T E C H N I C A L
R E P O R T

0511

PRODUCT-LIMIT ESTIMATORS OF THE SURVIVAL FUNCTION WITH LEFT OR RIGHT CENSORED DATA

PATILEA, V. and J.-M. ROLIN

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Product-limit estimators of the survival function with left or right censored data

Valentin Patilea*
CREST-ENSAI
Campus de Ker Lann
Rue Blaise Pascal BP 37203
35172 Bruz Cedex, France
Email: patilea@ensai.fr

Jean-Marie Rolin
Institut de Statistique
Université Catholique de Louvain
20 Voie du Roman Pays
1348 Louvain-la-Neuve, Belgium

Running Title: *PL estimators with left and right censoring*

Abstract

The problem of estimating the distribution of a lifetime when data may be left or right censored is considered. Two models are introduced and the corresponding product-limit estimators are derived. Strong uniform convergence and asymptotic normality are proved for the product-limit estimators on the whole range of the observations. A bootstrap procedure that can be applied to confidence intervals construction is proposed.

Key words: bootstrap, delta-method, left and right censoring, martingales, product-limit estimator, strong convergence, weak convergence
MSC 2000: 62N01, 62N02, 60G05

[0]Financial support from the IAP research network nr P5/24 of the Belgian State (Federal Office for Scientific, Technical and Cultural Affairs) is gratefully acknowledged.

1 Introduction

The goal of this paper is to propose and analyze models for lifetime data that may be left or right censored. Typically, a lifetime T is left or right censored if, instead of observing T we observe a finite nonnegative random variable Y, and a discrete random variable with values 0, 1 or 2. By definition, when $A = 0$, $Y = T$, when $A = 1$, $Y < T$ and, when $A = 2$, $Y > T$. Models for left or right censored data were proposed by Turnbull (1974), Samuelsen (1989) and Huang (1999). See also Gu and Zhang (1993), van der Laan and Gill (1999), Kim (2003).

Assume that the sample consists of n independent copies of (Y, A) and let F_T be the distribution of the lifetime of interest T. Using the plug-in (or substitution) principle, the nonparametric estimation of F_T is straight as soon as F_T can be expressed as an explicit function of the distribution of (Y, A). The existence of such a function requires a precise description of the censoring mechanism that is generally achieved by introducing 'latent' variables and by making assumptions on their distributions. In this paper, two latent models allowing for explicit inversion formula, that is closed-form function relating F_T to the distribution of (Y, A), are proposed.

In some sense, our first latent model lies between the classical right-censorship model and the current-status data model. It may be applied to the following framework. Consider a study where T the age at onset for a disease is analyzed. The individuals are examined only one time and they belong to one of the following categories: (i) evidence of the disease is present and the age at onset is known (from medical records, interviews with the patient or family members, ...); (ii) the disease is diagnosed but the age at onset is unknown or the accuracy of the information about this is questionable; and (iii) the disease is not diagnosed at the examination time. Let C denote the age of the individual at the examination time. In the first case the exact failure time T (age at onset) is observed, that is $Y = T$. In case (ii) the failure time T is left-censored by C and thus $Y = C$, $A = 2$. Finally, the onset time T is right-censored by C for the individuals who have not yet developed the disease; in this case $Y = C$, $A = 1$. If no observation as in (ii) occurs, we are in the classical right-censorship framework, while if no uncensored observation is recorded we have current-status data. Our first latent model can be applied, for instance, with the data sets analyzed by Turnbull and Weiss (1978) and Cupples *et alli* (1991, Table 1).

The second latent model proposed is closely related to the first one. It lies between the left-censorship model and the current-status data model. Consider the example of a reliability experiment where the failure time of a type of device is analyzed. A sample of devices is considered and a single inspection for each device in the sample is undertaken. Some of them already failed without knowing when (left censored observations). To increase the precision of the estimates, a proportion of the devices still working is selected randomly and followed until failure (uncensored observations). For the remaining working devices the failure time is right censored by the inspection time.

Let us point out that, without any model assumption, given a distribution for the observed variables (Y, A) with $Y \geq 0$ and $A \in \{0, 1, 2\}$, we can always apply our two inversion formulae. In this way we build two pseudo-true distribution functions of the lifetime of interest which are functionals of the observed distribution. If the experiment under observation is compatible with the hypothesis of one of our latent models, the true F_T can be exactly recovered from the observed distribution. Otherwise, we can only approximate the true lifetime distribution.

The paper is organized as follows. Section 2 introduces our two latent models through the equations relating the distribution of the observations to those of the latent variables. Solving these equations for F_T we deduce the inversion formulae. The product-limit estimators are obtained by applying the inversion formulae to the empirical distribution. Section 2 is ended with some remarks and comments on related models. In particular, it is shown that our first (resp. second) latent model can be extended to the case where T is a failure time in the presence of competing (resp. complementary) failure causes. Section 3 contains the asymptotic results for the first latent model (similar arguments apply for the second model). We prove strong uniform convergence for the product-limit estimator on the whole range of the observations. Our proof extends and simplifies the results of Stute and Wang (1993) and Gill (1994) provided in the case of the Kaplan-Meier estimator. Next, the asymptotic normality of our product-limit estimator is obtained. The variance of the limit Gaussian process being complicated, a bootstrap procedure for which the asymptotic validity is a direct consequence of the delta-method is proposed. In section 4 our first model is applied to the age of the first use of marijuana using California high school students data considered by Turnbull and Weiss (1978). The appendix contains some technical proofs.

2 The latent models

2.1 Model 1

The survival time of interest is T (e.g., the age at onset). Let C be a censoring time (e.g., the age of the individual at the examination time) and Δ be a Bernoulli random variable. Assume that the latent variables T, C and Δ are independent. The observations are independent copies of the variables (Y, A), with $Y \geq 0$ and $A \in \{0, 1, 2\}$. These variables are defined as follows:

$$\begin{cases} Y = T, \quad A = 0 \quad if \quad 0 \leq T \leq C \quad \text{and} \quad \Delta = 1; \\ Y = C, \quad A = 1 \quad if \quad 0 \leq C < T; \\ Y = C, \quad A = 2 \quad if \quad 0 \leq T \leq C \quad \text{and} \quad \Delta = 0. \end{cases}$$

We can also write

$$Y = \min(T, C) + (1 - \Delta)\max(C - T, 0) = C + \Delta \min(T - C, 0)$$

and $A = 2(1 - \Delta)\mathbf{1}_{\{T \leq C\}} + \mathbf{1}_{\{C < T\}}$, where $\mathbf{1}_A$ denotes the indicator function of the set A. With this censoring mechanism the lifetime T is observed, right censored or left censored. In view of the definitions of Y and A, note that if Δ is constant and equal to one (resp. zero), we obtain right censored (resp. current status) data. See, e.g., Huang and Wellner (1997) for a review on estimation with current status data.

Let F_T and F_C denote the distributions of T and C, respectively. Let $p = P(\Delta = 1)$. Define the observed subdistributions of Y as

$$H_k(B) = P(Y \in B,\ A = k), \qquad k = 0, 1, 2, \tag{2.1}$$

for any B Borel subset of $[0, \infty]$. As usually in survival analysis, the censoring mechanism defines a map Φ between the distributions of the latent variables and the observed distributions. For the censoring mechanism we consider, the relationship $(H_0, H_1, H_2) = \Phi(F_T, F_C, p)$ between the subdistributions of Y and the distributions of the latent variables T, C and Δ is the following:

$$\begin{cases} H_0(dt) = p\, F_C([t,\infty])\, F_T(dt) \\ H_1(dt) = F_T((t,\infty])\, F_C(dt) \\ H_2(dt) = (1-p)\, F_T([0,t])\, F_C(dt) \end{cases} . \tag{2.2}$$

Remark that when $p = 1$ (resp. $p = 0$) the equations (2.2) boil down to the equations of the classical independent right-censoring (resp. current status) model.

By plug-in applied with the empirical distribution, the nonparametric estimation of the distribution of T is straight as soon as the map Φ is invertible and F_T can be written as an explicit function of the observed subdistributions H_k, $k = 0, 1, 2$. The model considered allows us an explicit inversion formula for F_T. In order to derive this inversion formula, integrate the first and the second equation in (2.2) on $[t,\infty]$ and deduce

$$H_0([t,\infty]) + pH_1([t,\infty]) = pF_T([t,\infty])\, F_C([t,\infty]) . \tag{2.3}$$

For $t = 0$, it follows that

$$p = \frac{H_0([0,\infty])}{1 - H_1([0,\infty])} = \frac{H_0([0,\infty])}{H_0([0,\infty]) + H_2([0,\infty])}. \tag{2.4}$$

Recall that the hazard measure associated to a distribution F is $\Lambda(dt) = F(dt)/F([t,\infty])$. In our case, use (2.2)-(2.3) to deduce that the hazard function corresponding to F_T can be written as

$$\Lambda_T(dt) = \frac{H_0(dt)}{H_0([t,\infty]) + pH_1([t,\infty])}. \tag{2.5}$$

Finally, the distribution F_T can be expressed as

$$F_T((t,\infty]) = \prod_{[0,t]}(1 - \Lambda_T(ds)), \tag{2.6}$$

where $\prod$ is the product-integral (e.g., Gill and Johansen, 1990). Note that there is no explicit formula for F_T if $p = 0$ in equations (2.2), that is with current status data.

The inversion formula above applies only for $t \in I = \{t : H_0([t,\infty]) + pH_1([t,\infty]) > 0\}$. Obviously, we can have no information from data about $F_T((t,\infty])$ for t outside the interval I, unless $F_T(I) = 1$ in which case there is nothing else to know. If $F_T(I) < 1$, we make F_T a distribution on $[0,\infty]$ by considering that F_T has a supplementary mass $1 - F_T(I)$ at infinity.

Given the explicit relationship between the distribution of T and the observed subdistributions, to obtain the product-limit estimator of F_T, we simply replace H_k, $k = 0, 1, 2$ by their empirical counterparts. Consider a sample $\{(Y_i, A_i) : 1 \le i \le n\}$ and let $Z_1 \le ... \le Z_J$ be the distinct values in increasing order of Y_i. For any $j = 1, ..., J$ and $k = 0, 1, 2$, define

$$D_{kj} = \sum_{1\le i\le n} \mathbf{1}_{\{Y_i = Z_j,\ A_i = k\}} \qquad \text{and} \qquad \overline{N}_{kj} = \sum_{1\le i\le n} \mathbf{1}_{\{Y_i \ge Z_j,\ A_i = k\}} = \sum_{j\le l\le J} D_{kl}.$$

In view of (2.4), the estimator of p is

$$\widehat{p} = \frac{\overline{N}_{01}}{\overline{N}_{01} + \overline{N}_{21}},$$

while the estimator of the hazard measure is

$$\widehat{\Lambda}_T\left([0,t]\right) = \sum_{j:\, Z_j \leq t} \frac{D_{0j}}{\overline{N}_{0j} + \widehat{p}\,\overline{N}_{1j}}.$$

Finally, the product-limit estimator of F_T is a discrete (possibly sub)distribution $\widehat{F}_T$ with the mass concentrated at the points $Z_1 \leq ... \leq Z_J$ and such that

$$\widehat{F}_T((Z_j,\infty]) = \prod_{1\leq l \leq j} \{1 - \frac{D_{0l}}{\overline{N}_{0l} + \widehat{p}\,\overline{N}_{1l}}\}, \qquad 1 \leq j \leq J. \tag{2.7}$$

When $\overline{N}_{21} = 0$, $\widehat{F}_T$ is the Kaplan-Meier estimator for right-censored observations.

