Bernoulli **12**(5), 2006, 801–819

Product-limit estimators of the survival function for two modified forms of current-status data

VALENTIN PATILEA[1] and JEAN-MARIE ROLIN[2]

[1]*CREST-ENSAI, Campus de Ker Lann, Rue Blaise Pascal, BP 37203, 35172 Bruz Cedex, France. E-mail: patilea@ensai.fr*
[2]*Institut de Statistique, Université Catholique de Louvain, 20 Voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium. E-mail: rolin@stat.ucl.ac.be*

The problem of estimating the distribution of a lifetime that may be left or right censored is considered. Two data structures that extend the classical current-status data framework are introduced and the corresponding product-limit estimators are derived. The strong uniform convergence and asymptotic normality of the product-limit estimators are proved. A bootstrap procedure that can be applied to confidence intervals construction is proposed.

Keywords: bootstrap; current-status data; delta method; left and right censoring; martingales; product-limit estimator; strong convergence; weak convergence

1. Introduction

The goal of this paper is to propose and analyse data structures where a lifetime of interest may be left or right censored. Typically, a lifetime T is left or right censored if, instead of observing T we observe a finite non-negative random variable Y, and a discrete random variable A 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) and 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 straightforward if 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 data structures allowing for an explicit inversion formula, that is, a closed-form function relating F_T to the distribution of (Y, A), are proposed. The data structures considered extend the classical current-status data situation and represent, in some sense, a particular case of the left- or right-censoring model of Turnbull (1974). We shall use the term 'latent models', or simply 'models', to refer to these data structures.

1350–7265

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 cross-sectional study where T, the age at onset of an occult non-fatal disease, is analysed. The individuals are examined only once and 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 time of examination. It is supposed that when the disease is present the subsample of individuals for which the age at onset is known is *completely random*. Let C denote the age at which the individual is examined. 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, T is right censored by C for those 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 analysed by Turnbull and Weiss (1978) and Cupples *et al.* (1991: Table 1).

The second latent model proposed is related to the first and lies between the left-censorship model and the current-status data model. Consider the example of a reliability experiment on the failure time of a type of device. A random sample of devices is considered and a single inspection for each device in the sample is undertaken. Some of the devices have already failed at some unknown time (left-censored observations). To increase the precision of the estimates, a proportion of the devices still working is selected at random and *all* the selected devices are followed until failure (uncensored observations). For the remaining working devices the failure time is right censored by the inspection time.

We remark that, without any model assumption, given a distribution for the observed variables (Y, A) with $Y \geqslant 0$ and $A \in \{0, 1, 2\}$, one can always apply our two inversion formulae. In this way one constructs two pseudo-true distribution functions of the lifetime of interest which are functionals of the distribution of (Y, A). 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 distribution of (Y, A). Otherwise, in general, one can only approximate the true lifetime distribution.

Our paper is organized as follows. Section 2 introduces the two latent models through the equations relating the distribution of the observations to those of the latent variables. By 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 ends with some comments on related models. It is shown that our first (second) latent model can be extended to the case where T is a failure time in the presence of competing (complementary) failure causes. The relationship between our models and Turnbull's (1974) model is also discussed. Section 3 contains the asymptotic results for the first latent model (similar arguments apply for the second model). The strong uniform convergence for the product-limit estimator on the whole range of the observations is proved. Our proof extends and simplifies the strong uniform convergence result for the Kaplan–Meier estimator obtained by Stute and Wang (1993) and Gill (1994). 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, the first model is applied to data on California high school students 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 time of examination) 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 $Y \geqslant 0$ and $A \in \{0, 1, 2\}$. These variables are defined as follows:

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

We can also write

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

and $A = 2(1 - \Delta)\mathbf{1}_{\{T \leqslant C\}} + \mathbf{1}_{\{C < T\}}$, where $\mathbf{1}_S$ denotes the indicator function of the set S. With this censoring mechanism the lifetime T is observed, right censored or left censored. If Δ is constant and equal to one (zero), we obtain right-censored (current-status) data. See, for example, 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 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 usual in survival analysis, the censoring mechanism defines a map Φ between the distributions of the latent variables and the distribution of the observed lifetime. For the data structure described above, 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:

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

When $p = 1$ ($p = 0$), equations (2.2) boil down to the equations of the classical independent right-censoring (current-status) model.

