Modified EM-estimator of the Bivariate Survival Function

M.J. van der Laan
Department of Mathematics
University of Utrecht
Budapestlaan 6
Postbus 80010
3508 TA Utrecht
The Netherlands.

May 11, 1998

Abstract

Pruitt (1991b) proposed estimating a bivariate survival function for censored data by modifying the self-consistency equations of the EM-algorithm. Though not efficient, the estimator has very good practical performance. In this paper, we prove weak convergence at $\sqrt{n}$-rate, strong uniform consistency and a semiparametric bootstrap result for this implicitly defined estimator. The estimator is analyzed by application of the implicit function theorem. Because the estimator uses kernel density estimators, part of the analysis is concerned with proving root-n weak convergence of functionals of density estimators. We consider this problem in generality and show that the functional delta-method (Gill, 1989, van der Vaart and Wellner, 1993) can be refined to this case.

Key words: Maximum likelihood estimator, self-consistency, efficient estimator, implicit function theorem, weak convergence, functional delta-method, kernel density estimator.

1 Introduction

A wealth of proposals for estimation of the bivariate survival function in the presence of bivariate right censored data has been proposed. If the data is sampled from a continuous distribution function, then the nonparametric maximum likelihood (NPML) and thereby the self-consistency principle do not lead to a consistent estimator. Therefore most proposals are based on representations of the bivariate survival function in terms of the distribution of the censored data: among them Muñoz (1980), Campbell & Földes (1982), Tsai, Leurgans & Crowley (1986), Dabrowska (1988, 1989), the Volterra estimator of P.J. Bickel (see Dabrowska, 1988,1989), Burke (1988), Prentice & Cai (1992a, 1992b).

Dabrowska's multivariate product-limit estimator and the Prentice-Cai estimator, based on very nice representations of a multivariate survival function in terms of its conditional multivariate hazard measure, have a better practical performance than the Volterra and pathwise estimator and it is expected that they are also better than the other proposed explicit estimators (Bakker, 1990, Prentice & Cai, 1992b, Pruitt, 1993). These two estimators are smooth functionals of the empirical distributions of the data so that such results as consistency, asymptotic normality, correctness of the bootstrap, consistent estimation of the variance of the influence curve, LIL, all hold by application of the functional delta method as stated in van der Vaart, Wellner (1993) and Wellner (1993): see Gill, van der Laan, Wellner (1993).

All the estimators proposed above are ad hoc estimators which are not asymptotically efficient. The ad hoc nature of these estimators is also reflected by the fact that none of them is a true distribution (monotony is violated, Pruitt, 1991a).

Pruitt (1991b) proposed an interesting estimator which is a solution of a modification of the self-consistency equation. Pruitt modifies the self-consistency equation by replacing the hard terms, corresponding with the singly censored observations, by ad hoc estimates. So it is not a NPMLE, and not efficient, but it shares several of the appealing properties of a NPMLE and thereby of a self-consistent estimator. Pruitt (1991b) makes his estimator intuitively clear and proves its self-consistency properties. Uniform consistency, asymptotic normality and asymptotic validity of the bootstrap has not yet been proved, and is done here.

In the next section the bivariate censoring model and Pruitt's estimator will be defined in detail. Here we give the heuristic explanation of its estimator: Each observation in the bivariate censoring model (doubly, singly censored and uncensored) tells us that the survival time has fallen in a certain region: the survival time is known if the observation is uncensored, it is known to lie on a line for singly censored and in a quadrant for doubly censored observations. His estimator works now as follows: each observation gets mass $1/n$ which it has to redistribute over its associated region for the survival time. By using product limit estimators of univariate kernel density estimators the singly censored observation are told how to redistribute their mass $1/n$ over their associated lines. The uncensored observations give mass $1/n$ to the observed survival time. By solving modified EM-equations the mass $1/n$ of the doubly censored observations is redistributed self-consistently over their associated quadrants: i.e. a point t in the quadrant gets mass $1/n$ times the conditional density (computed as if the estimator itself is the true one), given the survival time lies in the quadrant. Consequently, his estimator is a distribution function and the mass $1/n$ corresponding with each observation is redistributed over the region where it belongs in a self-consistent or consistent way by listening to the other observations.

In this paper we consider a slightly different version of his estimator: we use edge corrected bivariate kernel density estimators while he smoothes in one direction.

There are several motivations for being interested in Pruitt's estimator. Simulations (Pruitt, 1993, and van der Laan, 1993a) show that his estimator is competitive with Dabrowska's and Prentice and Cai's estimators, while his estimator does not put negative mass on points (which is not true for Dabrowska's and Prentice and Cai's estimators). His estimator can be modified in several interesting ways; for example, instead of using product limit estimators of the conditional densities over lines (which we needed for telling the singly censored observations how to redistribute its mass) one might use the efficient SOR-MLE (based on a reduction of the data) in van der Laan (1993b) in order to obtain an estimate of the conditional density over the lines (see van der Laan, 1993c, for practical results with these estimators).

Pruitt's estimator uses kernel density estimators and therefore also depends on a bandwidth, but simulations show that his estimator is much *less sensitive* to the choice of the bandwidth than the SOR-MLE to the choice of the grid-width. This is intuitively clear because the bandwidth of Pruitt's estimator influences only the redistribution of mass $1/n$ over lines, while a change of the grid-width in the SOR-MLE changes all interactions between the regions generated by the observations and hence the estimator might change at all its support points. Pruitt's estimator is also much *less computer-intensive* than the SOR-MLE.

In the special submodel of the bivariate censoring model where one of the two survival times is always uncensored and the other is randomly censored, Pruitt's estimator is explicitly known and it is a NPMLE which by our results converges at root-n rate. Gill and van der Vaart (1993) have a general theory which shows efficiency (see Bickel et al., 1993) of NPMLE which are known to be root-n consistent. Unfortunately, their theory requires one cumbersome regularity condition which is expected to hold but which is hardly verifiable. All other conditions hold trivially (see van der Laan, 1993d). This submodel has an important application in regression analysis. Ritov (1992) proposes an efficient estimator for this submodel. For this submodel Pruitt's estimator is similar to Ritov's estimator. By explicitly writing out the influence curve of Pruitt's estimator one should be able to check that its asymptotic distribution is indeed the optimal one as given in Ritov (1992), but this goes beyond the scope of this paper. At least heuristically it is clear that Pruitt's estimator is efficient for this special submodel.

Finally, we have some remarks on points of technique: we use some novel methods which may well be useful in other analyses of M-estimators and analyses which involve density estimators. Pruitt's estimator is analyzed by applying the *implicit function theorem*. The implicit function theorem requires *invertibility of a derivative* of the modified self-consistency equation solved by Pruitt's estimator and a strong *differentiability* condition. We apply a general trick in order to get an equation with the required smoothness, see section 3. The invertibility proof (section 4) is highly non-trivial and might give techniques for proving invertibility of quite complicated operators of the form $I - A$ where A has a norm larger than 1: so where it is certainly non-trivial that the Neumann series $\sum_{i=1}^{\infty} A^i(h)$ converges. We also formulate a functional delta-method for functionals like $\int \phi(f_n) d\mu$, where f_n is a density estimator of f_0.

In this paper we will prove strong uniform consistency of S_n and weak convergence of the normalized difference $\sqrt{n}(S_n - S)$ of Pruitt's estimator of the survival function. The main work consists of proving weak convergence of the singly-censored terms in Pruitt's modified self-consistency equation (3), below, which involve density estimators, and proving the necessary conditions for the implicit function theorem for Banach spaces (Hildebrandt and Graves, 1927, Flett, 1980), to take care of the implicit character of equation (3) (its fourth term). The *organisation* of the paper is as follows. In section 2 the bivariate censoring model and the EM-algorithm are described in detail in order to be able to define and understand Pruitt's estimator and to write down the modified self-consistency equation solved by it. We also define that part of the equation which is explicitly known and denote it by Ψ_n. In section 3 we state the consistency and weak convergence theorem for Pruitt's estimator. After stating the theorem we firstly cover the essential ingredients for the technical differentiability parts of the proof and then give the outline of the proof based on the implicit function theorem and the functional delta-method. We end this section with the ingredients to be proved in the next sections (like weak convergence and consistency of the explicit term Ψ_n and invertibility of the derivative of the modified self-consistency equation). In section 4 we prove the first two ingredients; in particular the required invertibility of the derivative of the modified self-consistency-equation. It remains to cover the analysis of Ψ_n. This requires a functional delta-method for density estimators as proved in section 5. In section 6 this functional delta-method is applied in the analysis of Ψ_n. The probabilistic conditions of this delta-method are covered in generality in section 6.1 and 6.2 and the result is summarized in lemma 6.2. In section 6.3 we make clear how to apply this delta method to our specific term and in section 6.4 the differentiability condition is proved, which completes the proof of all four ingredients. Finally, in section 7 we state a weak consistency result for the semiparametric bootstrap and for a proof we refer to van der Laan (1991).

2 The bivariate censoring model and a modified EM-estimator

We do not use a special notation for vectors; if we do not mean a vector this will be made clear. So if we write t we mean $t = (t_1, t_2) \in \mathbb{R}^2$. If we write $\leq, \geq, <, >$ then this should hold componentwise for all components: so if $x \in \mathbb{R}^2$ then $x \leq y \Leftrightarrow x_1 \leq y_1,\ x_2 \leq y_2$.

Model. We will model the estimation of the bivariate survival function based on censored and non-censored observations as follows. T is a positive bivariate lifetime vector with bivariate distribution F_0 and survival function S_0; $F_0(t) \equiv Pr(T \leq t)$ and $S_0(t) \equiv Pr(T > t)$. C is a positive bivariate censoring vector with bivariate distribution G_0 and survival function H_0; $G_0(t) \equiv Pr(C \leq t)$ and $H_0(t) \equiv Pr(C > t)$. T and C are independent; $(T, C) \in \mathbb{R}^4$ has distribution $F_0 \times G_0$. (T_i, C_i), $i = 1, \ldots, n$ are n independent copies of (T, C). We only observe the following many to one mapping ϕ of (T_i, C_i);

$$Y_i \equiv (\widetilde{T}_i, D_i) \equiv \phi(T_i, C_i) \equiv (T_i \wedge C_i, I(T_i \leq C_i)),\ i = 1, \ldots, n$$

with components given by:

$$\widetilde{T}_{ij} = \min\{T_{ij}, C_{ij}\},\ D_{ij} = I(T_{ij} \leq C_{ij}),\ j = 1, 2.$$

So the minimum and indicator are taken componentwise, so that $\widetilde{T}_i \in [0, \infty)^2$ and $D_i \in \{0, 1\}^2$ are bivariate vectors. So the observations in the sample space $[0, \infty)^2 \times \{0, 1\}^2$ are $Y_i \sim P_{F_0, G_0} \equiv (F_0 \times G_0)(\phi^{-1})$. The probability measure P_{F_0, G_0} of the data has two unknown parameters F_0 and G_0. Each observation Y_i tells us that $(T_i, C_i) \in B(Y_i) \equiv \phi^{-1}(Y_i) \subset \mathbb{R}^2 \times \mathbb{R}^2$. All regions $B(Y)$ are of the form $B(Y)_1 \times B(Y)_2$ and therefore Y_i tells us that $T_i \in B(Y_i)_1 \equiv \{T : \phi(T, C) = Y_i \text{ for a C}\}$ and $C_i \in B(Y_i)_2$. If we say 'the region corresponding (or associated) with Y_i' we mean $B(Y_i)_1$.

The kind of region $B(Y_i)_1$ for T_i (point, vertical half-line, horizontal half-line, quadrant) generates a classification of the observations $Y_i = (\widetilde{T}_i, D_i)$ in 4 groups:

1. $(\widetilde{T}_i, D_i)$ with $D_i = (1, 1)$. The observation is called uncensored, and it tells you that $T_i \in B(Y_i)_1 = \{(\widetilde{T}_i)\}$. So $T_i = \widetilde{T}_i$.

2a,2b $(\widetilde{T}_i, D_i)$ with $D_i = (0, 1)$ or $D_i = (1, 0)$. The observation is called singly censored, and if $D_i = (0, 1)$ it tells you that $T_i \in B(Y_i)_1 = \{(\widetilde{T}_{i1}, \infty) \times \{\widetilde{T}_{i2}\}\}$ (horizontal half-line), or if $D_i = (1, 0)$ that $T_i \in B(Y_i)_1 = \{\{\widetilde{T}_{i1}\} \times (\widetilde{T}_{i2}, \infty)\}$ (vertical half-line).

3. $(\widetilde{T}_i, D_i)$ with $D_i = (0, 0)$. The observation is called doubly censored, and it tells you that $T_i \in B(Y_i)_1 = \{(\widetilde{T}_{i1}, \infty) \times (\widetilde{T}_{i2}, \infty)\}$ (upper quadrant).

We are concerned with estimation of F_0 using the information given by the observations $(\widetilde{T}_i, D_i)$, so that $T_i \in B(Y_i)_1$.

In our analysis of Pruitt's estimator we need the following assumptions on F_0 and G_0:

Assumptions.

1. We restrict functions to a rectangle $[0, \tau] \subset [0, \infty)^2$, $\tau = (\tau_1, \tau_2)$, where τ is chosen so that $G_0(\tau-) = \delta_1 < 1$, $G_0(\tau) = 1$ and G_0 has an atom at τ: $G_0(\{\tau\}) = \delta > 0$.

