ANNALES DE L'I. H. P., SECTION B

RICHARD D. GILL
MARK J. VAN DER LAAN
JON A. WELLNER

Inefficient estimators of the bivariate survival function for three models

Annales de l'I. H. P., section B, tome 31, nº 3 (1995), p. 545-597
<http://www.numdam.org/item?id=AIHPB_1995__31_3_545_0>

Ann. Inst. Henri Poincaré,

Vol. 31, n° 3, 1995, p. 545-597. *Probabilités et Statistiques*

Inefficient estimators of the bivariate survival function for three models

by

Richard D. GILL

Mathematical Institute, Utrecht University,
P.O. Box 80.010, 3508 TA Utrecht, Netherlands

Mark J. van der LAAN

Mathematical Institute, Utrecht University,
P.O. Box 80.010, 3508 TA Utrecht, Netherlands

and

Jon A. WELLNER

Department of Statistics, GN-22, B320 Padelford Hall,
University of Washington, Seattle, Washington 98195, U.S.A.

ABSTRACT. – Three explicit estimators of the bivariate survival functions for three types of data are analyzed by proving the analytical and probabilistic conditions necessary for application of the functional delta-method. It tells us that the estimators converge weakly at $\sqrt{n}$-rate to a Gaussian process and that for estimation of their asymptotical distribution the bootstrap works well. We also prove efficiency of the Dabrowska and Prentice-Cai estimator for the bivariate censoring model under independence.

Key words: Functional delta-method, compact differentiability, weak convergence, bootstrap, efficient estimator.

1. THREE MULTIVARIATE MODELS AND APPROACHES TO ESTIMATION

We begin by describing the three models and the resulting estimation problems upon which we will focus. In all three models the parameter to be estimated is the *bivariate survival function*. In this section we will describe the three *representations* of the bivariate survival function, as maps from the distribution function of the data, on which the three estimators for each model are based. The estimators are obtained by substituting the empirical distribution counterpart of the data into the representation.

Our aim is to prove that these estimators are *uniformly consistent* and that the estimators *converge weakly* in *supnorm* at root-n rate to a Gaussian process. Moreover, we also want to show that the *bootstrap* can be used to estimate the variance of these estimators and to obtain some *efficiency* results for these estimators.

The weak convergence and bootstrap can be proved by applying the *functional delta-method* (Gill, 1989, V. d. Vaart, Wellner, 1995). This means that we have to verify the (by the functional delta-method required) differentiability of the representation and the weak convergence of the empirical process which we plug in the representation. We were able to verify these conditions under essentially no conditions on the model. For a formal statement of our results see our final theorem in section 5. We also succeeded in proving that for the *bivariate censoring model* (our third model) the Dabrowska and Prentice-Cai estimator are efficient under independence. Practical simulations show that the asymptotic distribution is closely approached for surprisingly small samples (Bakker, 1990, Prentice-Cai, 1992*a*).

The organisation of the paper is as follows. In section 2 we will give the basic techniques as lemmas for obtaining the required *differentiability* result for the representations. In section 3 we will prove the differentiability results by applying these lemmas. In section 4 we will see how each representation leads to an estimator for each model by just substituting the empirical distribution of the data. In section 5 we verify the weak convergence of these empirical processes which provide us, by application of the functional delta method, with results which are summarized in our final theorem. Finally, in section 6 we prove that for the bivariate censoring model the Dabrowska and Prentice-Cai estimator are efficient under independence.

MODEL 1. – *Estimation of a bivariate distribution with known marginals.*

Suppose that $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent and identically distributed with distribution function F on $\mathbb{R}^2_{\geq 0}$, and suppose further that the marginal distribution functions F_X, F_Y of F are known: $F_X = F_X^0$ and $F_Y = F_Y^0$ where F_X^0, F_Y^0 are known one-dimensional marginal distributions.

Problem: Use the data and the knowledge of the marginal distributions to estimate the joint distribution F.

MODEL 2. – *The bivariate "three-sample" model.*

Suppose that $(X_{11}, Y_{11}), \ldots, (X_{1n_1}, Y_{1n_1})$ are independent and identically distributed with distribution function F on $\mathbb{R}^2_{\geq 0}$, that $X_{2,1}, \ldots, X_{2,n_2}$ are independent and identically distributed with distribution function $F_X = F(\cdot, \infty)$ on $\mathbb{R}$, and that $Y_{3,1}, \ldots, Y_{3,n_3}$ are independent and identically distributed with $F_Y = F(\infty, \cdot)$ on $\mathbb{R}$. Here the three samples are all independent. We can regard this as either a "missing data" model (the Y's are missing in sample 2 and the X's are missing in sample 3); or we can regard it as an "auxiliary samples" model in which in addition to the first sample of (X, Y) pairs we also have some auxiliary data (samples 2 and 3) concerning the marginal distributions.

Problem: Use all the data to estimate F.

MODEL 3. – *Bivariate random censoring.*

Suppose that $(S_1, T_1), \ldots, (S_n, T_n)$ are independent and identically distributed with distribution function F on $\mathbb{R}^2_{\geq 0} \equiv [0, \infty)^2$, that $(C_1, D_1), \ldots, (C_n, D_n)$ are independent and identically distributed with distribution function G on $\mathbb{R}^2_{\geq 0}$ independent of all of the (S, T)'s, and that we observe

$$(X_i, Y_i, \delta_i, \epsilon_i) \equiv (S_i \wedge C_i, T_i \wedge D_i, I[S_i \leq C_i], I[T_i \leq D_i]),\ i = 1, \ldots, n.$$

Problem: Use the observed data to estimate F.

For all three models we assumed that we have observations in $\mathbb{R}^2_{\geq 0}$. The estimators we propose are invariant under translations. Therefore our results can be generalized to data on $\mathbb{R}^2$.

We have the following relationships between these models. If n_2 and n_3 are very large relatively to n_1, then this model 2 approximates the known marginals model 1. If we allow G to have three atoms at $(0, \infty), (\infty, 0)$ and (∞, ∞) adding up to 1, then model 3 includes model 2 with n_1, n_2 and n_3 random.

Many other related models are of interest. For example, suppose that F has marginal distributions F_X and F_Y which are equal, but unknown: $F_X(x) = F_Y(x)$, $x \in \mathbb{R}$. All three models have generalizations to the case of k-variate F with $k \geq 3$, but of course there are many different generalizations of models 1 and 2 depending, for example, on which of the

$2^k - 2$ (multivariate) marginals are known in model 1. In this paper we restrict ourselves to the bivariate case. The analyzed estimators have natural generalizations to the k-variate case, and the k-variate analysis can be done by simply using k-variate analogues of the ingredients we use in the analysis for the case $k = 2$; for some of these, *see* Gill (1990).

Efficient estimation in models 1-3 has proved to be difficult. Bickel, Ritov, and Wellner (1989) have studied estimates of linear functionals of F for model 1 which are efficient under additional regularity conditions. Van der Laan (1992) proposed an implicit modified maximum likelihood estimator which depends on a bandwidth h_n (n is the number of observations) for model 3, which is proved to be efficient for $h_n \to 0$ slowly enough. The choice of the bandwidth in practice is still an open problem and practical tests remain to be done. Also here additional smoothness assumptions were needed. The estimator can be computed with the EM-algorithm. It is well known that this is an iterative algorithm which converges slowly. For information calculations for models 1 and 3, *see* Bickel, Klaassen, Ritov, and Wellner (1993) sections 6.3 and 6.6 and van der Laan (1992); for information calculations for model 2, *see* van der Vaart and Wellner (1990). It appeared that explicit information calculations are extremely difficult. Only in the special case of independence in model 3 we succeeded in obtaining an explicit expression for the information bound. These difficulties in constructing efficient estimators and that they only seem to work under additional regularity assumptions are a motivation for considering inefficient estimators.

Considerable work on construction of inefficient estimates in model 3 has been done: among them Burke (1988), Dabrowska (1988), (where also the Volterra estimator of P. J. Bickel is mentioned), Dabrowska (1989), Prentice and Cai (1992*a*), (1992*b*), Pruitt (1991*c*), van der Laan (1991). Many other explicit inefficient estimators have been proposed (see the reference list in Dabrowska, 1988). Bakker (1990) and Prentice and Cai (1992*b*) show that the Dabrowska and the Prentice-Cai estimators have a very good practical performance and beat the other proposed explicit estimators. Under independence they approximate surprisingly quickly the optimal variance. Pruitt's (1991*c*) estimator is an implicit estimator which solves an approximation to the self-consistency equations. This estimator uses kernel density estimators and consequently depends on a bandwidth. His estimator has a good practical performance (Pruitt, 1991*b*). Also here we have the problem of selecting the good bandwidth and additional smoothness assumptions on the model are necessary (van der Laan, 1991).