The nonparametric estimation of the distribution of T is straightforward if the map Φ is invertible and F_T can be written as an explicit function of the subdistributions H_k, as it suffices to apply the inverse of Φ to the empirical counterparts of H_k, $k = 0, 1, 2$. The model considered above allows for an explicit inversion formula for F_T. To derive this formula, integrate the first and the second equation in (2.2) on $[t, \infty]$ and deduce that

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

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 with a distribution F is $\Lambda(\mathrm{d}t) = F(\mathrm{d}t)/F([t, \infty])$. Use (2.2)–(2.3) to deduce that the hazard function corresponding to F_T can be written as

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

Finally, the distribution F_T can be expressed as

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

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, there is 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 supposing that $F_T(\{\infty\}) = 1 - F_T(I)$.

Given the explicit relationship between F_T and the subdistributions of Y, 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 $\{(Y_i, A_i) : 1 \leqslant i \leqslant n\}$ and let $Z_1 \leqslant \ldots \leqslant Z_J$ be the distinct values in increasing order of Y_i. For any $j = 1, \ldots, J$ and $k = 0, 1, 2$, define

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

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

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

while the estimator of the hazard measure is

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

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

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

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

Remark 1. In Model 1, like in most situations in survival analysis, there is a problem of identification in the sense that different joint distributions of (T, C, Δ) may generate the same subdistributions H_k, $k = 0, 1, 2$. A way to avoid this problem is to impose identification restrictions. The simplest one is to assume that the latent variables T, C and Δ are independent. As in the classical right-censorship model (see Fleming and Harrington 1991: Theorem 1.3.2), weaker assumptions than the independence of T, C and Δ may suffice for the identification of F_T. Nevertheless, any non-independence assumption on T, C and Δ is not testable using the observations since none of the pairs (T, C), (T, Δ) and (C, Δ) are completely observed.

Remark 2. By the second and third equations in (2.2), $(1 - p)H_1(\mathrm{d}t) + H_2(\mathrm{d}t) = (1 - p)F_C(\mathrm{d}t)$. Then equation (2.3) yields another solution of system (2.2):

$$F_T((t, \infty]) = \frac{1 - p}{p} \frac{H_0((t, \infty]) + pH_1((t, \infty])}{H_2((t, \infty]) + (1 - p)H_1((t, \infty])},$$

$t \geqslant 0$, with p as in (2.4). However, this solution may not be a survival function when H_0, H_1 and H_2 are replaced by their empirical counterparts.

Remark 3. Only the first two equations of system (2.2) have been used to derive formula (2.6). This means that the inversion formula proposed still recovers F_T from the subdistributions of Y even if the definition of H_2 is modified, for instance, to take into account that T lies in a subinterval of the positive half-line with right endpoint C.

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 $Y \geqslant 0$ and $A \in \{0, 1, 2\}$, where

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

The equations of this model are

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

When $p = 1$ ($p = 0$) equations (2.8) boil down to the equations of the classical independent left-censoring (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 that

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

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

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

Finally, the distribution F_T can be expressed as

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

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 $h \geqslant 0$ 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. In particular, Remarks 1–3 can be restated accordingly.

2.3. Extensions and related models

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

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

where H_{0a} (H_{0b}) is the subdistribution of the uncensored observations for which the failure cause is a (b). If H_0 denotes the subdistribution of the uncensored observations, $H_0(\mathrm{d}t) = p\,F_C([t, \infty])F_T(\mathrm{d}t)$ and thus (2.3) holds and p can be expressed as in (2.4). Consequently,

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

from which the expression for F_{T_a} can be derived. 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 (his 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, the 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. We remark that, contrary to our Model 1 (Model 2), in Huang's model one may observe exact failure times even if failure occurs after (before) the time of examination. 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 that introduced by Turnbull (1974). In Turnbull's model there are three latent lifetimes L (left censoring), T (lifetime of interest) and R (right censoring), with $L \leqslant R$. The observed variables are $Y = \max(L, \min(T, R)) = \min(\max(L, T), R)$ and A is defined as follows: $A = 0$ if $L < T \leqslant R$; $A = 1$ if $R < T$; and $A = 2$ if $T \leqslant L$. The equations for this model are