2. We assume that G_0 has a density g_0 w.r.t. the Lebesque measure on $[0, \tau)$ and that F_0 has a density f_0 w.r.t. the Lebesque measure on $[0, \tau + \epsilon]$ for certain $\epsilon > 0$. Furthermore we assume that $f_0, g_0 \in C^3[0, \tau]$.

3. Moreover, we assume that for a $\epsilon = (\epsilon_1, \epsilon_2)$ f_0 is strictly positive on $([0, \tau + \epsilon] \backslash [0, \tau])$.

Assumption 1 is not stringent because it can be easily accomplished by censoring observations which do not fall in the rectangle $[0, \tau]$ at the edge of $[0, \tau]$. Then the obtained data has a distribution P_{F_0, G_0^τ}, where G_0^τ lives on $[0, \tau]$ and has an atom at τ. In real life this artificial censoring means a small loss of information, but a gain in stability of the estimator. Let P_n be the empirical distribution function of Y_i, $i = 1, \ldots, n$. In order to describe Pruitt's estimator we firstly need to understand the NPMLE.

EM-algorithm. A NPMLE S_n of S_0 is computed as follows:

1. We construct an initial estimator as follows: Associate with each region $B(Y_i)_1$, $i = 1, \ldots, n$, mass $1/n$ and redistribute it as follows. If $D_i = (1,1)$ then $B(Y_i)_1$, i.e. a point corresponding with an uncensored observation, gets all the mass $1/n$. If $D_i \neq (1,1)$ and $B(Y_i)_1$ contains uncensored $T_j = B(Y_j)_1$, then redistribute the mass $1/n$ uniformly over all uncensored $T_j \in B(Y_i)_1$. If $D_i \neq (1,1)$ and $B(Y_i)_1$ contains no uncensored $T_j = B(Y_j)_1$, then put the mass on a finite number of chosen points in $B(Y_i)_1$. We denote the discrete density corresponding with these point-masses with f_n^1 and its corresponding survival function with S_n^1. Now, we also can compute the conditional probability $P_{S_n^1}(T = t \mid T \in B(Y_i)_1)$ for all $i = 1, \ldots, n$, which we will need in the next step.

2. Give each observation $Y_i = (\widetilde{T}_i, D_i)$ mass $1/n$ which it has to redistribute over its associated region $B(Y_i)_1$ as follows: each point t gets a fraction of the mass $1/n$ given by $1/n \times P_{S_n^1}(T = t \mid T \in B(Y_i)_1)$. So this gives a new estimator f_n^2:

$$f_n^2(t) = \sum_{i=1}^{n} \frac{1}{n} P_{f_n^1}(T = t \mid T \in B(Y_i)_1) = \int P_{f_n^1}(T = t \mid T \in B(y)_1)) dP_n(y).$$

   This step is natural because our goal is to give the T_i mass $1/n$, but because of the random censoring we have only the knowledge that $T_i \in B(Y_i)_1$, so we redistribute the $1/n$ over $B(Y_i)_1$ according to a good estimate of the conditional density over $B(Y_i)_1$. In terms of S_n^2 we have:

$$S_n^2(t) = \int P_{S_n^1}(T > t \mid T \in B(y)_1) dP_n(y) = \int P_{S_n^1}(T > t \mid y)) dP_n(y),$$

   where one has to notice that $P_{F,G}(T > t \mid Y) = P_F(T > t \mid Y)$ because G cancels.

3. Set $S_n^1 = S_n^2$ and go to step 2, again.

This algorithm is the EM-algorithm (Dempster, Laird and Rubin, 1977). With this algorithm one finds a solution of the self-consistency equation (Efron, 1967):

$$S_n(t) = \int P_{S_n}(T > t \mid y) dP_n(y).$$

Dempster, Laird and Rubin (1977) show that the algorithm increases the likelihood at each step. The loglikelihood is concave and therefore this implies that S_n will be a NPMLE.

Why does the EM-algorithm not give good estimators? Notice that the redistribution of the mass $1/n$ in the construction of the initial estimator is arbitrary if $B(Y_i)_1$ does not contain observed T_j. If $D_i = (1,0)$ or $D_i = (0,1)$, then these regions $B(Y_i)_1$ are lines. Therefore the $B(Y_i)_1$ with $\Delta_i \in \{(1,0),(0,1)\}$ will with probability one not contain any uncensored observations and therefore the initial estimator is arbitrary on these lines $B(Y_i)_1$. Moreover, this means that the conditional density $P_{f_n^1}(T = t \mid T \in \text{line})$ will not be changed at step k by the uncensored observations; the mass assigned to the uncensored observations by itself and other observations will not change the conditional density over the line, while the double and other singly censored observations do not tell much. So the singly censored observations do not get any information from the uncensored observations about how to redistribute their mass $1/n$ over their line. That is the reason why we do not have consistency of each NPMLE: *we* have to tell how the mass $1/n$ has to be redistributed over the lines corresponding with singly censored observations.

Pruitt's Estimator. Pruitt's estimator does the following: Find an estimate $\widehat{P_{F_0}}(T = t \mid T \in$

$B(\widetilde{T}_i, D_i)_1)$ of $P_{F_0}(T = t \mid T \in B(\widetilde{T}_i, D_i)_1)$ for $D_i \in \{(0,1),(1,0)\}$. Start the EM-algoritthm, but let the redistributions of mass $1/n$ over the line corresponding with the singly censored observations be replaced by redistributions of the mass $1/n$ over the lines according to the fixed estimate $\widehat{P_{F_0}}(T = t \mid T \in B(\widetilde{T}_i, D_i)_1)$ for $D_i \in \{(0,1),(1,0)\}$. In other words, the singly censored observations do redistribute their mass $1/n$ in a fixed way and the doubly censored observations still let the algorithm do the work and redistribute their mass self-consistently in the end.

Let's write down the modified self-consistency equation which S_n (Pruitt's estimator) solves. Firstly, define the subdistribution functions corresponding to the four different kind of observations:

$$\begin{aligned}
P_{11}(t) &\equiv P(\widetilde{T} \le t, D_1 = D_2 = 1) \\
P_{10}(t) &\equiv P(\widetilde{T} \le t, D_1 = 1, D_2 = 0) \\
P_{01}(t) &\equiv P(\widetilde{T} \le t, D_1 = 0, D_2 = 1) \\
P_{00}(t) &\equiv P(\widetilde{T} \le t, D_1 = D_2 = 0).
\end{aligned}$$

The empirical distributions of $P_{11}, P_{01}, P_{10}, P_{00}$ will be denoted with $P_{11}^n, P_{01}^n, P_{10}^n, P_{00}^n$. Then

$$P_n(\cdot, D) = P_{11}^n(\cdot)I(D = (1,1)) + P_{01}^n(\cdot)I(D = (0,1)) + P_{10}^n(\cdot)I(D = (1,0)) + P_{00}^n(\cdot)I(D = (0,0)).$$

The two conditional densities over the lines corresponding with the singly censored observations which appear in the self-consistency equation are given by:

$$\begin{aligned}
W_{1F_0}(t_1, y_1, y_2) &\equiv P_{F_0}(T_1 > t_1 \mid T_1 > y_1, T_2 = y_2) \qquad (1)\\
W_{2F_0}(t_2, y_1, y_2) &\equiv P_{F_0}(T_2 > t_2 \mid T_2 > y_2, T_1 = y_1).
\end{aligned}$$

Pruitt estimates them with two weighted product limit estimators $\widehat{W}_1, \widehat{W}_2$, respectively. We will define these product limit estimators in section 6 (see (17)).

We have the equation $S(t) = \int P_S(T > t \mid y)dP_{F,G}(y)$. In formulas, using at the second equality that G has support on $[0, \tau]$, it is given by:

$$\begin{aligned}
S(t) &= \int_{t_1}^{\tau_1}\int_{t_2}^{\tau_2} dP_{11}(y_1, y_2) \\
&\quad + \int_0^{\tau_1}\int_{t_2}^{\tau_2} W_{1F}(t_1, y_1, y_2)dP_{01}(y_1, y_2) \\
&\quad + \int_{t_1}^{\tau_1}\int_0^{\tau_2} W_{2F}(t_2, y_1, y_2)dP_{10}(y_1, y_2) \\
&\quad + \int_0^{\tau_1}\int_0^{\tau_2} \frac{S(t_1 \vee y_1, t_2 \vee y_2)}{S(y_1, y_2)} dP_{00}(y_1, y_2) \\
&\equiv \Psi(t) + \int_0^{\tau_1}\int_0^{\tau_2} \frac{S(t_1 \vee y_1, t_2 \vee y_2)}{S(y_1, y_2)} dP_{00}(y_1, y_2), \qquad (2)
\end{aligned}$$

where $S(t_1 \vee y_1, t_2 \vee y_2)/S(y_1, y_2) = P(T_1 > t_1, T_2 > t_2 \mid T_1 > y_1, T_2 > y_2)$ and

$$\Psi(t) \equiv \overline{P}_{11}(t) + \int_0^{\tau_1}\int_{t_2}^{\tau_2} W_1(t_1, y_1, y_2)dP_{01}(y_1, y_2) + \int_{t_1}^{\tau_1}\int_0^{\tau_2} W_2(t_2, y_1, y_2)dP_{10}(y_1, y_2).$$

If we replace $P_{F,G}$ by P_n, then we obtain the self-consistency equation. S_n (Pruitt's estimator) solves

$$S_n(t) = \tfrac{1}{n}\sum_{i=1}^n I(\widetilde{T}_i > t, D_{i1} = D_{i2} = 1) + \int_0^{\tau_1}\int_{t_2}^{\tau_2} \widehat{W}_1(t_1, y_1, y_2)dP_{01}^n(y_1, y_2)$$

$$+\int_{t_1}^{\tau_1}\int_0^{\tau_2}\widehat{W}_2(t_2,y_2,y_1)dP_{10}^n(y_1,y_2)+\int_0^{\tau_1}\int_0^{\tau_2}\frac{S_n(t_1\vee y_1,t_2\vee y_2)}{S_n(y_1,y_2)}dP_{00}^n(y_1,y_2)$$

$$\equiv \quad \Psi_n(t)+\int_0^{\tau_1}\int_0^{\tau_2}\frac{S_n(t_1\vee y_1,t_2\vee y_2)}{S_n(y_1,y_2)}dP_{00}^n(y_1,y_2). \qquad (3)$$

The only difference with the self-consistency equation is that in the self-consistency equation we have W_{1F_n}, W_{2F_n}, self-consistent redistribution of mass $1/n$, instead of $\widehat{W}_1$, $\widehat{W}_2$, redistribution of the mass $1/n$ according to a predetermined estimate. Ψ_n represents the empirical counterpart of Ψ.

3 Theorem and Outline of the Proof of Consistency and Weak Convergence

We will consider estimators, say X_n, as random (not necessarily measurable) elements of the cadlag function space $D[0,\tau]$ (functions defined on $[0,\tau]$), where cadlag is defined as in Neuhaus (1971). We will endow this space with the Borel sigma-algebra and the supremum norm and denote it then with $(D[0,\tau],B,\|\cdot\|_\infty)$. $(D[0,\tau],B,\|\cdot\|_\infty)$ 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 details in van der Vaart and Wellner (1993).

We will refer to the well known result that the normalized empirical distribution functions $\sqrt{n}(P_{00}^n-P_{00}),\ldots,\sqrt{n}(P_{11}^n-P_{11})$ converge weakly simultaneously in the Hoffmann-Jørgensen sense as elements of $(D[0,\tau],B,\|\cdot\|_\infty)$ (Dudley, 1966). In this section we will give the outline of the proof of the following theorem and state the ingredients to be proved in the subsequent sections.

Theorem 3.1 *Assume that the underlying densities f,g satisfy assumptions* **1**, **2** *and* **3** *made in the introduction. Assume that for $\widehat{W}_1,\widehat{W}_2$ (see 17) we use kernel density estimators with bandwidth $h_n=n^{-1/7}$ and a kernel K satisfying the assumptions as stated in lemma 6.2.*

Then $\|(S_n-S)\|_\infty \to 0$ in supnorm a.s. and $\sqrt{n}(S_n-S)$ converges weakly in $(D[0,\tau],B,\|\cdot\|_\infty)$.

Before we give the proof we discuss a few essential bivariate techniques which will be essential in our analysis.

3.1 Essential Ingredients of the Analysis.

We will need an *integration by parts* formula for bivariate functions. For this purpose let $[0,\tau]\subset \mathbb{R}^2$ be a fixed rectangle. Let $f:[0,\tau]\to\mathbb{R}$ 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 $\|f\|_{\mathrm{v}}$, 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

$$\|f\|_{\mathrm{v}} \equiv \sup_{\{A_{i,j}\}} \sum_{i,j} \mid f(A_{i,j}) \mid . \tag{4}$$

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

$$\|f\|_{\mathrm{v}}^* \equiv \max\left(\|f\|_\infty,\ \|f\|_{\mathrm{v}},\ \sup_u \|v \to f(u,v)\|_{\mathrm{v}},\ \sup_v \|u \to f(u,v)\|_{\mathrm{v}}\right) < \infty. \tag{5}$$

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 $\|f\|_{\mathrm{v}}^* < \infty$, $\|g\|_{\mathrm{v}} < \infty$.*

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

For this we refer to Gill, van der Laan and Wellner (1993) or for the k-variate case ($k \geq 2$) to Gill (1992). This provides us with the following corollary:

Corollary 3.1 *Let f and g be two bivariate cadlag functions ans suppose that $\|f\|_{\mathrm{v}}^* < \infty$. Then*

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

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 3.2 *Let $f : \mathbb{R}^2 \to \mathbb{R}$. If $\|f\|_{\mathrm{v}}^* < \infty$ and $f > \delta > 0$, then $\|1/f\|_{\mathrm{v}}^* < \infty$.*

The proof requires some combinatorial arguments following directly from the definition (4) of $\|\cdot\|_{\mathrm{v}}$ (it is sketched for general k in Gill, 1993a).