In this paper we focus on three inefficient estimators, but estimators (except the Volterra estimator) which have been shown to have a very good practical performance. We included the Volterra estimator because its representation is similarly derived as the Prentice-Cai representation, and the analysis of the Dabrowska and Volterra estimator gives the analysis of the Prentice-Cai estimator for free. The estimators are explicit and easy (quickly) to compute in contrary with efficient estimators which have to be computed with quite computer intensive algorithms (such as the EM-algorithm in van der Laan, 1992). The estimators are very smooth functions of the observations and therefore they are very robust: *i.e.* insensitive to small changes of the underlying distributions. Moreover for each rectangle $[0, \tau]$, where there is mass left for $[0, \tau]^c$, we do not need any assumptions on the model for obtaining the consistency, weak convergence and bootstrap results on $[0, \tau]$. Also the last two properties are certainly not shared with efficient estimators.

Our approach to estimation of F in these three models is as follows: we find representations of F as maps Φ from the distribution of the data, which can be estimated from the observed data, to F. The three particular representations which we study here are given by:

A. Dabrowska's (1988) representation.

B. The Volterra equation.

C. Prentice and Cai (1992*a*) representation.

We give a new proof of the Prentice and Cai (1992*a*) representation.

Notation and definition of $[0, \tau]$. – If we write $\leq, \geq, <, >$ then this should hold componentwise for both components: so if $x \in \mathbb{R}^2$ then $x \leq y \Leftrightarrow x_1 \leq y_1, x_2 \leq y_2$. We often will not use a special notation for the *bivariate* time-vector; if we do not mean a vector this will be made clear. If $F(t) = P(X \leq t)$ is a distribution function we will denote its survival function with $\overline{F}(t) = P(X > t)$. All functions we encounter are defined on a rectangle $[0, \tau] \subset \mathbb{R}^2_{\geq 0}$ where τ can be chosen arbitrarily large except that $\overline{F}(\tau-) > 0$ and $\overline{G}(\tau-) > 0$ is required. Finally, we define for a bivariate function $f : \mathbb{R}^2 \to \mathbb{R}$ $\|f\|_\infty \equiv \sup_{x \in [0,\tau]} \mid f(x) \mid$.

A. Dabrowska's representation

The representation, the estimator and L–measure were all introduced by Dabrowska (1988), but in a rather different way than we do here; we take the representation in terms of product-integrals as done in Gill, 1990 (*see* also Andersen, Borgan, Gill, Keiding, 1992). We define the following three hazard measures with their heuristic interpretation:

$$\begin{aligned}
\Lambda_{10}(du, v-) &= P(S \in [u, u+du) \mid (S,T) \geq (u,v)),\\
\Lambda_{01}(u-, dv) &= P(T \in [v, v+dv) \mid (S,T) \geq (u,v)),\\
\Lambda_{11}(du, dv) &= P(S \in [u, u+du),\ T \in [v, v+dv) \mid S \geq u,\ T \geq v).
\end{aligned}$$

Formally, we introduce a vector *hazard* function $\vec{\Lambda} : [0, \tau] \subset \mathbb{R}^2_{\geq 0} \to \mathbb{R}^3_{\geq 0}$ as follows: $\vec{\Lambda}(t) \equiv (\Lambda_{10}(t), \Lambda_{01}(t), \Lambda_{11}(t))$, $t \in \mathbb{R}^2_{\geq 0}$, where

$$\left.\begin{aligned}
\Lambda_{10}(t) &= \int_{[0,t_1]} \frac{1}{\overline{F}(u-, t_2)} F(du, t_2)\\
\Lambda_{01}(t) &= \int_{[0,t_2]} \frac{1}{\overline{F}(t_1, v-)} F(t_1, dv)\\
\Lambda_{11}(t) &= \int_{[0,t]} \frac{1}{\overline{F}(u-, v-)} F(du, dv).
\end{aligned}\right\} \tag{1}$$

One of the main advantages of model building in terms of hazards is that they are undisturbed by censoring and therefore in all three models we can get estimates of the integrated hazards by replacing them by their natural empirical counterparts.

For a bivariate distribution M (*i.e.* measure) $\prod_{(0,t]}(1 + dM)$ is the bivariate product integral over the rectangle $[0, t]$ (*see* Gill, Johansen, 1990, or our section 3.2). It is just like the univariate product integral defined as the limit of finite products over finite rectangular partitions of $[0, t]$. Now, the following representation can be proved:

$$\begin{aligned}
\overline{F}(t) &= \prod_{[0,t_1]} (1 - \Lambda_{10}(du, 0)) \prod_{[0,t_2]} (1 - \Lambda_{01}(0, dv)) \prod_{[0,t]} (1 - L(du, dv))\\
&\equiv \Gamma_1(\Lambda_{10}(\cdot, 0), \Lambda_{01}(0, \cdot), \Gamma_2(L)),
\end{aligned} \tag{2}$$

L is defined by

$$\begin{aligned}
L(t) &\equiv \int_{[0,t]} \frac{\Lambda_{10}(du, v-)\Lambda_{01}(u-, dv) - \Lambda_{11}(du, dv)}{(1 - \Lambda_{10}(\Delta u, v-))(1 - \Lambda_{01}(u-, \Delta v))}\\
&\equiv \Gamma_3(\Lambda_{10}, \Lambda_{01}, \Lambda_{11}),
\end{aligned} \tag{3}$$

and Γ_2 represents the bivariate product-integral mapping. With $\Lambda_{10}(\Delta u, v-) = \Lambda_{10}(u, v-) - \Lambda_{10}(u-, v-)$ we denote the jump of

$s \to \Lambda_{10}(s, v-)$ at u. Assume that for each v $(u \mapsto F(u,v)) \ll \mu_1$ and for each u $(v \mapsto F(u,v)) \ll \mu_2$ for certain (signed) measures μ_1 and μ_2. We define

$$\int F(du, v)G(u, dv) \equiv \int \frac{dF}{d\mu_1}(u,v)\frac{dG}{d\mu_2}(u,v)d\mu_1(u)d\mu_2(v).$$

These assumptions are easily verified for the hazard measures by choosing $\mu_1 = \overline{F}_1$ and $\mu_2 = \overline{F}_2$, the marginals of $\overline{F}$. We will see in section 4 that the empirical counterpart $\vec{\Lambda}_n$ of $\vec{\Lambda}$ is obtained by replacing in the representation of $\vec{\Lambda}$ $\overline{F}$ by an empirical survival function. Therefore, the assumptions are verified in exactly the same way by choosing μ_1 and μ_2 the marginals of this empirical survival function. We will do this in the proof of the final theorem in section 5.

Note that by (2) and (3) this gives a map Γ such that

$$\begin{aligned}\overline{F} &= \Gamma(\vec{\Lambda}) \equiv \Gamma_1(\Lambda_{10}(\cdot, 0), \Lambda_{01}(0, \cdot), \Gamma_2(L)) \\ &= \Gamma_1(\Lambda_{10}(\cdot, 0), \Lambda_{01}(0, \cdot), \Gamma_2\Gamma_3(\vec{\Lambda})).\end{aligned} \tag{4}$$

This representation can easily be heuristically verified, just as the one-dimensional Kaplan-Meier product-integral.

For models 1-3 there are natural empirical estimators of $\vec{\Lambda}$ which generalizes the famous *Nelson-Aalen* estimator from the one dimensional case; we will do this explicitly and in detail in section 4.

If we denote the estimate of $\vec{\Lambda}$ with $\vec{\Lambda}_n$, then the estimate of $\overline{F}$ based on Dabrowska's representation is simply

$$\overline{F}_n = \Gamma(\vec{\Lambda}_n).$$

This estimator was studied in the case of model 3 by Dabrowska (1988, 1989). Gill (1990) generalized the representation to dimension $k \geq 2$ and analyzed the estimator by applying the functional delta-method.

B. The Volterra equation

This equation is derived by extending the following argument for $k = 1$: let

$$\Lambda(t) \equiv \int_{[0,t]} \frac{1}{\overline{F}_-} dF \qquad \text{for} \quad t \geq 0$$

be the cumulative hazard function corresponding to F. Then

$$F(t) = \int_{[0,t]} \overline{F}_- d\Lambda$$

and consequently

$$\overline{F}(t) = 1 - \int_{[0,t]} \overline{F}_- d\Lambda.$$

For a given function Λ, this is a *homogeneous Volterra equation* for $\overline{F}$, where the solution is given by the *Peano series* (a special Neumann series) $\sum_{i=1}^{\infty} A^i(1)$, where $A(\overline{F}) = \int_{[0,\cdot]} \overline{F}_- d\Lambda$. In this case, $k = 1$, this is solved explicitly by the product-integral of Λ:

$$\overline{F}(t) = \prod_{[0,t]} (1 - d\Lambda(s)).$$

For theory about the univariate product-integral and in particular the equivalence between the univariate Peano series and the univariate product-integral we refer to Gill and Johansen (1990).