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

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 can be computed, for instance, by iterations based on the so-called self-consistency equation. Note that by imposing $F_C(\mathrm{d}t) = (1 - p)^{-1} F_L(\mathrm{d}t) = F_R(\mathrm{d}t)$ 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 proofs of asymptotic results below are much simpler and are derived under weaker conditions than in Turnbull's model (see Gu and Zhang 1993; Wellner and Zhan 1996).

3. Asymptotic results

In this section the strong uniform convergence and the asymptotic normality of the estimator of the distribution F_T in Model 1 are derived. Moreover, we propose a bootstrap procedure that can be used to construct confidence intervals for F_T. As in the previous sections, the distributions F_T and F_C need not be continuous. To simplify the notation, hereafter, the subscript T is suppressed when there is no possible confusion.

3.1. Strong uniform convergence

Let H_{nk} be the empirical counterparts of the subdistributions H_k, that is,

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

We wish to prove the strong uniform convergence of the distribution $\hat{F}$, that is,

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

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 $\sigma \in I$,*

$$\sup_{0 \leqslant t \leqslant \sigma} |\hat{\Lambda}([0, t]) - \Lambda([0, t])| \to 0 \quad \textit{as } n \to \infty, \qquad \textit{almost surely}.$$

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

To prove this theorem, first we consider the case where p replaces $\hat{p}$ in the definition of $\hat{\Lambda}$. In this case the functionals of the hazard function are reverse supermartingales in n, as is shown in the next lemma which extends a statement of Gill (1994). The proof of the lemma is relegated to the Appendix.

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

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

Define the σ-fields

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

Then, for all n,

$$\mathrm{E}[\Lambda_{n,p}(f) | \mathcal{B}_{n+1}] \leqslant \Lambda_{n+1,p}(f),$$

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

Proof of Theorem 3.1. The strong uniform convergence of $\hat{\Lambda}([0, t])$ when $t \in [0, \tau] \subset I$ can be obtained by the delta method (cf. Gill 1989, 1994; see also the proof of Theorem 3.4 below) from the almost sure uniform convergence of $H_{nk}([0, t])$, $k = 0, 1, 2$. For the last part of the theorem, write $\overline{H}_0(t-) = H_0([t, \infty])$ and $\overline{H}_1(t-) = H_1([t, \infty])$. 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

$$|\hat{\Lambda}((\tau, t_*)) - \Lambda((\tau, t_*))| \leqslant \left| \int_{(\tau,t_*)} \frac{H_{n0}(\mathrm{d}t)}{(\overline{H}_{n0} + \hat{p}\,\overline{H}_{n1})(t-)} - \int_{(\tau,t_*)} \frac{H_{n0}(\mathrm{d}t)}{(\overline{H}_{n0} + p\,\overline{H}_{n1})(t-)} \right| + \left| \int_{(\tau,t_*)} \frac{H_{n0}(\mathrm{d}t)}{(\overline{H}_{n0} + p\,\overline{H}_{n1})(t-)} - \int_{(\tau,t_*)} \frac{H_0(\mathrm{d}t)}{(\overline{H}_0 + p\,\overline{H}_1)(t-)} \right| =: A_1 + A_2.$$

A little algebra gives

$$A_1 \leqslant \frac{|\hat{p} - p|}{\hat{p}} \int_{(\tau,t_*)} \frac{H_{n0}(\mathrm{d}t)}{\overline{H}_{n0}(t-) + p\,\overline{H}_{n1}(t-)} = \frac{|\hat{p} - p|}{\hat{p}} \Lambda_{n,p}((\tau, t_*)).$$

Since $\hat{p} \to p$ 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 supermartingale. Now, by Doob's supermartingale convergence theorem, if $\sup_n \mathrm{E}[\Lambda_{n,p}((\tau, t_*))]$ 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 σ-field of asymptotic permutable events and is therefore a constant by the Hewitt–Savage 0–1 law. More precisely, the constant is equal to $\sup_n \mathrm{E}[\Lambda_{n,p}((\tau, t_*))]$. Hence, it remains to bound $\mathrm{E}[\Lambda_{n,p}((\tau, t_*))]$, $n \geqslant 1$. Note that $\Lambda_{n,p}((\tau, t_*)) \leqslant p^{-1}\Lambda_{n,1}((\tau, t_*))$. Next, it can be shown that

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

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

The strong uniform convergence of the distribution $\hat{F}$ follows.