Finally, our analysis makes strongly use of the following refined functional delta method:

Theorem 3.2 (Functional delta-method). *Let $\Phi : D_\phi \subset (D, \|\cdot\|) \to (E, \|\cdot\|_1)$, where $(D, \|\cdot\|)$ and $(E, \|\cdot\|_1)$ are normed vector spaces. Endow D and E with the Borel sigma-algebra.*

Suppose that $D_n, D_0, D_\phi \subset D$ are so that $F_n, F \in D_\phi$; $Z_n \equiv \sqrt{n}(F_n - F) \in D_n$; $Z \in D_0$, D_0 separable and

1. *$Z_n \stackrel{D}{\Longrightarrow} Z$ in $(D, \|\cdot\|)$, where Z is Borel measurable.*

2. *Φ satisfies the following differentiability property: if $h_n \equiv \sqrt{n}(G_n - G) \to h$, $G_n, G \in D_\phi$, with $h_n \in D_n$ and $h \in D_0$, then*

$$\sqrt{n}\left(\Phi(G + (1/\sqrt{n})h_n)) - \Phi(G)\right) - d\Phi(G)(h) \to 0 \tag{6}$$

for a certain continuous linear mapping $d\Phi(G) : D_0 \subset (D, \|\cdot\|) \to (E, \|\cdot\|_1)$.

Then

$$\sqrt{n}\,(\Phi(F_n) - \Phi(F)) \overset{D}{\Longrightarrow} d\Phi(F)(Z) \text{ in } (E, \|\cdot\|_1).$$

The delta-method formulated in this way, by using that the differentiability only has to be verified for sequences $h_n \in D_n$, is essentially more convenient than the usual formulation in terms of Hadamard-differentiability tangentially to a subspace: see Gill, 1989, Reeds, 1976, van der Vaart and Wellner (1993) and Wellner (1993). This theorem follows straightforwardly from the extended continuous mapping theorem as stated in van der Vaart and Wellner (1993). We will sometimes refer to the differentiability property (6) as compact differentiability, meaning (6) for an appropriate choice of D_n. For a discussion about the utility of the functional delta-method for analyzing estimators and also for obtaining its asymptotic variance we refer to Gill, van der Laan, Wellner (1993).

3.2 Outline of Proof of Consistency and Weak Convergence

We are now ready to give the proof of theorem 3.1. Equation (2) is given by:

$$S(t) = \Psi(t) + \int_0^{\tau_1}\int_0^{\tau_2} \frac{S(t \vee y)}{S(y)} dP_{00}(y). \tag{7}$$

If we consider P_{00} as fixed (known), then S can be considered as a solution of $K(S, \Psi) = 0$, where

$$K(S, \Psi)(t) \equiv \Psi(t) - S(t) + \int_0^{\tau_1}\int_0^{\tau_2} \frac{S(t \vee y)}{S(y)} dP_{00}(y). \tag{8}$$

Pruitt's estimator S_n is a solution of (3) which is the same equation but where $P_{11}, P_{01}, P_{10}, P_{00}$ are replaced by their empirical distributions and the singly censored conditional probabilities W_1 and W_2 are replaced by the weighted product limit estimators $\widehat{W}_1$ and $\widehat{W}_2$, respectively, defined by equation (17) in section 6. In formulas we have:

$$\begin{aligned} S_n(t) &= \Psi_n(t) + \int_0^{\tau_1}\int_0^{\tau_2} \frac{S_n(t \vee y)}{S_n(y)} dP_{00}^n(y) \\ &= \Psi_n^*(t) + \int_0^{\tau_1}\int_0^{\tau_2} \frac{S_n(t \vee y)}{S_n(y)} dP_{00}(y), \end{aligned}$$

where

$$\Psi_n^*(t) \equiv \Psi_n(t) + \int_0^{\tau_1}\int_0^{\tau_2} \frac{S_n(t \vee y)}{S_n(y)} d(P_{00}^n - P_{00})(y).$$

Consequently, S_n is an implicit solution of $K(S_n, \Psi_n^*) = 0$.

We will apply the implicit function theorem for Banach spaces (Hildebrandt and Graves, 1927, Flett, 1980, p. 205) to

$$K : (D([0,\tau]), \|\cdot\|_\infty)^2 \to (D([0,\tau]), \|\cdot\|_\infty) : \ (S, \Psi) \to K(S, \Psi).$$

It says:

Theorem 3.3 (Implicit function theorem). *Assume*

1. *K is a continuously Fréchet differentiable functional from an open subset W of $(D([0,\tau]), \|\cdot\|_\infty)^2$ into $(D([0,\tau]), \|\cdot\|_\infty)$, with $(S, \Psi) \in W$. Continuity of the derivative is defined as continuity with respect to the operator norm: If $\|x_n - x\|_\infty \to 0$, then* $\sup_{\|h\|_\infty = 1} \|dK(x_n)(h) - dK(x)(h)\|_\infty \to 0$.

2. The partial derivative $d_1K(S,\Psi) : (D([0,\tau]), \|\cdot\|_\infty) \to (D([0,\tau]), \|\cdot\|_\infty)$ *is invertible, and its inverse is continuous (i.e. it is an isomorphism).*

Then there are open neighborhoods U_0 *of* Ψ *and* V_0 *of* S *in* $(D([0,\tau]), \|\cdot\|_\infty)$ *such that for each* $\Psi' \in U_0$*, there is a unique* $S' \in V_0$ *such that* $K(S',\Psi') = 0$*. Moreover, if we define* Θ *by* $S' = \Theta(\Psi')$*, then for* U *and* V *small enough,* $\Theta(\cdot)$ *is a continuously Fréchet differentiable mapping from* U *into* V*. Its derivative is given by*

$$d\Theta(\Psi) = -\left(d_1K(\Theta(\Psi),\Psi)\right)^{-1} \circ d_2K(\Theta(\Psi),\Psi).$$

Because of the simple structure of K (P_{00} is fixed), continuous Fréchet differentiability of K is easy to verify, provided that $S > \epsilon > 0$ as is guaranteed by our assumptions.

All the work has to be done in the verification of 2. The partial derivative $d_1K(S,\Psi)$ of K with respect to S is given by: $-(I-A) : (D([0,\tau]), \|\cdot\|_\infty) \to (D([0,\tau]), \|\cdot\|_\infty)$, where

$$(I-A)(h)(t) = h(t) - \int_0^{\tau_1}\int_0^{\tau_2} \frac{h(t\vee y)S(y) - h(y)S(t\vee y)}{S^2(y)} dP_{00}(y).$$

In the next section it will be proved that $I-A$ is invertible and that its inverse $\sum_{n=0}^{\infty} A^n$ is a continuous operator (see theorem 4.1). In this proof it is important to notice that the integrand $(h(t\vee y)S(y) - h(y)S(t\vee y))/S^2(y)$ is zero at point τ. Therefore we only have to integrate over $\widetilde{B} \equiv [0,\tau]\backslash\{\tau\}$, and by assumption 1 we have that

$$\int_{\widetilde{B}} \frac{dP_{00}}{S} = G([0,\tau]) - G(\{\tau\}) = 1 - \delta < 1,$$

which we will need in the invertibility proof.

Consequently, we can apply the implicit function theorem. The implicit function theorem tells us that there exists a solution S_n' close to S where $S_n' = \Theta(\Psi_n^*)$. In section 5 we will prove that Ψ_n is uniformly consistent and that $\sqrt{n}(\Psi_n - \Psi)$ converges weakly as elements of $(D[0,\tau], B, \|\cdot\|_\infty)$. This does not immediately imply the same results for Ψ_n^* because it involves S_n. However, the following argument proves it.

The modified self-consistency equation (3) tells us that $S_n(t) > \overline{P}_{11}^n(t)$. By assumption 2 and 3 on f, g we have that $\overline{P}_{11}$ is uniformly bounded away from zero on $[0,\tau]$ and we know by Glivenko-Cantelli that $\overline{P}_{11}^n \to \overline{P}_{11}$. Consequently $S_n(t) > \delta > 0$ with probability tending to 1. Moreover, S_n is monotone (S_n only assigns positive mass) and $y \to S_n(t\vee y)/S_n(y)$ is bounded by 1 and, by lemma 3.2, is of uniformly (in n and t) bounded uniform sectional variation . Now, by using integration by parts we can bound $\int_0^{\tau_1}\int_0^{\tau_2} S_n(t\vee y)/S_n(y)d(P_{00} - P_{00}^n)(y)$ by a bounded constant (involving the latter variation) times the supremum norm of $P_{00} - P_{00}^n$, and consequently it follows that this term converges uniformly to zero with probability one, independently of the asymptotic behaviour of S_n. Therefore, if Ψ_n converges uniformly with probability one to Ψ then Ψ_n^* converges uniformly with probability one to Ψ, independently of the asymptotic behaviour of S_n.

Consequently the consistency of Ψ_n provides us with consistency of Ψ_n^* and therefore the continuous mapping theorem (see van der Vaart and Wellner, 1993) provides us with uniform consistency of $S_n = \Theta(\Psi_n^*)$ (Θ is Fréchet differentiable). Moreover, the continuous mapping theorem provides us also straightforwardly with the following: if S_n is uniformly consistent and $\sqrt{n}(\Psi_n - \Psi)$ converges weakly to a Gaussian process, then $\sqrt{n}(\Psi_n^* - \Psi)$ converges weakly to a Gaussian process, but another process. Now, the functional delta method theorem 3.2 applied to $\Theta(\Psi_n^*)$ provides us with the weak convergence of $\sqrt{n}(S_n - S)$ to a Gaussian process, namely $\sum A^k$ applied to the limiting distribution of $\sqrt{n}(\Psi_n^* - \Psi)$.

The implicit function theorem tells us that there exists a solution S_n' close to S where $S_n' = \Theta(\Psi_n^*)$ and the result derived above holds for this S_n'. Because $K(S, \Psi_n^*) = 0$ might have several solutions, the S_n which we compute with the EM-algorithm is not necessarily the $S_n' = \Theta(\Psi_n^*)$ given by the implicit function theorem. However, if we prove that each survival function S_n which solves $K(S_n, \Psi_n^*) = 0$ is consistent, then for n large enough we have $S_n = \Theta(\Psi_n^*)$. We will prove this in the next section (lemma 4.2).

We conclude that in the next sections the following four things have to be proved:

- $I - A$ is invertible, and has a continuous inverse (theorem 4.1).
- Each survival function S_n which solves $K(S_n, \Psi_n^*) = 0$ is consistent (lemma 4.2).
- $\|\Psi_n - \Psi\|_\infty$ converges with probability 1 to zero (section 6).
- $\sqrt{n}(\Psi_n - \Psi)$ converges weakly as elements of $(D[0,\tau], B, \|\cdot\|_\infty)$, jointly with the empirical process $\sqrt{n}(P_n - P_{F,G})$. This is also proved in section 6 by application of the results of section 5.

Notice that Ψ_n involves density estimators so that the third and fourth point do not follow from empirical process theory and are certainly not trivial. In order to carry through the analysis we need conditions on the kernel, bandwidth and smoothness assumptions on f and g as mentioned in theorem 3.1.

4 Invertibility of the derivative of the modified self-consistency equation.

Recall $\widetilde{B} \equiv [0,\tau] \backslash \{\tau\}$, and define the operator: $I - A : \ (D[0,\tau], \|\cdot\|_\infty) \to (D[0,\tau], \|\cdot\|_\infty)$ by

$$(I - A)(h)(t) = h(t) - \int_{\widetilde{B}} \left\{ \frac{h(t \vee y) S(y) - h(y) S(t \vee y)}{S^2(y)} \right\} dP_{00}(y).$$

As shown in the general proof, in order to apply the implicit function theorem to the equation $K(S, \Psi)$ we need to prove that the linear operator $(I - A)$ is an isomorphism.

Theorem 4.1 *The linear operator $I - A : (D[0,\tau], \|\cdot\|_\infty) \to (D[0,\tau], \|\cdot\|_\infty)$ as defined above is an isomorphism (i.e. a linear invertible mapping with continuous inverse). Its inverse is given by:*

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

Lemma 4.1 *If we say that f is non-increasing, then we mean: if $t \geq s$, so $t_1 \geq s_1$, $t_2 \geq s_2$, then $f(t) \leq f(s)$ (similarly for f increasing). If f is larger than or equal to zero on $[0,\tau]$ and f is increasing, then we denote that with $f \geq 0 \uparrow$. Define $\delta = \int_{\widetilde{B}} 1/S dP_{00} = \int_{\widetilde{B}} P(C_1 \in dy_1, C_2 \in dy_2) = P(C \in \widetilde{B})$. By assumption 1 ($G$ has an atom in the point $\{\tau\}$) we have $\delta < 1$.*

The linear operator A satisfies the following properties:

1) $h \geq 0 \uparrow \Rightarrow \ \|A(h)\|_\infty \leq \delta \|h\|_\infty$.

2) $h \geq 0 \uparrow \Rightarrow \ A(h) \geq 0 \uparrow$

and hence also

3) $h \geq 0 \uparrow \Rightarrow \ \|A^n(h)\|_\infty \leq \delta^n \|h\|_\infty$.

