

Nonparametric Estimation of the Bivariate Survival Function with Truncated Data.

Mark J. van der Laan
Dept. of Biostatistics, University of California at Berkeley

May 15, 2001

Abstract

Randomly left or right truncated observations occur when one is concerned with estimation of the distribution of time between two events and when one only observes the time if one of the two events falls in a fixed time-window, so that longer survival times have higher probability to be part of the sample than short survival times. In important AIDS-applications the time between seroconversion and AIDS is only observed if the person did not die before the start of the time-window. Hence, here the time of interest is truncated if another related time-variable is truncated. This problem is a special case of estimation of the bivariate survival function based on truncation by a bivariate truncation time, the problem covered in this paper; in the AIDS-application one component of the bivariate truncation time-vector is always zero. In this application the bivariate survival function is of interest itself in order to study the relation between time till AIDS and time between AIDS and death. We provide a quick algorithm for computation of the NPMLE. In particular, it is shown that the NPMLE is explicit for the special case when one of the truncation times is zero, as in the aids-application above. We prove that the NPMLE is consistent under the minimal condition that $\int dF/G < \infty$. Moreover, we prove asymptotic normality under a tail assumption at the origin which is an empirical analogue of $\int dF/G < \infty$. The condition holds in particular if the truncation distribution has an atom at zero. We provide an algorithm for estimation of its limiting variance. By simply plugging in one of the several proposals for estimation of the bivariate survival function based on right-censored data in the estimating equation we obtain an estimator based on right-censored randomly truncated data. Here, substitution of an estimator which handles the right-censoring efficiently leads to an efficient estimator.

1 Introduction.

We will start with an introduction to the univariate random truncation model. Suppose one is concerned with estimation of the distribution of the survival time T from AIDS till death. For this purpose one has available a database of AIDS-patients regularly visiting the hospital from 1978-1995. If we assume (for the moment) that there is no right-censoring, then for all these patients we are able to establish time at which they got AIDS and time at which they died; in other words, we will observe T . However, this is a clear case of biased sampling since patients with a short survival time T are less likely to be part of the data-base than patients with a long survival time; to be precise, if we define $C \equiv 1978 - T_{AIDS}$ if $T_{AIDS} < 1978$, where T_{AIDS} is the time at which the patient got AIDS, and $C = 0$ if $T_{AIDS} > 1978$, then a patient will only be part of the sample if $T \geq C$.

Hence the problem is to estimate the distribution of a random survival time T with survival function S , based on n i.i.d. random draws from the conditional distribution of (C, T) ,

given $T \geq C$ (left-truncation) or, given $T \leq C$ (right-truncation), where it is assumed that $C \sim G$ is independent of T . For the moment we will restrict our attention to left-truncation; results are trivially generalized to right-truncation. We will denote the observations with T'_i, C'_i , $i = 1, \dots, n$, where the $'$ is used to indicate that the observations are random draws from the conditional distribution of (C, T) , given $T \geq C$. The maximum likelihood estimator of the survival function S of T is the well known product limit estimator given by:

$$S_n(t) = \prod_{(0,t]} (1 - \Lambda_n(ds)),$$

where

$$\Lambda_n(ds) = \frac{\sum_{i=1}^n I(T'_i \in ds)}{\sum_{i=1}^n I(T'_i \geq s, C'_i \leq s)},$$

estimates the the hazard probability $\Lambda(ds) \equiv P(T \in ds \mid T \geq s)$. Asymptotic results of this estimator have been obtained by Woodrooffe (1985), Wang, Jewell and Tsai (1986), Keiding and Gill (1990) and van der Vaart (1991); under the assumption that $\int dF/G < \infty$ and $\int dG/S < \infty$ the estimator is asymptotically efficient. Moreover, if T'_i is right-censored by a censoring variable C_i^* , then we simply estimate Λ by:

$$\Lambda_n(ds) = \frac{\sum_{i=1}^n I(T'_i \in ds, \Delta_i = 1)}{\sum_{i=1}^n I(T'_i \wedge C_i^* \geq s, C'_i \leq s)},$$

where $\Delta_i = I(T'_i \leq C_i^*)$. In other words, the estimator is trivially generalized to right-censored truncated data.

Consider now the following application. In hemophilia AIDS-data sets the time of sero-conversion can be quite accurately determined since an hemophilia patient has to donate blood regularly. Hence these data sets are very good for estimation of the distribution of time T_1 between seroconversion and AIDS. However, again, a database will cover patients from, say 1978, till 1995, and hence a patient with a longer survival time will have a larger probability of being part of the sample than a patient with a short survival time. To be precise, let T_2 be the time between sero-conversion and death and let $C_2 = 1978 - T_{\text{sero}}$ if $T_{\text{sero}} < 1978$ and $C_2 = 0$ if $T_{\text{sero}} \geq 1978$. Then a patient will only be part of the sample if $T_2 \geq C_2$. In other words, we observe (T_1, T_2) and C_2 , given $T_2 \geq C_2$.

Hence the problem is to estimate the marginal distribution of T_1 based on n i.i.d. random draws from the conditional distribution of $(T_1, T_2), C_2$, given $T_2 \geq C_2$. Since T_1 is truncated by the event $T_2 \geq C_2$ instead of by a C_1 itself, this problem cannot be solved directly with the knowledge we have on the univariate truncation model. However, this problem is a special case of the following bivariate problem covered in this paper.

Let $C = (C_1, C_2) \sim G$, $T = (T_1, T_2) \sim F$ be independent bivariate random vectors. We observe n i.i.d. copies of (C, T) , given $T \geq C$; in other words, we only observe (C, T) if $T_1 \geq C_1$ and $T_2 \geq C_2$. We will refer to this data structure as bivariate truncated data. In this paper we are concerned with nonparametric maximum likelihood estimation of the bivariate survival function based on bivariate truncated data. This solves the application above by setting $C_1 = 0$ with probability 1 since it provides us with an estimate of the bivariate survival function of time T_1 between sero-conversion and AIDS and time T_2 between sero-conversion and death and hence, in particular, it provides us with the marginal distribution of T_1 .

The estimator is directly generalized to bivariate right-censored truncated data, using the rich amount of work carried out for estimation of the bivariate survival function based on right-censored data. Estimation of the bivariate survival function based on right-censored data has

been an extensively studied subject. In this problem the NPMLE is inconsistent so that several authors constructed representations of the bivariate survival functions in terms of quantities which can be directly estimated from the data. One of the nicest representations which resulted in good (better than the other explicit estimators) estimators are due to Dabrowska (1988,1989) and Prentice and Cai (1992a, 1992b); they represent the bivariate survival function as $S_1 S_2 R$, where R is a functional of three bivariate hazards and S_1, S_2 can be estimated with the Kaplan-Meier estimators. It is important to notice here (see Gürlér, 1994) that these representations do *not* (at least, not directly) provide us with consistent estimators of S based on bivariate truncated data. The problem is that for estimation of the marginals in these representations we need to estimate the hazard $P(T_1 \in ds)/P(T_1 \geq s)$ with n i.i.d. copies of $(T_1, T_2), (C_1, C_2)$, given $(T_1, T_2) \geq (C_1, C_2)$. This can only be done in the classic way (as we did above for the univariate random truncation problem) if T_1 and T_2 are independent. In other words, the product limit estimator for the marginal S_1 breaks down if T_1 is only observed if $T_1 \geq C_1$ and $T_2 \geq C_2$, where T_2 is related to T_1 .

In Gürlér (1994) a modification of Burke's estimate (Burke, 1988) of the bivariate survival function for right-censored data is developed which makes it applicable to the case where only one variable is subject to truncation. She shows that (also here) the approach cannot be generalized to bivariate truncation.

In the bivariate truncation model there is no evidence that the NPMLE does not work, in contrary to the bivariate right-censoring model; though, in the right-censoring model modifications of the NPMLE as the asymptotically efficient one in van der Laan (1995) and the one proposed by Pruitt (1991) are good candidates to use in practice. In this paper we will determine the set of empirical score equations for the NPMLE F_n , show how the solution can be quickly computed, and we will prove consistency and asymptotic normality of F_n by showing that the solution F_n is a smooth functional of the empirical distribution function. The univariate version of the estimating equation is solved by the product limit estimator. Using this, it will be shown that if $C_2 = 0$, i.e. we only have left-truncation on the first variable, then the NPMLE is explicit. So our AIDS-application above can be solved with an explicit estimator. The right-truncation and right-censored versions of our results will be given. A data-analysis using these bivariate survival function estimators will appear elsewhere. We will prove uniform consistency of F_n on a compact interval $[0, \tau]$ on which S is bounded away from zero under the minimal condition that $\int dF/G < \infty$. Because martingale arguments for bivariate processes are not available and that the NPMLE is implicit we need in the root- n proof a stronger version of the condition $\int_0^\infty dF/G$: see the theorem.