2.2 Model 2

As in Model 1, assume that T, C and Δ are independent. The observations are independent copies of the variables (Y, A), with $Y \geq 0$ and $A \in \{0,1,2\}$ where

$$\begin{cases} Y = T, & A = 0 \quad if \quad 0 \leq C \leq T \quad \text{and} \quad \Delta = 1; \\ Y = C, & A = 1 \quad if \quad 0 \leq C \leq T \quad \text{and} \quad \Delta = 0; \\ Y = C, & A = 2 \quad if \quad 0 \leq T < C. \end{cases}$$

We can also write $Y = \max(T,C) + (1-\Delta)\min(C-T,0) = C + \Delta\max(T-C,0)$. The equations of this model are

$$\begin{cases} H_0(dt) = p\,F_C\left([0,t]\right) F_T(dt) \\ H_1(dt) = (1-p)\,F_T\left([t,\infty]\right) F_C(dt) \\ H_2(dt) = F_T\left([0,t)\right) F_C(dt) \end{cases}. \tag{2.8}$$

Remark that when $p = 1$ (resp. $p = 0$) the equations (2.8) boil down to the equations of the classical independent left-censoring (resp. current status) model. This model also allows for an explicit inversion formula for F_T. By integration in the first and the third equation in (2.8), $H_0([0,t]) + pH_2([0,t]) = pF_T([0,t])F_C([0,t])$. Deduce

$$p = \frac{H_0([0,\infty])}{1 - H_2([0,\infty])}.$$

Recall that given a distribution F, the associated reverse hazard measure is $M(dt) = F(dt)/F([0,t])$. By equations (2.8) deduce that the reverse hazard function M_T associated to F_T can be written as

$$M_T(dt) = \frac{H_0(dt)}{H_0([0,t]) + pH_2([0,t])}.$$

Finally, the distribution F_T can be expressed as

$$F_T([0,t]) = \prod_{(t,\infty]} (1 - M_T(ds)).$$

The inversion formula applies on the interval $\{t : H_0([0,t]) + pH_2([0,t]) > 0\}$. Applying the inversion formula with the empirical subdistributions, we get the product-limit estimator of F_T. The details are omitted.

Note that if $\widetilde{T} = h(T)$ and $\widetilde{C} = h(C)$, with $h \geq 0$ a decreasing transformation, then $\widetilde{T}$, $\widetilde{C}$ and Δ are the variables of Model 1 applied to the left or right censored lifetime $h(Y)$. In other words, Model 2 is equivalent to Model 1, up to a time reversal transformation.

2.3 Extensions and related models

Model 1 can be easily extended in the following way: consider that $T = \min(T_a, T_b)$, with T_a (resp. T_b) the failure time due to cause a (resp. b). Assume that T_a and T_b are independent and independent of C and Δ. For simplicity, we only consider two failure causes, the extension to $k > 2$ competing failure causes being straight. Assume that if $T > C$, one only observes C and one knows that T is greater. When $T \leq C$ there are two cases: either C is observed and one only knows that the failure time T is less or equal to C, or T is observed and in this case one knows if the failure cause is a or b. The equations of the extended model are

$$\begin{cases} H_{0a}(dt) = p\, F_C([t,\infty])\, F_{T_b}([t,\infty])\, F_{T_a}(dt) \\ H_{0b}(dt) = p\, F_C([t,\infty])\, F_{T_a}((t,\infty])\, F_{T_b}(dt) \\ H_1(dt) = F_T((t,\infty])\, F_C(dt) \\ H_2(dt) = (1-p)\, F_T([0,t])\, F_C(dt) \end{cases},$$

where H_{0a} (resp. H_{0b}) is the subdistribution of the uncensored observations for which the failure cause is a (resp. b). If H_0 denotes the subdistribution of the uncensored observations, we have $H_0(dt) = p\, F_C([t,\infty])\, F_T(dt)$. Deduce that p can be expressed as in (2.4). Moreover, (2.3) holds. Consequently, if $\Lambda_{T_a}(dt) = F_{T_a}(dt)/F_{T_a}([t,\infty])$ is the hazard measure for T_a, we obtain

$$\Lambda_{T_a}(dt) = \frac{p\, F_C([t,\infty])\, F_{T_b}([t,\infty])\, F_{T_a}(dt)}{p\, F_C([t,\infty])\, F_T([t,\infty])} = \frac{H_{0a}(dt)}{H_0([t,\infty]) + pH_1([t,\infty])} \tag{2.9}$$

from which we deduce the expression of F_{T_a}. Model 2 can be extended in a similar way by considering $T = \max(T_a, T_b)$, with T_a and T_b the independent failure times corresponding to the complementary causes a and b, respectively.

Let us end this section with some comments on related models. Huang (1999) introduced a model for the so-called partly interval-censored data, Case 1; see also Kim (2003). In such data, for some subjects, the exact failure time of interest T is observed. For the remaining subjects, only the information on their current status at the examination time is available. Huang (1999) considered the nonparametric maximum likelihood estimator (NPMLE) of F_T. Unfortunately, NPMLE does not have an explicit form and therefore Huang needs strong assumptions for deriving its asymptotic properties and a numerical algorithm for the applications. Let us point out that, on contrary to our Model 1 (resp. Model 2), in Huang's model one may observe exact failure times even if failure occurs after (resp. before) the examination time. Moreover, in Huang's model one may still obtain a $\sqrt{n}-$consistent estimator of the distribution F_T if one simply considers the empirical distribution of the uncensored lifetimes. This is no longer true in our models.

Perhaps, the most popular model for left or right-censored data is the one introduced by Turnbull (1974); see also Gu and Zhang (1993). In Turnbull's model there are three latent lifetimes L (left-censoring), T (lifetime of interest) and R (right-censoring) with $L \leq R$. The observed variables are $Y = \max(L, \min(T, R)) = \min(\max(L, T), R)$ and A defines as follows: $A = 0$ if $L < T \leq R$; $A = 1$ if $R < T$; and $A = 2$ if $T \leq L$. The equations of this model are

$$\begin{cases} H_0(dt) = \{F_R([t,\infty]) - F_L([t,\infty])\}\ F_T(dt) \\ H_1(dt) = F_T((t,\infty])\ F_R(dt) \\ H_2(dt) = F_T([0,t])\ F_L(dt) \end{cases},$$

where H_k, $k = 0, 1, 2$ are defined as in (2.1) and F_T, F_L and F_R are the distributions of T, L and R, respectively. The NPMLE of the distribution of the failure time T is not explicit but it can be computed, for instance, by iterations based on the so-called self-consistency equation. Note that imposing $F_C(dt) = (1-p)^{-1}F_L(dt) = F_R(dt)$, one recovers the equations of Model 1. However, for the applications we have in mind, there is no natural interpretation for such a constraint in Turnbull's model. Moreover, we derive a product-limit estimator for our Model 1. Finally, the asymptotic results below (strong consistency, asymptotic normality and bootstrap consistency) are much simpler and they are obtained under weaker conditions than in Turnbull's model (see Gu and Zhang, 1993 and Wellner and Zhan, 1996).

3 Asymptotic results

In this section the strong uniform convergence and the asymptotic normality for the estimator of the distribution F_T in Model 1 are derived. Moreover, we propose a bootstrap procedure that can be used to build confidence intervals for F_T. As in the previous sections, the distributions F_T and F_C need not be continuous. For simpler notation, hereafter, the subscript T is suppressed when there is no possible confusion. We write $\widehat{F}$ (resp. F, $\widehat{\Lambda}$ and Λ) instead of $\widehat{F}_T$ (resp. F_T, $\widehat{\Lambda}_T$ and Λ_T).

3.1 Strong uniform convergence

Let H_{nk} be the empirical counterparts of the subdistributions H_k, $k = 0, 1, 2$, that is

$$H_{nk}([0,t]) = n^{-1} \sum_{i=1}^{n} \mathbf{1}_{\{Y_i \leq t,\ A_i = k\}}, \qquad k = 0, 1, 2.$$

It is known from empirical process theory that $\sup_{t\geq 0} |H_{nk}([0,t]) - H_k([0,t])| \to 0$, almost surely. We want to prove the strong uniform convergence of the distribution $\widehat{F}$, that is

$$\sup_{t\in I} \left| \widehat{F}([0,t]) - F([0,t]) \right| \to 0, \quad \text{as } n \to \infty, \quad \text{almost surely},$$

where $I = \{t : H_0([t,\infty]) + pH_1([t,\infty]) > 0\}$. First, the almost sure convergence of the hazard function is obtained.

Theorem 3.1 *Assume that $p \in (0,1]$ and let $t_* = \sup I$. For any $\sigma \in I$,*

$$\sup_{0\le t\le \sigma} \left| \widehat{\Lambda}\left([0,t]\right) - \Lambda\left([0,t]\right) \right| \to 0, \qquad \textit{as } n\to\infty, \ \ \textit{almost surely.}$$

Moreover, if $t_* \notin I$ and $\Lambda\left([0,t_*)\right) < \infty$, then $\widehat{\Lambda}\left([0,t_*)\right) \to \Lambda\left([0,t_*)\right)$, almost surely.

To prove this theorem, first we obtain the result when p replaces $\widehat{p}$ in the definition of $\widehat{\Lambda}$. In this case the functionals of the hazard function are reverse supermartingales in n, as it is shown in the next lemma which extends a statement of Gill(1994).

Lemma 3.2 *Let $p \in (0,1]$ and $f \ge 0$ be a Borel-measurable function. Let*

$$\Lambda_{n,p}(f) = \int_I \frac{f(t)H_{n0}(dt)}{H_{n0}([t,\infty]) + pH_{n1}([t,\infty])}.$$

Define the $\sigma-$fields

$$\mathcal{F}_n = \sigma(H_{n0}, H_{n1}, H_{n2},) \qquad \textit{and} \qquad \mathcal{B}_n = \bigvee_{n\le m<\infty} \mathcal{F}_m.$$

Then, for all n,

$$E[\Lambda_{n,p}(f) \mid \mathcal{B}_{n+1}] \le \Lambda_{n+1,p}(f),$$

that is $\Lambda_{n,p}(f)$, $n \ge 1$ is a positive reverse supermartingale.

Proof of Lemma 3.2. For simplicity, in this proof, let us write Λ_n instead of $\Lambda_{n,p}$. Define $N_n(t) = nH_{n0}([t,\infty]) + pnH_{n1}([t,\infty])$. Then we can write

$$\Lambda_n(f) = \sum_{1\le i\le n} f(Y_i)\mathbf{1}_{\{A_i=0\}} N_n(Y_i)^{-1}.$$

Next, notice that

$$\mathcal{B}_{n+1} = \mathcal{F}_{n+1} \vee \bigvee_{n+2\le m<\infty} \sigma(Y_m, A_m)$$

and therefore, by the i.i.d. property of the sample and elementary properties of the conditional independence (see, e.g., Florens, Mouchart and Rolin, 1990), we have

$$E[\Lambda_n(f) \mid \mathcal{B}_{n+1}] = nE[f(Y_1)\mathbf{1}_{\{A_1=0\}} N_n(Y_1)^{-1} \mid H_{n+1,0}, H_{n+1,1}, H_{n+1,2}].$$

The $\sigma-$field generated by $H_{n+1,0}$, $H_{n+1,1}$ and $H_{n+1,2}$ is the sub$-\sigma-$field of permutable events in the $\sigma-$field generated by $\{(Y_i, A_i) : 1 \le i \le n+1\}$. Hence

$$E[\Lambda_n(f) \mid \mathcal{B}_{n+1}] = \frac{n}{(n+1)!} \sum_{\tau\in P_{n+1}} f(Y_{\tau(1)})\mathbf{1}_{\{A_{\tau(1)}=0\}} N^{\tau}_{n+1}(Y_{\tau(1)})^{-1},$$

where P_{n+1} is the set of permutations of $n+1$ elements and

$$N^{\tau}_{n+1}(Y_{\tau(1)}) = N_{n+1}(Y_{\tau(1)}) - \mathbf{1}_{\{Y_{\tau(n+1)}\ge Y_{\tau(1)},\, A_{\tau(n+1)}=0\}} - p\mathbf{1}_{\{Y_{\tau(n+1)}\ge Y_{\tau(1)},\, A_{\tau(n+1)}=1\}}.$$