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

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

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

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

As pointed out by Gill (1994), to complete the proof the only difficulty appears when $I = [0, t_*)$ and $F([t_*, \infty]) > 0$. In this case, we write

$$\sup_{t\in I}|\hat{F}([0,\ t]) - F([0,\ t])| \leqslant \hat{F}((\tau,\ t_*)) + F((\tau,\ t_*)) + \sup_{t\in[0,\tau]}|\hat{F}([0,\ t]) - F([0,\ t])|.$$

Now, by definition, $\Lambda(\mathrm{d}t) = F(\mathrm{d}t)/F([t,\ \infty])$ and $\hat{\Lambda}(\mathrm{d}t) = \hat{F}(\mathrm{d}t)/\hat{F}([t,\ \infty])$. Deduce that $\Lambda([0,\ t_*)) < \infty$ and $\Lambda((\tau,\ t_*)) \downarrow 0$ as $\tau \uparrow t_*$. On the other hand, deduce that the measure $\hat{\Lambda}(\mathrm{d}t)$ is greater than the measure $\hat{F}(\mathrm{d}t)$. Thus, for any $\tau \in I$,

$$\sup_{t\in I}|\hat{F}([0,\ t]) - F([0,\ t])| \leqslant \hat{\Lambda}((\tau,\ t_*)) + F((\tau,\ t_*)) + \sup_{t\in[0,\tau]}|\hat{F}([0,\ t]) - F([0,\ t])|.$$

By Lemma 3.2, $\hat{\Lambda}((\tau,\ t_*)) \to \Lambda((\tau,\ t_*))$, almost surely. Since $\Lambda((\tau,\ t_*)) + F((\tau,\ t_*)) \downarrow 0$ as $\tau \uparrow t_*$, the proof is complete. □

Remark 4. The proofs of Lemma 3.2 and Theorem 3.1 also apply to the extension of Model 1 considered in Section 2.3. Deduce the strong uniform convergence on I of the nonparametric estimator of the distribution F_{T_a}.

Remark 5. 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

Here, the weak convergence of the process $\sqrt{n}(\hat{F} - F)$ with $\hat{F}$ from Model 1 is derived. Weak convergence (denoted by $\rightsquigarrow$) is as in Pollard (1984): $D[a,\ b]$, the space of cadlag functions on $[a,\ b]$, is endowed with the supremum norm and the ball σ-field. By the empirical central limit theorem (e.g., van de Vaart and Wellner 1996),

$$\sqrt{n}\{(\overline{H}_{0n},\ \overline{H}_{1n},\ \hat{p}) - (\overline{H}_0,\ \overline{H}_1,\ p)\} \rightsquigarrow (\overline{\mathbb{G}}_0,\ \overline{\mathbb{G}}_1,\ N) \tag{3.1}$$

in $(D[0,\ \infty])^2 \times \mathbb{R}$, where $\overline{H}_{kn}$ is the cadlag 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 $t,\ s \geqslant 0$, the vector $(\overline{\mathbb{G}}_0(t),\ \overline{\mathbb{G}}_1(t),\ N)$ has a zero-mean multivariate normal distribution.

Theorem 3.4. *Assume that* $p \in (0,\ 1]$ *and define* $U(t) = \sqrt{n}(\hat{F}([0,\ t]) - F([0,\ t]))$, $t \geqslant 0$.