The reason that the NPMLE for bivariate truncation does *not* fail in contrary to the NPMLE for bivariate right-censored data is that the truncation is monotone in the sense that we only observe T if $T > C$ and that we observe nothing if $T_1 < C_1, T_2 > C_2$. If T_1 is randomly truncated by a C_1 and T_2 is randomly truncated by a C_2 (i.e. we might observe T_1 , but not T_2 and vice versa), then one expects that the same modifications (i.e. smoothing) of the NPMLE as needed for bivariate right-censored data will be necessary. We are not aware of an application where the latter type of bivariate truncation occurs, but, for example, in problems where T_2 is a covariate which is always observed and T_1 is randomly truncated this would be the model to investigate.

2 Efficient Score equations for the NPMLE.

We will consider the NPMLE F_n of F as a solution of a set of score equations corresponding with one dimensional submodels through the NPMLE itself, where the scores of these one-dimensional models are orthogonal to the scores for the nuisance parameter G .

Let $\alpha \equiv P(C \leq T)$. The distribution of the observed (C', T') is given by:

$$P_{F,G}(C' \in dc, T' \in dt) = \frac{1}{\alpha} dF(t) dG(c) I(c \leq t).$$

Let $F \ll \mu_1$, $G \ll \mu_2$ and denote the corresponding densities with f and g , respectively. If we write $F_1 \ll_b F_2$ for two measures F_1, F_2 , then we mean that F_1 is absolutely continuous w.r.t. F_2 and that dF_1/dF_2 is bounded. For each $F_1 \ll_b F$ we define a line $f_{h_1}(\epsilon) = (1 + \epsilon h_1)f$ from F_1 to F , where $h_1 = (f_1 - f)/f \in L_0^2(F)$. Because h_1 is bounded it follows that $f_{h_1}(\epsilon)$ is also a well defined density for $\epsilon \in [-\delta, 1]$ for some $\delta > 0$.

Similarly, for each $G_1 \ll_b G$ we define a line $g_{h_2}(\epsilon) = (1 + \epsilon h_2)g$ from G_1 to G , where $h_2 = (g_1 - g)/g \in L_0^2(G)$. These lines imply a one-dimensional submodel $P_{f_{h_1}(\epsilon), g_{h_2}(\epsilon)}$ through $P_{f,g}$ with score $A_{F,G}(h_1) + B_{F,G}(h_2)$, where the score operator for F is given by:

$$A_{F,G} : L_0^2(F) \rightarrow L_0^2(P_{F,G}) : h_1 \rightarrow h_1(T') - P_{F,G}h_1$$

and the score operator for G is given by

$$B_{F,G} : L_0^2(G) \rightarrow L_0^2(P_{F,G}) : h_2 \rightarrow h_2(C') - P_{F,G}h_2.$$

Let P_n be the empirical distribution function. Let F_n, G_n be the NPMLE. By differentiating the one-dimensional loglikelihood

$$\epsilon \rightarrow \int \log \left(\frac{dP_{F_n, h_1(\epsilon), G_n, h_2(\epsilon)}(c, t)}{dP_{F_n, G_n}}(c, t) \right) dP_n(c, t),$$

corresponding with a one-dimensional submodel through P_{F_n, G_n} , it follows that the NPMLE F_n solves the score equations

$$\int A_{F_n, G_n}(h_1) dP_n = 0 \text{ and } \int B_{F_n, G_n}(h_2) dP_n = 0$$

for any $h_1 \in L_0^2(F_n)$, $h_2 \in L_0^2(G_n)$, h_1, h_2 bounded. One should always consider the NPMLE as an estimator of the so called efficient score equations which are obtained by subtracting the projection of $A_{F,G}(h)$ on the closure (in $L_0^2(P_{F,G})$) of the range of $B_{F,G}$ from $A_{F,G}(h)$. The main reason for this is that the efficient score is orthogonal to the nuisance tangent space of G which implies that its derivative w.r.t. G is zero and often it does not even depend on G (as is the case here and in most censoring models). To be formal, we define the *efficient score operator* $A_{F,G}^* : L_0^2(F) \rightarrow L_0^2(P_{F,G})$ as:

$$A_{F,G}^*(h) = A_{F,G}(h) - \Pi(A_{F,G}(h) | T_2(P_{F,G})),$$

where $T_2(P_{F,G})$ is the closure of the range of $B_{F,G}$ and Π is the projection operator.

The information for estimation of ϵ in $P_{F_{h_1}(\epsilon), G}$ in the model with G unknown is given by the variance of the efficient score (see Bickel, Klaassen, Ritov, Wellner, 1993) and consequently the generalized Cramér-Rao lower bound for estimation of $F(t) = \int_0^t dF_\epsilon(x) |_{\epsilon=0}$ is given by:

$$\sup_{h_1 \in L_0^2(F)} \left(\frac{\int_0^t h_1 dF}{\|A_{F,G}^*(h_1)\|_{P_{F,G}}} \right)^2. \quad (1)$$

It is straightforward to verify that the $L^2(P_{F,G})$ -projection of $A_{F,G}(h_1)$ on the closure $T_2(P_{F,G})$ of the range of $B_{F,G}$ is given by

$$\Pi(h_1 - P_{F,G}h_1 | T_2(P_{F,G})) = \Pi(h_1 | T_2(P_{F,G})) = E_F(h_1(X') | C') - P_{F,G}h_1.$$

Hence the efficient score operator $A_{F,G}^*$ for F is given by:

$$A_F^*(h_1)(T', C') = h_1(T') - E(h_1(T') | C') = h_1(T') - \frac{\int_{C'}^{\infty} h_1 dF}{S(C')},$$

where the integral is over $[C', \infty)$ and

$$S(x_1, x_2) \equiv 1 - F(x_1-, \infty) - F(\infty, x_2-) + F(x_1-, x_2-). \quad (2)$$

The definition of the bivariate survival function given by (2) is useful for us because throughout S and S_n will appear in denominators so that we do not have to use the notation $S(c-)$, but instead just use $S(c)$.

Similarly, we find that the efficient score operator for right-truncation is given by:

$$A_{F,r}^*(h_1)(T', C') = h_1(T') - E(h_1(T') | C') = h_1(T') - \frac{\int_0^{C'} h_1 dF}{F(C')}.$$

F_n solves

$$0 = P_n A_{F_n}^*(h_1) \equiv \int A_{F_n}^*(h_1)(t, c) dP_n(t, c) = \frac{1}{n} \sum_{i=1}^n A_{F_n}^*(h_1)(T'_i, C'_i) \quad (3)$$

for all $h_1 \in L_0^2(F_n)$. Similarly, the NPMLE F_n with right-truncation solves $P_n A_{F_n,r}^* = 0$. All our statements and derivations in this paper are similarly applied to F_n for right-truncated data. From now on we will restrict our attention to left-truncation.

In particular, the estimating equation (3) holds for $h_1 = I(T' \geq t) - S_n(t)$ for all t which provides us with:

$$\frac{1}{n} \sum_{i=1}^n I(T'_i > t) - \frac{1}{n} \sum_{i=1}^n \frac{S_n(C'_i \vee t)}{S_n(C'_i)} = 0.$$

Let G_n^e be the empirical of the C'_i , $i = 1, \dots, n$ and let F_n^e be the empirical of the T'_i , $i = 1, \dots, n$. Then the estimating equation is presented by:

$$\int \frac{S_n(c \vee t)}{S_n(c)} dG_n^e(c) = S_n^e(t). \quad (4)$$

By considering the left and right-hand side as measures in $t \in \mathbb{R}_{\geq 0}^2$ and computing the measure given to $(t, t + dt]$ we obtain the equality:

$$S_n(dt_1, dt_2) \int_0^{t_1} \int_0^{t_2} \frac{dG_n^e(c_1, c_2)}{S_n(c_1, c_2)} = S_n^e(dt_1, dt_2), \quad (5)$$

where $S_n(dt_1, dt_2)$ stands for the pointmass

$$S_n(t_1 + dt_1, t_2 + dt_2) - S_n(t_1 + dt_1, t_2) - S_n(t_1, t_2 + dt_2) + S_n(t_1, t_2)$$

which S_n gives to point (t_1, t_2) .