(By definition, $0/0 = 0$.) There are $(n-1)!$ permutations such that $\tau(1) = i$ and $\tau(n+1) = j$ and therefore,

$$E[\Lambda_n(f) \mid \mathcal{B}_{n+1}] = \frac{1}{n+1} \sum_{1 \le i \le n+1} f(Y_i) \mathbf{1}_{\{A_i = 0\}} \times \sum_{1 \le j \ne i \le n+1} \left[N_{n+1}(Y_i) - \mathbf{1}_{\{Y_j \ge Y_i,\, A_j = 0\}} - p \mathbf{1}_{\{Y_j \ge Y_i,\, A_j = 1\}} \right]^{-1}. \tag{3.1}$$

Now,

$$\begin{aligned} &\sum_{1 \le j \ne i \le n+1} \left[N_{n+1}(Y_i) - \mathbf{1}_{\{Y_j \ge Y_i,\, A_j = 0\}} - p \mathbf{1}_{\{Y_j \ge Y_i,\, A_j = 1\}} \right]^{-1} \\ &\quad = \sum_{1 \le j \ne i \le n+1} \left[\frac{\mathbf{1}_{\{Y_j < Y_i\}} + \mathbf{1}_{\{Y_j \ge Y_i, A_j = 2\}}}{N_{n+1}(Y_i)} + \frac{\mathbf{1}_{\{Y_j \ge Y_i, A_j = 0\}}}{N_{n+1}(Y_i) - 1} + \frac{\mathbf{1}_{\{Y_j \ge Y_i, A_j = 1\}}}{N_{n+1}(Y_i) - p} \right] \\ &\quad = \frac{n+1}{N_{n+1}(Y_i)} - R_i, \end{aligned}$$

where

$$\begin{aligned} R_i &= \frac{1}{N_{n+1}(Y_i)} \\ &\quad + \sum_{1 \le j \ne i \le n+1} \left[\frac{\mathbf{1}_{\{Y_j \ge Y_i, A_j = 0\}} + \mathbf{1}_{\{Y_j \ge Y_i, A_j = 1\}}}{N_{n+1}(Y_i)} - \frac{\mathbf{1}_{\{Y_j \ge Y_i, A_j = 0\}}}{N_{n+1}(Y_i) - 1} - \frac{\mathbf{1}_{\{Y_j \ge Y_i, A_j = 1\}}}{N_{n+1}(Y_i) - p} \right]. \end{aligned}$$

Use the inequality

$$\frac{a+b+1}{a+1+pb} \ge \frac{a}{a+pb} + \frac{b}{a+1+pb-p}$$

with $a = \sum_{j \ne i} \mathbf{1}_{\{Y_j \ge Y_i, A_j = 0\}}$ and $b = \sum_{j \ne i} \mathbf{1}_{\{Y_j \ge Y_i, A_j = 1\}}$ and deduce $R_i \ge 0$. (Notice that $N_{n+1}(Y_i) = a + 1 + bp$ because the observations Y_i involved in equation (3.1) are such that $A_i = 0$.) Therefore,

$$E[\Lambda_n(f) \mid \mathcal{B}_{n+1}] \le \sum_{1 \le i \le n+1} f(Y_i) \mathbf{1}_{\{A_i = 0\}} N_{n+1}(Y_i)^{-1} = \Lambda_{n+1}(f),$$

that is $\Lambda_n(f)$, $n \ge 1$ is a reverse supermartingale. ■

Proof of Theorem 3.1. The strong uniform convergence of $\widehat{\Lambda}([0,t])$ when $t \in [0,\tau] \subset I$ can be obtained by delta-method (cf. Gill (1989, 1994); see also the proof of Theorem 3.6 below) from the almost sure uniform convergence of $H_{nk}([0,t])$, $k = 0,1,2$. For the last part of the theorem, denote with $\overline{H}_0\,(t-)$ and $\overline{H}_1\,(t-)$ the quantities $H_0([t,\infty])$ and $H_1([t,\infty])$, respectively. Let $\overline{H}_{n0}\,(t-)$ and $\overline{H}_{n1}\,(t-)$ be the empirical counterparts of $\overline{H}_0\,(t-)$ and $\overline{H}_1\,(t-)$, respectively. Fix some $\tau < t_*$ and write

$$\begin{aligned} \left| \widehat{\Lambda}((\tau,t_*)) - \Lambda((\tau,t_*)) \right| &\le \left| \int_{(\tau,t_*)} \frac{H_{n0}(dt)}{(\overline{H}_{n0} + \widehat{p}\,\overline{H}_{n1})(t-)} - \int_{(\tau,t_*)} \frac{H_{n0}(dt)}{(\overline{H}_{n0} + p\,\overline{H}_{n1})(t-)} \right| \\ &\quad + \left| \int_{(\tau,t_*)} \frac{H_{n0}(dt)}{(\overline{H}_{n0} + p\,\overline{H}_{n1})(t-)} - \int_{(\tau,t_*)} \frac{H_0(dt)}{(\overline{H}_0 + p\,\overline{H}_1)\,(t-)} \right| \\ &= : A_1 + A_2. \end{aligned}$$

By little algebra,

$$A_1 \leq \frac{|\widehat{p}-p|}{\widehat{p}} \int_{(\tau,t_*)} \frac{H_{n0}(dt)}{\overline{H}_{n0}(t-)+p\,\overline{H}_{n1}(t-)} = \frac{|\widehat{p}-p|}{\widehat{p}} \Lambda_{n,p}((\tau,t_*)),$$

Since $\widehat{p} \to p$, almost surely, we obtain $A_1 \to 0$, almost surely, given that $\Lambda_{n,p}((\tau,t_*))$ converges almost surely to a finite constant. Use Lemma 3.2 with $f = \mathbf{1}_{(\tau,t_*)}$ to deduce that $\Lambda_{n,p}((\tau,t_*))$ is a reverse martingale. Now, by Doob's supermartingale convergence theorem, as soon as $\sup_n E[\Lambda_{n,p}((\tau,t_*))]$ is finite, the functional $\Lambda_{n,p}((\tau,t_*))$ converges almost surely to some integrable limit. It is not difficult to see that the limit is in the $\sigma-$field of asymptotic permutable events and is therefore a constant by the Hewitt-Savage 0-1 law. The constant is equal to $\sup_n E[\Lambda_{n,p}((\tau,t_*))]$ and thus $\Lambda_{n,p}((\tau,t_*))$ converges almost surely and in expectation to this quantity. Consequently, to obtain $A_1 \to 0$, almost surely, it remains to bound the sequence $E[\Lambda_{n,p}((\tau,t_*))]$, $n \geq 1$. Note that $\Lambda_{n,p}((\tau,t_*)) \leq p^{-1}\Lambda_{n,1}((\tau,t_*))$. It can be shown (see Lemma A.1 in the appendix) that

$$E\left(\Lambda_{n,1}((\tau,t_*))\right) = \int_{(\tau,t_*)} P\left\{\overline{H}_{n0}(u-) + \overline{H}_{n1}(u-) > 0\right\} \Lambda_1(du) < \Lambda_1((\tau,t_*)),$$

where $\Lambda_1(du) = \left\{\overline{H}_0(u-) + \overline{H}_1(u-)\right\}^{-1} H_0(du)$. Since $\Lambda_1(du) \leq \Lambda(du)$, deduce that, for any τ, $\sup_n E[\Lambda_{n,p}((\tau,t_*))] \leq p^{-1}\Lambda([0,t_*)) < \infty$. Hence, $A_1 \to 0$, almost surely, uniformly in τ. Concerning A_2, note that $A_2 \leq \Lambda_{n,p}((\tau,t_*)) + \Lambda((\tau,t_*))$ and $\Lambda_{n,p}((\tau,t_*))$ converges to a constant smaller than $p^{-1}\Lambda_1((\tau,t_*))$, almost surely. Since $\Lambda_1((\tau,t_*)) \downarrow 0$ as $\tau \uparrow t_*$, deduce that $\widehat{\Lambda}([0,t_*)) \to \Lambda([0,t_*))$, almost surely. ■

The strong uniform convergence of the distribution $\widehat{F}$ follows without any additional assumption.

Theorem 3.3 *Assume that $p \in (0,1]$. Then*

$$\sup_{t\in I}\left|\widehat{F}([0,t]) - F([0,t])\right| \to 0, \qquad \text{as} \quad n\to\infty, \quad \text{almost surely.}$$

Proof. ¿From the first part of Theorem 3.1 and using delta-method, for any $\tau \in I$,

$$\sup_{t\in[0,\tau]}\left|\widehat{F}([0,t]) - F([0,t])\right| \to 0, \quad \text{as } n\to\infty, \quad \text{almost surely.} \tag{3.2}$$

Recall that $t_* = \sup I$. To complete the proof, follow Gill (1994) and distinguish three cases: a) $I = [0,t_*]$ (hence, $t_* < \infty$); b) $I = [0,t_*)$ and $F([t_*,\infty]) = 0$; and c) $I = [0,t_*)$ and $F([t_*,\infty]) > 0$. In case a) there nothing more to prove. In case b),

$$\sup_{t\in I}\left|\widehat{F}([0,t]) - F([0,t])\right| \leq F((\tau,\infty]) + \sup_{t\in[0,\tau]}\left|\widehat{F}([0,t]) - F([0,t])\right|.$$

Since $F((\tau,\infty]) \downarrow 0$ as $\tau \uparrow t_*$, we deduce the strong uniform convergence $\widehat{F}$ in case b). Finally, in case c)

$$\sup_{t\in I}\left|\widehat{F}([0,t]) - F([0,t])\right| \leq \widehat{F}((\tau,t_*)) + F((\tau,t_*)) + \sup_{t\in[0,\tau]}\left|\widehat{F}([0,t]) - F([0,t])\right|.$$

Since $F((\tau, t_*)) \to 0$ as $\tau \uparrow t_*$, it remains to bound $\widehat{F}((\tau, t_*))$ when $F([t_*, \infty]) > 0$. By definition, $\Lambda(dt) = F(dt)/F([t, \infty])$ and $\widehat{\Lambda}(dt) = \widehat{F}(dt)/\widehat{F}([t, \infty])$. Deduce that $\Lambda([0, t_*)) < \infty$ and $\Lambda((\tau, t_*)) \downarrow 0$ as $\tau \uparrow t_*$. On the other hand, deduce that the measure $\widehat{\Lambda}(dt)$ is greater than the measure $\widehat{F}(dt)$. We have

$$\sup_{t \in I} \left| \widehat{F}([0,t]) - F([0,t]) \right| \leq \widehat{\Lambda}((\tau, t_*)) + F((\tau, t_*)) + \sup_{t \in [0,\tau]} \left| \widehat{F}([0,t]) - F([0,t]) \right|.$$

Use Lemma 3.2 and deduce that, for any $\tau \in I$, the left-hand side in the last display converges to $\Lambda((\tau, t_*)) + F((\tau, t_*))$, almost surely. Since $\Lambda((\tau, t_*)) + F((\tau, t_*)) \downarrow 0$ as $\tau \uparrow t_*$, the proof is complete. ■

Remark 1. The proofs of Lemma 3.2 and Theorem 3.1 apply also to the extension of Model 1 considered in subsection 2.3. Deduce the strong uniform convergence on I of the nonparametric estimator of the distribution F_{T_a}.

Remark 2. With $p = 1$ one recovers the strong uniform convergence result for the Kaplan-Meier estimator obtained by Stute and Wang (1993); see also Gill (1994). Our alternative proof is simpler, especially the arguments used for Lemma 3.2.

3.2 Asymptotic normality

Let us study the weak convergence of the process $\sqrt{n}(\widehat{F} - F)$ where $\widehat{F}$ is the product-limit estimator of Model 1. In this case, $\widehat{\Lambda}$ does no longer have a martingale structure (in t) as in the case of the Nelson-Aalen estimator, that is when $p = 1$. However, a continuous time *submartingale* property for $\Lambda_{n,p}$ defined in Lemma 3.2 can be obtained. (Recall that $\Lambda_{n,p}$ is defined as $\widehat{\Lambda}$ but with $\widehat{p}$ replaced by p.) This suffices us to extend the techniques of Gill (1983) and to use them in combination with the functional delta-method in order to establish the weak convergence of $\sqrt{n}(\widehat{F} - F)$ to a Gaussian process. Weak convergence is denoted by $\rightsquigarrow$ and it is in the sense considered by Pollard (1984), that is $D[a, b]$ the space of càdlàg functions on $[a, b]$ is endowed with the supremum norm and the ball σ−field.