Proof. 1) is trivially proved. Assume that $h \in D([0,\tau])$ is non-decreasing and $h \geq 0$, then it is easy to see that $A(h)$ is also non-decreasing: if $t \geq s$, then

$$\begin{aligned} A(h)(t) - A(h)(s) &= \int \frac{(h(t\vee y) - h(s \vee y))\, S(y) + h(y)\,(S(s\vee y) - S(t \vee y))}{S^2(y)} dP_{00}(y) \\ &\geq 0, \end{aligned}$$

where we use that $h(t\vee y) - h(s\vee y) \geq 0$ and $S(s\vee y) - S(t\vee y) \geq 0$. Now, rewrite the numerator of the integrand of A as follows:

$$h(t\vee y)S(y) - h(y)S(t\vee y) = (h(t\vee y) - h(y))\, S(y) + h(y)\,(S(y) - S(t\vee y))\,.$$

So we have $A(h) \geq 0$. This shows that: if $h\geq 0 \uparrow$, then $A(h) \geq 0 \uparrow$, which proves 2). By 1) we have that if $h \geq 0$, then $\|A(h)\|_\infty \leq \delta \|h\|_\infty$. Therefore, if $h\geq 0 \uparrow$, then $\|A^n(h)\|_\infty \leq \delta^n \|h\|_\infty$, which proves 3).□

Proof (Theorem). Define

$$\begin{aligned} A_1(h) &\equiv \int_{\widetilde{B}} \left(\frac{h(t\vee y)S(y)}{S^2(y)}\right) dP_{00}(y) \\ A_2(h) &\equiv \int_{\widetilde{B}} \left(\frac{h(y)S(t\vee y)}{S^2(y)}\right) dP_{00}(y) \\ A(h) &\equiv A_1(h) - A_2(h). \end{aligned}$$

One should notice that if for an $h \in D([0,\tau])$ the series $T(h) \equiv \sum_{n=0}^\infty A^n(h)$ converges, then $(I - A)\left(\sum_{n=0}^\infty A^n\right)(h) = \left(\sum_{n=0}^\infty A^n\right)(I - A)(h) = h$. Assume that for $h \geq 0$: $\|T(h)\|_\infty \leq M\|h\|_\infty$. Then the same inequality holds for $h \leq 0$. Consider now a general $h \in D([0,\tau])$ with $h = hI(h>0) + hI(h \leq 0) \equiv h_1 + h_2$, $h_1, h_2 \in D([0,\tau])$. Then

$$\|T(h)\|_\infty \leq \|T(h_1)\|_\infty + \|T(h_2)\|_\infty \leq M(\|h_1\|_\infty + \|h_2\|_\infty) \leq 2M\|h\|_\infty.$$

So, then T is a well defined bounded linear operator, which proves the theorem. So it remains to prove that if $h \geq 0$, then $\|T(h)\|_\infty \leq M\|h\|_\infty$.

Here follows the proof of this. Let $h \geq 0$ be fixed. For a constant c we have that $A_1(c) = \delta c$, where $\delta \equiv \int_{\widetilde{B}} 1/S dP_{00} < 1$ (see lemma above). Using this tells us that:

$$A(h) = A_1(h) - A_2(h) = A_1\left(h - \|h\|_\infty\right) - \left(A_2(h) - \delta\|h\|_\infty\right). \tag{9}$$

We have for each $h \in D([0,\tau])$: $\|A_i(h)\|_\infty \leq \delta\|h\|_\infty$, $i = 1, 2$. We have that $A_2(h) - \delta\|h\|_\infty \leq 0$ and because S is non-increasing $A_2(h) - \delta\|h\|_\infty$ is non-increasing (recall $h \geq 0$). Now, by applying property 3) of lemma 4.1 to $A_2(h) - \delta\|h\|_\infty \leq 0 \downarrow$ we have:

$$\|A^n\left(A_2(h) - \delta\|h\|_\infty\right)\|_\infty \leq \delta^n \|\left(A_2(h) - \delta\|h\|_\infty\right)\|_\infty \leq \delta^{n+1}\|h\|_\infty, \tag{10}$$

Notice also that $A_1\left(h - \|h\|_\infty\right) \leq 0$. Now, we are ready to prove with induction that the following statement $P(n)$ is true for all $n \in \mathbb{N}$:

$$P(n),\ n \in \mathbb{N}: \text{ If } h \geq 0, \text{ then } \|A^n(h)\|_\infty \leq n\delta^n\|h\|_\infty.$$

$P(1)$ is trivially true. Assume $P(n)$ is true. We will prove $P(n+1)$.

$$\begin{aligned} \|A^{n+1}(h)\|_\infty &\leq \|A^n A_1\left(\|h\|_\infty - h\right)\|_\infty + \|A^n\left(A_2(h) - \delta\|h\|_\infty\right)\|_\infty \\ &\quad \text{(by (9) and the triangle inequality, respectively.)} \\ &\leq n\delta^n \|A_1\left(\|h\|_\infty - h\right)\|_\infty + \delta^{n+1}\|h\|_\infty \\ &\quad \text{(by } P(n) \text{ and (10), respectively).} \\ &\leq n\delta^{n+1}\|h\|_\infty + \delta^{n+1}\|h\|_\infty = (n+1)\delta^{n+1}\|h\|_\infty. \end{aligned}$$

So with induction we proved: if $h \geq 0$, then $\|A^n(h)\|_\infty \leq n\delta^n \|h\|_\infty$. By completeness of the Banach space $(D[0,\tau], \|\cdot\|_\infty)$ it follows by proving the Cauchy property for $T(h)$ that $T(h)$ exists. So if $h \geq 0$, then

$$\|T(h)\|_\infty \leq \sum_{n=0}^{\infty} \|A^n(h)\|_\infty \leq \|h\|_\infty \sum_{n=0}^{\infty} n\delta^n = \|h\|_\infty \frac{\delta}{(1-\delta)^2}.$$

This completes the proof of theorem 4.1. $\square$

We will now prove consistency for each survival function S_n which solves $K(S_n, \Psi_n^*) = 0$. We have

$$S_n(t) = \Psi_n^*(t) + \int \frac{S_n(t \vee y)}{S_n(y)} dP_{00}(y)$$

and

$$S_0(t) = \Psi_0(t) + \int \frac{S_0(t \vee y)}{S_0(y)} dP_{00}(y).$$

Subtracting these two equations provides us with:

$$\begin{aligned}(S_n - S_0)(t) \quad &= \quad (\Psi_n^* - \Psi_0)(t) + \int (S_n - S_0)(t \vee y) \frac{dP_{00}(y)}{S_0(y)} \\ &\quad - \int (S_n - S_0)(y) \frac{S_n(t \vee y)}{S_n(y) S_0(y)} dP_{00}(y). \qquad (11)\end{aligned}$$

Denote

$$\begin{aligned}A_{S_n}(S_n - S_0)(t) \quad &\equiv \quad \int (S_n - S_0)(t \vee y) \frac{dP_{00}(y)}{S_0(y)} - \int (S_n - S_0)(y) \frac{S_n(t \vee y)}{S_n(y) S_0(y)} dP_{00}(y) \\ &\equiv \quad A_1(S_n - S_0)(t) - A_2^{S_n}(S_n - S_0)(t),\end{aligned}$$

where A_1 is the same as defined in the proof of theorem 4.1 and $A_2^{S_n}$ is slighly different from the operator A_2. Now, (11) reduces to:

$$(I - A_{S_n})(S_n - S_0) = (\Psi_n^* - \Psi_0).$$

Therefore for consistency of S_n it suffices to prove that $\sum_{k=0}^{\infty} A_{S_n}^k$ is a bounded linear operator (uniformly in n). However, because $A_2^{S_n}$ has all the properties which we needed from A_2 (as the reader can verify for himself) we can do exactly the same proof as the proof of theorem 4.1 and we also get the same bound $\delta/(1-\delta)^2$ of the norm of $(I - A_{S_n})^{-1}$. The only condition we need is that $S_n > \epsilon > 0$ on $[0,\tau]$ which holds for n large enough (because $S_n > \overline{P}_{11}^n$ and $\overline{P}_{11}^n \to \overline{P}_{11} > \delta > 0$, by assumption 1). This proves that $\|S_n - S_0\|_\infty \to 0$ a.s.

In the same way it is proved that $K(S, \Psi_n^*) = 0$ has a unique solution among the survival functions $S > 0$ on $[0,\tau]$. This provides us with the following lemma:

Lemma 4.2 *Recall the assumptions on the model. Each survival function S_n which solves $K(S_n, \Psi_n^*) = 0$ is strongly uniformly consistent and if $S_n > 0$ on $[0,\tau]$, then S_n is also the unique survival function solution of $K(S, \Psi_n^*) = 0$ in the class of survival functions S with $S > 0$ on $[0,\tau]$.*

5 Functional delta-method for functionals of density estimators.

Consider the problem of estimation of a functional $\Phi(F) = \Gamma(f) \in (D[0,\tau], \|\cdot\|_\infty)$, where $f \equiv dF/d\mu : \mathbb{R}^d \to \mathbb{R}$ is the density of a d-variate distribution F w.r.t. to the Lebesque measure μ, using i.i.d. observations $X_1, \ldots, X_n$, $X_i \sim F$. We can estimate $\Gamma(f)$ with $\Gamma(f_n)$ where $f_n(x) = (nh^d)^{-1}\sum_{i=1}^n K((x - X_i)/h)$ is the usual d-variate kernel density estimator (Silverman, 1986) with a bandwidth $h = h(n) \to 0$. In this section we show how we can use theorem 3.2 in order to obtain a functional delta method theorem for the analysis of functionals of density estimators.

Lemma 5.1 *Assume that:*
1. $\|f_n - f_0\|_\infty \to 0$ *a.s.*
2. *Define* $\widetilde{F}_n(x) \equiv \int_0^x f_n(x)dx$ *and let* F_n *be the empirical distribution function of* X_i, $i = 1, \ldots, n$. *Denote the limiting distribution of* $\sqrt{n}(F_n - F)$ *with* Z *(i.e. the* F*-Brownian bridge), where* Z *is a Borel measurable Gaussian process concentrated on a separable subset* D_0 *of* $(D[0,\tau], B, \|\cdot\|_\infty)$. *Assume that* $\widetilde{Z}_n \equiv \sqrt{n}(\widetilde{F}_n - F) \stackrel{D}{\Longrightarrow} Z$ *in* $(D[0,\tau], B, \|\cdot\|_\infty)$.
3. $\limsup_n \|f_n\|_{\mathrm{v}}^* < M < \infty$ *a.s.*

Assume now that Φ *satisfies the following purely analytical property: For each sequence* $\widetilde{Z}_n \equiv \sqrt{n}(\widetilde{F}_n - F) \to Z$ *in supnorm for* $Z \in D_0$, $\|f_n - f\|_\infty \to 0$ *and* $\|f_n\|_{\mathrm{v}}^* = O(1)$, *we have:*

$$\sqrt{n}(\Phi(\widetilde{F}_n) - \Phi(F)) - d\Phi(F)(Z) \to 0$$

in supnorm for a continuous linear mapping $d\Phi(F) : (D[0,\tau], \|\cdot\|_\infty) \to (D[0,\tau], \|\cdot\|_\infty)$.

Then

$$\sqrt{n}\left(\Phi(\widetilde{F}_n) - \Phi(F)\right) \stackrel{D}{\Longrightarrow} d\Phi(F)(Z) \text{ in } (D[0,\tau], B, \|\cdot\|_\infty). \tag{12}$$

The proof of this lemma is nothing else than an application of theorem 3.2 applied to $\Phi : (D[0,\tau], \|\cdot\|_\infty) \to (D[0,\tau], \|\cdot\|_\infty)$ with a good choice for D_n so that we only have to verify the differentiability property for sequences $\widetilde{F}_n$ for which $f_n \to f$ and $\|f_n\|_{\mathrm{v}}^* < M < \infty$. It works as follows.