Apparently, the support of the NPMLE is uniquely determined, namely F_n puts mass solely on T'_i , $i = 1, \dots, n$. If the NPMLE exists, then it solves this equation. Proving existence of the NPMLE is standard and straightforward; we know its support so that it is defined as the one which maximizes the loglikelihood over a vector of pointmasses, existence follows now from continuity and compactness arguments, see e.g. van der Laan (1993, chapter 3).

Just as in the univariate version of this equation terms cancel out towards each other leading to (5). Notice that if we replace P_n by $P_{F,G}$ and S_n by S , then indeed the equation holds; i.e. the estimating equation is unbiased as it should be. Notice also that because $\alpha dG^e(c) = P(T \geq c)dG(c)$ the survival function S_n in the denominator is indeed defined as in (2). In terms of $F_n(t)$ this equation says:

$$F_n(t_1, t_2) = \int_0^{t_1} \int_0^{t_2} \frac{F_n^e(ds_1, ds_2)}{\int_0^{s_1} \int_0^{s_2} \frac{dG_n^e(c_1, c_2)}{S_n(c_1, c_2)}}. \quad (6)$$

The univariate version of the equation (5) is:

$$S_n(dt) \int_0^t \frac{dG_n^e(c)}{S_n(c)} = S_n^e(dt), \quad (7)$$

which is solved by the product limit estimator as can be verified, as it should since the product limit estimator is the NPMLE in the univariate problem. In this case S_n can be quickly computed by using this equation: Firstly, $S_n(dt)$ puts only mass at the observed T'_i , $i = 1, \dots, n$. Now, we can order the observations T'_i from small to large. Since $S_n(c) = 1$ for $c \leq T'_{(1)}$ the equation provides us directly with $S_n(dT'_{(1)})$. This provides us also with S_n between 0 and T_2 . Hence the equation provides us directly with $S_n(dT'_2)$ and we can proceed in this way till we arrive at the last observation $T'_{(n)}$. In other words, in n trivial steps we can compute the the n jumps of the survival function. A similar approach fails for the general bivariate equation since we do not have a total ordering on $\mathbb{R}_{\geq 0}^2$. However, the following iterative algorithm can be expected to converge exponentially fast as will be made clear in section 3 where we show that the derivative of the equation can be exponentially fast inverted in the same iterative way:

$$S_n^{k+1}(dt) = \frac{S_n^e(dt)}{\int_{(0,t]} \frac{G_n^e(dc)}{S_n^k(c)}}. \quad (8)$$

So one just chooses an initial estimator $S_n^0(dt)$ which puts mass on the observed T'_i and we iterate the equation till convergence is established.

2.1 Special case: one truncation time is zero.

Suppose now that $C_2 = 0$ with probability 1. Then the estimating equation (5) becomes:

$$S_n(dt_1, dt_2) \int_0^{t_1} \frac{dG_n^e(c_1)}{S_n(c_1, 0)} = S_n^e(dt_1, dt_2). \quad (9)$$

Integrating both sides over t_2 provides us with:

$$S_n(dt_1, 0) \int_0^{t_1} \frac{dG_n^e(c_1)}{S_n(c_1, 0)} = S_n^e(dt_1, 0).$$

However, this is just the efficient score equation (7) for the univariate left-truncation model and hence $S_n(t_1, 0)$ is simply the product-limit estimator based on the marginal sample from (C'_1, T'_1) . This is not a surprising result since intuitively T'_2 does not help in estimating the marginal distribution of T_1 ; formally this follows by the fact that the marginal of T_1 factorizes out in the likelihood. Substitution of the product limit estimator in the bivariate equation (9) tells us that

$$S_n(dt_1, dt_2) = \frac{S_n^e(dt_1, dt_2)}{\int_0^{t_1} \frac{dG_n^e(c_1)}{S_{PL}(c_1, 0)}}, \quad (10)$$

where S_{PL} stands for the product limit estimator. In other words, we have:

$$F_n(t_1, t_2) = \frac{1}{n} \sum_{i=1}^n \frac{I(T'_{1i} \leq t_1, T'_{2i} \leq t_2)}{1/n \sum_{j=1}^n I(C_{1j} \leq T'_{1i}) / S_{PL}(C_{1j}, 0)}.$$

Consequently, for this special case the NPMLE S_n of the bivariate survival function S is explicit. This shows that our AIDS-application can be explicitly solved: here we want to have the marginal $F_n(\infty, t_2)$:

$$F_n(\infty, t_2) = \int_0^\infty \frac{F_n^e(ds_1, t_2)}{\int_0^{s_1} \frac{dG_n^e(c_1)}{S_{PL}(c_1, 0)}}. \quad (11)$$

2.2 Right-censored randomly truncated data.

In many practical situations one will have right-censored truncated data. In other words, we observe $(\tilde{T}_i, \Delta_i, C_i)$, where $\tilde{T}_i = T_i \wedge C_i^*$ and $\Delta_i = I(T_i \leq C_i^*)$, where the bivariate random variable C_i^* is independent of the bivariate T_i and bivariate C_i ; here T_i, C_i are observations from (T, C) , given $T \geq C$. We can estimate S^e , based on \tilde{T}_i, Δ_i , with Dabrowska's estimator or with the modified NPMLE of van der Laan (1995) or others. Now, we substitute this estimate \hat{S}^e for S_n^e in the estimating equation (5). If C_i is also right-censored, then we can substitute an estimate \hat{G}^e for G_n^e as well. This provides us with an estimator S_n^{cens} of S based on right-censored truncated data. In our analysis below we show that S_n is a smooth functional of G_n^e, S_n^e and by the fact that it is a NPMLE we have that S_n is asymptotically efficient. Hence S_n^{cens} is also a smooth functional of G_n^e, \hat{S}^e so that, by the functional delta-method, the consistency and weak convergence results for \hat{S}^e immediately translate to S_n^{cens} . Moreover, if \hat{S}^e is asymptotically efficient, then S_n^{cens} is asymptotically efficient; this follows by a result of van der Vaart (1991) which says that a compactly differentiable functional of an efficient estimator is efficient.

3 Analysis of the NPMLE.

We have for all $t \in \mathbb{R}_{\geq 0}^2$

$$U(F, P_{F,G})(t) \equiv F(t) - \int_{(0,t]} \frac{F^e(ds)}{\int_{(0,s]} dG^e(c)/S(c)} = 0. \quad (12)$$

Equation (6) tells us that the NPMLE F_n solves $U(F_n, P_n)(t) = 0$ for all t , where P_n is the empirical distribution function of $P_{F,G}$.

This equation involves two singularities. In the end tail we have the singularity $1/S$. Moreover, notice that the denominator $\int_{(0,s]} dG^e(c)/S(c)$ converges to zero if $s \rightarrow 0$ as $G(s)$. Therefore we will have to control this singularity with an assumption like the classic univariate assumption $\int dF/G < \infty$ (see Woodroffe, 1985).

In order to control the first singularity we assume in the analysis that the support of F is restricted to a rectangle $[0, \tau] \subset [0, \infty)^2$, $\tau = (\tau_1, \tau_2)$, where τ is chosen so that $S(\tau-) > \delta > 0$ (or just $S(\tau) > \delta > 0$ if S is defined as in (2)), which holds for example if $F(\{\tau\}) > 0$. As shown below this assumption happens to be no assumption for estimation of $F(t)$ with $S(t) > 0$, but it is used to go through an analysis where F_n is considered as a whole random function.

The analytically important implication is that now S in $U(F, P)$ is uniformly bounded away from zero on $[0, \tau]$; recall that S in the denominators in U is defined by $S(\tau) = 1 - F(\tau_1-, \infty) -$

$F(\infty, \tau_2-) + F(\tau_1-, \tau_2-)$ (see 2). Moreover, it guarantees that $S_n^e(\{\tau\}) > \delta > 0$ with probability tending to 1 so that equation (5) tells us that

$$S_n(\{\tau\}) = \frac{S_n^e(\{\tau\})}{\int_0^\tau dG_n^e/S_n} \geq S_n(\tau)S_n^e(\{\tau\}).$$

Hence the bivariate hazard (at τ) $F_n(\{\tau\})/(1 - F_n(\tau-)) > \delta > 0$ with probability tending to 1 which implies that F_n has an atom at τ which is bounded away from zero (uniformly in n). Consequently, the denominator $S_n(c)$ in $U(F_n, P_n)$ is uniformly bounded away from zero on $[0, \tau]$. This will control the end-tail singularity in the analysis below.