(*a*) *Let* τ *be a point in* I. *Then* $U \rightsquigarrow \mathbb{G}$ *in* $D[0,\ \tau]$, *where* $\mathbb{G}$ *is the Gaussian process*

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

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* (3.1). *The first integral is defined by integration by parts.*

(*b*) *If* $t_* \notin I$, *but*

$$\int_{[0,t_*)} \frac{H_0(\mathrm{d}t)}{\{\overline{H}_0(t-) + p\overline{H}_1(t-)\}^2} < \infty, \tag{3.2}$$

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

When $t_* \notin I$, condition (3.2) is equivalent to

$$F_T([t_*, \infty]) > 0 \quad \text{and} \quad \int_{[0,t_*)} \frac{F_T(\mathrm{d}t)}{F_C([t, \infty])} < \infty. \tag{3.3}$$

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 in $D[0, t_*]$ still holds when $p < 1$ and only the second part of (3.3) is satisfied remains an open question.

The proof of part (a) of Theorem 3.4 is postponed to the Appendix. It is based on the delta method and the weak convergence in (3.1). The proof of (b) is based on the martingale arguments of Gill (1983) but modified in order to take into account that in Model 1 when $p < 1$ the estimator $\hat{\Lambda}$ no longer has an obvious martingale structure (in t), as the Nelson–Aalen estimator does. The details of the proof of (b) can be found in Patilea and Rolin (2004) and, for the sake of brevity, are omitted.

3.3. Bootstrapping the product-limit estimator

Theorem 3.4 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 having to handle this process in applications. Here, a bootstrap sample is obtained by simple random sampling with replacement from the set of observations. Let H_k^*, $k = 0, 1, 2$, denote the bootstrap versions of the subdistributions of the observed lifetime. Apply equations (2.4)–(2.6) to obtain the bootstrap estimator $\hat{F}^*$. The following theorem states 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 is based on the uniform Hadamard differentiability of the maps involved in the inversion formula of Model 1.

Theorem 3.5. *Let* $\tau \in I$ *and let* $\widetilde{\mathbb{G}}(t)$ *be the limit of* $\sqrt{n}\{\hat{F}([0, t]) - F([0, t])\}/F((t, \infty])$ *in* $D[0, \tau]$, *as obtained from Theorem 3.4. Then, the process*

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

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

4. Application

By way of illustration, we apply Model 1 to the California high school students data (Table 1). The data are part of a study conducted by the Stanford-Palo Alto Peer Counseling Program; see Hamburg *et al.* (1975). In this study, 191 high schools boys were asked ‘When did you first use marijuana?’. The answers were the exact ages, ‘I have used it but

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

cannot recall just when the first time was' and 'I never used it'. The latent variable T is the age at first use of marijuana. It is supposed that the observations are independent and identically distributed (i.i.d.) Turnbull and Weiss (1978) (see also Klein and Moeschberger 1997: Chapter 5) used the double-censorship model of Turnbull (1974) with this data set to estimate the distribution of T. 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 time of the study. It is assumed that T and C are independent and the missing values of T are missing independently of T and the age of the student at interview. Condition $F_T([t_*, \infty]) > 0$ in (3.2) means that some students will never use marijuana.

The distribution $\hat{F}_T$ obtained with Model 1 is reported in Table 2. The value of $\hat{p}$ is 0.893. Moreover, we provide the estimator F_n^{TW} obtained in Turnbull's model, as reported in Klein and Moeschberger (1997: 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 $\hat{F}_T$ and F_n^{TW} due to the small number of left-censored observations. The fact that, in some sense, Model 1 is a special case of Turnbull's model explains the closeness between $\hat{F}_T$ and F_n^{TW}.

Pointwise confidence intervals and confidence bands for the survival probability are rather

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

Z_j	$\hat{F}_T((Z_j, \infty])$	$F_n^{\mathrm{TW}}((Z_j, \infty])$	$F_n^{\mathrm{KM}}((Z_j, \infty])$
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

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 bootstrap pointwise confidence intervals for the survival function estimated by $\hat{F}_T$. The level is $1-\alpha = 0.95$ and the confidence interval for $F((t, \infty])$ is defined as

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

where $q^*_\alpha(t)$ is the bootstrap approximation of the α-quantile of the distribution of $\sqrt{n}\{\hat{F}((t, \infty]) - F((t, \infty])\}/F((t, \infty])$. The number of bootstrap samples used was 5000.