Let us introduce the right-continuous filtration

$$\mathcal{G}_t = \bigvee_{i=1}^{n} \bigvee_{k=0}^{2} \sigma\left(\mathbf{1}_{\{A_i = k,\, Y_i \leq s\}}; 0 \leq s \leq t\right), \quad 0 \leq t \leq \infty, \tag{3.3}$$

on the probability space $(\Omega, \mathcal{F}, P)$ where the random variables are defined. Consider

$$N(t) = \sum_{i=1}^{n} \mathbf{1}_{\{Y_i \leq t,\, A_i = 0\}}, \qquad t \geq 0,$$

a counting process adapted to the filtration $(\mathcal{G}_t)$. The compensator of N is defined by

$$A(dt) = E[N(dt) \mid \mathcal{G}_{t-}],$$

see, e.g., Fleming and Harrington (1991), chapter 1. Simple computations give

$$A(dt) = \overline{Y}(t-) \frac{P(A = 0,\, Y \in dt)}{P(Y \geq t)},$$

where $\overline{Y}(t-) = \sum_{i=1}^{n} \mathbf{1}_{\{Y_i \geq t\}}$. This together with standard arguments give us:

Lemma 3.4 *Let $\overline{H}(t-) = P(Y \geq t)$, $t \geq 0$. The stochastic process M defined by*

$$M(t) = N(t) - \int_{[0,t]} \frac{\overline{Y}(s-)}{\overline{H}(s-)} H_0(ds)$$

is a squared integrable, zero-mean $\mathcal{G}_t$-martingale on $[0, \infty]$. The associated predictable quadratic variation process is

$$\langle M \rangle(t) = \int_{[0,t]} \left(1 - \frac{\Delta H_0(s)}{\overline{H}(s-)}\right) \frac{\overline{Y}(s-)}{\overline{H}(s-)} H_0(ds),$$

where $\Delta H_0(s) = H_0(\{s\})$.

Recall that $\overline{H}_k(t-) = H_k([t, \infty])$, $k = 0, 1, 2$. Define the hazard function

$$\Lambda_p^*([0,t]) = \int_{[0,t]} \frac{\mathbf{1}_{\{\overline{Y}(s-)>0\}} \{1 - [\overline{H}_2(s-)/\overline{H}(s-)]^{\overline{Y}(s-)}\}^2}{\overline{H}_0(s-) + p\,\overline{H}_1(s-)} H_0(ds), \qquad t \geq 0.$$

Note that $\Lambda_p^*(dt) = 0$ when $t > Y_{max}$, where $Y_{max} = \max_i Y_i$. The proof of the following lemma is given in the appendix.

Lemma 3.5 *Consider the filtration $(\mathcal{G}_t)$ defined in (3.3). Let $X(t)$, $t \geq 0$ be a nonnegative, bounded, $\mathcal{G}_t$-predictable process. Then, the process*

$$W(t) = \int_{[0,t]} X(s) \left(\Lambda_{n,p}(ds) - \Lambda_p^*(ds)\right), \qquad t \geq 0,$$

is a $\mathcal{G}_t$-submartingale and $E[W(t)] \geq 0$.

Define the stochastic process

$$Z_p(t) = \sqrt{n} \frac{F_{n,p}([0,t]) - F_p^*([0,t])}{F_p^*((t,\infty])}, \qquad t \geq 0, \tag{3.4}$$

where $F_{n,p}$ (resp. F_p^*) is the distribution corresponding to $\Lambda_{n,p}$ (resp. Λ_p^*). It is well-known that the process Z_p is a (squared integrable) martingale on $[0, \tau]$, for any $\tau < t_* = \sup I$, provided that $p = 1$ (see Gill, 1983). When $p \in (0,1)$, the process Z_p is a submartingale on $[0, \tau]$, $\forall \tau < t_*$. This is implied by the identity

$$\frac{F_{n,p}([0,t]) - F_p^*([0,t])}{F_p^*((t,\infty])} = \int_{[0,t]} \frac{F_{n,p}([s,\infty])}{F_p^*((s,\infty])} \left(\Lambda_{n,p}(ds) - \Lambda_p^*(ds)\right)$$

(which is a direct consequence of Duhamel's equation; see Gill, 1994) and Lemma 3.5.

To prove the asymptotic normality of $\widehat{F}$, notice that, by empirical central limit theorem (e.g., van de Vaart and Wellner, 1996),

$$\sqrt{n} \left\{ \left(\overline{H}_{0n}, \overline{H}_{1n}, \widehat{p}\right) - \left(\overline{H}_0, \overline{H}_1, p\right) \right\} \rightsquigarrow \left(\mathbb{G}_0, \mathbb{G}_1, N\right) \tag{3.5}$$

in $(D[0,\infty])^2 \times \mathbb{R}$, where $\overline{H}_{kn}$ is the càdlàg process $\overline{H}_{nk}(t) = H_{nk}((t,\infty])$, $k = 0,1,2$. The process $(\overline{\mathbb{G}}_0, \overline{\mathbb{G}}_1)$ is a tight, zero mean Gaussian process and, for any $t, s \geq 0$, the vector $(\overline{\mathbb{G}}_0(t), \overline{\mathbb{G}}_1(t), N)$ has a zero-mean multivariate normal distribution. Moreover,

$$E\left\{\overline{\mathbb{G}}_0(t)\,\overline{\mathbb{G}}_0(s)\right\} = \overline{H}_0(t \vee s) - \overline{H}_0(t)\,\overline{H}_0(s), \tag{3.6}$$

$$E\left\{\overline{\mathbb{G}}_0(t)\,\overline{\mathbb{G}}_1(s)\right\} = -\overline{H}_0(t)\,\overline{H}_1(s), \qquad E\left\{\overline{\mathbb{G}}_1(t)\,\overline{\mathbb{G}}_1(s)\right\} = \overline{H}_1(t \vee s) - \overline{H}_1(t)\,\overline{H}_1(s)$$

and

$$E\left\{\overline{\mathbb{G}}_0(t)\,N\right\} = \frac{\overline{H}_0(t)(1-p)}{H_{02}([0,\infty])}, \qquad E\left\{\overline{\mathbb{G}}_1(t)\,N\right\} = 0, \qquad E\left\{N^2\right\} = \frac{p(1-p)}{H_{02}([0,\infty])},$$

where $H_{02} = H_0 + H_2$ and $t \vee s = \max(t,s)$.

Theorem 3.6 *Assume that $p \in (0,1]$ and define $U(t) = \sqrt{n}(\widehat{F}([0,t]) - F([0,t]))$, $t \geq 0$.*
a) Let τ be a point in I. Then, $U \rightsquigarrow \mathbb{G}$ in $D[0,\tau]$, where $\mathbb{G}$ is the Gaussian process

$$\mathbb{G}(t) = -F((t,\infty])\left\{\int_{[0,t]} \frac{d\overline{\mathbb{G}}_0(s)}{\overline{H}_0(s) + p\overline{H}_1(s)} + \int_{[0,t]} \frac{\overline{\mathbb{G}}_2(s-)}{\overline{H}_0(s-) + p\overline{H}_1(s-)} d\Lambda(s)\right\},$$

and $\overline{\mathbb{G}}_2 = \overline{\mathbb{G}}_0 + p\overline{\mathbb{G}}_1 + N\,\overline{H}_1$ with $\overline{\mathbb{G}}_0$ and $\overline{\mathbb{G}}_1$ the limit processes in (3.5). The first integral is defined by integration by parts.
b) If $t_* \notin I$, but

$$\int_{[0,t_*)} \frac{H_0(dt)}{\{\overline{H}_0(t-) + p\overline{H}_1(t-)\}^2} < \infty, \tag{3.7}$$

then $\mathbb{G}$ can be extended to a Gaussian process on $[0,t_*]$ and $U \rightsquigarrow \mathbb{G}$ in $D[0,t_*]$.

The proof of the weak convergence is postponed to the appendix. Note that when $t_* \notin I$, condition (3.7) is equivalent to

$$F_T([t_*,\infty]) > 0 \qquad \text{and} \qquad \int_{[0,t_*)} \frac{F_T(dt)}{F_C([t,\infty])} < \infty. \tag{3.8}$$

See Chen and Lo (1997, section 1) for a discussion on similar conditions in the case of the Kaplan-Meier estimator. Whether the weak convergence still holds when $p < 1$ and only the second part of (3.8) is satisfied remains an open question.

3.3 Bootstrapping the product-limit estimator

Theorem 3.6 may be used to obtain confidence intervals and confidence bands for F. However, the law of the process $\mathbb{G}(t)/F((t,\infty])$ being complicated, one may prefer a bootstrap method in order to avoid handling this process in practical applications. Here, a bootstrap sample is obtained by simple random sample with replacement from the set of observations. Let H_k^*, $k = 0,1,2$ denote the bootstrap versions of the observed subdistributions. Apply equations (2.4) to (2.6) to obtain the bootstrap estimator $\widehat{F}^*$. The following theorem state that the bootstrap works almost surely for our product-limit estimator on any interval $[0,\tau]$ such that $H_0([\tau,\infty]) + pH_1([\tau,\infty]) > 0$. This result, for which the proof is skipped, is a simple corollary of Theorem 3.9.13 of van der Vaart and Wellner (1996) (see also Theorem 4 of Gill, 1989) and it is based on the uniform Hadamard differentiability of the maps involved in the inversion formula of Model 1.

Theorem 3.7 *Let $\tau \in I$ and let $\widetilde{\mathbb{G}}(t)$ be the limit of $\sqrt{n}\{\widehat{F}([0,t]) - F([0,t])\}/F((t,\infty])$ in $D[0,\tau]$, as obtained from Theorem 3.6. Then, the process*

$$\sqrt{n}\{\widehat{F}^*([0,t]) - \widehat{F}([0,t])\}/\widehat{F}((t,\infty])$$

converges to $\widetilde{\mathbb{G}}$ in $D[0,\tau]$, almost surely.

Using this result and the uniform convergence of $\widehat{F}$ to F, one can derive pointwise confidence intervals and confidence bands for F. For instance, a bootstrap $(1-\alpha)$ −level confidence interval for $F([0,t])$ can be defined as

$$\left[\widehat{F}([0,t]) - q^*_{1-\alpha/2}(t)\, n^{-1/2}\widehat{F}((t,\infty]), \quad \widehat{F}([0,t]) - q^*_{\alpha/2}(t)\, n^{-1/2}\widehat{F}((t,\infty])\right],$$

where $q^*_\alpha(t)$ is the bootstrap approximation of the α−quantile of the distribution of $\sqrt{n}\{\widehat{F}([0,t]) - F([0,t])\}/F((t,\infty])$.

4 Applications

Let us apply our first model to the California high school students data presented in Table 1. The data is part of a study conducted at Stanford-Palo Alto Peer Counselling Program; see Hamburg *et al.* (1975). In this study, 191 California high schools boys were asked "When did you first use marijuana?" The answers were the exact ages, "I have used it but I can not recall just when the first time was" and "I never used". The latent variable T is the age at the first use of marijuana.

Table 1. First use of marijuana: Z_j are the distinct observed values of the lifetimes Y_i and $D_{kj} = \sum_i \mathbf{1}_{\{Y_i = Z_j,\, A_i = k\}}$, $k = 0, 1, 2$.

Z_j	10	11	12	13	14	15	16	17	18	> 18
D_{0j} (uncensored)	4	12	19	24	20	13	3	1	0	4
D_{1j} (right-censored)	0	0	2	15	24	18	14	6	0	0
D_{2j} (left-censored)	0	0	0	1	2	3	2	3	1	0

Turnbull and Weiss (1978) (see also and Klein and Moeschberger, 1997, chapter 5) analyzed this sample using the doubly censorship model of Turnbull (1974). However, there is no natural interpretation for two censoring times, that is the left and right-censoring lifetimes L and R, with this data set. On contrary, Model 1 can be easily interpreted as follows: $\Delta = 1$ if the student recalls the value of T and $\Delta = 0$ otherwise; the variable C is the age of the student at the study time. Condition $F_T([t_*,\infty]) > 0$, see equation (3.7), means that some high schools boys will never use marijuana.