This assumption is accomplished by artificially pulling back each T_i' and C_i' which does not fall in the rectangle $[0, \tau]$ to the closest point on the edge of $[0, \tau]$; notice that this does not change the order $C_i' \leq T_i'$. For estimation of $F(t_1, t_2)$ with $t < \tau$ this reduction of the data does *not* change the NPMLE, because equation (6) tells us that $F_n(t)$ does only depend on the data through F_n^e and G_n^e on $[0, t]$ which are *not* changed by the artificial censoring! In other words, our derived results uniformly on $[0, \tau]$ imply the same results for the (unchanged) NPMLE on $[0, \tau]$ without the artificial censoring. Our theorem below shows our results as implied for the original NPMLE: We just assume that $S(\tau) > \delta > 0$ instead of $S(\tau-) > \delta > 0$. Then for some bivariate $\epsilon > 0$ we can do the artificial censoring (as explained above) at the edge of the rectangle $[0, \tau + \epsilon]$ so that $S(\tau + \epsilon-) > \delta > 0$ as required in the analysis. By the argument given above the artificial censoring does not change the NPMLE on $[0, \tau]$ so that our proved results on the artificially censored F_n on $[0, \tau + \epsilon]$ prove results on the original NPMLE on $[0, \tau]$.

The second singularity at zero happens to be a real one in the sense that it affects also the NPMLE $S_n(t)$ at t not close to zero. Because of the implicitness of F_n and the fact that no martingale arguments are available for bivariate processes we will need a more stringent assumption than $\int dF/G < \infty$, as will be discussed in detail below.

Our analysis is following a standard M-estimator approach as highlighted in van der Vaart (1992). This general method works as follows. Because $U(F_n, P_n) = U(F, P) = 0$ we have

$$U(F_n, P) - U(F, P) = -(U(F_n, P_n) - U(F_n, P)). \quad (13)$$

Let $(D[0, \tau], \|\cdot\|_\infty, \mathcal{B})$ be the space of bivariate cadlag functions as defined in Neuhaus (1972), i.e. bivariate real valued functions f which are right-continuous and for which the left-hand limits $f(s-, t-)$, $f(s-, t)$, $f(s, t-)$ exist, endowed with the supremum-norm and the Borel-sigma-algebra. We consider estimators, say X_n , as random (not necessarily measurable) elements of this space.

We will first prove uniform consistency of F_n in subsection 2. Here we will only need that $\int dF/G < \infty$ and dG/dF has bounded supnorm on $[0, \tau]$.

Since F appears in $U(F, P)$ only as a function it is straightforwardly verified in subsection 3 that $F \rightarrow U(F, P)$ is Fréchet-differentiable: for any sequence F_n s.t. $\|F_n - F\|_\infty \rightarrow 0$ we have

$$\frac{1}{\|F_n - F\|_\infty} (U(F_n, P) - U(F, P) - d_1U(F, P)(F_n - F)) \rightarrow 0$$

w.r.t. the supnorm, where $d_1U(F, P)$ is a linear mapping which will be precisely specified. Also here we only need that $\int_0^\tau dF/G < \infty$.

In subsection 5 we show that \sqrt{n} times the right-hand side, which we will denote with $-Z_n$, converges weakly in $(D[0, \tau], \|\cdot\|_\infty, \mathcal{B})$ to a Gaussian process, hereby using the uniform consistency of S_n , the supnorm-weak convergence of $\sqrt{n}(P_n - P)$ and the smoothness of $U(F, P)$ in P . Because all mass of F lies on $[0, \tau]$ S depends only through F on $[0, \tau]$ and hence the

uniform consistency of $F_n(S_n)$ suffices here. This supnorm-weak convergence analysis will be carried out in the weak convergence subsection 3.

Here the proof can only be carried out under the following conditions: we need that the class of functions \mathcal{F} , where \mathcal{F} consists of bivariate monotone (i.e. distribution) functions bounded by $1/G$ is a F^e -Donsker class (see van der Vaart, Wellner, 1995). It is well known that if \mathcal{F} has a uniformly bounded envelope, then it is a Donsker class. Our envelope $1/G$ has a singularity at 0. Because of this singularity a Donsker class proof requires a tightness argument at 0. Van der Vaart and Wellner (1995) prove that \mathcal{F} , where \mathcal{F} is the set of *univariate* monotone functions with envelope g , $Pg^{2+\epsilon} < \infty$, is a P -Donsker class. Their result can be generalized to bivariate functions which proves that if for some $\epsilon > 0$

$$\int \frac{dF^e}{G^{2+\epsilon}} = \int \frac{dF}{G^{1+\epsilon}} < \infty,$$

then \mathcal{F}/G is a F^e -Donsker class.

Secondly, we need that

$$\int_c^\tau \frac{dF_n^e}{GG_n^e} \leq M(c) \tag{14}$$

with probability tending to 1, where the class of bivariate monotone functions with envelope $M(\cdot)$ should be a G^e -Donsker class. The generalization of the univariate result in van der Vaart and Wellner (1995) shows that this holds if for some $\epsilon > 0$ $\int M(c)^{2+\epsilon} dG^e(c) < \infty$.

Notice that

$$\int \frac{dF^e}{GG^e} = \int \frac{GdF}{G \int_0^s SdG} \leq M \int dF/G$$

because $1/S < M$ for some $M < \infty$. In other words, our condition (14) is the empirical counter part of the all the time needed tail condition that $\int dF/G < \infty$. We are not aware of a result in empirical process theory which covers the result (27). If G has an atom at 0, then (27) holds with $M(c) = M < \infty$, but it is clear that a weaker condition should suffice here. Notice that if for a fraction of the subjects the startpoints from where we start measuring T_1 and T_2 fall in the observed time-window, i.e. $P(T_{AIDS} > 1978) > 0$ in the aids-application discussed in the introduction, then G has indeed an atom at $\{0\}$ and thus even the atom-assumption is very realistic in practice.

By the usual kind of argument (see e.g. van der Vaart, 1992) for M-estimators it follows now that

$$d_1U(F, P)(\sqrt{n}(F_n - F)) = -Z_n + o_P(1).$$

It remains to prove that $d_1U(F, P)$ has a bounded inverse; then $\sqrt{n}(F_n - F) = -d_1U(F, P)^{-1}(Z_n) + o_P(1)$ so that the weak convergence of Z_n implies, by the continuous mapping theorem, weak convergence of $\sqrt{n}(F_n - F)$. This will be proved in subsection 4 by using that the derivative is of the type $I - A$, I identity operator, and exploiting that the operator A has a nice Volterra structure so that the Neumann-series of A converges exponentially fast. Again, here we only need $\int_0^\tau dF/G < \infty$ and dG/dF has bounded supnorm on $[0, \tau]$.

Our result states supnorm weak convergence of $\sqrt{n}(S_n - S)$ as random elements in a function space endowed with the supnorm. $(D[0, \tau], \|\cdot\|_\infty, \mathcal{B})$ is a non-separable space. In this case the Borel-sigma algebra is very large and therefore X_n will usually not be measurable. On the other hand, for all known applications the limit random variable X_0 lies in a separable (sub)space and thereby will be measurable w.r.t. the Borel sigma-algebra, except for some pathological cases.

Because we are only concerned with the asymptotic behavior of X_n , only “asymptotic measurability” should be relevant. Indeed there exists a powerful weak convergence theory for non-separable spaces without giving up the Borel sigma-algebra, but giving up that X_n induces a distribution on the Borel-sigma algebra. Weak convergence of X_n to X_0 in this modern sense is defined as in the traditional definition of Billingsley (1968), except that expectations and probabilities for X_n are replaced by *outer expectations* and *outer probabilities*. This weak convergence theory is due to Hoffmann-Jørgensen (1984) and Dudley (1985) following an evolution from Dudley (1966) and Wichura (1968) and is presented in full detail in van der Vaart and Wellner (1995). If Z_n converges weakly to Z in $(D[0, \tau], \|\cdot\|_\infty, \mathcal{B})$ we will denote this with $Z_n \xrightarrow{D} Z$. We refer to the well known result that the empirical process $\sqrt{n}(P_n - P_{F,G})$ indexed by the indicators $\{I(0, t) : t \in [0, \tau]\}$ converges weakly as random elements of $(D[0, \tau], \|\cdot\|_\infty, \mathcal{B})$ to a Gaussian process with the same covariance structure as $\sqrt{n}(P_n - P_{F,G})$.