The intervals plotted in Figure 1 are rather wide in the right tail of F_T. By analogy with the confidence intervals and bands proposed for the Kaplan–Meier estimator (e.g., Akritas 1986; Gill 1994), many other confidence intervals and bands for F_T together with their bootstrap versions can be constructed with Model 1. They are based on the asymptotic behaviour of a weighted process $\sqrt{n}\, W(\hat{F}_T - F_T)$ with $W(t)$, $t \geqslant 0$, a suitable weight function that depends upon the asymptotic variance of $\sqrt{n}(\hat{F}_T - F_T)$. The main difficulty in constructing such improved intervals and bands in Model 1 is the complicated expression for the asymptotic variance of $\sqrt{n}(\hat{F}_T - F_T)$. Perhaps a rough approximation of this variance by that obtained in the case $p = 1$ could be a satisfactory compromise. This issue requires extensive empirical investigation which we leave for future work.

Let us end this application with some comments suggested by a referee and motivated by the discrete nature of the data in Table 1. When observations are recorded at a finite set of time points, the study of the asymptotics of $\hat{F}_T$ defined in (2.7) is much simplified as it relies on the asymptotics of finite-dimensional vectors of observed frequencies. However, the asymptotic variance of $\hat{F}_T$ remains quite intricate. Finally, the discrete recording of the lifetimes may introduce a bias in the estimation of F_T. To eliminate this type of

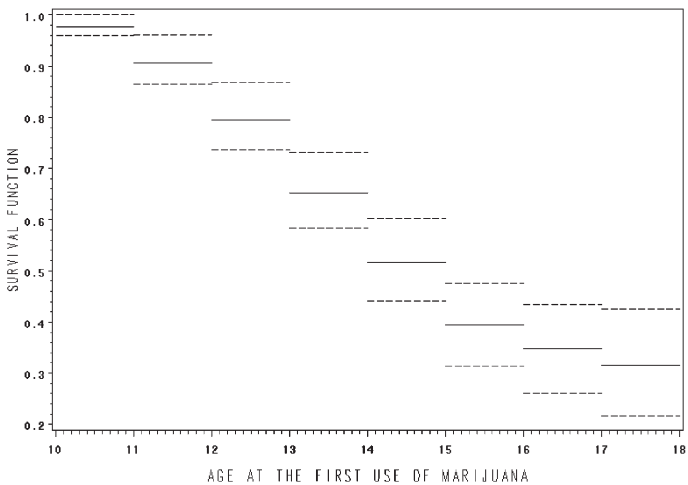

Figure 1. Pointwise confidence intervals for the survival function for the age at the first use of marijuana.

discretization bias, which is quite common in survival analysis, one may suppose that some lifetimes are not observed but lie in an interval. The analysis of such models could be much more complicated and is beyond our present scope.

Appendix

Proof of Lemma 3.2. To simplify the notation, in this proof we write Λ_n instead of $\Lambda_{n,p}$. Define $N_n(t) = nH_{n0}([t, \infty]) + pnH_{n1}([t, \infty])$ and write

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

Next, notice that

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

and therefore, by the i.i.d. property of the sample and elementary properties of the conditional independence (see Florens *et al.*, 1990), we have

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

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

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

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

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

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

$$\mathrm{E}[\Lambda_n(f)|\mathcal{B}_{n+1}] = \frac{1}{n+1}\sum_{1\leqslant i\leqslant n+1} f(Y_i)\mathbf{1}_{\{A_i=0\}} \times \sum_{1\leqslant j\neq i\leqslant n+1} [N_{n+1}(Y_i) - \mathbf{1}_{\{Y_j\geqslant Y_i, A_j=0\}} - p\mathbf{1}_{\{Y_j\geqslant Y_i, A_j=1\}}]^{-1}. \tag{A.1}$$