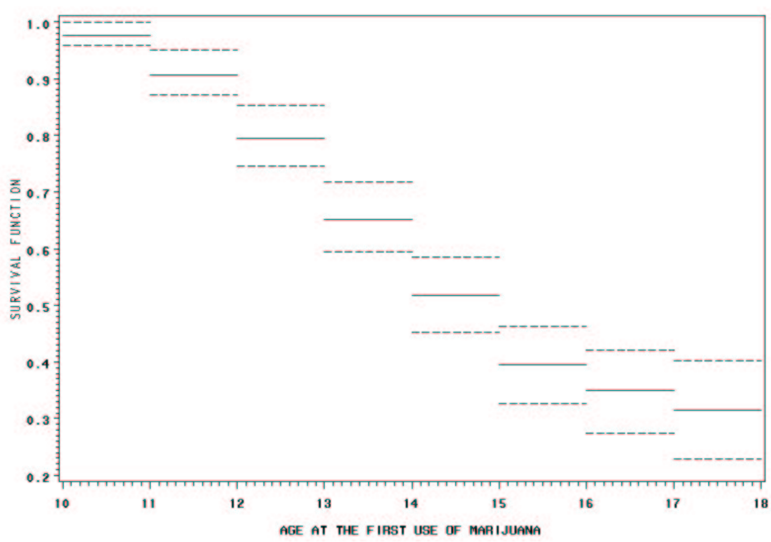
Table 2. Survival function estimates for the first use of marijuana.

Z_j	$\widehat{F}_T((Z_j,\infty])$	$F_n^{TW}((Z_j,\infty])$	$F_n^{KM}((Z_j,\infty])$
10	0.977	0.977	0.978
11	0.906	0.906	0.911
12	0.795	0.794	0.804
13	0.652	0.651	0.669
14	0.517	0.516	0.539
15	0.394	0.392	0.420
16	0.349	0.345	0.375
17	0.315	0.308	0.341
18	0.315	0.308	0.341
> 18	0.000	0.000	0.000

In Table 2 we report the estimates $\widehat{F}_T$ obtained from our Model 1. The value of $\widehat{p}$ is 0.893. Moreover, we provide the estimator F_n^{TW} obtained by iteration in the self-consistency equation of Turnbull's model, as reported in Klein and Moeschberger (1997), page 129. The Kaplan-Meier estimator F_n^{KM} based only on uncensored and right-censored observations is also presented. Note that F_n^{KM} is quite close to $\widehat{F}_T$ and F_n^{TW} and this is due to the small number of left-censored observations. The fact that the equations of Model 1 represent a special case of Turnbull's model equations explains the closeness between $\widehat{F}_T$ and F_n^{TW}.

Pointwise confidence intervals and confidence bands for the survival probability are difficult to obtain in Turnbull's model (see Wellner and Zhan, 1997, section 6). Such confidence regions are easily obtained by bootstrapping in our Model 1. In Figure 1 we provide the pointwise confidence intervals for the survival function estimated by the product-limit estimator $\widehat{F}_T$. A number of 5000 bootstrap samples were used.

Figure 1: Pointwise confidence intervals for the survival function for the age at the first use of marijuana



A Appendix

In this section, K denotes a positive constant, not necessarily the same at each appearance and possibly depending on p.

Lemma A.1 *Let $f \geq 0$ be a Borel-measurable function. Let $\overline{H}_{01}(t-) = H_0([t,\infty]) + H_1([t,\infty])$ and $\overline{H}_{n01}(t) = H_{n0}([t,\infty]) + H_{n1}([t,\infty])$. Define*

$$\Lambda_{n,1}(f) = \int_I \frac{f(t)}{\overline{H}_{n01}(t)} H_{n0}(dt) \qquad \text{and} \qquad \Lambda_1(f) = \int_I \frac{f(t)}{\overline{H}_{01}(t)} H_0(dt),$$

where $I = \{t : \overline{H}_{01}(t) > 0\}$. Then

$$E(\Lambda_{n,1}(f)) = \int_I f(t) P\left\{\overline{H}_{n01}(t-) > 0\right\} \Lambda_1(dt) = \int_I f(t) \left[1 - \left\{\overline{H}_{01}(t-)\right\}^n\right] \Lambda_1(dt).$$

Proof. Define the measures $N_n = n\,(H_{n0}+H_{n1})$ and $N_{0n} = nH_{n0}$. The empirical hazard measure $\Lambda_{n,1}$ may be written under the integral form

$$\Lambda_{n,1}\left((t,t+s]\right) = \int_{(t,t+s]} N_n\left([u,\infty]\right)^{-1} N_{0n}\left(du\right)$$

that can be approximated as follows. Let $t_{m,k} = t + (k/2^m)\, s$, with $0 \leq k \leq 2^m$, and

$$S_m\left(u\right) = \sum_{1\leq k\leq 2^m} \frac{\mathbf{1}_{\left\{N_n\left((t_{m,k-1},\infty]\right)>0\right\}}}{N_n\left((t_{m,k-1},\infty]\right)} \mathbf{1}_{(t_{m,k-1},\, t_{m,k}]}\left(u\right)$$

and note that

$$\lim_{m\to\infty} S_m\left(u\right) = \frac{\mathbf{1}_{\{N_n([u,\infty])>0\}}}{N_n\left([u,\infty]\right)} \mathbf{1}_{(t,\, t+s]}\left(u\right).$$

Since $S_m\left(u\right) \leq 1$, by dominated convergence theorem,

$$\begin{aligned}
\Lambda_{n,1}\left((t,t+s]\right) &= \lim_{m\to\infty} \int_{(t,t+s]} S_m\left(u\right) N_{0n}\left(du\right) \\
&= \lim_{m\to\infty} \sum_{1\leq k\leq 2^m} \frac{N_{0n}\left((t_{m,k-1},\, t_{m,k}]\right)}{N_n\left((t_{m,k-1},\infty]\right)} \mathbf{1}_{\left\{N_n\left((t_{m,k-1},\infty]\right)>0\right\}}.
\end{aligned}$$

On the other hand, define $A_n\left((t,t+s]\right) = \int_{(t,t+s]} \mathbf{1}_{\{N_n([u,\infty])>0\}} \Lambda_1\left(du\right)$ and

$$T_m\left(u\right) = \sum_{1\leq k\leq 2^m} \frac{\mathbf{1}_{\left\{N_n\left((t_{m,k-1},\infty]\right)>0\right\}}}{H_{01}\left((t_{m,k-1},\infty]\right)} \mathbf{1}_{(t_{m,k-1},\, t_{m,k}]}\left(u\right).$$

Clearly, $\lim_{m\to\infty} T_m\left(u\right) = \overline{H}_{01}^{\,-1}\left(u-\right) \mathbf{1}_{\{N_n([u,\infty])>0\}} \mathbf{1}_{(t,\, t+s]}\left(u\right)$. Therefore, if $t+s \in I$, by dominated convergence theorem,

$$\begin{aligned}
A_n\left((t,t+s]\right) &= \lim_{m\to\infty} \int_{(t,t+s]} T_m\left(u\right) H_0\left(du\right) \\
&= \lim_{m\to\infty} \sum_{1\leq k\leq 2^m} \frac{H_0\left((t_{m,k-1},\, t_{m,k}]\right)}{H_{01}\left((t_{m,k-1},\infty]\right)} \mathbf{1}_{\left\{N_n\left((t_{m,k-1},\infty]\right)>0\right\}}.
\end{aligned}$$

Now, by a well-known property of the multinomial law, the law of $N_{0n}\left((t_{m,k-1},\, t_{m,k}]\right)$ given $N_n\left((t_{m,k-1},\infty]\right)$ is a binomial with number of trials $N_n\left((t_{m,k-1},\infty]\right)$ and parameter $H_0\left((t_{m,k-1},\, t_{m,k}]\right)/H_{01}\left((t_{m,k-1},\infty]\right)$. Use again the dominated convergence theorem and deduce that

$$E\left[\Lambda_{n,1}\left((t,t+s]\right)\right] = E\left[A_n\left((t,t+s]\right)\right].$$

When $t_* = \sup I$ does not belong to I, consider $t+s$ increasing to t_* and use the monotone convergence theorem to deduce $E\left[\Lambda_{n,1}\left((t,t_*)\right)\right] = E\left[A_n\left((t,t_*)\right)\right]$. Finally, by the monotone class theorem we obtain the stated result. ■

Lemma A.2 *Consider (N_1, N_2, N_3) a random vector with a multinomial distribution $\mathcal{M}(m; q_1, q_2, q_3)$, where m is a positive integer and $q_1+q_2+q_3 = 1$. Let $c > 0$. Then*

$$E\left[\frac{\mathbf{1}_{\{N_1+N_2>0\}}}{N_1+cN_2}\right] \geq \frac{(1-q_3^m)^2}{m(q_1+cq_2)}.$$

Proof. The inequality is a direct consequence of Jensen's inequality. ∎

Proof of Lemma 3.5. Let $J(s-) = 1$ if $\overline{H}_{n0}(s-) + \overline{H}_{n1}(s-) > 0$ and $J(s-) = 0$ otherwise. Let us write

$$\begin{aligned} W(t) &= \int_{[0,t]} \frac{X(s)J(s-)}{\overline{H}_{n0}(s-) + p\,\overline{H}_{n1}(s-)} \left[H_{n0}(ds) - \frac{\overline{Y}(s-)/n}{\overline{H}(s-)} H_0(ds) \right] \\ &\quad + \left\{ \int_{[0,t]} \frac{X(s)J(s-)}{\overline{H}_{n0}(s-) + p\,\overline{H}_{n1}(s-)} \frac{\overline{Y}(s-)/n}{\overline{H}(s-)} H_0(ds) - \int_{[0,t]} X(s)\Lambda_p^*(ds) \right\} \\ &=: W_1(t) + W_2(t). \end{aligned}$$

The process W_1 is a squared integrable, zero-mean $\mathcal{G}_t$-martingale, since it is obtained as the integral of the bounded predictable process

$$\frac{1}{n} \frac{X(t)J(s-)}{\overline{H}_{n0}(t-) + p\,\overline{H}_{n1}(t-)}$$

with respect to the martingale M defined in Lemma 3.4. To complete the proof it suffices to show that W_2 is a submartingale. It is easy to see that, for any t, $E(|W_2(t)|) \leq KH_0([0,t])$ with K a constant depending only on n, p and the process X. To prove the submartingale condition, let us first consider $X(t) \equiv 1$. Fix $t' < t$. By Fubini's theorem, successive conditioning, the properties of the multinomial process and Lemma A.2,

$$\begin{aligned} & E\left[\int_{(t',\,t]} \frac{J(s-)}{\overline{H}_{n0}(s-) + p\,\overline{H}_{n1}(s-)} \frac{\overline{Y}(s-)/n}{\overline{H}(s-)} H_0(ds) \mid \mathcal{G}_{t'} \right] \qquad \text{(A.1)} \\ &= \int_{(t',\,t]} E\left[\frac{\overline{Y}(s-)}{\overline{H}(s-)} E\left[\frac{J(s-)}{n\,\overline{H}_{n0}(s-) + p\,n\,\overline{H}_{n1}(s-)} \mid \overline{Y}(s-) \right] \mid \mathcal{G}_{t'} \right] H_0(ds) \\ &\geq \int_{(t',\,t]} E\left[\mathbf{1}_{\{\overline{Y}(s-)>0\}} \{1 - [\overline{H}_2(s-)/\overline{H}(s-)]^{\overline{Y}(s-)}\}^2 \mid \mathcal{G}_{t'} \right] \frac{H_0(ds)}{\overline{H}_0(s-) + p\,\overline{H}_1(s-)}. \end{aligned}$$