Theorem 3.1 *Let $S(x_1, x_2) = 1 - F(\infty, x_2) - F(x_1, \infty) + F(x_1, x_2)$ the the survival function of F . Let τ be such that $S(\tau) > \delta > 0$ and assume that $F = F_d + F_c$, where F_d is purely discrete and F_c is continuous. Moreover, assume that $\int_0^\tau dF/G < \infty$ and dG/dF is uniformly bounded on $[0, \tau]$. Then F_n is uniformly consistent on $[0, \tau]$.*

Assume now also that the class of functions \mathcal{F} , where \mathcal{F} consists of bivariate monotone functions with envelope $1/G$ is a F^e -Donsker class. Moreover, assume that

$$\int_c^\tau \frac{dF_n^e}{GG_n^e} \leq M(c) \tag{15}$$

with probability tending to 1, where $M(\cdot)$ is such that the class of bivariate monotone functions with envelope $M(\cdot)$ is G^e -Donsker. Both conditions hold if for some $\epsilon > 0$

$$\int \frac{dF}{G^{1+\epsilon}} < \infty \text{ and } \int M(c)^{2+\epsilon} dG^e(c) < \infty.$$

Then $\sqrt{n}(F_n - F)$ converges weakly as random elements of $(D[0, \tau], \|\cdot\|_\infty, \mathcal{B})$ to a Gaussian random element. In particular, $\sqrt{n}(F_n - F)(t)$ is asymptotically normal and efficient for every $t \in [0, \tau]$.

The efficiency of F_n is a consequence of the asymptotic normality and the fact that F_n solves the score equations; see Gill, van der Vaart (1993) and van der Vaart (1992).

The condition (15) seems only verifiable in practice if G has an atom at 0. However, it shows the minimal condition under which our proof works. We will now state the direct corollary of the theorem for the case that G has an atom at zero:

Corollary 3.1 *Let $S(x_1, x_2) = 1 - F(\infty, x_2) - F(x_1, \infty) + F(x_1, x_2)$ the the survival function of F . Let τ be such that $S(\tau) > \delta > 0$ and assume that $S = S_d + S_c$, where S_d is purely discrete and S_c is continuous. Moreover, assume that $\int_0^\tau dF/G < \infty$ and dG/dF is uniformly bounded on $[0, \tau]$. Then S_n is uniformly consistent on $[0, \tau]$.*

If also $G(\{0\}) > 0$, then $\sqrt{n}(S_n - S)$ converges weakly as random elements of $(D[0, \tau], \|\cdot\|_\infty, \mathcal{B})$ to a Gaussian random element. In particular, $\sqrt{n}(S_n - S)(t)$ is asymptotically normal and efficient for every $t \in [0, \tau]$.

3.1 Essential ingredients for the consistency proof.

For the consistency proof we will need an *integration by parts* formula and a notion of bounded variation for bivariate functions, as has been done in Gill, van der Laan, Wellner (1995). This subsection summarizes these ingredients in order to make the paper self-contained.

Let $[0, \tau] \subset \mathbb{R}^2$ be a fixed rectangle. Let $f : [0, \tau] \rightarrow \mathbb{R}$ be a real valued *bivariate* function on $[0, \tau]$. The *generalized difference of f over (a, b)* is defined as

$$f(a, b) \equiv f(b_1, b_2) - f(a_1, b_2) - f(b_1, a_2) + f(a_1, a_2).$$

The *variation norm of f* , which will be denoted with $\|f\|_{\mathbf{v}}$, is defined as the supremum over all *lattice* (rectangular) partitions of $[0, \tau]$ of the sum of the absolute values of the generalized differences of f over the elements of the partition; let $\{A_{i,j}\}$ be a collection of disjoint rectangles forming a lattice-partition of $[0, \tau]$, then

$$\|f\|_{\mathbf{v}} \equiv \sup_{\{A_{i,j}\}} \sum_{i,j} |f(A_{i,j})|. \quad (16)$$

If $\|f\|_{\mathbf{v}} < \infty$, then we say that f is of *bounded variation*. We will say that $f : [0, \tau] \rightarrow \mathbb{R}$ is of *bounded uniform sectional variation* if

$$\|f\|_{\mathbf{v}}^* \equiv \max \left(\|f\|_{\infty}, \|f\|_{\mathbf{v}}, \sup_u \|v \rightarrow f(u, v)\|_{\mathbf{v}}, \sup_v \|u \rightarrow f(u, v)\|_{\mathbf{v}} \right) < \infty. \quad (17)$$

Let $(D[0, \tau], \|\cdot\|_{\infty})$ be the space of bivariate real valued functions defined on $[0, \tau]$ which are right-continuous with left-hand limits (cadlag), as defined in Neuhaus (1971), endowed with the supremum norm. If f is a bivariate cadlag function on $[0, \tau]$ which is of bounded variation, then it generates a signed measure on the Borel sigma-algebra on $[0, \tau]$ (see Hildebrandt, 1963, p. 108). Moreover, we have the following integration by parts formula:

Lemma 3.1 (Integration by parts). *Let $f, g \in D[0, \tau]$ and $\|f\|_{\mathbf{v}}^* < \infty, \|g\|_{\mathbf{v}} < \infty$.*

$$\begin{aligned} \int_0^s \int_0^t f(u, v) g(du, dv) &= \int_0^s \int_0^t g([(u, s) \times (v, t)]) f(du, dv) \\ &+ \int_0^s g([(u, s) \times (0, t)]) f(du, 0) + \int_0^t g([(0, s) \times [v, t]]) f(0, dv) \\ &+ f(0, 0) g([(0, s) \times (0, t)]). \end{aligned}$$

For this we refer to Gill, van der Laan and Wellner (1995) or for the k -variate case ($k \geq 2$) to Gill (1993). This provides us with the following lemma:

Lemma 3.2 *Let f and g be two bivariate cadlag functions and suppose that $\|f\|_{\mathbf{v}}^* < \infty$. Then*

$$\int_{[0, \tau]} f dg \leq 16 \|f\|_{\mathbf{v}}^* \|g\|_{\infty}.$$

Here, if g is not of bounded variation, then the left-hand side is defined by integration by parts. The following lemma is useful:

Lemma 3.3 *Let $f : \mathbb{R}^2 \rightarrow \mathbb{R}$. If $\|f\|_{\mathbf{v}}^* < \infty$ and $f > \delta > 0$, then $\|1/f\|_{\mathbf{v}}^* < \infty$.*

The proof requires some combinatorial arguments following directly from the definition (16) of $\|\cdot\|_{\mathbf{v}}$ (it is sketched for general k in Gill, 1993).

A useful trick, which is just the equivalent of the product rule for differentiating discrete functions, is *telescoping* the difference of two products of terms as a sum of products containing one difference a time:

$$\prod_{i=1}^k a_i - \prod_{i=1}^k b_i = \sum_{i=1}^k \prod_{j=1}^{i-1} a_j (a_i - b_i) \prod_{j=i+1}^k b_j.$$

3.2 Consistency.

We will first prove consistency of F_n on $[0, \tau]$. For the consistency proof we consider F_n as the solution of (4)

$$H(S_n, P_n)(t) \equiv \int \frac{S_n(c \vee t)}{S_n(c)} dG_n^e(c) - S_n^e(t) = 0.$$

We will assume that S has only a finite number of discontinuity points, i.e. point masses, and is continuous everywhere else. This implies the same for S^e . Since S_n has the same support as S_n^e it follows that $S_n = S_n^d + S_n^c$, where S_n^d is discrete on the pointmasses of S^e and S_n^c has the same support as S_n^e , excluding the pointmass points. Moreover, S_n^d will be uniformly bounded away from zero; this is shown in the same way as we showed above that $S_n(\{\tau\})$ is uniformly bounded away from zero. By Helly's selection theorem S_n^c has a subsequence which converges pointwise to a distribution S_∞^c on $[0, \tau]$ at each continuity point of S_∞ . Hence by continuity of S^e S_n^c converges pointwise to S_∞^c at each point. It is a well known fact that a sequence of monotone functions which converges pointwise to a continuous limit converges uniformly. By Bolzano-Weierstrass S_n^d (it is just a vector of pointmasses) has a convergent subsequence which converges to a S_∞^d . Consequently, S_n has a convergent subsequence which converges uniformly to a S_∞ which has the same support as S .

