1} = \frac{1}{M}(K'+2)(K'+1)(K')(M-w)(u+w-v)^{K'-1}$$
$$= \frac{1}{M}K(K-1)(K-2)(M-w)(u+w-v)^{K-3}. \tag{37}$$

Finally, combining (36) and (37) we obtain $r_3(u,v,w)$ which is absolutely continuous with respect to λ^3 and is given as

$$r_3 = K(K-1)\{(u+1-v)^{K-2} - (u+w-v)^{K-2} + (K-2)(M-1)(u+w-v)^{K-3}\}/M \tag{38}$$

for $0 \leqslant u < v < w \leqslant 1$.

We compute now r_2, corresponding to the case where $L(Z)$ coincides with $R(X)$, so that $w = v$. For this case we need to calculate the common density of $(Y_j, Y_{j+1}, Z, Y_{j+2})$ as Y_{j+1} coincides with Y_k and Z falls into the interval (Y_{j+1}, Y_{j+2}). So,

$$r_{Y_j,Y_{j+1},Y_{j+2},Z}(u,v,w,z) = \frac{1}{M}\frac{K!}{(j-1)!(K-j-2)!}u^{j-1}(1-w)^{K-j-2}$$

and

$$r_{Y_j,Y_{j+1}}(u,v) = \frac{1}{M}\frac{K!}{(j-1)!(K-j-2)!}u^{j-1}\int_v^1 (1-w)^{K-j-2}(w-v)\,\mathrm{d}w$$

and consequently,

$$r_{Y_j,Y_{j+1}}(u,v)=\frac{1}{M}\frac{K!}{(j-1)!(K-j)!}u^{j-1}(1-v)^{K-j}.$$

Now we need to sum over j from 1 to $K-2$. By also making the changes $K'=K-1$ and $j'=j-1$ we get

$$\frac{1}{M}K\sum_{j'=0}^{K'-2}\frac{(K')!}{j'!(K'-j')!}u^{j'}(1-v)^{K'-j'}=K\{(u+1-v)^{K-1}-(K-1)u^{K-2}(1-v)-u^{K-1}\} \tag{39}$$

for $0\leqslant u<v\leqslant 1$.

Again, we have to consider the special case where Y_{j+1} coincides with Y_K and therefore Z falls within the interval (Y_K,M). The density $r_{Y_{K-1},Y_K,Z}(u,v,z)$ is given as

$$\frac{1}{M}\frac{K!}{(K-2)!}u^{K-2}$$

and consequently,

$$r_{Y_{K-1},Y_K}(u,v)=\frac{1}{M}K(K-1)u^{K-2}(M-v) \tag{40}$$

for $0\leqslant u<v\leqslant 1$.

Finally, combining (39) and (40) we obtain $r_2(u,v)$ which is absolutely continuous with respect to λ^2 and is given as

$$r_2(u,v)=K\{(u+1-v)^{K-1}-u^{K-1})+(K-1)(M-1)u^{K-2}\}/M \tag{41}$$

for $0\leqslant u<v\leqslant 1$. □

The space

$$\mathscr{G}=\{r(u,v,w)\text{ defined by (38) and (41)}\} \tag{42}$$

consists of only one element r, which is known, since K is fixed.

Note that all of the appropriate assumptions required for the consistency of the estimator of f hold for the case of this example. First, by Eqs. (38), (41) we have

$$r_3(u,v,w)\leqslant K(K-1)[1+(K-2)(M-1)],$$

$$r_2(u,v)\leqslant K(1+(K-1)(M-1)).$$

Thus $r^*(u,v,w)=[(v-u)/w]r(u,v,w)\leqslant r(u,v,w)$ is bounded by $r_0=K^2(1+(K-1)(M-1))$ and our density r satisfies the required condition. It is also clear by the definition of r that it satisfies the identifiability condition (25).

References

Alioum, A., Commenges, D., 1996. A proportional hazards model for arbitrarily censored and truncated data. Biometrics 52, 512–524.

Finkelstein, D.M., Moore, D.F., Schoenfeld, D.A., 1993. A proportional hazards model for truncated aids data. Biometrics 49, 731–740.

Frydman, H., 1994. A note on nonparametric estimation of the distribution function from interval-censored and truncated observations. J. Roy. Statist. Soc. Ser. B 56, 71–74.

Huber, C., Vonta, F., 2004. Frailty models for arbitrarily censored and truncated data. Lifetime Data Analysis 10, 369–388.

Huber, C., Solev, V., Vonta, F., 2005. Estimation of density for arbitrarily censored and truncated data. In: Nikulin, M., Commenges, D., Huber, C. (Eds.), Probability, Statistics and Modelling in Public Health. Springer, Berlin, pp. 246–265.

Stein, E.M., 1993. Harmonic Analysis. Princeton University Press, Princeton.

Turnbull, B.W., 1976. The empirical distribution function with arbitrary grouped, censored and truncated data. J. Roy. Statist. Soc. 38, 290–295.

van der Vaart, A.W., 1998. Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

van der Vaart, A.W., Wellner, J.A., 2000. Weak Convergence and Empirical Processes (With Applications to Statistics). Springer Series in Statistics. Springer, Berlin.

Wong, W.H., Shen, N.X., 1995. Probability inequalities for likelihood ratios and convergence rates of Sieve MLEs. Ann. Statist. 23 (2), 339–362.