Firstly, notice that $\|f_n - f\|_\infty \to 0$ a.e. is equivalent with: for each $\epsilon > 0$ $P(\lim_{N\to\infty}\sup_{n>N}\|f_n - f\|_\infty > \epsilon) = 0$. By Fatou's lemma this implies that for all $\epsilon > 0$ $\lim_{N\to\infty} P(\sup_{n>N}\|f_n - f\|_\infty > \epsilon) = 0$. This implies that there exist sequences $\epsilon_n \to 0$, $\delta_n \to 0$ and $N(\delta_n) \in \mathbb{N}$ so that

$$P\left(\sup_{n>N(\delta_n)}\|f_n - f\|_\infty < \epsilon_n\right) > 1 - \delta_n. \tag{13}$$

Now, we will define the D_n in theorem 3.2. Let $\mathcal{F}$ be the set of all distribution functions $F_1 : \mathbb{R}^d \to \mathbb{R}$ which are absolute continuous w.r.t. the Lebesque measure. Define now

$$D_n \equiv \{\sqrt{n}(F_1 - F) : F_1 \in \mathcal{F},\ \|f_1 - f\|_\infty < \epsilon_n, \|f_1\|_{\mathrm{v}}^* \leq M\}$$

and

$$\widetilde{Z}_n^* \equiv \widetilde{Z}_n I(\widetilde{Z}_n \in D_n),$$

where we mean with $I(\widetilde{Z}_n \in D_n)$ that if $\widetilde{Z}_n \notin D_n$, then $I(\widetilde{Z}_n \in D_n) = 0$. Consequently $\widetilde{Z}_n^* \in D_n$. By $\limsup_n \|f_n\|_{\mathrm{v}}^* = O(1)$ and (13) we have that for each $\epsilon > 0$, there exists a $N(\epsilon)$ so that

$\widetilde{Z}_n \in D_n$ for all $n > N(\epsilon)$ with probability $1-\epsilon$. Therefore, $\widetilde{Z}_n \stackrel{D}{\Longrightarrow} Z$ implies $\widetilde{Z}_n^* \stackrel{D}{\Longrightarrow} Z$. Now, apply theorem 3.2 to $\Phi(\widetilde{F}_n^*)$, where $\widetilde{F}_n^* = F + 1/\sqrt{n}\widetilde{Z}_n^*$. This provides us with:

$$\sqrt{n}\left(\Phi(\widetilde{F}_n^*) - \Phi(F)\right) \stackrel{D}{\Longrightarrow} d\Phi(F)(Z).$$

Because $Z_n^* = Z_n$ with probability tending to 1 we have that

$$\|\sqrt{n}\left(\Phi(\widetilde{F}_n^*) - \Phi(F)\right) - \sqrt{n}\left(\Phi(\widetilde{F}_n) - \Phi(F)\right)\|_\infty = o_P(1).$$

The required weak convergence follows now from the general fact that

$$X_n \stackrel{D}{\Longrightarrow} X,\ Y_n = o_P(1) \Rightarrow Z_n \equiv X_n + Y_n \stackrel{D}{\Longrightarrow} X,$$

which completes the proof of the lemma.

The lemma can be immediately generalized to all kinds of properties of the sequences $\widetilde{F}_n$ which we plug in, as long as these properties hold with probability tending to 1. The lemma will be applied in the analysis of $\sqrt{n}(\Psi_n - \Psi)$ in the next section. Here, the probabilistic conditions of the lemma will be analyzed in generality.

6 Weak convergence of the explicit part.

We will apply the refined functional delta-method lemma 5.1 in the analysis of $\sqrt{n}(\Psi_n - \Psi)$. We will see in the next subsection that $\int W_1 dP_{01}$ has a representation in terms of two distribution functions F_N and F_Y of the data, and of course a symmetric version of this statement holds for $\int W_2 dP_{10}$ (say F_N', F_Y'). So we can represent Ψ in terms of distribution functions of the data for which we have a joint weak convergence result for its empirical counterpart, namely $P \equiv (F_N, F_Y, F_N', F_Y', P_{11}, P_{01}, P_{10}, P_{00})$. In order to get Ψ_n one replaces these distributions by their empirical versions: so $\Psi = \Psi(P)$ and $\Psi_n = \Psi(P_n)$ where we know that $\sqrt{n}(P_n - P) \stackrel{D}{\Longrightarrow} Z$ for a certain Gaussian process Z. The refined delta method lemma 5.1 states now that in order to prove weak convergence of $\sqrt{n}(\Psi_n - \Psi)$ it is enough to show that this representation satisfies the characterization of compact differentiability for all sequences $Z_n = \sqrt{n}(P_n - P) \in D_n$, where D_n is chosen so that the empirical process $Z_n \in D_n$ with probability tending to 1.

Define the following normalized estimators:

$$\begin{aligned}
U_{01}^n(t) &\equiv \sqrt{n}(P_{01}^n - P_{01})(t)\\
U_{11}^n(t) &\equiv \sqrt{n}(P_{11}^n - P_{11})(t)\\
U_{10}^n(t) &\equiv \sqrt{n}(P_{10}^n - P_{10})(t).
\end{aligned}$$

Then we can rewrite $\sqrt{n}(\Psi_n - \Psi)$ in terms of these normalized empirical differences:

$$\begin{aligned}
\sqrt{n}(\Psi_n - \Psi)(t) = U_{11}^n(t)&\\
&+ \int_0^{\tau_1}\int_{t_2}^{\tau_2} W_1(t_1,y_1,y_2)dU_{01}^n(y) + \sqrt{n}\int_0^{\tau_1}\int_{t_2}^{\tau_2}\left(\widehat{W}_1 - W_1\right)(t_1,y_1,y_2)dP_{01}(y)\\
&+ \int_{t_1}^{\tau_1}\int_0^{\tau_2} W_2(t_2,y_2,y_1)dU_{uc}^n(y) + \sqrt{n}\int_{t_1}^{\tau_1}\int_0^{\tau_2}\left(\widehat{W}_2 - W_2\right)(t_2,y_2,y_1)dP_{10}(y)\\
&+ \int_0^{\tau_1}\int_{t_2}^{\tau_2}\left(\widehat{W}_1 - W_1\right)(t_1,y_1,y_2)dU_{01}^n(y) + \int_{t_1}^{\tau_1}\int_0^{\tau_2}\left(\widehat{W}_2 - W_2\right)(t_2,y_2,y_1)dU_{10}^n(y).
\end{aligned}$$

We will now verify the purely analytical characterization of compact differentiability. Assume that $\sqrt{n}(P_n - P)$ converges in supremum norm to Z. In order to prove the characterization of compact differentiability we need to prove that the first, second, third, fourth and fifth term converge in supremum norm, and that the last terms converge to zero in supremum norm. The third term will be analyzed in the next subsection. In that analysis one has to keep continually in mind that if we consider weak convergence of the normalized empirical processes which occur in this term that these should be taken jointly with the other normalized empirical processes! (we will not remind the reader again of this fact). The fifth term is of exactly the same structure. The first term is trivial. For the convergence of the second and fourth term we apply integration by parts lemma 3.1 so that the integrals become integrals with respect to W_1 and W_2 and that U_{01}^n and U_{10}^n appear as functions. This can be done because W_1 and W_2 are of bounded uniform sectional variation uniformly in $t \in [0, \tau]$, by assumption 2 about f, g and $S(\tau) > 0$. This proves the convergence of the second and fourth integral. In the next section we will see that $\widehat{W}_1, \widehat{W}_2$ are continuous functionals of strongly uniformly consistent estimators. This gives that $\widehat{W}_1 - W_1$ and $\widehat{W}_2 - W_2$ converge uniformly to zero almost everywhere. Furthermore, we will show that $\widehat{W}_1$ and $\widehat{W}_2$ are of bounded uniform sectional variation uniformly in n. Therefore the last two terms are of the form:

Lemma 6.1 (Helly-Bray). *Let* $f_n, Z_n, Z \in (D[0, \tau], \| \cdot \|_\infty)$. *Assume* $\|f_n\|_\infty \to 0$, $\|f_n\|_{\mathrm{v}}^* < M < \infty$, $\|Z_n - Z\|_\infty \to 0$. *Then* $\int f_n dZ_n \to 0$.

Proof. These terms are shown to converge to zero as follows. $\int f_n dZ_n = \int f_n d(Z_n - Z) + \int f_n dZ$. By corollary 3.1 we can bound the first integral by $16 \|Z_n - Z\|_\infty \|f_n\|_v^*$ which converges to zero. For the second term we apply the Helly- Bray technique which works as follows: approach Z by a function (e.g. step function) Z_m in supnorm, where Z_m is of bounded variation. Then $\int f_n dZ = \int f_n d(Z - Z_m) + \int f_n dZ_m$. Apply integration by parts to the first term, then it follows that this term converges to zero if $m \to \infty$. Bound the second term by the supnorm of f_n times the variation norm of Z_m. Then this term converges to zero if $m \to \infty$ slowly enough w.r.t. n. By straightforward lim sup argument, this proves that $\int f_n dZ$ converges to zero.

For this Helly Bray technique we refer to (Loève (1955) page 181, Gill (1989) for the one dimensional case, Gill (1992) and Gill, van der Laan, Wellner (1993).□

This proves the convergence to zero of the last two terms. Now, we have verified the required differentiability of $\Psi(P)$ at P. Application of the functional delta method provides us now with weak convergence of $\sqrt{n}(\Psi_n - \Psi)$.

Similarly, but easier, it is shown that the strong uniform consistency of $\widehat{W}_1, \widehat{W}_2$ and P_{01}^n, P_{10}^n, P_{11}^n provides us with the strong uniform consistency of Ψ_n.

It remains to analyze the third term. We will do this by application of the functional delta-method for functionals of density estimators as stated in lemma 5.1 in the preceding section. For this we need uniformly consistent density estimators on $[0, \tau]$ which are of bounded uniform sectional variation and the integrated density estimator should be asymptotically equivalent with the empirical distribution function. These will be constructed in the next two subsections. The uniform consistency on $[0, \tau]$ requires an edge-correction at the edge of $[0, \tau]$. We will study this in the next subsection.

6.1 Uniformly consistent edge-corrected bivariate density estimators.

If we have a density which is uniformly continuous on $\mathbb{R}^2$, then necessary and sufficient conditions for strong uniform consistency of the kernel density estimator f_n with kernel K and bandwidth

h_n are: $h_n \to 0$, $(nh_n^2)/\log n \to \infty$ as $n \to \infty$, K measurable w.r.t. Lebesque measure, $\int \mid K(t) \mid dt < \infty$, $\int K(t)dt = 1$ (see Bertrand-Retali, 1974, 1978). We will also assume that K has compact support within $[-1,1]^2$. Schuster (1985) shows how to get uniformly consistent estimators of univariate densities with support $[c,d]$, by using symmetrization techniques which brings one back to the problem of estimating a uniformly continuous density on $\mathbb{R}$. In van der Laan (1991) this method is generalized to the two dimensional case. Here we only discuss a method introduced by Richard Gill. From now on we will work with a symmetric kernel which satisfies the conditions mentioned above.

Gill's method. This method requires that f also puts mass outside the rectangle $[0,\tau]$. Let f_n be the kernel density estimator with bandwidth h. Let $\vec{h} = (h,h)$. Now, define the edge corrected kernel density estimator f_n^* as follows: f_n^* equals f_n on $[\vec{h}, \tau - \vec{h}]$, and if $x \notin [\vec{h}, \tau - \vec{h}]$, then $f_n^*(x)$ gets the value of $f_n(x')$, where x' is the closest point to x on the edge of $[\vec{h}, \tau - \vec{h}]$. We will prove that f_n^* is uniformly consistent.

We have:

$$\sup_{x \in [0,\tau]} \mid f_n^*(x) - f(x) \mid \leq \sup_{x \in [\vec{h}, \tau - \vec{h}]} \mid f_n(x) - f(x) \mid + \sup_{x \in [\vec{h}, \tau - \vec{h}]^c} \mid f_n(x') - f(x) \mid,$$

where the complement is taken within $[0,\tau]$. Consider the first term. f_n uses here only data on $[0,\tau]$. Therefore, if there is mass outside $[0,\tau]$, then the data is indistinguishable from a uniformly continuous density which equals f on $[0,\tau]$, but nicely bends down to zero outside $[0,\tau]$. Now, by Bertrand-Retali's result f_n is uniformly consistent on $[\vec{h}, \tau - \vec{h}]$, which proves that the first term converges to zero.

For the second term we have:

$$\mid f_n(x') - f(x) \mid \leq \mid f_n(x') - f(x') \mid + \mid f(x') - f(x) \mid .$$

So the supremum over $[\vec{h}, \tau - \vec{h}]^c$ of the first term converges to zero by the uniform consistency of f_n on $[\vec{h}, \tau - \vec{h}]$ and the supremum of the second term converges to zero by the uniform continuity of f.

6.2 Bivariate kernel density estimators: simultaneous uniform consistency, consistency of derivative and asymptotic normality of the integrated kernel density estimator.

Let X_i, $i = 1, \ldots, n$ be n i.i.d. copies of a bivariate random variable $X = (X_1, X_2) \sim f$, where f is a bivariate continuous density on $[0,\tau]$ w.r.t. the Lebesque measure. Let

$$f_n(x_1, x_2) = \frac{1}{nh^2} \sum_{i=1}^{n} K\left(\frac{x_1 - X_{1i}}{h}, \frac{x_2 - X_{2i}}{h}\right)$$

be the bivariate kernel density estimator with kernel K.

We will now find conditions on f, K and h which provides us with a kernel density estimator which is uniformly consistent, which is of bounded uniform sectional variation uniformly in n and for which the integrated density estimator is asymptotically equivalent with the empirical distribution. We know already the conditions for uniform consistency (see above): for the bandwidth we need $(nh^2)/\log(n) \to \infty$. We will find the conditions for each of the two remaining properties and then combine them in a lemma.