Let $S_{n(k)}$ be this convergent subsequence. Because $S_{n(k)}$ is uniformly bounded away from zero for n large enough and S_n is of uniformly (in n) bounded sectional variation (it is a distribution function) it follows by lemma 3.3 and 3.2 that

$$\int \frac{S_{n(k)}(c \vee t)}{S_{n(k)}(c)} d(G_{n(k)}^e - G^e)(c) \leq C \|G_{n(k)}^e - G^e\|_\infty,$$

for a $C < \infty$. Empirical process theory tells us that $\|G_n^e - G^e\|_\infty = O_P(1/\sqrt{n})$ and $\|S_n^e - S^e\|_\infty = O_P(1/\sqrt{n})$. Hence we have that $\|H(S_{n(k)}, P_{n(k)}) - H(S_{n(k)}, P)\|_\infty = O_P(1/\sqrt{n(k)})$. This shows that $H(S_{n(k)}, P) = O_P(1/\sqrt{n(k)})$. The uniform consistency of $S_{n(k)}$ to S_∞ implies trivially that $\|H(S_{n(k)}, P)(t) - H(S_\infty, P)(t)\|_\infty \rightarrow 0$ which proves that $H(S_\infty, P)(t) = 0$ for all t . It remains to show that $H(S, P_{F_0, G_0}) = 0$ implies $S = S_0$, where we can use that $S > \delta > 0$ on $[0, \tau]$; that implies that $S_\infty = S_0$ and hence that each subsequence of S_n has a uniformly consistent subsequence, all having the same limit S_0 , which implies that the NPMLE S_n is uniformly consistent. We have that $H(S, P_0) = 0$ is equivalent with $U(S, P_0) = 0$. We have

$$\begin{aligned} 0 &= U_t(F, P_0) - U_t(F_0, P_0) \\ &= (F - F_0)(t) + \int_0^t \int_0^s \frac{(F - F_0)(c)}{S(c)} dG_0(c) \frac{dF_0(s)}{a_S(s)}, \end{aligned} \quad (18)$$

where $a_S(s) = \int_0^s dG_0^e(c)/S(c)$. Consider now the linear operator $I - A_{S,0} : (D[0, \tau], \|\cdot\|_\infty) \rightarrow (D[0, \tau], \|\cdot\|_\infty)$ defined by:

$$(I - A_{S,0})(h)(t) = h(t) + \int_0^t \int_0^s \frac{h(c)}{S(c)} dG_0(c) \frac{dF_0(s)}{a_S(s)}.$$

Equation (18) tells us that $(I - A_{S,0})(S - S_0) = 0$. If we can prove that the linear operator $I - A_{S,0}$ is 1-1, then that shows that $S = S_0$. This operator is a nice Volterra operator which can be proved to be invertible in exactly the same way as we will prove this in subsection 4 for the similar operator $I - A$ (the derivative of $S \rightarrow U(S, P)$) defined in (22). Replace in this proof $d\Lambda_1$ by dG_0/S and $d\Lambda_2$ by dF_0/a_S . In the proof in subsection 4 we needed that $d\Lambda_1/d\Lambda_2$ is uniformly

bounded on $[0, \tau]$ and we need that $\Lambda_2(\tau) < \infty$. So we need that dG_0/dF_0 is uniformly bounded and that $\int_0^\tau dF_0/a_S < \infty$. We have that

$$a_S(s) = \int_0^s dG_0^e(c)/S(c) = \int_0^s \frac{S_0(c)}{S(c)} dG_0(c) \geq \delta G_0(s)$$

because $S > \delta > 0$. Hence we only need $\int_0^\tau dF_0/G_0 < \infty$ and dG_0/dF_0 is uniformly bounded on $[0, \tau]$.

3.3 Differentiability of the estimating equation.

Define $a(s) \equiv \int_0^s dG^e(c)/S(c)$ and $a_n(s) \equiv \int_0^s dG^e(c)/S_n(c)$. We have that

$$U_t(F_n, P) - U_t(F, P) = (F_n - F)(t) + \int_0^t \frac{\int_0^s \frac{(S_n - S)(c)}{S_n S(c)} dG^e(c)}{a_n(s)a(s)} F^e(ds). \quad (19)$$

Substitution of $\alpha dG^e(c) = S(c)dG(c)$ and $\alpha dF^e(s) = G(s)F(ds)$ tells us that this equals

$$(F_n - F)(t) + \alpha \int_0^t \frac{\int_0^s \frac{(S_n - S)(c)}{S_n(c)} dG(c)}{a_n(s)} F(ds). \quad (20)$$

Hence

$$d_1 U_t(F, P)(F_n - F) = (F_n - F)(t) + \alpha \int_0^t \frac{\int_0^s \frac{(S_n - S)(c)}{S(c)} dG(c)}{a(s)} F(ds). \quad (21)$$

Let $F_n = F + \epsilon_n h$. In order to show Frechet-differentiability we need to prove that the supnorm of the remainder $\frac{U(F_n, P) - U(F, P)}{\epsilon_n} - d_1 U(F, P)(h)$ converges to zero uniformly in h with $\|h\|_\infty \leq 1$. By telescoping the remainder can be expressed as a sum of two terms, one with the difference $1/a_n - 1/a$ and one with a difference $1/S_n - 1/S$. The first term is given by:

$$\int_0^t \int_0^s h(c) dG(c) \frac{\int_0^s \frac{(S_n - S)(c)}{S_n S(c)} dG^e(c)}{a(s)a_n(s)} F(ds).$$

Here $S_n - S = \epsilon_n h$ and h can be bounded in supnorm by 1. So we can bound this term by:

$$\epsilon_n \int_0^t \frac{\int_0^s \frac{dG(c)}{S_n(c)}}{a_n(s)} \int_0^s \frac{dG(c)}{S_n(c)} \frac{F(ds)}{G(s)}.$$

We have that $S_n > \delta > 0$ with probability tending to 1. We also assumed that $\int F(ds)/G(s) < \infty$. This shows trivially that this term is bounded by $M\epsilon_n$ for some $M < \infty$ with probability tending to 1. The second term of the remainder is given by:

$$\int_0^t \int_0^s h(c) \frac{(S - S_n)(c)}{S_n S(c)} dG(c) \frac{F(ds)}{a(s)}.$$

In the same way it follows that if $S, S_n > \delta > 0$ and $\int F(ds)/G(s) < \infty$, then this term is bounded by $M\epsilon_n$ for some $M < \infty$. This proves that $F \rightarrow U(F, P)$ is supnorm Frechet differentiable at a S with $S > \delta$ on $[0, \tau)$ and $\int_0^\tau dF/G < \infty$.

3.4 Invertibility of the derivative.

Recall that $S(c)$ in (12) is given by

$$1 - F(c_{1-}, \tau_{2-}) - F(\tau_{1-}, c_{2-}) + F(c_{1-}, c_{2-}).$$

Now, it is easily verified that the derivative $d_1U(F, P)$ of $F \rightarrow U(F, P)$ at $(F, P_{F,G})$ in the direction h is given by

$$(I-A)(h)(t) \equiv h(t) - \alpha \int_{(0,t]} \int_{(0,s]} (h(c_{1-}, c_{2-}) - h(c_{1-}, \tau_{2-}) - h(\tau_{1-}, c_{2-})) \frac{dG(c_1, c_2)}{S(c_1, c_2)} \frac{dF(s_1, s_2)}{G(s_1, s_2)}. \quad (22)$$

Here $(I - A)$ is a mapping from $(D[0, \tau], \|\cdot\|_\infty)$ to $(D[0, \tau], \|\cdot\|_\infty)$. Define

$$d\Lambda_1(c) \equiv \frac{dG(c)}{S(c)} \text{ and } d\Lambda_2(s) \equiv \frac{dF(s)}{G(s)}.$$

Notice that $A(h) = B(h - h(\cdot, \tau_2) - h(\tau_1, \cdot))$, where

$$B(h)(t) = \alpha \int_{(0,t]} \int_{(0,s]} h(c-) d\Lambda_1(c) d\Lambda_2(s).$$

We can represent B as:

$$B(h)(t) = \int_{(0,t]} \int_{(0,s]} h(c-) \frac{d\Lambda_1}{d\Lambda_2}(c) d\Lambda_2(c) d\Lambda_2(s),$$

where we assume that $G \ll F$ on $[0, \tau]$ so that $d\Lambda_1/d\Lambda_2$ is well defined on $[0, \tau]$. Notice that

$$\frac{d\Lambda_1}{d\Lambda_2} = \frac{dG}{dF} \frac{G}{S}.$$

Because $S > \delta > 0$ on $[0, \tau]$ this is uniformly bounded if dG/dF is uniformly bounded on $[0, \tau]$. Furthermore, we assume that

$$\Lambda_2(\tau) = \int_0^\tau d\Lambda_2 = \int dF/G < \infty.$$

Now, the operator B has the well known Volterra structure (see Gill, Johansen, 1990), Kantorovick and Akilov, 1982, page 396). Because of this structure we have:

$$\|B^k(h)\|_\infty \leq \|d\Lambda_1/d\Lambda_2\|_\infty^k \Lambda_2(\tau)^k \frac{1}{2k!} \|h\|_\infty.$$