Now

$$\sum_{1\leqslant j\neq i\leqslant n+1}[N_{n+1}(Y_i)-\mathbf{1}_{\{Y_j\geqslant Y_i,A_j=0\}}-p\mathbf{1}_{\{Y_j\geqslant Y_i,A_j=1\}}]^{-1}$$

$$=\sum_{1\leqslant j\neq i\leqslant n+1}\left[\frac{\mathbf{1}_{\{Y_j<Y_i\}}+\mathbf{1}_{\{Y_j\geqslant Y_i,A_j=2\}}}{N_{n+1}(Y_i)}+\frac{\mathbf{1}_{\{Y_j\geqslant Y_i,A_j=0\}}}{N_{n+1}(Y_i)-1}+\frac{\mathbf{1}_{\{Y_j\geqslant Y_i,A_j=1\}}}{N_{n+1}(Y_i)-p}\right]$$

$$=\frac{n+1}{N_{n+1}(Y_i)}-R_i,$$

where

$$R_i=\frac{1}{N_{n+1}(Y_i)}$$

$$+\sum_{1\leqslant j\neq i\leqslant n+1}\left[\frac{\mathbf{1}_{\{Y_j\geqslant Y_i,A_j=0\}}+\mathbf{1}_{\{Y_j\geqslant Y_i,A_j=1\}}}{N_{n+1}(Y_i)}-\frac{\mathbf{1}_{\{Y_j\geqslant Y_i,A_j=0\}}}{N_{n+1}(Y_i)-1}-\frac{\mathbf{1}_{\{Y_j\geqslant Y_i,A_j=1\}}}{N_{n+1}(Y_i)-p}\right].$$

Use the inequality

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

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

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

that is, $\Lambda_n(f)$, $n\geqslant 1$, is a reverse supermartingale. □

Lemma A.1. *Let* $f\geqslant 0$ *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*

$$\Lambda_{n,1}(f)=\int_I\frac{f(t)}{\overline{H}_{n01}(t)}H_{n0}(\mathrm{d}t)\quad\textit{and}\quad\Lambda_1(f)=\int_I\frac{f(t)}{\overline{H}_{01}(t)}H_0(\mathrm{d}t),$$

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

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

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 in the integral form

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

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

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

and note that

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

Since $S_m(u) \leqslant 1$, by the dominated convergence theorem,

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

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

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

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

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

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

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

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

Proof of Theorem 3.4. The inversion formula of Model 1 can be thought of as the composition of three mappings

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

where φ_2 is the map $(x,\, y) \mapsto -\int_{[0,\cdot]} (1/y_-)\mathrm{d}x$ and φ_3 is the product-integral mapping $z \mapsto \prod_{[0,\cdot]} (1 - \mathrm{d}z)$. The notation y_- means that we consider the left limits of y. The Hadamard derivative of the map φ_1 at $(\overline{H}_0,\, \overline{H}_1,\, p)$ is given by $(\alpha, \beta, c) \mapsto$

$(\alpha, \alpha + p\beta + c\,\overline{H}_1)$. By the delta method (Gill 1989; van de Vaart and Wellner 1996: Section 3.9) applied with φ_1,

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

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 $(\overline{\mathbb{G}}_0, \overline{\mathbb{G}}_2)$ is a tight, zero-mean Gaussian process with covariance structure

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

Let τ be a point in the interval $I = \{t : \overline{H}_0(t-) + p\overline{H}_1(t-) > 0\}$. Let $\int |\mathrm{d}A|$ denote the total variation of the cadlag function $t \mapsto A(t)$. The map φ_2 is Hadamard-differentiable on a domain of the type $\{(A, B) : \int |\mathrm{d}A| \leqslant M, B \geqslant \varepsilon\}$ 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 $(\overline{H}_{0n}, \overline{H}_{0n} + \hat{p}\overline{H}_{1n})$ is contained in this domain with probability tending to one for $M \geqslant 1$ and sufficiently small ε. The derivative of φ_2 at $(\overline{H}_0, \overline{H}_0 + p\overline{H}_1)$ is given by

$$(\gamma, \eta) \mapsto -\int \frac{1}{(\overline{H}_0 + p\overline{H}_1)_-}\,\mathrm{d}\gamma - \int \frac{\eta}{(\overline{H}_0 + p\overline{H}_1)_-^2}\,\mathrm{d}H_0.$$