For $k = 2$ the argument generalizes as follows. For F on $[0, \infty)^2$, as defined above

$$\Lambda_{11}(t) \equiv \int_{[0,t]} \frac{1}{\overline{F}_-} dF,$$

where $\overline{F}(x) \equiv P(X > x)$. This implies that

$$F(t) = \int_{[0,t]} \overline{F}_- d\Lambda_{11}. \tag{5}$$

It remains only to relate F to $\overline{F}$ and the marginal distributions: let F_1 and F_2 denote the marginal distributions of F. Then since

$$F_1(t_1) + F_2(t_2) - F(t) + \overline{F}(t) = 1,$$

(5) yields

$$\begin{aligned} \overline{F}(t) &= 1 - F_1(t_1) - F_2(t_2) + \int_{[0,t]} \overline{F}_- d\Lambda_{11} \\ &\equiv \Psi(t) + \int_{[0,t]} \overline{F}_- d\Lambda_{11}, \end{aligned} \tag{6}$$

where $\Psi(t) = 1 - F_1(t_1) - F_2(t_2)$ involves only the marginal distributions F_1 and F_2. Regarded as an equation for $\overline{F}$ given fixed functions Ψ and Λ, (6) is an *inhomogeneous Volterra equation* with a unique solution $\Phi_1(\Psi, \Lambda_{11})$ [Gill and Johansen (1990), Kantorovich and Akilov (1982), p. 396]. This can be seen as follows. Represent the equation as $(I - A_\Lambda)(\overline{F}) = \Psi$ where $A_\Lambda(\overline{F}) = \int_{[0,t]} \overline{F}_- d\Lambda$. It is easy to check that this structure takes care that

$$\|A_\Lambda^k(\Psi)\|_\infty \leq \frac{\|\Psi\|_\infty \|\Lambda\|_\infty^k}{k!},$$

where one has to notice that by definition of τ the supnorm (over $[0, \tau]$) $\|\Lambda\|_\infty$ is bounded.

Consequently, $\sum_{k=0}^{\infty} A_\Lambda^k$ is a bounded operator:

$$\left\| \sum_{k=0}^{\infty} A_\Lambda^k(h) \right\|_\infty \leq \|h\|_\infty \exp(\|\Lambda\|_\infty).$$

This proves that $\overline{F}$ is given by the Neumann series of $A_{\Lambda_{11}}$:

$$\overline{F} = \sum_{k=0}^{\infty} A_{\Lambda_{11}}^k(\Psi). \tag{7}$$

Because $A_{\Lambda_{11}}$ depends only on Λ_{11} and

$$\begin{aligned} \Psi(t) &= 1 - \prod_{(0,t_1]} (1 - \Lambda_{10}(ds_1, 0)) - \prod_{(0,t_2]} (1 - \Lambda_{10}(0, ds_2)) \\ &\equiv \Phi_2(\Lambda_{10}(\cdot, 0), \Lambda_{01}(0, \cdot)) \end{aligned}$$

(7) defines a map

$$\overline{F} = \Phi(\vec{\Lambda}) \equiv \Phi_1(\Psi, \Lambda_{11}) = \Phi_1(\Phi_2(\Lambda_{10}(\cdot, 0), \Lambda_{01}(0, \cdot)), \Lambda_{11}). \tag{8}$$

It is not clear from (7) that F does also continuously depend on Λ_{11}, but we will prove in section 3, as in Gill and Johansen (1990), that the bivariate Peano series, and thereby Φ_1, satisfies the characterization of *weakly continuously compactly differentiable* at (Ψ, Λ).

Finally, it should be noticed that because of the exponential convergence of the terms $A_{\Lambda_{11}}^k$ to zero, (7) provides us also with an exponentially fast algorithm for finding a solution of the Volterra equation for known (Ψ, Λ_{11}). Now, the Volterra estimator of $\overline{F}$ is given by:

$$\overline{F}_n \equiv \Phi(\vec{\Lambda}_n).$$

C. Prentice-Cai representation

We give a new proof of the Prentice-Cai representation (*see* also Prentice and Cai, 1992*a*). For still another proof, *see* Wellner (1993). For this we need the following differentiability rules for $U : \mathbb{R} \to \mathbb{R}$ and $V : \mathbb{R} \to \mathbb{R}$:

$$d(UV) = U_- dV + dUV$$
$$d\left(\frac{1}{U}\right) = \frac{dU}{UU_-}.$$

If we apply these one dimensional rules to the sections $u \to F(u,v)$ and $v \to F(u,v)$ of a bivariate function F, then we denote these with d_1 and d_2, respectively. We apply these two one dimensional rules to each of the two variables of $R \equiv \overline{F}/\overline{F}_1\overline{F}_2$ in turn in order to express $dR = d_{12}R = d_1(d_2(R))$ as follows:

$$dR = R_- d\widetilde{L}$$

for certain measure $\widetilde{L}$. Define the well known univariate hazards $\Lambda_1(ds_1) \equiv \Lambda_{10}(ds_1, 0)$, $\Lambda_2(ds_2) \equiv \Lambda_{01}(0, ds_2)$. The reader can easily verify the following result [when applying the product rule to $\overline{F}/\overline{F}_i$ we give the left continuous version to $\overline{F}$ instead of one of the $\overline{F}_i$, $i = 1, 2$, and we denote $F_{(-1)}(s_1, s_2) \equiv F(s_1-, s_2)$, $F_{(-2)}(s_1, s_2) \equiv F(s_1, s_2-)$):

$$\begin{aligned}
dR &= d_{12}R \\
&= \frac{d_{12}\overline{F}}{\overline{F}_1\overline{F}_2} - \frac{d_2\overline{F}_2 d_1\overline{F}_{(-2)}}{\overline{F}_2\overline{F}_{2-}\overline{F}_1} - \frac{d_1\overline{F}_1 d_2\overline{F}_{(-1)}}{\overline{F}_1\overline{F}_{1-}\overline{F}_2} + \frac{d_1\overline{F}_1 d_2\overline{F}_2\overline{F}_-}{\overline{F}_1\overline{F}_{1-}\overline{F}_2\overline{F}_{2-}} \\
&= R_- \frac{\overline{F}_{1-}\overline{F}_{2-}}{\overline{F}_1\overline{F}_2}\left(\frac{d_{12}\overline{F}}{\overline{F}_-} - \frac{d_2\overline{F}_2}{\overline{F}_{2-}}\frac{d_1\overline{F}_{(-2)}}{\overline{F}_-}\right. \\
&\qquad\qquad \left. - \frac{d_1\overline{F}_1}{\overline{F}_{1-}}\frac{d_2\overline{F}_{(-1)}}{\overline{F}_-} + \frac{d_1\overline{F}_1}{\overline{F}_{1-}}\frac{d_2\overline{F}_2}{\overline{F}_{2-}}\right) \\
&= R_-\left(\frac{\left\{\begin{array}{l}\Lambda_{11}(ds) - \Lambda_2(ds_2)\Lambda_{10}(ds_1, s_2-) \\ \quad -\Lambda_1(ds_1)\Lambda_{01}(s_1-, ds_2) + \Lambda_1(ds_1)\Lambda_2(ds_2)\end{array}\right\}}{(1-\Lambda_1(\Delta s_1))(1-\Lambda_2(\Delta s_2))}\right) \\
&\equiv R_- d\widetilde{L}.
\end{aligned}$$

At the third equality notice that $1/(1-\Lambda_1(\Delta s_1))(1-\Lambda_2(\Delta s_2)) = \overline{F}_{1-}\overline{F}_{2-}/\overline{F}_1\overline{F}_2$. Integrating the left and right-hand side over the rectangle $(0, t]$ provides us with:

$$R(t) = 1 + \int_{(0,t]} R(s-)\widetilde{L}(ds). \tag{9}$$

Here one has to notice that $R(t_1, 0) = R(0, t_2) = 1$ for all t. This is a homogeneous Volterra equation with a unique solution given by the Peano series of $\widetilde{L}$ which we will write out below.