Deduce that $E[W_2(t) \mid \mathcal{G}_{t'}] \geq W_2(t')$, for any $t' < t$. The same conclusion can be obtained when $X(t) = L_a \mathbf{1}_{(a,b]}(t)$, with L_a is $\mathcal{G}_a$-measurable nonnegative random variable, and when $X(t) = L_0 \mathbf{1}_{\{0\}}(t)$, where $L_0 \geq 0$ is $\mathcal{G}_0$-measurable. Use the Monotone Class Theorem to complete the proof of the general case (see, e.g., Fleming and Harrington (1991), section 1.5, for the details). ∎

Lemma A.3 *Consider the case where $t_* \notin I$. Assume that condition (3.7) of Theorem 3.6 holds and let $Y_{max} = \max_i Y_i$. Then, $\sqrt{n}F_T((Y_{max}, t_*)) \to 0$, in probability.*

Proof. (See Ying, 1989) Let $u_n^\varepsilon = \inf\{t : \sqrt{n}F_T((t, t_*)) \leq \varepsilon\}$, $\varepsilon > 0$. Then

$$\begin{aligned} P\left(\sqrt{n}F_T((Y_{max}, t_*)) > \varepsilon\right) \leq P\left(Y_{max} < u_n^\varepsilon\right) &= H([0, u_n^\varepsilon))^n \\ &\leq \left[1 - \left\{\overline{H}_0(u_n^\varepsilon-) + p\overline{H}_1(u_n^\varepsilon-)\right\}\right]^n \\ &\leq \left[1 - pF_T([u_n^\varepsilon, t_*))\,F_C([u_n^\varepsilon, \infty])\right]^n \\ &\leq \left[1 - p\frac{\varepsilon^2}{n} \frac{F_C([u_n^\varepsilon, \infty])}{F_T([u_n^\varepsilon, t_*))}\right]^n \to 0, \end{aligned}$$

since

$$0 \le \frac{F_T\left([u_n^\varepsilon, t_*)\right)}{F_C\left([u_n^\varepsilon, \infty]\right)} \le \int_{[u_n^\varepsilon, t_*)} \frac{F_T(dt)}{F_C([t,\infty])} \to 0.$$

■

Lemma A.4 *For a function f on $[a,b]$, let $\|f\|_a^b$ denote $\sup_{a\le t\le b}|f(t-)|$. Let $T_{01} = \max_i Y_i \mathbf{1}_{\{A_i\neq 2\}}$ and $Q(t) = \{\overline{H}_0(t) + p\overline{H}_1(t)\}/\{\overline{H}_{n0}(t) + p\overline{H}_{n1}(t)\}$, $t < T_{01}$. Then, for any $\beta \in (0,1)$, $P\left\{\|Q\|_0^{T_{01}} \ge 1/\beta\right\} \le 2e(1/\beta)e^{-1/\beta}$ and $P\left[\|\overline{Y}/\{n\overline{H}\}\|_0^{Y_{max}} \ge 1/\beta\right] \le e/\beta$, where $\overline{Y}(t-) = \sum_i \mathbf{1}_{\{Y_i \ge t\}}$ and $\overline{H}(t-) = P(Y \ge t)$.*

Proof. Let $T_k = \max_i Y_i \mathbf{1}_{\{A_i = k\}}$, $k = 0, 1$. The event $\left\{\|Q\|_0^{T_{01}} \ge 1/\beta\right\}$ is included in the union $\left\{\|\overline{H}_0/\overline{H}_{n0}\|_0^{T_0} \ge 1/\beta\right\} \cup \left\{\|\overline{H}_1/\overline{H}_{n1}\|_0^{T_1} \ge 1/\beta\right\}$. For each of the last two events apply Lemma 2.7 of Gill (1983) with the random variable $Y\mathbf{1}_{\{A=k\}}$, $k = 0, 1$. For the second probability bound, follow the arguments in Remark 1(i) of Wellner (1978). ■

Proof of Theorem 3.6. *a)* The inversion formula of Model 1 can be thought as the composition of three mappings

$$\left(\overline{H}_0, \overline{H}_1, p\right) \overset{\varphi_1}{\longmapsto} \left(\overline{H}_0, \overline{H}_0 + p\overline{H}_1\right) \overset{\varphi_2}{\longmapsto} \Lambda \overset{\varphi_3}{\longmapsto} F \tag{A.2}$$

where φ_2 is the map $(x, y) \mapsto -\int_{[0,\cdot]}(1/y_-)dx$ and φ_3 is the product-integral mapping $z \mapsto \prod_{[0,\cdot]}(1 - dz)$. The notation y_- means that we consider the left-limits of y. The Hadamard derivative of the map φ_1 at $\left(\overline{H}_0, \overline{H}_1, p\right)$ is given by $(\alpha, \beta, c) \mapsto \left(\alpha, \alpha + p\beta + c\,\overline{H}_1\right)$. By delta-method (Gill, 1989, van de Vaart and Wellner, 1996, section 3.9) applied with φ_1,

$$\sqrt{n}\left\{\left(\overline{H}_{0n}, \overline{H}_{0n} + \widehat{p}\overline{H}_{1n}\right) - \left(\overline{H}_0, \overline{H}_0 + p\overline{H}_1\right)\right\} \rightsquigarrow \left(\overline{\mathbb{G}}_0, \overline{\mathbb{G}}_2\right),$$

in $(D[0,\infty])^2$, where $\overline{\mathbb{G}}_2 = \overline{\mathbb{G}}_0 + p\overline{\mathbb{G}}_1 + N\,\overline{H}_1$. The process $\left(\overline{\mathbb{G}}_0, \overline{\mathbb{G}}_2\right)$ is a tight zero mean Gaussian process with covariance structure given by (3.6) and

$$E\left\{\overline{\mathbb{G}}_0(t)\,\overline{\mathbb{G}}_2(s)\right\} = \overline{H}_0(t\vee s) - \overline{H}_0(t)\,\overline{H}_0(s) + \overline{H}_0(t)\,\overline{H}_1(s)\left[\frac{(1-p)}{H_{02}([0,\infty])} - p\right]$$

$$\begin{aligned}
E\left\{\overline{\mathbb{G}}_2(t)\,\overline{\mathbb{G}}_2(s)\right\} &= \overline{H}_0(t\vee s) - \overline{H}_0(t)\,\overline{H}_0(s) \\
&\quad + \left\{\overline{H}_0(t)\,\overline{H}_1(s) + \overline{H}_0(s)\,\overline{H}_1(t)\right\}\left[\frac{1-p}{H_{02}([0,\infty])} - p\right] \\
&\quad + p^2\left\{\overline{H}_1(t\vee s) - \overline{H}_1(t)\,\overline{H}_1(s)\right\} \\
&\quad + \overline{H}_1(t)\,\overline{H}_1(s)\,\frac{p(1-p)}{H_{02}([0,\infty])}.
\end{aligned}$$

Let τ be a point in the interval $I = \left\{t : \overline{H}_0(t-) + p\overline{H}_1(t-) > 0\right\}$. Let $\int |dA|$ denote the total variation of the càdlàg function $t \mapsto A(t)$. The map φ_2 is Hadamard-differentiable on a domain of the type $\left\{(A, B) : \int |dA| \le M,\ B \ge \varepsilon\right\}$ for given M and $\varepsilon > 0$, at every point (A, B) such that $1/B$ is of bounded variation. If t is restricted to $[0, \tau]$, then $\left(\overline{H}_{0n}, \overline{H}_{0n} + \widehat{p}\overline{H}_{1n}\right)$ is contained in this domain with probability tending to one for $M \ge 1$ and sufficiently small ε. The derivative of φ_2 at $\left(\overline{H}_0, \overline{H}_0 + p\overline{H}_1\right)$ is given by

$$(\gamma, \eta) \mapsto -\int\left\{1/\left(\overline{H}_0 + p\overline{H}_1\right)_-\right\}d\gamma - \int\left\{\eta/\left(\overline{H}_0 + p\overline{H}_1\right)_-^2\right\}dH_0.$$

(The integrals with respect to functions which are not of bounded variation have to be understood via partial integration.) Use again the delta-method, this time with φ_2, and deduce that $\sqrt{n}(\widehat{\Lambda}-\Lambda) \rightsquigarrow \mathbb{G}_3$ in $D[0,\tau]$, where

$$\mathbb{G}_3 = -\int \frac{d\overline{\mathbb{G}}_0}{\left(\overline{H}_0 + p\overline{H}_1\right)_-} - \int \frac{\overline{\mathbb{G}}_{2-}}{\left(\overline{H}_0 + p\overline{H}_1\right)_-^2} dH_0. \tag{A.3}$$

The process $\mathbb{G}_3$ is a tight zero mean Gaussian process with the variance provided in Lemma A.5 below. Finally, apply delta-method with φ_3 and deduce that $\sqrt{n}(\widehat{F}-F) \rightsquigarrow \mathbb{G}$ in $D[0,\tau]$, where

$$\begin{aligned}
\mathbb{G}(t) &= F((t,\infty)) \int_{[0,t]} \frac{F([s,\infty])}{F((s,\infty])} d\mathbb{G}_3 \\
&= -F((t,\infty]) \int_{[0,t]} \frac{d\overline{\mathbb{G}}_0(s)}{\overline{H}_0(s) + p\overline{H}_1(s)} - F((t,\infty]) \int_{[0,t]} \frac{\overline{\mathbb{G}}_2(s-)}{\overline{H}_0(s-) + p\overline{H}_1(s-)} d\Lambda(s).
\end{aligned}$$

(See also Lemma 3.9.30 in van der Vaart and Wellner (1996) for the derivative of the product integration.) The process $\mathbb{G}$ is a tight zero mean Gaussian process. Its covariance can be obtained by direct but tedious calculations and therefore we do not provide it here.

b) Concerning the weak convergence on $D[0,t_*]$ when $t_* \notin I$, note that the variance of $\mathbb{G}_3$ converges to a finite limit when $t \uparrow t_*$, provided that assumption 3.7 holds. We want to extend the definition of $\mathbb{G}_3$ to $[0,t_*]$ by taking the limits along the paths when t grows to t_*. To prove that $\mathbb{G}_3$ indeed has a limit almost surely as $t \uparrow t_*$, first, deduce the variance of $\mathbb{G}_3(s) - \mathbb{G}_3(t)$ using calculations as in the proof of Lemma A.5 below. Next, use an exponential inequality for the increments of a Gaussian process [see, e.g., van der Vaart and Wellner (1996), appendix A.2.2] to suitably bound the probability $P\{\sup_{s\in[t,t_*)} |\mathbb{G}_3(s) - \mathbb{G}_3(t)| \geq \varepsilon\}$. Finally, proceed as Gill (1983), page 52-3, that is use Borel-Cantelli lemma to deduce that $\mathbb{G}_3$ converges almost surely. Since $\mathbb{G} = \varphi_3'(\Lambda)(\mathbb{G}_3)$ and $\varphi_3'(\Lambda) : D[0,t_*] \to D[0,t_*]$ is a continuous linear map, $\mathbb{G}$ is a Gaussian process in $D[0,t_*]$. Now, to prove the weak convergence of U in $D[0,t_*]$, it suffices to show that

$$\lim_{\tau\uparrow t_*} \limsup_{n\to\infty} P\Big(\sup_{\tau\leq t\leq t_*} |U(t) - U(\tau)| > \varepsilon\Big) = 0, \qquad \forall \varepsilon > 0, \tag{A.4}$$

that is a "tightness at t_*" condition for U. See Pollard (1984), page 70; see also Gill (1983). (Alternatively, one may prove this tightness property and obtain weak convergence in $D[0,t_*]$ for $\sqrt{n}(\widehat{\Lambda}-\Lambda)$. Next, the delta-method ensures the weak convergence for U. Nevertheless, analyzing U provides more insight on the difficulty to relax assumption (3.8) in the case $p < 1$.) The process U can be decomposed

$$U = \sqrt{n}(\widehat{F} - F_{n,p}) + \sqrt{n}(F_{n,p} - F_p^*) + \sqrt{n}(F_p^* - F) =: U_1 + U_2 + U_3.$$