Pointwise consistency of derivative of density. Assume that K has compact support and that $K^{1,1}(x) \equiv (d^2/dx_1dx_2)K(x_1,x_2)$ is continuous on $[0,\tau]$. Firstly, we will find conditions which guarantee that $\limsup \|f_n\|_{\mathrm{v}}^* = O(1)$ a.s. By the triangle inequality we have

$$\mid f_n^{1,1}(x) - f^{1,1}(x) \mid \leq \mid f_n^{1,1}(x) - Ef_n^{1,1}(x) \mid + \mid Ef_n^{1,1}(x) - f^{1,1}(x) \mid, \tag{14}$$

where

$$f_n^{1,1}(x_1,x_2) \equiv \frac{d^2}{dx_1dx_2} f_n(x_1,x_2) = \frac{1}{nh^4}\sum_{i=1}^n K^{1,1}\left(\frac{x_1 - X_{1i}}{h}, \frac{x_2 - X_{2i}}{h}\right).$$

Firstly, we will study the second term. We have that

$$\begin{aligned} E\left(f_n^{1,1}(x_1,x_2)\right) &= \frac{1}{h^4} E\left(K^{1,1}\left(\frac{x_1 - X_{1i}}{h}, \frac{x_2 - X_{2i}}{h}\right)\right) \\ &= \frac{1}{h^4}\int K^{1,1}\left(\frac{x_1 - y_1}{h}, \frac{x_2 - y_2}{h}\right) f(y_1,y_2)dy_1dy_2 \\ &= \frac{1}{h^2}\int K^{1,1}(z_1,z_2) f\left(x_1 - hz_1, x_2 - hz_2\right) dz_1dz_2. \end{aligned}$$

We will say that $f \in C^k[0,\tau]$ if f has k derivatives in both coordinates and $f^{k,k}$ is continuous. If $f \in C^k[0,\tau]$, then we have the following Taylor expansion:

$$f\left(x_1 + hz_1, x_2 + hz_2\right) = \sum_{i\geq 0, j\geq 0, i+j\leq k} \frac{(hz_1)^i (hz_2)^j}{i!j!} f^{i,j}(x_1,x_2) + o(h^k).$$

Assume that $K^{1,1}$ satisfies the following *orthogonality conditions.*

$$\int K^{1,1}(z_1,z_2) z_1^i z_2^j dz_1dz_2 = 0,\ i \geq 0,\ j \geq 0,\ 1 \leq i+j \leq k. \tag{15}$$

Then we have

$$\begin{aligned} Ef_n^{1,1}(x_1,x_2) &= f^{1,1}(x_1,x_2) + o\left(\frac{1}{h^2} h^k \int \mid K^{1,1}(z_1,z_2) \mid dz_1dz_2\right) \\ &= f^{1,1}(x_1,x_2) + o(h^{k-2}), \end{aligned} \tag{16}$$

using that $\int \mid K^{1,1}(z_1,z_2) \mid dz_1dz_2 < \infty$ because $K \in C^1[0,\tau]$. So if $f \in C^2[0,\tau]$, then $\mid Ef_n^{1,1}(x_1,x_2) - f^{1,1}(x_1,x_2) \mid = o(1)$, uniformly in $x \in [0,\tau]$.

Let's now consider the first term of (14). We have:

$$\mathrm{Var}\left(f_n^{1,1}(x_1,x_2)\right) = \frac{1}{nh^8}\mathrm{Var}\left(K^{1,1}\left(\frac{x_1 - X_{1i}}{h}, \frac{x_2 - X_{2i}}{h}\right)\right)$$

and

$$\mathrm{Var}\left(K^{1,1}\left(\frac{x_1 - X_{1i}}{h}, \frac{x_2 - X_{2i}}{h}\right)\right) \leq E\left(K^{1,1}\left(\frac{x_1 - X_{1i}}{h}, \frac{x_2 - X_{2i}}{h}\right)\right)^2 = O(h^2),$$

where we use that we only have to integrate over a square with width h. So we conclude that if $f \in C^2[0,\tau]$, then

$$\sup_{x\in[0,\tau]} \mathrm{Var}\left(\mid f_n^{1,1}(x) - f^{1,1}(x) \mid\right) = O\left(\frac{1}{nh^6}\right) + o(1).$$

This tells us that if we choose h so that $h_n n^{1/6} \to \infty$, then $f_n^{1,1}(x) \to f^{1,1}(x)$ a.s. for all x. Moreover, by the triangle inequality, $\mathrm{E}\left(\int \mid f_n^{1,1} - Ef_n^{1,1} \mid (x)dx\right) \to 0$ and (16) it follows also that

$$\limsup_{n\to\infty} \int \mid f_n^{1,1}(x) \mid dx = \int \mid f^{1,1}(x) \mid dx < \infty \text{ a.s.}$$

By doing similar calculation for the sections this provides us with: $\limsup \|f_n\|_{\mathrm{v}}^* < \infty$.

Weak convergence of the integrated kernel density estimator. Let F_n be the empirical distribution function and $k_n(z) \equiv 1/h^2 K(z)$. Then

$$\widetilde{F}_n(x) \equiv \int_0^x f_n(y)dy = F_n * k_n(x), \text{ the convolution of } F_n \text{ and } k_n.$$

Now,

$$\sqrt{n}(\widetilde{F}_n - F) = \sqrt{n}\,(F_n * k_n - F * k_n) + \sqrt{n}\,(F * k_n - F)\,.$$

The first term is an empirical process indexed by smoothed indicators which clearly form a Donsker class and therefore empirical process theory provides us immediately with weak convergence of the first term. The second term is the bias of which we have to take care. We have ($\int K(z)dz = 1$)

$$\begin{aligned}(F * k_n - F)(x) &= \int F(x-y)\frac{1}{h^2}K\left(\frac{y}{h}\right)dy - F(x) \\ &= \int \left(F(x_1 - hz_1, x_2 - hz_2) - F(x)\right) K(z_1, z_2)dz_1dz_2.\end{aligned}$$

Assume $F \in C^k[0,\tau]$ and that K satisfies (15). Then we have that

$$(F * k_n - F)(x) = o(h^k).$$

In other words we have to choose k so that $\sqrt{n}o(h^k) \to 0$. However, for the bounded variation condition we needed that h_n converges to zero slower than $n^{-1/6}$. So we need that $F \in C^4[0,\tau]$ and hence that $f \in C^3[0,\tau]$. This proves the following lemma:

Lemma 6.2 *Let X_i be n i.i.d. copies of a bivariate $X \sim f$, where $f \in C^3[0,\tau]$. Let f_n be a bivariate kernel density estimator with kernel K and bandwidth h_n, as defined above and let $\widetilde{F}_n = \int_0^x f_n(y)dy$ be the integrated kernel density estimator.*

Assume that $K \in C^1[0,\tau]$, K satisfies the orthogonality conditions (15) for $k = 4$ and $\int K(t)dt = 1$.

- *If $h_n \to 0$, $nh_n^2/\log(n) \to \infty$, then f_n is uniformly consistent.*
- *If $h_n n^{1/6} \to \infty$, then $f_n^{1,1}(x) \to f^{1,1}(x)$ a.s. for all $x \in [0,\tau]$ and $\limsup_{n\to\infty} \|f_n\|_{\mathrm{v}}^* \le M < \infty$ a.s.*
- *If $\sqrt{n}o(h_n^4) \to 0$, then the integrated density estimator is asymptotically equivalent with the empirical distribution function.*

Consequently, if $h_n = n^{-1/7}$, then $\|f_n - f\|_\infty \to 0$ a.s, $\limsup_n \|f_n\|_{\mathrm{v}}^* = O(1)$ a.s. and $\sqrt{n}(\widetilde{F}_n - F) \stackrel{D}{\Longrightarrow} Z$ where Z is the F-Brownian bridge.

If we choose a K of the form $K(z_1, z_2) = g(z_1)g(z_2)$ for a certain differentiable $g : \mathbb{R} \to \mathbb{R}$, then it is trivial to construct a kernel which satisfies the orthogonality conditions (15) for $k = 4$.

Remark. In our application we have that f_n is an edge-corrected kernel density estimator. In the preceding section we already showed that under the same assumptions as with the uncorrected kernel density estimator it will be uniformly consistent on $[0,\tau]$. It is straightforward to verify that we can also apply the lemma to the edge corrected f_n (using the symmetrization technique as in van der Laan, 1991) or an higher order correction (using Gill's method) at the edges. We will not go into these technical details.

6.3 Application of the functional delta-method for density estimators.

Here, we will prove weak convergence of $\sqrt{n}\int_0^{\tau_1}\int_{t_2}^{\tau_2}\left(\widehat{W}_1 - W_1\right)(t_1, y_1, y_2)dP_{01}(y_1, y_2)$ as random elements of the cadlag function space $D([0,\tau])$ endowed with the supremum norm and the Borel sigma-algebra, by application of the delta-method lemma 5.1. Let K be a kernel satisfying the properties as mentioned in Lemma 6.2. We will now define $\widehat{W}_1$. Firstly, we define

$$
\begin{aligned}
F_N(t_1,t_2) &\equiv P(\widetilde{T}_1 \le t_1, \widetilde{T}_2 \le t_2, D_1 = 1, D_2 = 1)\\
F_Y(t_1,t_2) &\equiv P(\widetilde{T}_1 > t_1, \widetilde{T}_2 \le t_2, D_2 = 1).
\end{aligned}
$$

So F_N is the subdistribution of the doubly uncensored observations and F_Y is the subdistribution of the in the second coordinate uncensored observations. Moreover, we define their derivatives w.r.t. the second coordinate.

$$
\begin{aligned}
N(y_1,y_2) &\equiv \frac{d}{dy_2}F_N(y_1,y_2)\\
Y(y_1,y_2) &\equiv \frac{d}{dy_2}F_Y(y_1,y_2).
\end{aligned}
$$

These are densities which appear in W_1. We will estimate them with kernel density estimators. In fact, we will use edge corrected kernel density estimators N_n, Y_n which converges uniformly on the rectangle $[0,\tau]$. This means that we make the usual kernel density estimator constant at distance h from the edge of $[0,\tau]$, as discussed in section 6.1. In order not to complicate the notation we will suppress this fact in the notation. We denote these density estimators with N_n and Y_n:

$$
\begin{aligned}
N_n(y_1,y_2) &\equiv \int_0^{y_1}\frac{1}{nh^2}\sum_{i=1}^n K\left(\frac{u-\widetilde{T}_{1i}}{h}, \frac{y_2-\widetilde{T}_{2i}}{h}\right) I(D_{i1}=1, D_{i2}=1)du\\
Y_n(y_1,y_2) &\equiv \int_{y_1}^{\tau_1}\frac{1}{nh^2}\sum_{i=1}^n K\left(\frac{u-\widetilde{T}_{1i}}{h}, \frac{y_2-\widetilde{T}_{2i}}{h}\right) I(D_{i2}=1)du.
\end{aligned}
$$

We define $S(\cdot \mid y_2)$ as the survival function of T_1 given $T_2 = y_2$. Therefore $S(\cdot \mid y_2)$ has the well known product limit representation in terms of its conditional hazard $\int_0^{t_1} N(du, y_2)/Y(u-, y_2)$ (Gill and Johansen, 1990):

$$
\begin{aligned}
S(y_1 \mid y_2) &\equiv P(T_1 > y_1 \mid T_2 = y_2) = \prod_{(0,y_1]}\left(1 - \frac{P(T_1 \in du \mid T_2 = y_2)}{P(T_1 \ge u \mid T_2 = y_2)}\right)\\
&= \prod_{(0,y_1]}\left(1 - \frac{N(du,y_2)}{Y(u-,y_2)}\right).
\end{aligned}
$$

Of course, a natural estimator of S is obtained by replacing N, Y by N_n, Y_n:

$$
S_n(y_1 \mid y_2) \equiv \prod_{(0,y_1]}\left(1 - \frac{N_n(du,y_2)}{Y_n(u-,y_2)}\right).
$$

We can now define $\widehat{W}_1$:

$$\begin{aligned} W_1(t_1, y_1, y_2) &\equiv \frac{S(t_1 \vee y_1 \mid y_2)}{S(y_1 \mid y_2)} \\ \widehat{W}_1(t_1, y_1, y_2) &\equiv \frac{S_n(t_1 \vee y_1 \mid y_2)}{S_n(y_1 \mid y_2)}. \end{aligned} \tag{17}$$

Because we are also concerned with the probabilistic behaviour of the integrated density estimators, which play the role of $\widetilde{F}_n$ in lemma 5.1, we also need notation for them.

$$\begin{aligned} \widetilde{F}_{Yn}(t_1, t_2) &\equiv \int_0^{t_2} Y_n(t_1, y_2) dy_2 \\ \widetilde{F}_{Nn}(t_1, t_2) &\equiv \int_0^{t_2} N_n(t_1, y_2) dy_2. \end{aligned}$$

So these are smoothed empirical distribution functions F_{Nn}, F_{Yn} of F_N and F_Y, respectively. Now, we define the following mappings:

$$\begin{aligned} \Phi_1 : D([0,\tau])^2 &\to D([0,\tau]) : \\ (F_N, F_Y) &\mapsto \int_0^{y_1} \frac{N(du, y_2)}{Y(u, y_2)} \equiv \Lambda(y_1 \mid y_2) \\ \Phi_2 : D([0,\tau]) &\to D([0,\tau]) : \\ \Lambda &\mapsto \int_0^{\tau_1} \int_{t_2}^{\tau_2} \prod_{[y_1, t_1 \vee y_1]} (1 - \Lambda(du \mid y_2)) \, dP_{01}(y_1, y_2) \\ &= \int_0^{\tau_1} \int_{t_2}^{\tau_2} W_1(t_1, y_1, y_2) dP_{01}(y_1, y_2). \end{aligned}$$

Define $\Lambda_n(y_1 \mid y_2) \equiv \int_0^{y_1} N_n(du, y_2)/Y_n(u-, y_2)$. Finally, we can define the Φ to which we apply the refined functional delta method lemma 5.1.