By definition of R and the well known product integral representation of the univariate $\overline{F}_i$, $i = 1, 2$, this provides us with the following representation for the bivariate survival function:

$$\begin{aligned}\overline{F}(t) &= \prod_{[0,t_1]} (1 - \Lambda_{10}(du, 0)) \prod_{[0,t_2]} (1 - \Lambda_{01}(0, dv)) R(t) \\ &\equiv \Theta_1(\Lambda_{10}(\cdot, 0), \Lambda_{01}(0, \cdot), R)\end{aligned} \tag{10}$$

where R is the unique solution of (9), just the Neumann series $\sum_{k=0}^{\infty} A_{\widetilde{L}}^k(1)$ as given in (7), given by the Peano series:

$$\begin{aligned}R &= 1 + \sum_{n=1}^{\infty} \int \cdots \int_{0 \le u^1 < u^2 < \ldots < u^n \le t} \prod_{j=1}^{n} \widetilde{L}(du^j) \\ &\equiv \Theta_2(\widetilde{L}).\end{aligned}$$

Define

$$\beta(s) \equiv (1 - \Lambda_{10}(\triangle s_1, 0))(1 - \Lambda_{01}(0, \triangle s_2)).$$

Above, we derived the following representation of $\widetilde{L}$ in terms of the hazards $\vec{\Lambda}$

$$\begin{aligned}\widetilde{L}(t) &= \int_{(0,t]} \frac{1}{\beta(s)} \{\Lambda_{11}(ds) - \Lambda_{10}(ds_1, s_2-)\Lambda_{01}(0, ds_2) \\ &\quad - \Lambda_{10}(ds_1, 0)\Lambda_{01}(s_1-, ds_2) + \Lambda_{10}(ds_1, 0)\Lambda_{01}(0, ds_2)\} \\ &\equiv \Theta_3(\Lambda_{10}(\cdot, 0), \Lambda_{01}(0, \cdot), \vec{\Lambda}).\end{aligned}$$

Of course, we can write $\Theta_3(\vec{\Lambda})$, but because of model 1 (marginals known) we want to distinguish between hazards which are known in model 1 and hazards which are not known in model 1.

Note that this gives a map

$$\begin{aligned}\overline{F} &= \Theta(\vec{\Lambda}) = \Theta_1(\Lambda_{10}(\cdot, 0), \Lambda_{01}(0, \cdot), \Theta_2(\widetilde{L})) \\ &= \Theta_1(\Lambda_{10}(\cdot, 0), \Lambda_{01}(0, \cdot), \Theta_2\Theta_3(\Lambda_{10}(\cdot, 0), \Lambda_{01}(0, \cdot), \vec{\Lambda})).\end{aligned} \tag{11}$$

Again, the estimate of $\overline{F}$ based on the Prentice and Cai representation is simply

$$\overline{F}_n = \Theta(\vec{\Lambda}_n).$$

Prentice and Cai (1992a) motivated this representation through a connection between $\tilde{L}$ and the covariance of univariate counting process martingales. Moreover, they proved almost sure consistency of the resulting estimator for model 3 via continuity of Θ.

Remark. – Firstly, the Volterra estimator is based on the idea to express $d\overline{F}$ in $\overline{F}_- dA$ for a certain measure A which makes $\overline{F}$ a solution of an inhomogeneous Volterra equation, while in Prentice-Cai's representation we do the same with $d(\overline{F}/\overline{F}_1\overline{F}_2)$ which leads in this case to a homogeneous Volterra equation. The Volterra uses an estimate of only one bivariate hazard Λ_{11}, while the Dabrowska and the Prentice-Cai representations involve other functions L and $\tilde{L}$ which describe the covariance structure. Furthermore, notice the similarity in the structure of L and $\tilde{L}$.

Elaborating on the functional delta-method and results

Our approach to studying the estimators, which we will denote by $\overline{F}_n^V, \overline{F}_n^D, \overline{F}_n^{PC}$ will be to study the maps Φ, Γ and Θ which define them (analytically). In section 2 and 3 we show that these satisfy the characterization of weakly continuous Hadamard differentiability with respect to the supremum norm-metric for the sequences which can occur in practice. In section 4 we represent $\vec{\Lambda}$ as maps from the distribution function of the data to $\vec{\Lambda}$ itself for all three models. By applying the functional delta-method to these representations we prove the required weak convergence and validity of the bootstrap for $\vec{\Lambda}_n$. Now, application of the functional delta-method to the mappings Φ, Θ and Γ provides us with with consistency, weak convergence, and asymptotic validity of the bootstrap for $\overline{F}_n^V, \overline{F}_n^D, \overline{F}_n^{PC}$. These results are summarized in Theorem 5.1.

This method of analyzing provides us with optimal results for the estimators in the sense that we essentially do not need any conditions. The only improvements can be made by extending these results to the whole plane and by investigating the rate at which the normalized estimators converge to its linearization in terms of the empirical processes we plug in.

This analysis of an estimator, say $\Phi(\vec{\Lambda}_n)$, separates the analysis in a purely analytical (differentiability of Φ) and a purely probabilistic [weak convergence of $Z_n = \sqrt{n}(\vec{\Lambda}_n - \vec{\Lambda})$] part. After establishing purely analytical properties of components of Φ one can deduce several results for different sampling methods, models (e.g. our models 1-3) or statistics without repeating the analysis. The supnorm might be considered as a quite naive choice in order to get an optimal weak convergence result, but the supnorm is easy to use, to interpret, and has a natural generalization to higher dimensions.

After establishing the differentiability of the functionals which appear in the representation Γ and Φ we got the differentiability of Prentice representation for free: by the chain-rule a differentiability result for a functional can be used for establishing differentiability for any composition of several mappings (like the three representations) which involves this functional. Because of this property of the analysis, other proposed explicit estimators can immediately be put in our framework.

2. LEMMAS

In this section we give the basic techniques as lemmas for obtaining the (by the functional delta-method) required differentiability result for the representations. Moreover, we will give two illustrations which show how these techniques easily lead to differentiability of relevant functionals.

Definitions and notation. – All functions we encounter are considered as elements of the space of bivariate cadlag functions on $[0, \tau]$ (Neuhaus, 1972) endowed with the *supnorm*.

We will denote this space with $(D[0, \tau], \|\cdot\|_\infty)$. If F_n converges in supnorm to F, then we will denote this with $F_n \to F$ (this does not lead to problems because we only use supnorm convergence). The *variation norm* of a function F is defined as usual as the supremum over all partitions of rectangles of the sum of the absolute values of generalized differences over the partition elements (*i.e.* rectangles) of F, and it will be denoted with $\|F\|_v$. The *uniform sectional variation norm* of a bivariate function F is defined as the supremum of the variations of the sections $u \to F(u,v)$, $v \to F(u,v)$, $(u,v) \in [0,\tau]$, and the variation of F itself. It will be denoted with $\|F\|_v^*$. If a cadlag function is of bounded variation, then it generates a signed measure. If for a F we write $F(du,v), F(u,dv), F(du,dv)$, we mean the one dimensional measures generated by the sections $u \mapsto F(u,v), v \mapsto F(u,v)$ and the two dimensional measure generated by $(u,v) \mapsto F(u,v)$, respectively, and it will be automatically assumed that these sections and the function itself are of bounded variation. Finally, integrals like $\int F(du,v)G(u,dv)$ are defined as

$\int \frac{dF}{d\mu_1}(u,v)\frac{dG}{d\mu_2}(u,v)d\mu_1(u)d\mu_2(v)$, assuming that for each v the mapping $(u \to F(u,v)) \ll \mu_1$ and for each u the mapping $(v \to F(u,v)) \ll \mu_2$ for certain (signed) measures μ_1 and μ_2. This assumption is satisfied by a simple choice of measures μ_1, μ_2 in all our applications as will be shown in the proof of the final theorem in section 5 (*see* also illustration 2 at the end of this section).

LEMMA 2.1 (Telescoping). – *Let* a_i, $i = 1, \ldots, k$, b_i, $i = 1, \ldots, k$ *be real numbers.*

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

This can be easily verified and it holds also for matrices (*see* Gill and Johansen, 1990). It is a very useful lemma for proving that differences of two products converge to zero and that is what we often have to do in the differentiability proofs.

In our applications we want to be able to define integrals $\int F dH$ when H is of *unbounded* variation. This can be done by applying *integration by parts* so that H appears as function.

LEMMA 2.2 (Integration by Parts).

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

Notice that with these formulas we can also define these integrals for H of unbounded variation. Also, we can bound $\int F dH$ by $16\|H\|_\infty \|F\|_v^*$.

Notice that if F is zero at the edges of $[0,\tau]$, then only the first term at the right hand-side is non-zero.