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

$$\mathbb{G}_3 = -\int \frac{\mathrm{d}\overline{\mathbb{G}}_0}{(\overline{H}_0 + p\overline{H}_1)_-} - \int \frac{\overline{\mathbb{G}}_{2-}}{(\overline{H}_0 + p\overline{H}_1)_-^2}\,\mathrm{d}H_0. \tag{A.3}$$

The process $\mathbb{G}_3$ is a tight, zero-mean Gaussian process. Its variance is given in Lemma A.5 of Patilea and Rolin (2004). Finally, apply the delta method with φ_3 and deduce that $\sqrt{n}(\hat{F} - F) \rightsquigarrow \mathbb{G}$ in $D[0, \tau]$, where $\mathbb{G}$ is a tight, zero-mean Gaussian process defined by

$$\mathbb{G}(t) = F((t,\infty])\int_{[0,t]} \frac{F([s,\infty])}{F((s,\infty])}\,\mathrm{d}\mathbb{G}_3$$

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

(See also van der Vaart and Wellner 1996: Lemma 3.9.30 for the derivative of the product integration.) The variance of $\mathbb{G}$ can be obtained by direct but tedious calculations and is thus omitted. □

References

Akritas, M.G. (1986) Bootstrapping the Kaplan–Meier estimator. *J. Amer. Statist. Assoc.*, **81**, 1032–1039.

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.

Cupples, L.A., Risch, N., Farrer, L.A. and Myers, R.H. (1991) Estimation of morbid risk and age at onset with missing information. *Amer. J. Hum. Genetics*, **49**, 76–87.

Fleming, T.R. and Harrington, P. (1991) *Counting Processes and Survival Analysis.* New York: Wiley.

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

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

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

Gill, R. (1994) Lectures on survival analysis. In P. Bernard (ed.), *Lectures on Probability Theory: École d'Été de Probabilités de Saint-Flour XXII – 1992*, Lecture Notes in Math. 1581, pp. 115–241. Berlin: Springer-Verlag.

Gill, R. and Johansen, S. (1990) A survey of product-integration with a view towards application in survival analysis. *Ann. Statist.*, **18**, 1501–1555.

Gu, M.G. and Zhang, C.-H. (1993) Asymptotic properties of self-consistent estimators based on doubly censored data. *Ann. Statist.*, **21**, 611–624.

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. *Amer. J. Psychiatry*, **132**, 1155–1163.

Huang, J. (1999) Asymptotic properties of nonparametric estimation based on partly interval-censored data. *Statist. Sinica*, **9**, 501–519.

Huang, J. and Wellner, J.A. (1997) Interval censored survival data: A review of recent progress. In D.Y. Lin and T.R. Fleming (eds), *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, pp. 123–169. New York: Springer-Verlag.

Kim, J.S. (2003) Maximum likelihood estimation for the proportional hazards models with partly interval-censored data. *J. Roy. Statist. Soc. Ser. B*, **65**, 489–502.

Klein, J.P. and Moeschberger, M.L. (1997) *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer-Verlag.

Patilea, V. and Rolin, J.-M. (2004) Product-limit estimators of the survival function with left or right censored data. Working Paper 2004–36, CREST-INSEE, Paris.

Pollard, D. (1984) *Convergence of Stochastic Processes*. New York: Springer-Verlag.

Samuelsen, S.O. (1989) Asymptotic theory for non-parametric estimators from doubly censored data. *Scand. J. Statist.*, **16**, 1–21.

Stute, W. and Wang, J. (1993) The strong law under random censorship. *Ann. Statist.*, **21**, 1591–1607.

Turnbull, B.W. (1974) Nonparametric estimation of a survivorship function with doubly censored data. *J. Amer. Statist. Assoc.*, **69**, 169–173.

Turnbull, B.W. and Weiss, L. (1978) A likelihood ratio statistic for testing goodness of fit with randomly censored data. *Biometrics*, **34**, 367–375.

van der Laan, M.J. and Gill, R.D. (1999) Efficiency of NPMLE in nonparametric missing data models. *Math. Methods Statist.*, **8**, 251–276.

Van der Vaart, A.W. and Wellner, J.A. (1996) *Weak Convergence and Empirical Processes*. New York: Springer-Verlag.

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

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.

Received January 2005 and revised February 2006