$$\Phi \equiv \Phi_2 \circ \Phi_1 \ : D([0,\tau])^2 \to D([0,\tau]). \tag{18}$$

We will prove weak convergence of

$$\sqrt{n} \left(\Phi(\widetilde{F}_{Nn}, \widetilde{F}_{Yn}) - \Phi(F_N, F_Y) \right)(t_1, t_2) = \sqrt{n} \int_0^{\tau_1} \int_{t_2}^{\tau_2} \left(\widehat{W}_1 - W_1 \right)(t_1, y_1, y_2) dP_{01}(y_1, y_2)$$

by applying lemma 5.1 to Φ at the point (F_N, F_Y). We apply lemma 5.1 with (F_n is the empirical df. of F):

$$\begin{aligned} F &= (F_N, F_Y) \\ F_n &= (F_{Nn}, F_{Yn}) \\ \widetilde{F}_n &= (\widetilde{F}_{Nn}, \widetilde{F}_{Yn}) \\ f_n &= \left(\frac{d}{dy_1} N_n(y_1, y_2), \frac{d}{dy_1} Y_n(y_1, y_2) \right) \\ f &= \left(\frac{d}{dy_1} N(y_1, y_2), \frac{d}{dy_1} Y(y_1, y_2) \right). \end{aligned}$$

According to lemma 5.1 we can separate the analysis in a purely probabilistic part and a purely analytical part. For the probabilistic conditions we just have to apply lemma 6.2 with the $f, f_n, F, \widetilde{F}_n$ above. This provides us with the following proposition:

Proposition 6.1 *Assume $f, g \in C^3[0,\tau]$, $K \in C^1[0,\tau]$, K satisfies the orthogonality conditions (15) for $k=4$ and $\int K(t)dt = 1$. Moreover assume that $h_n = n^{-1/7}$. Then*
1. *$m_n(y_1,y_2) \equiv (d/dy_1 N_n(y_1,y_2), d/dy_1 Y_n(y_1,y_2))$ is a uniformly consistent estimator of $m \equiv (d/dy_1 N(y_1,y_2), d/dy_1 Y(y_1,y_2))$.*
2. *$m_n^{1,1}(x) \to m^{1,1}(x)$ a.s. for all $x \in [0,\tau]$ and $\limsup_{n\to\infty} \|m_n\|_{\mathrm{v}}^* = O(1)$ a.s.*
3. *$\widetilde{M}_n(\cdot) \equiv \int_0^{(\cdot)} m_n(x)dx$ is asymptotically equivalent with the empirical distribution $M_n = (F_{Nn}, F_{Yn})$ of $M \equiv (F_N, F_Y)$.*

Notice that if $f, g \in C^3[0,\tau]$, then $m \in C^3[0,\tau]$. Moreover notice that m_n is the kernel density estimator of m. Therefore, this proposition is an immediate corollary of lemma 6.2.

6.4 Analytical part of the analysis.

We will now prove the differentiability property for Φ as stated in lemma 5.1. We can consider sequences N_n, Y_n with: $N_n \to N$, $Y_n \to Y$, $\|N_n\|_{\mathrm{v}}^* = O(1)$, $\|Y_n\|_{\mathrm{v}}^* = O(1)$, $H_Y^n \equiv \sqrt{n}(\widetilde{F}_Y^n - F_Y) \to H_Y$, $H_N^n \equiv \sqrt{n}(\widetilde{F}_N^n - F_N) \to H_N$, where $F_Y(y_1,y_2) = \int_0^v Y(y_1,v)dv$ and similarly for F_N, $\widetilde{F}_Y^n$ and $\widetilde{F}_N^n$. Firstly, we will consider the difference $W_{1n} - W_1$.

Recall the definition of $\Lambda(du \mid y_2) = N(du,y_2)/Y(u-,y_2)$ and its empirical version $\Lambda_n(du \mid y_2) = N_n(du,y_2)/Y_n(u-,y_2)$. We can consider $S(\cdot \mid y_2)$ also as a measure: $S((a,b] \mid y_2) \equiv P(T_1 \in (a,b] \mid T_2 = y_2)$. Recalling its product integral representation w.r.t. Λ it follows that:

$$\begin{aligned} S\left((a,b] \mid y_2\right) &= \prod_{(a,b]} (1 - \Lambda(du \mid y_2)) \\ S_n\left((a,b] \mid y_2\right) &= \prod_{(a,b]} (1 - \Lambda_n(du \mid y_2)). \end{aligned}$$

Now, we have that

$$\begin{aligned} (W_{1n} - W_1)(t_1,y_1,y_2) &= \prod_{[y_1, t_1 \vee y_1]} (1 - \Lambda_n(du \mid y_2)) - \prod_{[y_1, t_1 \vee y_1]} (1 - \Lambda_1(du \mid y_2)) \\ &= S_n\left([y_1, t_1 \vee y_1] \mid y_2\right) - S\left([y_1, t_1 \vee y_1] \mid y_2\right). \end{aligned}$$

Duhamel's equation for the product integral (see Gill and Johansen, 1990, equation 42) tells us that:

$$(S_n - S)\left([y_1, t_1 \vee y_1] \mid y_2\right) = \int_{y_1}^{t_1 \vee y_1} (\Lambda_n - \Lambda)(du \mid y_2) S_n\left([y_1,u) \mid y_2\right) S\left((u, t_1 \vee y_1] \mid y_2\right).$$

So we need to study the difference $(\Lambda_n - \Lambda)(du \mid y_2) = (\Phi_1(N_n,Y_n) - \Phi_1(N,Y))(du,y_2)$. We have that

$$(\Lambda_n - \Lambda)(du \mid y_2) = D\Phi_1^n(du,y_2) + R_1^n(du,y_2),$$

where

$$\begin{aligned} D\Phi_1^n(du,y_2) &\equiv \frac{(N_n - N)(du,y_2)}{Y(u,y_2)} + \frac{N(du,y_2)\,(Y - Y_n)(u,y_2)}{Y^2(u,y_2)} \\ R_1^n(du,y_2) &\equiv \frac{(N_n - N)(du,y_2)\,(Y - Y_n)(u,y_2)}{Y^2(u,y_2)} + \frac{N_n(du,y_2)\,(Y - Y_n)^2(u,y_2)}{YY_n^2(u,y_2)}. \end{aligned}$$

This is trivially verified.

Furthermore, we have by using the multiplicativity of the product integral that

$$\begin{aligned} S_n\left([y_1,u)\mid y_2\right)S\left((u,t_1\vee y_1]\mid y_2\right) &= (S_n-S)\left([y_1,u)\mid y_2\right)S\left((u,t_1\vee y_1]\mid y_2\right) \\ &\quad +S([y_1,u)\mid y_2)S\left((u,t_1\vee y_1]\mid y_2\right) \\ &= W_1(t_1,y_1,y_2)+R_2^n(u,t_1,y_1,y_2), \end{aligned}$$

where

$$R_2^n(u,t_1,y_1,y_2)\equiv (S_n-S)\left([y_1,u)\mid y_2\right)S((u,t_1\vee y_1]\mid y_2).$$

So we have that

$$(W_{1n}-W_1)(t_1,y_1,y_2)=\int_{y_1}^{t_1\vee y_1}\left(D\Phi_1^n(du,y_2)+R_1^n(du,y_2)\right)\left(W_1(t_1,y_1,y_2)+R_2^n(u,t_1,y_1,y_2)\right).$$

This provides us with:

$$\begin{aligned} &\sqrt{n}\left(\Phi(N_n,Y_n)-\Phi(N,Y)\right)(t) \\ =\;& \sqrt{n}\int_0^{\tau_1}\int_{t_2}^{\tau_2}(W_{1n}-W_1)(t_1,y_1,y_2)dP_{01}(y_1,y_2) \\ =\;& \sqrt{n}\int_0^{\tau_1}\int_{t_2}^{\tau_2}\int_{y_1}^{t_1\vee y_1}\left(D\Phi_1^n(du,y_2)+R_1^n(du,y_2)\right)\left(W_1(t_1,y_1,y_2)+R_2^n(u,t_1,y_1,y_2)\right)dP_{01}(y) \\ =\;& \sqrt{n}\int_0^{t_1}\int_{t_2}^{\tau_2}\int_{y_1}^{t_1}D\Phi_1^n(du,y_2)W_1(t_1,y_1,y_2)dP_{01}(y_1,y_2)+\text{ Remainder.} \end{aligned} \tag{19}$$

The expression (19) consists of four terms of which three involve R_1^n or R_2^n or both. These three terms form the remainder and are shown to converge to zero (below). Firstly, we will be concerned with the main term which is linear in N_n-N and Y_n-Y.

Because N,Y,N_n,Y_n,W_1 are of bounded uniform sectional variation, integrals with respect to N,Y,N_n,Y_n,W_1 are well defined. The terms are of the form $\int(N_n-N)Gdx$ for a function G which involves Y,N in numerator and Y in denominator. Because N,Y are of bounded uniform sectional variation and Y is uniformly bounded away from zero by assumption 3 (that f,g are strictly positive on $[0,\tau+\epsilon]\backslash[0,\tau]$, see introduction) this G is of bounded uniform sectional variation (lemma 3.2). Therefore we can apply integration by parts so that the terms become of the form $\int(\widetilde{F}_N^n-F_N)dG$ (and similar one dimensional integrals over sections) and hence we can linearize the expression above in H_Y^n,H_N^n.

After having linearized in $H_Y^n,\ H_N^n$ and using that H_Y^n and H_N^n converge in supremum norm to H_Y,H_N, it is trivial to see that this term converges in supremum norm to an expression linear in H_N and H_Y: just bound terms of the form $\int(H_N^n-H_N)dG\le\|H_N^n-H_N\|_\infty\|G\|_{\mathrm{v}}^*$ and use that we know that $\|H_N^n-H_N\|_\infty$ converges to zero and that G is of bounded uniform sectional variation because N and Y are. This expression is written down in van der Laan (1991) and we will denote it with $d\Phi(N,Y)(H_N,H_Y)$.

Remainder. Because N_n,N,Y_n,Y are of uniformly bounded uniform sectional variation and Y_n is uniformly bounded away from zero (because Y is uniformly bounded away from zero and $Y_n\to Y$) $y\to R_2^n(u,t_1,y)$ is of bounded uniform sectional variation uniformly in n and (u,t). The product integral is a continuous functional in (N,Y) with respect to the supremum norm (see Gill and Johansen 1990). Therefore R_2^n converges uniformly to zero. Here we use for the first time that $N_n-N\to 0$ and $Y_n-Y\to 0$.

All three terms are in fact dealt with in the same way. These terms have the following structure:

$$
\begin{aligned}
&\sqrt{n}\int_0^{\tau_1}\int_{t_2}^{\tau_2}\int_{y_1}^{t_1}(Y_n-Y)^2(s,y_2)m_n(s,y_1,y_2)dsdy_1dy_2\\
&=\sqrt{n}\int_0^{\tau_1}\int_{y_1}^{t_1}\left((F_Y^n-F_Y)(s,\cdot)(Y_n-Y)(s,\cdot)m_n(s,y_1,\cdot)\right)((t_2,\tau_2])dsdy_1\\
&\quad-\sqrt{n}\int_0^{\tau_1}\int_{y_1}^{t_1}\int_{t_2}^{\tau_2}(F_Y^n-F_Y)(s,y_2)\left((Y_n-Y)(s,\cdot)m_n(s,y_1,\cdot)\right)(dy_2)dsdy_1\\
&\longrightarrow 0 \text{ w.r.t. the supnorm.}
\end{aligned}
$$

Here m_n involves N_n, N, Y, Y_n, which are of uniformly bounded uniform sectional variation. The convergence to zero in supnorm of the first term is trivial. The second integral is of the structure $\int Z_n dF_n$, where $\|Z_n - Z\|_\infty \to 0$ (for $\|H_Y^n - H_Y\|_\infty \to 0$) and $\|F_n\|_{\mathrm{v}}^* = O(1)$ (for $\|(Y_n, m_n)\|_{\mathrm{v}}^* = O(1)$) and $\|F_n - F\|_\infty \to 0$ (for $\|Y_n - Y\|_\infty \to 0$). Therefore convergence to zero follows from the Helly-Bray lemma 6.1 with $Z_n \equiv \sqrt{n}(F_Y^n - F_Y)$ and $F_n(\cdot) \equiv (Y_n - Y)(s,\cdot)m_n(s,y_1,\cdot)$.

Consider the term $\sqrt{n}\int D\Phi_1^n(N_n - N, Y_n - Y)R_2^n dP_{01}$. Just as above we linearize this in H_N^n and H_Y^n. Then we obtain a term of the form $\int Z_n dF_n$ where $\|Z_n - Z\|_\infty \to 0$ (for $\|H_N^n - H_N\|_\infty \to 0$), $\|F_n\|_{\mathrm{v}}^* = O(1)$ (for $\|R_2^n\|_{\mathrm{v}}^* = O(1)$) and $\|F_n - F\|_\infty \to 0$ (for $\|R_2^n - R_2\|_\infty \to 0$). The Helly-Bray lemma 6.1 tells us now that this term converges to zero.

Consider now the term $\sqrt{n}\int R_1^n W_1 dP_{01}$. Again, by doing integration by parts we can linearize in H_N^n (or H_Y^n) and these will be integrated with respect to a measure which involves $d(Y_n - Y)$ (or $d(N_n - N)$). Therefore this term is again of the form $\int Z_n dF_n$ where $\|Z_n - Z\|_\infty \to 0$ (for $\|H_N^n - H_N\|_\infty \to 0$), $\|F_n\|_{\mathrm{v}}^* = O(1)$ (for $\|(N_n, Y_n, N, Y)\|_{\mathrm{v}}^* = O(1)$) and $\|F_n - F\|_\infty \to 0$ (for $\|(N_n, Y_n) - (N, Y)\|_\infty \to 0$). The Helly-Bray lemma 6.1 tells us now that this term converges to zero.