Proof. – We refer to (Gill, 1990) for the general $\mathbb{R}^k$ case. It works as follows. For the first integral, substitute

$$F(u,v) = \int_{(0,u]\times(0,v]} F(du',dv') + \int_{(0,u]} F(du',0) + \int_{(0,v]} F(0,dv') + F(0,0)$$

and for the second integral substitute

$$F(du,v) = \int_{(0,v]} F(du,dv') + F(du,0)$$

and apply Fubini's theorem. □

This integration by parts formula is a special case of the following general integration by parts formula by letting $A_s = (0,s]$, but it is one which takes account of mass given to the edge of the rectangle $[0,\tau]$: Let $A_s \subset \mathbb{R}^2$ be a measurable set indexed by $s \in \mathbb{R}^2$. Then we have:

$$\int I_{A_s}(u)F(u)dH(u) = \int I_{A_s}(u) \int_{(-\infty,u]} F(du')dH(u)$$
$$= \int \left(\int I_{A_s}(u) I_{(-\infty,u]}(u')dH(u) \right) F(du').$$

Assume that F_1, F_2 are of bounded variation and that $F_i(s_1,s_2)$ or $F_i(s_1-,s_2)$ or $F_i(s_1,s_2-)$ or $F_i(s_1-,s_2-)$ is cadlag, $i = 1,2$. Then F_1 and F_2 generate signed measures. We have

$$\int F_1(u)F_2(u)dH(u) = \int F_1(u)d\left(\int_0^u F_2(v)dH(v) \right). \tag{12}$$

So by two times applying lemma 2.2 to

$$\int F_1(u)F_2(u)dH(u) = \int F_1(u)d\left(\int_0^u F_2(v)dH(v) \right)$$

we can do integration by parts so that H appears as function and F_1, F_2 as measures.

The following lemma is trivially checked, but useful.

LEMMA 2.3 (d-$\triangle$ interchange). – *We have:*

$$\int\int F(\triangle s, \triangle t)G(ds,dt) = \int\int F(ds,dt)G(\triangle s,\triangle t)$$
$$\int\int F(ds,t)G(\triangle s,dt) = \int\int F(\triangle s,t)G(ds,dt).$$

Recall the denominator in the mappings L and $\widetilde{L}$ in Γ and Θ, which are of the form $1/\{(1-a)(1-b)\}$, where a, b are are only nice functions in one coordinate, and therefore certainly do not generate a measure. Therefore it is not clear how we can integrate w.r.t. this denominator. The following lemma will take care of this problem.

LEMMA 2.4 (Denominator splitting). – *Let a_1, a_2 be real numbers. Then the following holds:*

$$\frac{1}{(1-a_1)(1-a_2)} = \frac{1}{1-a_1} + \frac{a_2}{(1-a_1)(1-a_2)}$$

or

$$\frac{1}{(1-a_1)(1-a_2)} = 1 + \frac{a_1}{1-a_1} + \frac{a_2}{1-a_2} + \frac{a_1 a_2}{(1-a_1)(1-a_2)}.$$

In general, we have:

$$\frac{1}{\prod_{i=1}^{n}(1-a_i)} = 1 + \sum_i \frac{a_i}{1-a_i} + \sum_{i,j,i\neq j} \frac{a_i a_j}{(1-a_i)(1-a_j)} + \ldots + \frac{\prod_{i=1}^{n} a_i}{\prod_{i=1}^{n}(1-a_i)}.$$

This follows from the identity $\dfrac{1}{1-a_i} = 1 + \dfrac{a_i}{1-a_i}$.

Now, we are able to define the following terms with integration by parts as follows:

COROLLARY 2.1. – *Define $\beta(u,v) \equiv (1-\Lambda_{10}(\Delta u, v))(1-\Lambda_{01}(u,\Delta v))$. We have:*

$$\int\!\!\int \frac{H(du,v)\Lambda(u,dv)}{\beta(u,v)} = \int\!\!\int H(du,v)\frac{\Lambda(u,dv)}{1-\Lambda_{01}(u,\Delta v)} + \int\!\!\int \frac{H(\Delta u,v)\Lambda_{10}(du,v)\Lambda(u,dv)}{\beta(u,v)}$$

$$\begin{aligned}\int\!\!\int \frac{H(du,dv)}{\beta(u,v)} &= \int\!\!\int H(du,dv) \\ &+ \int\!\!\int \frac{H(\Delta u,\Delta v)\Lambda_{10}(du,v)\Lambda_{01}(u,dv)}{\beta(u,v)} \\ &+ \int\!\!\int H(du,\Delta v)\frac{\Lambda_{01}(u,dv)}{1-\Lambda_{01}(u,\Delta v)} \\ &+ \int\!\!\int H(\Delta u,dv)\frac{\Lambda_{10}(du,v)}{1-\Lambda_{10}(\Delta u,v)}.\end{aligned}$$

H plays the role of a function of unbounded variation (Brownian bridge) and $\Lambda, \Lambda_{10}, \Lambda_{01}$ are cadlag functions of bounded variation. Notice that all terms at the right-hand side of the equalities where H appears as measure are of the form $\int F dH$ where F generates a finite measure. Therefore, for all these terms we can apply the integration by parts formulas of Lemma 2.2 in order to make H appear as function.

Again, this corollary is simple to prove by applying denominator splitting and d–Δ-interchange. In the differentiability proof of the L-mapping we have to be able to bound the terms above in the supnorm of H. It is now clear that this can be done with the integration by parts formulas. We will see that this is the whole story of the differentiability proofs: we use denominator-splitting and d–Δ-interchange in order to get an integral $\int F dH$, where F generates a measure (so it must be of bounded variation) and then we apply our integration by parts formulas in order to bound these terms in the supnorm of H.

We did not deal, yet, with an integral of the form $\int H dF_n$, $\|F_n\|_\infty \to 0$, $\|H\|_v = \infty$, which we want to show to converge to zero. Here, one cannot do integration by parts in order to bound this in the supnorm of F_n, because H is not of finite variation. The next ingredient takes care of this, the so called Helly-Bray technique:

LEMMA 2.5 (Helly-Bray lemma). – *If $H \in (D[0,\tau], \|\cdot\|_\infty)$ is of unbounded variation, then we can approximate H with a sequence H_m where $\|H_m\|_v \leq M(m) < \infty$ and $\|H - H_m\|_\infty \to 0$. This gives us the following bound:*

$$\left\| \int H dF \right\|_\infty \leq \|H - H_m\|_\infty \|F\|_v + 16\|F\|_\infty M(m).$$

For H_m one can (e.g.) take the step function equal to H on a grid π^m. We did the substitution $H = (H - H_m) + H_m$, integration by parts and bounding terms like $\int F dH_m$ by $\|F\|_\infty \|H_m\|_v$. The bound in Lemma 2.5 is useful because it proves that integrals of the form $\int H dF_n$ converge to zero when $\|F_n\|_\infty \to 0$, even if $\|H\|_v = \infty$, provided that $\|F_n\|_v < \infty$ (just let $m \to \infty$ slowly enough).

ILLUSTRATION 1. – We will illustrate how these lemmas easily provide us with the characterization of compact differentiability of $\Phi : (F, G) \to \int F dG$ at a point (F, G) with F and G cadlag functions of bounded variation for sequences F_n, G_n of uniformly (in n) bounded variation: if $Y_n \equiv \sqrt{n}(F_n - F) \rightarrow_u Y$, $Z_n \equiv \sqrt{n}(G_n - G) \rightarrow_u Z$, then

$$\sqrt{n}(\Phi(F_n, G_n) - \Phi(F, G)) - d\Phi(F, G)(Y, Z) \rightarrow_u 0$$

for a certain continuous linear mapping $d\Phi(F, G) : (D[0, \tau])^2 \rightarrow \mathbb{R}$. We have by telescoping:

$$\sqrt{n}(\Phi(F_n, G_n) - \Phi(F, G)) = \int Y_n dG + \int F_n dZ_n.$$

So if we subtract from this its supposed limit

$$d\Phi(F, G)(Y, Z) = \int Y dG + \int F dZ,$$

then we get (again) by telescoping:

$$\int (Y_n - Y) dG + \int (F_n - F) dZ + \int F_n d(Z_n - Z)$$

where the last two integrals are defined by integration by parts. The first integral can immediately be bounded by $\|Y_n - Y\|_\infty \|G\|_v \rightarrow 0$. The second integral converges to zero by the Helly-Bray lemma. For the third integral we can do integration by parts with respect to F_n and thereby bound this term by $c\|Z_n - Z\|_\infty \|F_n\|_v \rightarrow 0$.

ILLUSTRATION 2. – We will give an illustration of how these lemmas are used to prove convergence to zero of quite complicated terms which we will encounter in our analysis of the Dabrowska's estimator. Consider the term $\int (1/\beta_n(u, v) - 1/\beta(u, v)) H(du, dv)$, where H is of unbounded variation, and $\beta_n(u, v), \beta(u, v)$ is the denominator of L as defined above corresponding to $(\Lambda_{10}^n, \Lambda_{01}^n), (\Lambda_{10}, \Lambda_{01})$ and $(\Lambda_{10}^n, \Lambda_{01}^n)$ converges in supnorm to $(\Lambda_{10}, \Lambda_{01})$. We will show that this term converges to zero if $\Lambda_{10}, \Lambda_{01}, \Lambda_{10}^n, \Lambda_{01}^n$ have the following four properties :

1) $\beta > \delta > 0$ on $[0, \tau]$ for certain $\delta > 0$.