Since the case $p = 1$ was studied by Gill (1983) and Ying (1989), for the rest of the proof, consider $p \in (0,1)$. For any $\tau < t$,

$$\begin{aligned}
U_1(t) - U_1(\tau) &= \sqrt{n}\left\{\widehat{F}((\tau,t]) - F_{n,p}((\tau,t])\right\} \\
&= \sqrt{n}\int_{(\tau,t]} \widehat{F}([s,\infty])\left[\widehat{\Lambda}(ds) - \Lambda_{n,p}(ds)\right] \\
&\qquad + \sqrt{n}\int_{(\tau,t]} \left[\widehat{F}([s,\infty]) - F_{n,p}([s,\infty])\right] \Lambda_{n,p}(ds) \\
&= : A_1((\tau,t]) + B_1((\tau,t]).
\end{aligned}$$

Easy calculations yield

$$|A_1((\tau,t])| \leq \sqrt{n}\frac{|\widehat{p}-p|}{p}\int_{(\tau,t]} \widehat{F}\left([s,\infty]\right)\widehat{\Lambda}(ds) = \sqrt{n}\frac{|\widehat{p}-p|}{p}\widehat{F}\left((\tau,t]\right).$$

On the other hand, use Duhamel's equation to write

$$\left|\widehat{F}\left([s,\infty]\right) - F_{n,p}\left([s,\infty]\right)\right| \leq \frac{|\widehat{p}-p|}{p}F_{n,p}\left([s,\infty]\right)\int_{[0,s)}\frac{\widehat{F}\left([v,\infty]\right)}{F_{n,p}\left((v,\infty]\right)}\widehat{\Lambda}(dv).$$

Deduce that

$$|B_1((\tau,t])| \leq \sqrt{n}\frac{|\widehat{p}-p|}{p}\Lambda_{n,p}\left((\tau,t]\right).$$

Use the first condition in (3.8), Lemma 3.2 and Theorem 3.3 to deduce that

$$\lim_{\tau\uparrow t_*}\limsup_{n\to\infty} P(\sup_{\tau\leq t\leq Y_{max}} |U_1\left(t\right) - U_1\left(\tau\right)| > \varepsilon) = 0, \qquad \forall\varepsilon > 0.$$

Next, we have

$$\begin{aligned} U_3\left(t\right)-U_3\left(\tau\right) &= \sqrt{n}\int_{(\tau,t]}\{F_p^*\left([s,\infty]\right)-F\left([s,\infty]\right)\}\Lambda_p^*(ds)+\sqrt{n}\int_{(\tau,t]}F\left([s,\infty]\right)\left[\Lambda_p^*(ds)-\Lambda(ds)\right] \\ &=: A_3((\tau,t]) + B_3((\tau,t]). \end{aligned}$$

First, note that when $t_* \notin I$ and (3.8) holds, there exists some constant $\tilde{K} < 1$ such that

$$\overline{H}_2(t-) \leq \tilde{K}\overline{H}(t-), \quad \forall t \geq 0. \tag{A.5}$$

Indeed, by equations (2.2), $F_C([t_*,\infty]) = H_2([t_*,\infty]) = 0$ and for $t < t_*$,

$$H_2\left(dt\right) = \left(1-p\right)F_T\left([0,t]\right)F_C\left(dt\right) \leq K\left(1-p\right)F_C\left(dt\right) = \left(1-p\right)KH_1\left(dt\right) + KH_2\left(dt\right),$$

with $K = F([0,t_*)) < 1$. Thus, (A.5) holds with $\tilde{K}^{-1} = 1 + (1-K)/\{K(1-p)\}$. Now,

$$\begin{aligned} |B_3((\tau,t])| &\leq \sqrt{n}\left\{2\int_{(\tau,t]}\mathbf{1}_{\{\overline{Y}(s-)>0\}}[\overline{H}_2(s-)/\overline{H}(s-)]^{\overline{Y}(s-)}F(ds) + \int_{(\tau,t]}\mathbf{1}_{\{\overline{Y}(s-)=0\}}F(ds)\right\} \\ &\leq \int_{(\tau,t]}\mathbf{1}_{\{\overline{Y}(s-)>0\}}\frac{g(\overline{Y}(s-))}{\left[\overline{Y}(s-)/n\right]^{1/2}}F(ds) + \sqrt{n}F\left((Y_{max},t_*)\right) \\ &\leq K\int_{(\tau,t]}\frac{\mathbf{1}_{\{\overline{Y}(s-)>0\}}}{\left[\overline{Y}(s-)/n\right]^{1/2}}F(ds) + \sqrt{n}F\left((Y_{max},t_*)\right) \end{aligned}$$

where $g(m) = 2\sqrt{m}\,\tilde{K}^{\,m}$, $m = 1,2...$, and K is some constant such that $g(m) \leq K$, $\forall m$. By Lemma A.3, $\sqrt{n}F\left((Y_{max},t_*)\right) \to 0$, in probability. Next, use Lemma 2.7 of Gill (1983) to replace $\overline{Y}(s-)/n$ by $\overline{H}(s-)$ in the denominator of the last integral above. Since

$$\int_{(\tau,t]}\frac{\mathbf{1}_{\{\overline{Y}(s-)>0\}}}{\left[\overline{H}(s-)\right]^{1/2}}F_T(ds) \leq (1-p)^{-1}\int_{(\tau,t]}\frac{F_T(ds)}{F_C([s,\infty])}$$

and $B_3((\tau,t_*]) = B_3((\tau,t_*))$, deduce that

$$\lim_{\tau\uparrow t_*}\limsup_{n\to\infty} P(\sup_{\tau\leq t\leq t_*} |B_3\left((\tau,t]\right)| > \varepsilon) = 0, \qquad \forall\varepsilon > 0. \tag{A.6}$$

For A_3, use Duhamel's equation and calculations as for B_3 to deduce that for some $K > 0$

$$|A_3((\tau, t])| \leq K \int_{(\tau,t]} \int_{[0,s)} \frac{\mathbf{1}_{\{\overline{Y}(v-)>0\}}}{\left[\overline{Y}(v-)/n\right]^{1/2}} F(dv)\ \Lambda_p^*(ds),$$

By Lemma 2.7 of Gill (1983) and assumption (3.7), condition (A.6) is satisfied by A_3.

Concerning U_2, we can write

$$\sup_{t\in[\tau,t_*)} |U_2(t) - U_2(\tau)| \leq F_p^*(\tau, t_*)|Z_p(\tau)| + \sup_{t\in[\tau,t_*)} F_p^*((t,\infty])\,|Z_p(t) - Z_p(\tau)|\,, \tag{A.7}$$

with $Z_p(t)$ defined in (3.4). First, we prove the "tightness at t_*" condition for Z_p. Recall that $J(s-) = \mathbf{1}_{\{\overline{H}_{n0}(s-)+\overline{H}_{n1}(s-)>0\}}$. By Duhamel's equation, for any $\tau < t \leq t_*$,

$$\begin{aligned} Z_p(t) - Z_p(\tau) &= \sqrt{n}\int_{(\tau,t]} \frac{F_{n,p}([s,\infty])}{F_p^*((s,\infty])} \frac{J(s-)/n}{\overline{H}_{n0}(s-) + p\,\overline{H}_{n1}(s-)} M(ds) \\ &\quad -\sqrt{n}\int_{(\tau,t]} \frac{F_{n,p}([s,\infty])}{F_p^*((s,\infty])} B(s-)\frac{H_0(ds)}{\overline{H}(s-)} \\ &=: A_2((\tau,t]) - B_2((\tau,t]), \end{aligned}$$

where M is the martingale defined in Lemma 3.4 and

$$B(s-) = \frac{\mathbf{1}_{\{\overline{Y}(s-)>0\}}\left[1 - \left\{\overline{H}_2(s-)/\overline{H}(s-)\right\}^{\overline{Y}(s-)}\right]^2 \overline{H}(s-)}{\overline{H}_0(s-) + p\,\overline{H}_1(s-)} - \frac{J(s-)\overline{Y}(s-)/n}{\overline{H}_{n0}(s-) + p\,\overline{H}_{n1}(s-)}.$$

For A_2 proceed as in Gill (1983), that is apply Lenglart's inequality (see also Fleming and Harrington (1991), Corollary 3.4.1). For this purpose, note that, by definition and assumption (3.7), the process $t \mapsto F_{np}([t,\infty])/F_p^*((t,\infty])$ is a bounded, $\mathcal{G}_t-$predictable process. Deduce that, for each $t' < t_*$ and for any $\varepsilon, \eta > 0$,

$$\begin{aligned} P\left[\sup_{\tau\leq t\leq t'\wedge Y_{max}} |A_2((\tau,t])| > \varepsilon\right] &\leq \frac{\eta}{\varepsilon^2} \\ + P&\left[\int_{(\tau,t'\wedge Y_{max}]} \left\{\frac{F_{n,p}([s,\infty])}{F_p^*((s,\infty])}\right\}^2 \frac{J(s-)\left\{1 - \Delta H_0(s)/\overline{H}(s-)\right\}}{\left\{\overline{H}_{n0}(s-) + p\,\overline{H}_{n1}(s-)\right\}^2} \frac{\overline{Y}(s-)/n}{\overline{H}(s-)} H_0(ds) > \eta\right]. \end{aligned} \tag{A.8}$$

Since assumption (3.7) implies $F_p^*([t_*,\infty]) > 0$, it suffices to bound the last probability above when the ratio between $F_{n,p}([s,\infty])$ and $F_p^*((s,\infty])$ is replaced by a constant. In this case, use Lemma A.4 and deduce that condition (A.6) is also satisfied by A_2. Finally, it remains to analyze $B_2((\tau,t])$. Note that $B_2((\tau,t])$ is a submartingale in $t \geq \tau$. For checking this property follow the steps in the display (A.2) and use the conditional independence between $\overline{H}_{n0}(s-) + p\,\overline{H}_{n1}(s-)$ and $F_{n,p}([s,\infty])/F_p^*((s,\infty])$ given $\overline{Y}(s-)$. Now, we can apply Birnbaum-Marshall inequality for $t \to B_2((\tau,t])$ (cf. Birnbaum and Marshall, 1961). For this we need to bound $t \mapsto E\left[B_2^2((\tau,t])\right]$. For any $t \leq t_*$,

$$E[B_2^2((\tau,t])] \leq K\mu((\tau,t_*)) \int_{(\tau,t_*)} n\,E[B^2(s-)]\frac{H_0(ds)}{\overline{H}(s-)},$$

with $\mu\left((\tau,t_*)\right)=\int_{(\tau,t_*)}\overline{H}(s-)^{-1}H_0(ds)$ and K a constant. Note that

$$\begin{aligned}|B(s-)| &\leq \frac{\overline{H}(s-)}{\overline{H}_0(s-)+p\overline{H}_1(s-)}+\frac{J(s-)}{\overline{H}_0(s-)+p\overline{H}_1(s-)}\left[\overline{H}(s-)-n^{-1}\overline{Y}(s-)\right]\\ &\quad +J(s-)\overline{Y}(s-)n^{-1}\left[\{\overline{H}_0(s-)+p\overline{H}_1(s-)\}^{-1}-\{\overline{H}_{n0}(s-)+p\overline{H}_{n1}(s-)\}^{-1}\right].\end{aligned}$$

By Lemma A.2, inequality (A.5) and the properties of the binomial law,

$$nE\left[B^2(s-)\right]\leq\frac{K}{\overline{H}_0(s-)+p\overline{H}_1(s-)}$$

for some $K>0$. Deduce that $t\mapsto E\left[B_2^2\left((\tau,t]\right)\right]$ is bounded by a constant. By Birnbaum-Marshall's inequality and assumption (3.7), condition (A.6) is satisfied also by B_2. Next, deduce from above that $\limsup_{\tau\uparrow t_*}E\left[Z^2\left(\tau\right)\right]<\infty$. Since $F_p^*((\tau,t_*))\to 0$ as $\tau\uparrow t_*$, deduce from (A.7) the tightness condition for U_2. Conclude that condition (A.4) is satisfied. ■

Lemma A.5 *Let $\mathbb{G}_3$ be defined as in equation (A.3). Then*

$$\begin{aligned}E\left\{\mathbb{G}_3\left(t\right)\mathbb{G}_3\left(s\right)\right\} &= \int_{[0,t\wedge s]}\frac{\{1-\Delta\Lambda\left(u\right)\}d\Lambda(u)}{\overline{H}_{01,p}\left(u-\right)}\\ &\quad+\int_{[0,t]}\int_{[0,s]}\frac{\{c(p)g_1(u-,v-)-p\left(1-p\right)g_2(u-,v-)\}\,dH_0(u)dH_0(v)}{\left\{\overline{H}_{01,p}\left(u-\right)\right\}^2\left\{\overline{H}_{01,p}\left(v-\right)\right\}^2},\end{aligned}$$

with $g_1(u,v)=\overline{H}_0\left(u\right)\overline{H}_1\left(v\right)+\overline{H}_0\left(v\right)\overline{H}_1\left(u\right)+p\overline{H}_1\left(u\right)\overline{H}_1\left(v\right)$, $g_2(u,v)=\overline{H}_1\left(u\vee v\right)$, $c(p)=(1-p)/H_{02}([0,\infty])$ *and* $t\wedge s=\min(t,s)$.