The term $\sqrt{n}\int R_1^n R_2^n dP_{01}$ is proved to converge to zero in the same way as $\sqrt{n}\int R_1^n W_1 dP_{01}$. This completes the differentiability proof.

We conclude that once we have arranged that we only have to verify the differentiability property in lemma 5.1 for sequences with a consistent density which is of uniformly (in n) bounded uniform sectional variation, then integration by parts and the Helly-Bray technique are the only ingredients one needs in order to prove that the remainder converges in supremum norm to zero. We have proved the analytical condition of lemma 5.1. Application of lemma 5.1 provides us now with weak convergence of $\sqrt{n}\int(W_{1n} - W_1)dP_{01}$ to the Gaussian process $d\Phi(N,Y)(H_N, H_Y)$.

7 Weak Consistency of the Semiparametric Bootstrap

In principle one can obtain from the previous analysis the influence curve of Pruitt's estimator S_n: $\sum_{n=0}^\infty A^n(IC(\Psi,t))$, where $IC(\Psi,t)$ stands for the influence curve of Ψ at the point $t=(t_1,t_2)$, and A is the operator analyzed in section 4. However, in view of the expected complexity, a bootstrap approach is more useful in practice. We are interested if we can use a bootstrap result in order to estimate the variance of the limiting distribution, say Z, of $\sqrt{n}(S_n - S)$. The observations $(\widetilde{T}_i, D_i),\ i = 1,\ldots,n$ are distributed with probability measure $P_{F,G}$ indexed by the unknown underlying distributions functions $F,\ G$ with properties as given in the introduction. Assume $\widetilde{F}_n,\ \widetilde{G}_n$ is a sequence of estimators of $F,\ G$, based on n i.i.d. $(\widetilde{T}_i, D_i) \sim P_{F,G},\ i = 1,\ldots,n$. Let $\widetilde{S}_n^\#$ be Pruitt's estimator based on n i.i.d. observations $(\widetilde{T}_i, D_i)^\#$ sampled from the known probability measure $P_{\widetilde{F}_n,\widetilde{G}_n}$. Let $(\Omega^\infty, \Sigma^\infty, P^\infty)$ be the infinite but countable cartesian product of $([0,\tau], \text{Borel}, P_{F,G})$ endowed with the product probability measure $P^\infty = P_{F,G} \times P_{F,G} \ldots$, and let $\omega^\infty \in \Omega^\infty$ be the sequence of the i.i.d. observations $(\widetilde{T}_i, D_i) \sim P_{F,G}$.

Let Z_n be a sequence of random elements of $(D[0,\tau], B, \|\cdot\|_\infty)$ converging weakly to Z, and let P_n and P be their corresponding probability measures. Assume d is any metric on the space of probability measures on $(D[0,\tau], B, \|\cdot\|_\infty)$ which metrizes weak convergence in the sense that

$$P_n \stackrel{D}{\Longrightarrow} P \iff d(P_n, P) \to 0.$$

We denote bootstrap distributions with $P_n^{\#}$. We want to know what assumptions about $\widetilde{F}_n$ and $\widetilde{G}_n$ are sufficient for achieving that $\widetilde{S}_n^{\#}$ has the same limiting behaviour as S_n: so if $Z_n \equiv \sqrt{n}(S_n - S) \stackrel{D}{\Longrightarrow} Z$ then we would like to have $Z_n^{\#} \equiv \sqrt{n}(\widetilde{S}_n^{\#} - \widetilde{S}_n) \stackrel{D}{\Longrightarrow} Z$ given ω^∞ for P^∞-almost each ω^∞. This result is called almost sure consistency of the bootstrap, which is equivalent with $d(P_n^{\#}, P) \to 0$ P^∞ a.s., where $Z_n^{\#} \sim P_n$ and $Z \sim P$. In this section we are able to get this result for an almost sure representation of $\widetilde{F}_n$ and $\widetilde{G}_n$. Back in real life this means that we have this result in probability instead of almost surely: $d(P_n^{\#}, P) \stackrel{P_\infty}{\to} 0$. We will now state the bootstrap theorem.

Theorem 7.1 (Weak consistency of the semiparametric bootstrap) *See definitions above. Assume $\widetilde{F}_n$, $\widetilde{G}_n$ satisfy assumption 1-3, as given in the introduction, uniformly in n. Furthermore, assume that $\sqrt{n}(\widetilde{F}_n - F) \stackrel{D}{\Longrightarrow} Z_F$ and $\sqrt{n}(\widetilde{G}_n - G) \stackrel{D}{\Longrightarrow} Z_G$ and $\|\tilde{f}_n - f\|_\infty \stackrel{P^\infty}{\to} 0$, $\|\tilde{g}_n - g\|_\infty \stackrel{P^\infty}{\to} 0$, where $\tilde{f}_n$, $\tilde{g}_n$ are the densities of $\widetilde{F}_n, \widetilde{G}_n$. We know that $Z_n \equiv \sqrt{n}(S_n - S) \stackrel{D}{\Longrightarrow} Z$. Denote the distributions of Z_n with Q_n and of Z with Q. Let $Q_n^{\#}$ be the distribution of $Z_n^{\#} \equiv \sqrt{n}(\widetilde{S}_n^{\#} - \widetilde{S}_n)$ given ω_∞ (so $\widetilde{S}_n$ is given). Then we have that $d(Q_n^{\#}, Q) \stackrel{P^\infty}{\to} 0$.*

For the proof we refer to van der Laan (1991).

Remark. For $\widetilde{F}_n$ and $\widetilde{G}_n$ we can take appropriately smoothed Dabrowska estimators. This can be seen as follows: Consider the Dabrowska representation of the survival function S, and derive from this representation a representation of the underlying density f. Then if we establish the expected continuity w.r.t. to $\|\cdot\|_\infty$ of this density representation as functional of the densities of the subsurvival functions, it follows that if we replace in the Dabrowska representation the empirical subsurvival functions by integrated strongly uniformly consistent kernel density estimators that the density $\tilde{f}_n$ of this smoothed Dabrowska estimator $\widetilde{S}_n$ of the survival function is a strongly uniformly consistent estimator of f, which is more than required in this theorem. We will illustrate this with the univariate Kaplan-Meier estimator (Dabrowka's representation is just a generalization of this).

Example. For the univariate survival function we have $S_0(t) = \prod_0^t(1 - \frac{N(ds)}{Y(s)})$, where N and Y can be estimated by integrated uniformly consistent kernel density estimators N_n and Y_n, respectively, where the bandwidth is chosen so that $\sqrt{n}(N_n - N)$ and $\sqrt{n}(Y_n - Y)$ are asymptotically normal (so $\frac{N(ds)}{Y(s)}$ is just a smoothed version of the famous Nelson-Aalen estimator). We have $f_0(t) = \prod_0^t(1 - \frac{N(ds)}{Y(s)})\frac{N(dt)}{Y(t)}$. Now, consider the estimator $f_n(t) = \prod_0^t(1 - \frac{N_n(ds)}{Y_n(s)})\frac{N_n(dt)}{Y_n(t)}$ and $\tilde{S}_n(t) = \prod_0^t(1 - \frac{N_n(ds)}{Y_n(s)})$. The continuity of the product integral (Gill, Johansen, 1990) provides us with $\|f_n - f_0\|_\infty \to 0$ in probability. Compact differentiability of the product integral (Gill, Johansen, 1990), provides us with weak convergence of $\sqrt{n}(\tilde{S}_n - S_0)$.

Recalling the Dabrowska representation of the bivariate survival function, it is now clear that we can imitate this proof if we have continuity and compact differentiability of the Dabrowska representation. This is proved in Gill, van der Laan, Wellner (1993): here we need compact differentiability of a quite complicated mapping which represents the covariance structure between T_1 and T_2 and of the bivariate product integral.

References

Bakker, D.M. (1990). *Two nonparametric estimators of the survival function of bivariate right censored observations*, Report BS-R9035, Centre for mathematics and computer science, Amsterdam.

Bertrand-Retali, M. (1974). Convergence uniforme d'un estimateur d'une densité de probabilité dans R^s. *Comptes Rendus de l'Ac. Sci. Paris* **278** 451–453.

Bertrand-Retali, M. (1978). Convergence uniforme d'un estimateur de la densité par la méthode du noyau. *Rev. Roumaine Math. Pures & Appl.* **23** 361–385.

Bickel, P.J., Klaassen, C.A.J., Ritov, Y., Wellner, J.A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* John Hopkins University Press. To appear.

Burke, M.D. (1988). Estimation of a bivariate survival function under random censorship, *Biometrika* **75**, 379–382.

Campbell, G., Földes, A. (1982). Large-sample properties of non-parametric bivariate estimators with censored data, pp. 103–122 in: *Nonparametric Statistical Inference*, B.V. Gnedenko, M.L. Puri and I. Vincze, eds., North Holland, Amsterdam.

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.

Dempster, A.P, Laird, N.P, and Rubin, D.B. (1977), Maximum likelihood from incomplete data via the EM-algorithm, *J. Roy. Statist. soc. sec*, **39** 1–38.

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.

Efron, B. (1967). The two sample problem with censored data, *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **4** 831–883.

Flett, T.M. (1980). *Differential Analysis*, Cambridge University Press, Cambridge.

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. (1991). *Non-and Semi-parametric Maximum Likelihood Estimators and the von Mises Method (Part 2)*, preprint nr. 664, Dept of Maths., Utrecht Univ.

Gill, R.D. & Johansen, S. (1990). A survey of product integration with a view towards application in survival analysis, *Ann. Statist.* **18**.

Gill, R.D. (1992). Multivariate survival analysis, *Theory Prob. Appl.* **37**, 19–36 and 307–328. (in Russian; English translation to appear).

Gill, R.D. (1993), *Lectures on survival analysis (part I)*, Preprint nr. 764, Department of Mathematics, Utrecht, the Netherlands.

Gill, R.D., van der Laan, M.J. and Wellner, J.A. (1993), *Inefficient estimators of the bivariate* *survival function for three models*, Preprint nr. 767, Department of Mathematics, Utrecht, the Netherlands.

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.

Hildebrandt, T.J. (1963), *Introduction to the theory of integration*, Academic Press, New York and London.

Hildebrandt, T.H., & Graves, L.M. (1927). Implicit functions and their differentials in general analysis, *Trans. Amer. Math. Soc.* **29** 127–153.

Hoffmann-Jörgensen, J. (1984). *Stochastic processes on Polish Spaces*, Unpublished manuscript.

van der Laan, M.J. (1991). *Analysis of Pruitt's estimator of the bivariate survival function*, Preprint nr. 648, Dept of Maths., Utrecht Univ.

van der Laan, M.J. (1992). *Efficient Estimator of the Bivariate Survival Function for Right Censored Data*, Technical Report No. 337, Department of Statistics, University of California, Berkeley.

van der Laan, M.J. (1993a) *Large sample simulations in the bivariate censoring model I.* Unpublished manuscript, Dept of Maths., Utrecht Univ.

van der Laan, M.J. (1993b), *Efficient estimator of the bivariate survival function and repairing NPMLE*, Preprint nr. 788, Department of Mathematics, Utrecht, the Netherlands.

van der Laan, M.J. (1993c) *Large sample simulations in the bivariate censoring model II.* Unpublished manuscript, Dept of Maths., Utrecht Univ.

van der Laan, M.J. (1993d), *Modified EM-estimator of the bivariate survival function*, Preprint nr. 768, Department of Mathematics, Utrecht, the Netherlands.

Loève, M. (1955). *Probability theory*, van Nostrand, New York.

Muñoz, A. (1980). *Non-parametric estimation from censored bivariate observations*, Technical Report, Stanford University.

Neuhaus, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *Ann. Math. Statist.* **42** 1285–1295.

Pollard, D. (1984). *Convergence of stochastic processes*, Springer, New York.

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 bivarate survival data and a bivariate survival function. In Klein, J.P. and Goel, P.K., editors, *Procedings NATO advanced studies workshop on survival analysis and related topics.* Karger, Basel.

Pruitt, R.C. (1991a). On negative mass assigned by the bivariate Kaplan-Meier estimator. *Ann. Stat.* **19**, 443–453.

Pruitt, R.C. (1991b). *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.

Reeds, J.A. (1976). *On the definition of von Mises functionals* (Ph.D. thesis), Dept. of Statistics, Harvard University, Research Report S-44.

Ritov, Y. (1992). *Estimating the survival function in presence of covariates,* to appear in Ann. Statist.

Schuster, E.F. (1985). Incorporating support constraints into nonparametric estimators of densities, *Commun. Statist.-Theor. Meth.*, **14(5)** 1123–1136.

Silverman, B.W. (1986). *Density estimation for statistics and data analysis*, Chapman and Hall, London.

Tsay, W-Y., & Leurgans, S. & Crowley, J. (1986). Nonparametric estimation of a bivariate survival function in the presence of censoring. *Ann. Statist.* **14** 1351-1365.

van der Vaart, A.W. (1988). *Statistical Estimation in Large Parameter Spaces.*, CWI tract **44**, Centre for Mathematics and Computer Science, Amsterdam.

Van der Vaart, A.W. & Wellner, J.A. (1993). *Weak Convergence and Empirical Processes.* IMS Lecture Notes-Monograph Series (to appear).

Wellner, J.A. (1993). The delta-method and the bootstrap. Tech. Report, Dept. of Statistics, University of Washington, Seattle.

Wichura, M.J. (1968), On the weak convergence of non-Borel probabilities on a metric space, *Ph.D. dissertation*, Columbia University.