2) There exists a sequence of uniformly in n finite (signed) measures μ_{2n} so that $\Lambda_{10}^n(u, dv) \ll \mu_{2n}(dv)$ for all u. Similarly for $\Lambda_{10}, \Lambda_{01}, \Lambda_{01}^n$.

3) There exists a sequence of uniformly in n finite (signed) measures μ_{1n} so that $\Lambda_{10}^n(du, v) \ll \mu_{2n}(du)$ for all u. Similarly for $\Lambda_{10}, \Lambda_{01}, \Lambda_{01}^n$.

4) $\|\Lambda_{10}^n(du, v)/\mu_{1n}(du)\|_\infty < M$ and $\|\Lambda_{10}^n(u, dv)/\mu_{2n}(dv)\|_\infty < M$ for certain $M < \infty$ (uniform boundedness of the Radon-Nykodym derivatives). Similarly for $\Lambda_{10}, \Lambda_{01}, \Lambda_{01}^n$.

In our applications the assumptions 2-4 are easily verified by a simple choice of $\mu_{1n}, \mu_1, \mu_{2n}, \mu_2$ and by definition of $[0, \tau]$ assumption 1 will hold trivially. This will be done in the proof of the final theorem in section 5.

This term involves all the above techniques. Apply the *denominator splitting* lemma to rewrite $1/\beta_n(u, v) - 1/\beta(u, v)$. This gives

$$\begin{aligned}
&\frac{1}{\beta_n(u,v)} - \frac{1}{\beta(u,v)}\\
&\quad = \left(\frac{\Lambda_{10}^n(\Delta u, v)}{1 - \Lambda_{10}^n(\Delta u, v)} - \frac{\Lambda_{10}(\Delta u, v)}{1 - \Lambda_{10}(\Delta u, v)}\right)\\
&\quad + \left(\frac{\Lambda_{01}^n(u, \Delta v)}{1 - \Lambda_{01}^n(u, \Delta v)} - \frac{\Lambda_{01}(u, \Delta v)}{1 - \Lambda_{01}(u, \Delta v)}\right)\\
&\quad + \left(\frac{\Lambda_{10}^n(\Delta u, v)\Lambda_{01}^n(u, \Delta v)}{\beta_n(u,v)} - \frac{\Lambda_{10}(\Delta u, v)\Lambda_{01}(u, \Delta v)}{\beta(u,v)}\right).
\end{aligned}$$

Then the integral is the sum of three integrals which we will denote with A, B and C respectively. The first term A is given by:

$$\begin{aligned}
&\int \left(\frac{\Lambda_{10}^n(\Delta u, v)}{1 - \Lambda_{10}^n(\Delta u, v)} - \frac{\Lambda_{10}(\Delta u, v)}{1 - \Lambda_{10}(\Delta u, v)}\right) H(du, dv)\\
&\quad = \int \left(\frac{\Lambda_{10}^n(\Delta u, v)}{1 - \Lambda_{10}^n(\Delta u, v)} - \frac{\Lambda_{10}(\Delta u, v)}{1 - \Lambda_{10}(\Delta u, v)}\right) (H - H_m)(du, dv)\\
&\quad + \int \left(\frac{\Lambda_{10}^n(\Delta u, v)}{1 - \Lambda_{10}^n(\Delta u, v)} - \frac{\Lambda_{10}(\Delta u, v)}{1 - \Lambda_{10}(\Delta u, v)}\right) H_m(du, dv)\\
&\quad = \int \left(\frac{\Lambda_{10}^n(du, v)}{1 - \Lambda_{10}^n(\Delta u, v)} - \frac{\Lambda_{10}(du, v)}{1 - \Lambda_{10}(\Delta u, v)}\right) (H - H_m)(\Delta u, dv)\\
&\quad + \int \left(\frac{\Lambda_{10}^n(\Delta u, v)}{1 - \Lambda_{10}^n(\Delta u, v)} - \frac{\Lambda_{10}(\Delta u, v)}{1 - \Lambda_{10}(\Delta u, v)}\right) H_m(du, dv).
\end{aligned}$$

We did the substitution $H = H - H_m + H_m$ (Helly-Bray) and applied d–Δ-interchange. Consider the first term, say $A1$.

A1. Here, we can apply the second integration by parts formula of Lemma 2.2. Then one of the terms we get is the following:

$$\begin{aligned}
&\int (H - H_m)(\Delta u, [v, \tau_2]) \frac{1}{(1 - \Lambda_{10}^n(\Delta u, v))^2} \Lambda_{10}^n(du, v)\Lambda_{10}^n(\Delta u, dv)\\
&\quad \leq \frac{1}{(\inf_{(u,v)\in[0,\tau]} | 1 - \Lambda_{10}^n(\Delta u, v) |)^2} \|H - H_m\|_\infty\\
&\quad \times \int | \Lambda_{10}^n(du, v)\Lambda_{10}^n(\Delta u, dv) |\\
&\quad \leq C\|H - H_m\|_\infty \int | \Lambda_{10}^n(du, v)\Lambda_{10}^n(\Delta u, dv) |,
\end{aligned}$$

where we used assumption $\beta_n > \delta > 0$ on $[0, \tau]$ for certain $\delta > 0$ in the last line, which follows from assumption 1 and the uniform convergence of β_n to β. The other terms one gets after applying integration by parts are dealt in the same way. By assumption 2-4 we have:

$$\begin{aligned}
&\int \mid \Lambda_{10}^n(du, v)\Lambda_{10}^n(\Delta u, dv) \mid \\
&= \int \left| \frac{\Lambda_{10}^n(du, v)}{\mu_{1n}(du)} \frac{\Lambda_{10}^n(u, dv)}{\mu_{2n}(dv)} \right| \mid \mu_{1n}(du)\mu_{2n}(dv) \mid \\
&\leq M^2 \|\mu_{1n}\|_v \|\mu_{2n}\|_v \leq M',
\end{aligned}$$

for certain $M' < \infty$. So if Λ_{10}^n satisfies assumptions 1-4, then $C\|H - H_m\|_\infty \int \mid \Lambda_{10}^n(du, v)\Lambda_{10}^n(\Delta u, dv) \mid \to 0$ for $m \to \infty$. The other terms are dealt similarly using the assumptions 1-4 for Λ_{10} and Λ_{10}^n.

A2. The second term can be bounded by the supnorm of $\Lambda_{10}^n(\Delta u, v)/(1 - \Lambda_{10}^n(\Delta u, v)) - \Lambda_{10}(\Delta u, v)/(1 - \Lambda_{10}(\Delta u, v))$ (which converges to zero) times the variation norm of H_m.

So if we let $m = m(n) \to \infty$ slowly enough for $n \to \infty$, then both terms $A1$ and $A2$ converge to zero.

The second term B is dealt similarly. Now, we will deal with the third term C. Firstly, by *telescoping* we can rewrite:

$$\begin{aligned}
&\frac{\Lambda_{10}^n(\Delta u, v)\Lambda_{01}^n(u, \Delta v)}{\beta_n(u, v)} - \frac{\Lambda_{10}(\Delta u, v)\Lambda_{01}(u, \Delta v)}{\beta(u, v)} \\
&= \left(\frac{1}{\beta_n(u, v)} - \frac{1}{\beta(u, v)} \right) \Lambda_{10}(\Delta u, v)\Lambda_{01}(u, \Delta v) \\
&\quad + \frac{1}{\beta_n(u, v)} (\Lambda_{10}^n(\Delta u, v) - \Lambda_{10}(\Delta u, v))\Lambda_{01}(u, \Delta v) \\
&\quad + \frac{1}{\beta_n(u, v)} \Lambda_{10}^n(\Delta u, v)(\Lambda_{01}^n(u, \Delta v) - \Lambda_{01}(u, \Delta v)).
\end{aligned}$$

We have to integrate these terms with respect to H. We set $H = (H - H_m) + H_m$ (here an application of *Helly-Bray*-lemma starts). By using the *d–Δ-interchange* trick we can transform all three terms with $H - H_m$ into integrals where $H - H_m$ appears as a function: e.g.

$$\begin{aligned}
&\int (H - H_m)(du, dv) \frac{1}{\beta_n(u, v)} (\Lambda_{10}^n(\Delta u, v) - \Lambda_{10}(\Delta u, v))\Lambda_{01}(u, \Delta v) \\
&= \int (H - H_m)(\Delta u, \Delta v) \frac{1}{\beta_n(u, v)} (\Lambda_{10}^n(du, v) - \Lambda_{10}(du, v))\Lambda_{01}(u, dv).
\end{aligned}$$