Denote the right-hand side with $c(k)\|h\|_\infty$. $A(h)$ is computed by summing over 3 terms obtained by applying B to h and its marginals. For computing $A^2(h)$ we have to apply B to $A(h)$ which means that we obtain 9 terms $B^2(v)$ where v ranges over h , marginals of h or marginals of the marginals, which are constants $h(\tau)$. Proceeding in this way we conclude that $A^k(h)$ is a sum of 3^k terms of the form $B^k(v)$, with v ranging over h , marginals of h and v being equal to $h(\tau)$. Hence we have:

$$\|A^k(h)\|_\infty \leq 3^k c(k) \|h\|_\infty,$$

where $3^k c(k)$ converges to zero exponentially fast; $c(k)$ kills any polynomial power. Since $(I - A)(h) = 0$ implies $h = A^k(h)$ this implies that $I - A$ is 1-1. Moreover, this bound implies that $\sum_{k=0}^\infty A^k$ is a bounded operator and hence that

$$(I - A)^{-1} = \sum_{k=0}^\infty A^k.$$

This proves that d_1U has a bounded inverse given by this Neuman series. Notice that all we needed to assume was that dG/dF is uniformly bounded and that $\int dF/G < \infty$.

Convergence of algorithm. This explains why one should expect this type of convergence for our proposed algorithm for computing the NPMLE (see 8);

$$F_n^{k+1}(t) = \int_{(0,t]} \frac{dF_n^e(s)}{\int_{(0,s]} dG_n^e(c)/S_n^k(c)}.$$

This is shown by writing out what this iteration means in terms of $F_n^{k+1} - F_n^k$; it is easily shown that one obtains that $F_n^{k+1} - F_n^k = A_{k,k-1}(F_n^k - F_n^{k-1})$, where $A_{k,k-1}$ is a linear operator which depends on F_n^{k-1}, F_n^k in such a way that if we set $F_n^{k-1} = F_n^k = F$, then $A_{k,k-1} = A$; in other words, it has the Volterra structure. Since $F_n^{k+1} - F_n^k = A(F_n^k - F_n^{k-1})$ implies $\|F_n^{k+1} - F_n^k\|_\infty \leq c(k)\|F_n^1 - F_n^0\|_\infty$ and hence implies exponential fast convergence, one can expect that this data driven algorithm will converge exponentially fast.

3.5 Weak convergence of the empirical process part.

Our task is to show that $Z_n = \sqrt{n}(U(F_n, P_n) - U(F_n, P)) \xrightarrow{D} Z$ in $(D[0, \tau], \|\cdot\|_\infty, \mathcal{B})$, using the uniform consistency of S_n , the supnorm-weak convergence of $\sqrt{n}(P_n - P)$ and the smoothness of $U(F, P)$ in P . Empirical process theory is concerned with weak convergence of an empirical process $(\sqrt{n}(P_n - P)(f) : f \in \mathcal{F})$ in $\ell^\infty(\mathcal{F})$, which is referred to as the empirical process indexed by \mathcal{F} . We say that \mathcal{F} is a Donsker class if this process converges weakly. If \mathcal{F} is a Donsker class, then the empirical process is tight which implies that

$$\sqrt{n}(P_n - P)(f_n) \rightarrow_P 0 \text{ if } \|f_n\|_P \rightarrow_P 0. \quad (23)$$

We will use (23) in our weak convergence proof.

Consider the second term in U as a functional U^* in (F, F^e, G^e) . Let $a_n(s) \equiv \int_0^s dG^e(c)/S_n(c)$ and $a_n^n(s) = \int_0^s dG_n^e(c)/S_n(c)$. Define the empirical processes $h_2^n = \sqrt{n}(F_n^e - F^e)$ and $h_3^n = \sqrt{n}(G_n^e - G^e)$. We know that $h_2^n \xrightarrow{D} h_2$ and $h_3^n \xrightarrow{D} h_3$ (jointly) in $(D[0, \tau], \|\cdot\|_\infty, \mathcal{B})$ for two Gaussian processes h_2, h_3 . By telescoping it follows that we have:

$$\sqrt{n}(U^*(S_n, F_n^e, G_n^e) - U^*(S_n, F^e, G^e)) = \int_{(0,t]} \frac{dh_2^n(s)}{a_n(s)} - \int_{(0,t]} \frac{1}{a_n^n a_n(s)} \int_0^s 1/S(c) dh_3^n(c) dF_n^e(s). \quad (24)$$

To prove weak convergence of the first term we will show that

$$\int_{(0,t]} \left(\frac{1}{a_n(s)} - \frac{1}{a(s)} \right) dh_2^n(s) \rightarrow 0 \text{ in probability.} \quad (25)$$

Then it remains to show that the i.i.d empirical process $\int_0^t 1/a(s) dh_2^n(s)$ converges weakly. The latter is the empirical process $\sqrt{n}(F_n^e - F^e)$ indexed by $\mathcal{F}_1 \equiv \{I_{(0,t]}(s)/a(s) : t \in [0, \tau]\}$. The mentioned generalization of the univariate result in van der Vaart and Wellner (1995) shows that if for some $\epsilon > 0$

$$\int \left(\frac{1}{a(s)} \right)^{2+\epsilon} dF^e(s) = \int \frac{dF(s)}{G(s)^{1+\epsilon}} < \infty, \quad (26)$$

then \mathcal{F}_1 is a F^e -Donsker class. We will now show (25). For proving (25) it suffices to prove that $1/a_n$ and $1/a$ fall in a Donsker class with probability tending to 1 and that $1/a_n - 1/a$ converges to zero in probability in $L^2(F^e)$. Firstly, notice that $1/a_n$ and $1/a$ fall in the class of bivariate monotone decreasing functions. Notice now that because $S > \delta > 0$ we have

$$\frac{G(s)}{a_n(s)} \leq \frac{1}{\delta}.$$

Hence $1/a_n, 1/a$ fall in a class of functions bivariate monotone functions with envelope $1/G$. The generalization of the univariate result in van der Vaart and Wellner (1995) shows that this is a F^e -Donsker class if for some $\epsilon > 0$ $\int 1/G^{2+\epsilon} dF^e < \infty$. This means that we need to assume (26).

Finally, we have

$$\left| \frac{1}{a_n} - \frac{1}{a} \right| (s) \leq M \|S_n - S\|_\infty \frac{1}{G(s)}.$$

Hence $\int (1/a_n - 1/a)^2 dF^e \rightarrow 0$ if

$$\int \frac{dF}{G} < \infty.$$

We will now prove weak convergence of the second term in (24) in the same way. By Fubini's theorem we can rewrite the second term as:

$$\int_0^t \frac{1}{S(c)} \left(\int_c^t \frac{1}{a_n^n(s)a_n(s)} dF_n^e(s) \right) dh_n^3(c).$$

Consider the function

$$f_n(c) \equiv \int_c^t \frac{1}{a_n^n(s)a_n(s)} dF_n^e(s) \text{ and } f = \int_c^t 1/a^2(s) dF^e(s).$$

As with the first term we want to show that f_n, f fall with probability tending to 1 in a Donsker class. Because $1/S$ is uniformly bounded this implies that $1/S f_n$ falls in a Donsker class (see van der Vaart, Wellner, 1995): if a Donsker class is multiplied with a fixed bounded function, then one obtains a new Donsker class. Notice that f_n falls in the class of bivariate monotone decreasing functions. We have the following bound on $f_n(c)$:

$$f_n(c) \leq M \int_c^t \frac{dF_n^e}{GG_n^e}.$$

Assume that

$$\int_c^\tau \frac{dF_n^e}{GG_n^e} \leq M(c) \tag{27}$$

with probability tending to 1. The generalization of the univariate result in van der Vaart and Wellner (1995) shows that the class of bivariate monotone decreasing functions with envelope $M(\cdot)$ is G^e -Donsker if for some $\epsilon > 0$ $\int M(c)^{2+\epsilon} dG^e(c) < \infty$.

It remains to show that $\int (f_n - f)^2 dG \rightarrow 0$ in probability. Because $\sup_{c \geq \epsilon} |f_n - f| (c) \rightarrow 0$ we have

$$\int_\epsilon^\tau (f_n - f)^2 dG \rightarrow 0.$$

Now, use that $(f_n - f)^2(c) \leq M^2(c)$ and $\int M^2(c) dG(c) < \infty$ to show that $\int_0^\epsilon (f_n - f)^2 dG \rightarrow 0$ if $\epsilon \rightarrow 0$. The usual standard argument proves now that $\int (f_n - f)^2 dG \rightarrow 0$.