Proof. Let $\overline{H}_{01,p}=\overline{H}_0+p\overline{H}_1$. We have

$$\begin{aligned}E\left\{\mathbb{G}_3\left(t\right)\mathbb{G}_3\left(s\right)\right\} &= E\left[\int_{[0,t]}\frac{d\overline{\mathbb{G}}_0\left(u\right)}{\overline{H}_{01,p}\left(u-\right)}\int_{[0,s]}\frac{d\overline{\mathbb{G}}_0\left(v\right)}{\overline{H}_{01,p}\left(v-\right)}\right]\\ &\quad+E\left[\int_{[0,t]}\frac{d\overline{\mathbb{G}}_0\left(u\right)}{\overline{H}_{01,p}\left(u-\right)}\int_{[0,s]}\frac{\overline{\mathbb{G}}_2\left(v-\right)}{\left\{\overline{H}_{01,p}\left(v-\right)\right\}^2}dH_0(v)\right]\\ &\quad+E\left[\int_{[0,t]}\frac{\overline{\mathbb{G}}_2\left(u-\right)}{\left\{\overline{H}_{01,p}\left(u-\right)\right\}^2}dH_0(u)\int_{[0,s]}\frac{d\overline{\mathbb{G}}_0\left(v\right)}{\overline{H}_{01,p}\left(v-\right)}\right]\\ &\quad+E\left[\int_{[0,t]}\frac{\overline{\mathbb{G}}_2\left(u-\right)}{\left\{\overline{H}_{01,p}\left(u-\right)\right\}^2}dH_0(u)\int_{[0,s]}\frac{\overline{\mathbb{G}}_2\left(v-\right)}{\left\{\overline{H}_{01,p}\left(v-\right)\right\}^2}dH_0(v)\right]\\ &=:I+II+III+IV.\end{aligned}$$

Next,

$$I=\int_{[0,t]}\int_{[0,s]}\frac{d\left\{\overline{H}_0\left(u\vee v\right)-\overline{H}_0\left(u\right)\overline{H}_0\left(v\right)\right\}}{\overline{H}_{01,p}\left(u-\right)\overline{H}_{01,p}\left(v-\right)}=\int_{[0,t\wedge s]}\frac{d\Lambda(u)}{\overline{H}_{01,p}\left(u-\right)}-\Lambda\left([0,t]\right)\Lambda\left([0,s]\right),$$

$$\begin{aligned}II &= \int_{[0,t]}\int_{[0,s]}\frac{d_uE\left\{\overline{\mathbb{G}}_0\left(u\right)\overline{\mathbb{G}}_2\left(v-\right)\right\}dH_0(v)}{\overline{H}_{01,p}\left(u-\right)\left\{\overline{H}_{01,p}\left(v-\right)\right\}^2}\\ &= -\int_{[0,t\wedge s]}\frac{\Lambda([v,t])d\Lambda(v)}{\overline{H}_{01,p}\left(v-\right)}+\Lambda\left([0,t]\right)\Lambda\left([0,s]\right)-c(p)\Lambda\left([0,t]\right)\int_{[0,s]}\frac{\overline{H}_1\left(v-\right)d\Lambda(v)}{\overline{H}_{01,p}\left(v-\right)},\end{aligned}$$

$$III = \int_{[0,t]}\int_{[0,s]} \frac{d_v E\left\{\overline{\mathbb{G}}_2(u-)\,\overline{\mathbb{G}}_0(v)\right\} dH_0(u)}{\overline{H}_{01,p}(v-)\left\{\overline{H}_{01,p}(u-)\right\}^2}$$

$$= -\int_{[0,t\wedge s]} \frac{\Lambda([u,s])d\Lambda(u)}{\overline{H}_{01,p}(u-)} + \Lambda([0,t])\,\Lambda([0,s]) - c(p)\Lambda([0,s])\int_{[0,t]} \frac{\overline{H}_1(u-)\,d\Lambda(u)}{\overline{H}_{01,p}(u-)},$$

and

$$IV = \int_{[0,t]}\int_{[0,s]} \frac{E\left\{\overline{\mathbb{G}}_2(u-)\,\overline{\mathbb{G}}_2(v-)\right\} dH_0(u)dH_0(v)}{\left\{\overline{H}_{01,p}(u-)\right\}^2\left\{\overline{H}_{01,p}(v-)\right\}^2}$$

$$= \int\int \frac{\overline{H}_{01,p}((u\vee v)-)\,dH_0(u)dH_0(v)}{\left\{\overline{H}_{01,p}(u-)\right\}^2\left\{\overline{H}_{01,p}(v-)\right\}^2}$$

$$+c(p)\iint \frac{\left\{\overline{H}_0(u-)\,\overline{H}_1(v-)+\overline{H}_0(v-)\,\overline{H}_1(u-)+p\overline{H}_1(u-)\,\overline{H}_1(v-)\right\} dH_0(u)dH_0(v)}{\left\{\overline{H}_{01,p}(u-)\right\}^2\left\{\overline{H}_{01,p}(v-)\right\}^2}$$

$$-p(1-p)\int\int \frac{\overline{H}_1((u\vee v)-)\,dH_0(u)dH_0(v)}{\left\{\overline{H}_{01,p}(u-)\right\}^2\left\{\overline{H}_{01,p}(v-)\right\}^2} - \Lambda([0,t])\,\Lambda([0,s]).$$

Since

$$\int\int \frac{\overline{H}_{01,p}((u\vee v)-)\,dH_0(u)dH_0(v)}{\left\{\overline{H}_{01,p}(u-)\right\}^2\left\{\overline{H}_{01,p}(v-)\right\}^2} = \int_{[0,t\wedge s]} \frac{\Lambda([v,t])d\Lambda(v)}{\overline{H}_{01,p}(v-)} + \int_{[0,t\wedge s]} \frac{\Lambda((u,s])d\Lambda(u)}{\overline{H}_{01,p}(u-)},$$

deduce the formula for $E\left\{\mathbb{G}_3(t)\,\mathbb{G}_3(s)\right\}$. ■

References

[1] Birnbaum, Z.W., and Marshall, A.W. (1961). Some multivariate Chebyshev inequalities with extensions to continuous parameter processes. *Ann. Math. Statist.* **32**, 687-703.

[2] Chen, K., and Lo, S.H. (1997). On the rate of uniform convergence of the product-limit estimator: strong and weak laws. *Ann. Statist.* **25**, 1050-1087.

[3] Cupples, L.A., Risch, N., Farrer, L.A., and Myers, R.H. (1991). Estimation of Morbid Risk and Age at Onset with Missing Information. *Am. J. Hum. Genet.* **49**, 76-87.

[4] Fleming, T.R., and Harrington, P. (1991). *Counting Processes and Survival Analysis.* Wiley, New-York.

[5] Florens, J.P., Mouchart, M., and Rolin, J.M. (1990). *Elements of Bayesian Statistics.* Marcel Dekker, New-York.

[6] Gill, R. (1983). Large sample behavior of the product-limit estimator on the whole line. *Ann. Statist.* **11**, 49-58.

[7] Gill, R. (1989). Non- and semi-parametric maximum likelihood and the von-Mises method (part I). *Scand. J. Statist.* **16**, 97-128.

[8] GILL, R. (1994). Lectures on Survival Analysis (Ecole d'été de Probabilités de Saint-Flour XXII 1992). *Lecture Notes in Mathematics* (ed. P. Bernard), **1581**, 115-241, Springer Verlag, New-York.

[9] GILL, R., AND JOHANSEN, S. (1990). A survey of product-integration with a view towards application in survival analysis. *Ann. Statist.* **18**, 1501-1555.

[10] GU, M.G., AND ZHANG, C.-H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *Ann. Statist.* **21**, 611-624.

[11] HAMBURG, B.A., KRAEMER H.C., AND JAHNKE, W. (1975). A Hierarchy of drug use in adolescence: behavioral and attitudinal correlates of substantial drug use. *American J. Psychiatry* **132**, 1155-1163.

[12] HUANG, J. (1999). Asymptotic Properties of Nonparametric Estimation Based on Partly Interval-Censored Data. *Statist. Sinica* **9**, 501-519.

[13] HUANG, J., AND WELLNER, J. A. (1997). Interval censored survival data: A review of recent progress. In *Proceedings of the First Survival Symposium in Biostatistics: Survival Analysis,* D.Y. Lin and T.R. Fleming, Eds., pp 123-169, Springer-Verlag, New-York.

[14] KIM, J.S. (2003). Maximum likelihood estimation for the proportional hazards models with partly interval-censored data. *J. R. Stat. Soc. Ser B* **65**, 489-502.

[15] KLEIN, J.P., AND MOESCHBERGER, M.L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer-Verlag, New-York.

[16] VAN DER LAAN, M.J., AND GILL, R.D. (1999). Efficiency of the sieved-NPMLE in CAR-Censored Data Models. *Math. Methods Statist.* **8**, 251-276.

[17] POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer Verlag, New-York.

[18] ROLIN, J.-M. (2001). Nonparametric Competing Risks Models: Identification and Strong Consistency. DP 0115, Institut de Statistique, Louvain-la-Neuve, http://www.stat.ucl.ac.be.

[19] SAMUELSEN, S. O. (1989). Asymptotic Theory for Non-parametric Estimators from Doubly Censored Data. *Scand. J. Statist.* **16**, 1-21.

[20] STUTE, W., AND WANG, J. (1993). The Strong Law Under Random Censorship. *Ann. Statist.* **21**, 1591-1607.

[21] TURNBULL, B.W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *J. Amer. Statist. Assoc.* **69**, 169-173.

[22] TURNBULL, B.W., AND WEISS, L. (1978). A Likelihood Ratio Statistics for Testing Goodness of Fit with Randomly Censored Data. *Biometrics* **34**, 367-375.

[23] VAN DER VAART, A.W., AND WELLNER, J.A. (1996). *Weak convergence and empirical processes.* Springer Verlag, New-York.

[24] WELLNER, J.A. (1978). Limit Theorem for the Ration of the Empirical Distribution Function to the True Distribution Function. *Z. Wahrsch. verw. Gebiete* **45**, 73-88.

[25] Wellner, J.A. and Zhan, Y. (1996). Bootstrapping Z-estimators. Technical Report 308, University of Washington, Dept. of Statistics.

[26] Wellner, J.A. and Zhan, Y. (1997). A Hybrid Algorithm for Computation of the Nonparametric Maximum Likelihood Estimator From Censored Data. *J. Amer. Statist. Assoc.* **92**, 945-959.

[27] Ying, Z. (1989). A Note on the Asymptotic properties of the product-limit estimator on the whole line. *Statis. Prob. Letters* **7**, 311-314.