So if assumption 1-4 holds, then as we did above we can bound this term by

$$c\|H - H_m\|_\infty M^2(\|\mu_{1n}\|_v\|\mu_2\|_v + \|\mu_1\|_v\|\mu_2\|_v) \leq M'\|H - H_m\|_\infty.$$

Similarly, we have this bound for the other terms with $H - H_m$. The three terms with H_m we can directly bound by

$$\|(1/\beta_n(u,v) - 1/\beta(u,v))\|_\infty M(m),$$

$$\|(\Lambda_{10}^n(\Delta u, v) - \Lambda_{10}(\Delta u, v))\|_\infty M(m),$$

$$\|(\Lambda_{01}^n(u, \Delta v) - \Lambda_{01}(u, \Delta v))\|_\infty M(m),$$

where $M(m)$ stands for a constant times the variation norm of H_m. So we can conclude that we have the following bound:

$$\left\| \int H(du,dv)\left(\frac{\Lambda_{10}^n(\Delta u, v)\Lambda_{01}^n(u,\Delta v)}{\beta_n(u,v)} - \frac{\Lambda_{10}(\Delta u, v)\Lambda_{01}(u,\Delta v)}{\beta(u,v)} \right)\right\|_\infty$$
$$\leq c\|H - H_m\|_\infty + \epsilon_n M(m),$$

where ϵ_n converges to zero. Let now $m \to \infty$ slowly enough to obtain that this bound converges to zero (here ends *Helly-Bray*). This proves the statement.

In general all terms we will encounter in the differentiability proofs are dealt in the following way:

Telescoping

STEP 1. – Firstly, we do telescoping in order to rewrite a difference of two products as a sum of single differences: $\int A_n B_n - \int AB = \int (A_n - A)B + \int A_n(B_n - B)$. Consider one term (e.g.) $\int (A_n - A)B$. Here, we know that $A_n \to A$, but A_n can appear as a measure in one or two coordinates: $\int (A_n - A)(du,dv)B(u,v)$ or $\int (A_n - A)(du,v)B(u,dv)$ or the easiest case $\int (A_n - A)(u,v)B(du,dv)$.

Goal

STEP 2. – We want to bound the term $\int (A_n - A)B$, where we usually have that $A_n - A$ appears as a measure, in the supnorm of $A_n - A$ which is known to converge to zero. Therefore our goal is to get this term in a form so that we can apply integration by parts with respect to B.

Denominator-splitting, d–$\triangle$-interchange

STEP 3.

Case 0. – If $A_n - A$ appears as function we do not even have to apply integration by parts and we are ready.

Case 1. – If B generates a measure of bounded variation or is it a product of such functions (of bounded variation but some left and some right continuous) we can bound the term in the supnorm of $A_n - A$ by applying the integration by parts formula of Lemma 2.2.

Case 2. – If B is of unbounded variation, we substitute $B = (B - B_m) + B_m$ and we now want to bound the term with $B - B_m$ in the supnorm of $B - B_m$ and the term with B_m in the variation norm of B_m (Helly-Bray Lemma 2.5). We go back to step 3.

Case 3. – If B involves the denominator β we firstly apply the denominator trick Lemma 2.4 and $d - \triangle$-interchange Lemma 2.3 as in Corollary 2.1 in order to rewrite the term to a term of Case 0 or 1.

3. DIFFERENTIABILITY OF THE DABROWSKA, VOLTERRA, AND PRENTICE AND CAI REPRESENTATIONS OF F

In this section our goal is to establish Hadamard differentiability of the Volterra, Dabrowska, and Prentice-Cai representations of F. In fact, we will prove that each of these representations is continuously Hadamard differentiable, thereby paving the way for validity of the bootstrap in each case.

NOTATION AND ASSUMPTIONS ON SEQUENCES. – For any symbol which occurs as argument of the analyzed mapping, say Λ, Λ_n and $\Lambda_n^{\#}$ are sequences which both converge in supnorm to Λ and moreover it will be automatically assumed that they are of bounded variation uniformly in n.

Λ_n plays the role of the estimator of Λ using the original data and $\Lambda_n^{\#}$ plays the role of the same estimator, but using a bootstrap sample of the original data.

3.1. The Volterra Representation

We do the proof of the Volterra representation before the proof of the bivariate product integral (as part of the Dabrowska representation), because the proof is easier to generalize from the univariate case and we will still be able to refer to the global differentiability proof.

Consider the inhomogeneous Volterra equation

$$\overline{F}(t) = \Psi(t) + \int_{[0,t]} \overline{F}(s-)d\Lambda_{11}(s). \tag{13}$$

We consider this equation as an implicit equation for $\overline{F}$ for given functions Ψ and Λ_{11}. For any measure α on $\mathbb{R}^2$ set:

$$P((s,t];\alpha) = 1 + \sum_{n=1}^{\infty} \int_{s<u^1<\ldots<u^n\leq t} \alpha(du^1)\ldots\alpha(du^n). \tag{14}$$

$P(\cdot;\alpha) \equiv P_\alpha$ is the Peano series corresponding to α. The following propositions will be proved below in a separate subsection. The proofs are similar to the proofs given in Gill and Johansen (1990) as they already remarked on p. 1531. The non-homogeneous Volterra equation has a unique solution in terms of $P(\cdot;\Lambda_{11})$:

PROPOSITION 3.1. – *If* $\overline{F}$ *satisfies* (13), *then*

$$\overline{F}(t) = \Psi(t) + \int_{0<s\leq t} \Psi(s-)P((s,t];\Lambda_{11})d\Lambda_{11}(s).$$

Repeated substitution of the Volterra equation into itself and interchange of the order of integration make the claim intuitively clear. Here are two propositions giving useful properties of the Peano series P.

PROPOSITION 3.2 (Kolmogorov equations). – *The Peano series* $P \equiv P_\alpha$ *defined by* (14) *satisfies*

$$P_\alpha(s,t] = 1 + \int_{s<u\leq t} P_\alpha(s,u)\alpha(du) = 1 + \int_{s<u\leq t} P_\alpha(u,t]\alpha(du).$$

PROPOSITION 3.3 (Duhamel equation). – *If* α *and* β *are two measures on* $\mathbb{R}^2$ *with corresponding Peano series* P_α *and* P_β, *then*

$$P_\beta(s,t] - P_\alpha(s,t] = \int_{s<u\leq t} P_\alpha(s,u)P_\beta(u,t](\beta-\alpha)(du). \tag{15}$$

With the Duhamel equation one can show the following differentiability result for the Peano series. For all propositions and theorems recall our assumptions on the sequences Λ_n.

PROPOSITION 3.4 (weak continuous compact differentiability of P_α in supremum norm). – *Assume*

$$h_n^\# \equiv \sqrt{n}(\alpha_n^\# - \alpha_n) \to h \quad in \quad (D[0,\tau], \|\cdot\|_\infty).$$

Then, with $P_n^\# \equiv P(\cdot;\alpha_n^\#)$, $P_n \equiv P(\cdot;\alpha_n)$

$$\sqrt{n}(P_n^\# - P_n)(s,t] \to \dot{P}h(s,t] \quad uniformly\ in \quad (s,t) \in [0,\tau]^2 \tag{16}$$

where $\dot{P}$ *is given by*

$$\dot{P}h(s,t] = \int_{s<u\leq t} P(s,u)P(u,t]dh(u). \tag{17}$$

If h is of unbounded variation this is defined by (repeated) integration by parts (see *Lemma 2.2*).

CONSISTENCY. – In general, notice that this differentiability result for a mapping A certainly implies continuity of A; if $F_n \to F$ then $A(F_n) \to A(F)$. Therefore our differentiability results will also provide us with almost sure uniform consistency of our estimators.

Now, we have the tools to prove the weak continuous differentiability property of the Volterra representation $\Phi_1(\Psi, \Lambda_{11})$.

THEOREM 3.1 (weak continuous differentiability of Φ_1). – *Suppose that*

$$t_n^{-1}(\Psi_n^\# - \Psi_n) \to \alpha \quad in \quad D[0,\tau]$$
$$t_n^{-1}(\Lambda_{11}^{n\#} - \Lambda_{11}^n) \to \beta \quad in \quad D[0,\tau].$$

Then

$$t_n^{-1}(\Theta(\Psi_n^\#, \Lambda_{11}^{n\#}) - \Theta(\Psi_n, \Lambda_{11}^n)) \to d\Theta(\Psi, \Lambda_{11})(\alpha, \beta) \quad in\ D[0,\tau], \tag{18}$$

where $d\Theta(\Psi, \Lambda_{11})(\cdot,\cdot)$ *is a continuous linear functional defined on* $D[0,\tau]$.