This proves that

$$Z_n \equiv \sqrt{n}(U^*(S_n, F_n^e, G_n^e) - U^*(S_n, F^e, G^e)) \xrightarrow{D} Z \equiv \int_{(0,t]} \frac{dh_2(s)}{a(s)} - \int_{(0,t]} \frac{1}{a^2(s)} \int_0^s 1/S dh_3(c) dF^e(s). \tag{28}$$

In particular, this implies that for a fixed t , the left-hand side of (13) is asymptotically normal with mean zero and variance equal to the variance of the corresponding influence curve (just substitute for h_2, h_3 the empirical process based on one observation)

$$IC(F, P)(T', C') \equiv - \left(\frac{I(T' \leq t)}{a(T')} - \int_{(0,t]} \frac{1}{a^2(s)} \frac{I(C' \leq s)}{S(C')} dF^e(s) \right). \tag{29}$$

Denote the two terms on the right-hand side of (29) with $J_1(t)$ and $J_2(t)$. We can bound the variance of $J_1 + J_2$ with $2EJ_1^2(t) + 2EJ_2^2(t)$. Notice now that

$$EJ_1^2(t) = \alpha \int_0^t dF/G$$

and

$$\begin{aligned} EJ_2^2(t) &= \alpha E \int_0^t \int_0^s \frac{G_n^e(dv)}{S(v)} \frac{F(ds)}{G(s)} \\ &= \alpha \int_0^t \frac{G(dv)}{S(v)} \left[\int_v^t \frac{F(ds)}{G(s)} \right]^2. \end{aligned}$$

Consequently, since $S > \delta > 0$ the variance of $IC(F, P)$ is bounded if $\int dF/G < \infty$.

4 Construction of confidence bands.

Our proof shows that $F_n(t)$ is asymptotically linear with influence curve given by $IC^* \equiv d_1U(F, P)^{-1}(IC)$, where IC is defined in (29):

$$\sqrt{n}(F_n(t) - F(t)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n d_1U(F, P)^{-1}(IC(F, P, \cdot))(C'_i, T'_i) + o_P(1).$$

In particular, this implies that $\sqrt{n}(F_n - F)(t)$ is asymptotically normal with mean zero and variance equal to the variance of $IC^*(F, G, t)(T', C')$. We showed above that the variance of IC is bounded if $\int dF/G < \infty$. This shows also that the variance of IC^* is bounded if $\int dF/G < \infty$. The result on the derivative $d_1U = I - A$ teaches us that we can find the solution of the equation $h - A(h) = \kappa$ by starting with an initial h^0 and iterate as follows:

$$h^{k+1} = \kappa + A(h^k),$$

and that this is an exponentially fast algorithm. This provides us with an algorithm for computing $IC^*(F, G, t)$ for a given (F, G) and t .

The variance of $IC^*(F, G, t)$ can now be estimated with

$$\frac{1}{n} \sum_{i=1}^n IC^*(F_n, G_n, t)^2(T'_i, C'_i),$$

where F_n, G_n is the NPMLE of F, G . This estimate of the variance provides us with an asymptotic confidence interval for $F(t)$.

Alternatively, one could use the bootstrap (i.e. resampling from the original sample) for construction of confidence intervals. The asymptotic validity of the bootstrap follows from the fact that F is a compactly differentiable functional of P ; see Gill (1989).

Acknowledgement. The referee pointed out the problem with the singularity at zero and had useful comments on the presentation of the paper. I am very grateful to him for his important contribution.

References

- Bickel, P.J, Klaassen, C.A.J, Ritov, Y. & Wellner, J.A. (1993), *Efficient and adaptive inference in semiparametric models*, Johns Hopkins University Press, Baltimore.
- Dabrowska, D.M. (1988). Kaplan Meier Estimate on the Plane, *Ann. Statist.* **16** 1475–1489.
- Dabrowska, D.M. (1989). Kaplan Meier Estimate on the Plane: Weak Convergence, LIL, and the Bootstrap, *J. Multivar. Anal.* **29** 308–325.
- Dudley, R.M. (1966). Weak convergence of probabilities on nonseparable metric spaces, and empirical measures on Euclidean spaces, *Illinois J.Math.* **10** 109–126.
- Dudley, R.M. (1985). An extended Wichura theorem, definitions of Donsker classes, and weighted empirical distributions. *Lecture notes in Math.* **1153**, 141–178. Springer Verlag, New York.
- Gill, R.D. (1980), *Censoring and Stochastic Integrals*, Mathematical Centre Tracts 124, Mathematisch Centrum, Amsterdam.
- Gill, R.D. (1989), Non-and Semi-parametric Maximum Likelihood Estimators and the von Mises Method (Part 1), *Scand. J. Statist.* **16**, 97–128.
- Gill, R.D. & Johansen, S. (1990), A survey of product integration with a view towards application in survival analysis, *Ann. Statist.* **18**, 1501–1555.
- Gill, R.D. & Keiding, N. (1990), Random truncation models and Markov processes, *Ann. Statist.* **18**, 582–602.
- Gill, R.D. (1992). Multivariate survival analysis, *Theory Prob. Appl.* **37**, 19–36 and 307–328. (in Russian; English translation to appear).
- R.D. Gill, M.J. van der Laan and J.A. Wellner (1995), Inefficient estimators of the bivariate survival function for three models, to appear in *Annales de L'I.H.P. Probabilites et Statistiques*.
- Gill, R.D. and van der Vaart, A.W. (1993), *Non-and Semi-parametric Maximum Likelihood Estimators and the von Mises Method (Part 2)*, Submitted to *Scand. J. Statist*, preprint nr. 664, Dept of Mathematics, Utrecht, the Netherlands.
- Gürler, Ü (1994), Bivariate distribution function when a component is randomly truncated, truncated data with applications, *Discussion Paper 9407*, Institut de Statistique, Université Catholique de Louvain, Belgium.
- Hildebrandt, T.J. (1963), *Introduction to the theory of integration*, Academic Press, New York and London.
- Hoffmann-Jørgensen, J. (1984). *Stochastic processes on Polish Spaces*, Unpublished manuscript.
- Kantorovich, L.V., Akilov, G.P. (1982) *Functional Analysis*. translated by Howard L. Silcock, New York, Pergamon Press.
- van der Laan, M.J. (1993) *Efficient and Inefficient Estimation in Semiparametric Models*, thesis, will be published as C.W.I.-tract, ISBN 90-393-0339-8.
- van der Laan, M.J. (1995a) *Nonparametric estimation of a bivariate survival function with right-censored data*, submitted to *Statistica Neerlandica*.
- van der Laan, M.J. (1995b), *Efficient estimator of the bivariate survival function and repairing NPMLE*, submitted for publication in *Ann. Statist.*.
- Loève, M. (1955). *Probability theory*, van Nostrand, New York.
- Neuhaus, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *Ann. Math. Statist.* **42** 1285–1295.
- Prentice, R.L., Cai, J. (1992a). Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika* **79**, 495–512.
- Prentice, R.L., Cai, J. (1992b). A covariance function for bivariate survival data and a bivariate

- survival function. In Klein, J.P. and Goel, P.K., editors, *Proceedings NATO advanced studies workshop on survival analysis and related topics*. Karger, Basel.
- Pruitt, R.C. (1991). *Strong consistency of self-consistent estimators: general theory and an application to bivariate survival analysis*. Technical Report nr. 543, University of Minnesota.
- Pruitt, R.C. (1993). *Small sample comparisons of five bivariate survival curve estimators*. *J. Statist. Comput. Simul.*, **45**, 147–167.
- A.W. van der Vaart (1991), On differentiable functionals, *Ann. Statist.* **19**, 178–204.
- A.W. van der Vaart (1993), Efficiency of infinite dimensional M-estimators, *Statistica Neerlandica*.
- A.W. van der Vaart and J.A. Wellner (1995), *Weak convergence and empirical processes*. IMS Lecture Notes-Monograph Series (to appear).
- M.C. Wang, N.P. Jewell and W.Y. Tsai (1986), Asymptotic properties of the product limit estimate under random truncation, *Ann. Statist.* **14**, 1597–1605.
- M. Woodroffe (1985), Estimating a distribution function with truncated data, *Ann. Statist.* **13**, 163–177. Correction: *Ann. Statist.* **15** (1987), 883.
- Wichura, M.J. (1968), On the weak convergence of non-Borel probabilities on a metric space, *Ph.D. dissertation*, Columbia University.