Proof of Theorem 3.1. – For convenience denote Λ_{11} with Λ.

$$P_n^\#(s,t] \equiv P((s,t];\Lambda_n^\#)$$
$$P_n(s,t] \equiv P((s,t];\Lambda_n),$$

and write $\overline{F}_n^\# = \Phi_2(\Psi_n^\#, \Lambda_n^\#)$, $\overline{F}_n = \Phi_2(\Psi_n, \Lambda_n)$ and $\overline{F} = \Phi_2(\Psi, \Lambda)$. By equation (13)

$$\overline{F}_n^\#(t) = \Psi_n^\#(t) + \int_{s\leq t} \Psi_n^\#(s-)P_n^\#(s,t]d\Lambda_n^\#(s) \tag{19}$$

and

$$\overline{F}_n(t) = \Psi_n(t) + \int_{s\leq t} \Psi_n(s-)P_n(s,t]d\Lambda_n(s) \tag{20}$$

so that subtraction yields (by telescoping)

$$\begin{aligned} t_n^{-1}(\overline{F}_n^{\#}(t) - \overline{F}_n(t)) =& t_n^{-1}(\Psi_n^{\#}(t) - \Psi_n(t)) \\ &+ \int_{s\leq t} t_n^{-1}(\Psi_n^{\#} - \Psi_n)(s-)P_n^{\#}(s,t]d\Lambda_n^{\#}(s) \\ &+ \int_{s\leq t} \Psi_n(s-)t_n^{-1}(P_n^{\#} - P_n)(s,t]d\Lambda_n^{\#} \\ &+ \int_{s\leq t} \Psi_n(s-)P_n(s,t]t_n^{-1}(d\Lambda_n^{\#} - d\Lambda_n)(s) \\ =& I_n + II_n + III_n + IV_n. \end{aligned}$$

$I_n \to I$ by hypothesis. Our goal is to show that II_n, III_n, IV_n converge to their supposed limits II, III, IV. Firstly, one should notice that the supposed limits are well defined: for example $IV = \int_{s\leq t} \Psi(s-)P(s,t]d\beta(s)$ is defined by repeated integration by parts (Lemma 2.2). By telescoping we have:

$$\begin{aligned} II_n - II =& \int_{s\leq t} (\alpha_n^{\#} - \alpha)(s-)P(s,t)d\Lambda(s) \\ &+ \int_{s\leq t} \alpha_n^{\#}(s-)(P_n^{\#} - P)(s,t]d\Lambda(s) \\ &+ \int_{s\leq t} \alpha(s-)P_n^{\#}(s,t]d(\Lambda_n^{\#} - \Lambda)(s). \\ &+ \int_{s\leq t} (\alpha_n^{\#} - \alpha)(s-)P_n^{\#}(s,t]d(\Lambda_n^{\#} - \Lambda)(s). \end{aligned}$$

Because Λ is of bounded variation the first two terms can directly be bounded by a constant times the supnorm of $(\alpha_n^{\#} - \alpha)$ and $(P_n^{\#} - P)(s,t]$, respectively. $(\alpha_n^{\#} - \alpha)$ converges to zero by hypothesis and $(P_n^{\#} - P)(s,t]$ converges to zero by Proposition 3.4. Similarly, using that $\Lambda_n, \Lambda_n^{\#}$ are of bounded variation uniformly in n, we prove that the fourth term converges to zero by bounding it in the supnorm of $(\alpha_n^{\#} - \alpha)$. For the third term we have to apply the Helly-Bray Lemma 2.5 with $H(s) = \alpha(s)P_n^{\#}(s,t]$ and $F(s) = \Lambda_n^{\#} - \Lambda$, because α is of unbounded variation. So here we need that $s \to P_n^{\#}(s,t]$ is of bounded variation uniformly in n. This follows from the bounded variation of $\Lambda_n^{\#}$ as shown in the proof of Proposition 3.4.

The convergence of III_n, IV_n to their supposed limits is proved, similarly: only integration by parts and Helly-Bray are needed. This completes the proof. □

3.1.1. *Proofs of Propositions*

Proof of Proposition 3.2 (Kolmogorov equations). – For convenience, we define the region which appears in each term of the Peano series: $B_n(s,t] \equiv \{(u^1,\ldots,u^n) \in (\mathbb{R}^2)^n : s < u^1 < \ldots < u^n \leq t\}$. Now,

$$P_\alpha(s,u) = 1 + \sum_{n=1}^{\infty} \int_{B_n(s,u)} \alpha(du^1)\cdots\alpha(du^n), \tag{21}$$

so

$$\begin{aligned}\int_{s<u\leq t} P_\alpha(s,u)\alpha(du) =& \int_{s<u\leq t} 1\alpha(du)\\ &+ \sum_{n=1}^{\infty} \int_{B_{n+1}(s,t]} \alpha(du^1)\cdots\alpha(du^n)\alpha(du)\\ =& \sum_{n=1}^{\infty} \int_{B_n(s,t]} \alpha(du^1)\cdots\alpha(du^n)\\ =& P_\alpha(s,t] - 1.\end{aligned}$$

The backward equation is similarly proved. □

Proof of Proposition 3.3 (Duhamel equation). – Consider the following $m+n$-fold integral:

$$\int_{B_{m+n}(s,t]} \alpha(du^1)\cdots\alpha(du^m)\beta(du^{m+1})\cdots\beta(du^{m+n}). \tag{22}$$

By splitting the integration on u^m we can write this as:

$$\begin{aligned}&\int_{s<u^m\equiv u\leq t}\left\{\int_{B_{m-1}(s,u^m)} \alpha(du^1)\cdots\alpha(du^{m-1})\right\}\\ &\quad\times\left\{\int_{B_n(u^m,t]} \beta(du^{m+1})\cdots\beta(du^{m+n})\right\}\alpha(du).\end{aligned} \tag{23}$$

Similarly, splitting the integration on u^{m+1}, we can also write it as:

$$\begin{aligned}&\int_{s<u^{m+1}\equiv u\leq t}\left\{\int_{B_m(s,u^{m+1})} \alpha(du^1)\cdots\alpha(du^m)\right\}\\ &\quad\times\left\{\int_{B_n(u^{m+1},t]} \beta(du^{m+2})\cdots\beta(du^{m+n})\right\}\beta(du).\end{aligned} \tag{24}$$

Since (23) equals (24) for all m and n, we can sum up the resulting identity on m and n to obtain

$$\int_{s<u\leq t} P_\alpha(s,u)\{P_\beta(u,t]-1\}\alpha(du)$$
$$= \int_{s<u\leq t} \{P_\alpha(s,u)-1\}P_\beta(u,t]\beta(du). \tag{25}$$

Combining (25) with the Kolmogorov equations yields the Duhamel equation. □

Proof of Proposition 3.4 (weak continuous differentiability of P_α). – By the Duhamel equation we have:

$$t_n^{-1}\{P_n^\# - P_n\}(s,t] = \int_{s<u\leq t} P_n(s,u)P_n^\#(u,t]dh_n^\#(u). \tag{26}$$

The difference with its supposed limit is given by (telescoping)

$$\int_{s<u\leq t} (P_n - P)(s,u)P(u,t]dh(u)$$
$$+ \int_{s<u\leq t} P_n(s,u)(P_n^\# - P)(u,t]dh(u)$$
$$+ \int_{s<u\leq t} P_n(s,u)P_n^\#(u,t]d(h_n^\# - h)(u).$$

Firstly, notice that all three terms are defined by repeated integration by parts (Corollary 12), which can be done because $s \to P_n(s,t]$ (and $P_n^\#, P$) are of bounded variation uniformly in n (*see* below). The first and second term converge to zero by the Helly-Bray Lemma 2.5 and the third term can be bounded by the supnorm of $h_n^\# - h$ by applying integration by parts (Lemma 2.2). In all three bounds the variation norm of $P_n, P_n^\#, P$ considered as functions $s \to P_n(s,t]$ appear which are uniformly bounded. This is seen as follows. From the definition (14) of the Peano series it follows directly that $\|P_\alpha\|_\infty \leq \exp(\|\alpha\|_\infty)$ [*see* (7)]. Then by the Kolmogorov equation we have:

$$\|P_\alpha\|_v \leq \|P_\alpha\|_\infty \|\alpha\|_v \leq \exp(\|\alpha\|_\infty).$$

So if $\|\alpha\|_v < M$, then $\|P_\alpha\|_v$ is bounded. This proves the bounded variation of $P_n, P_n^\#, P$ by assumption on α_n. This completes the proof. □

3.2. The Dabrowska Representation

3.2.1. *The L-mapping*

We will state the differentiability result for the by far most complicated mapping L in the Dabrowska representation.

PROPOSITION 3.5. – *Denote* $\vec{\Lambda} = (\Lambda_{10}, \Lambda_{01}, \Lambda_{11})$. *Let* $\Gamma_3 = \Gamma_{31} - \Gamma_{32}$, *where*

$$\Gamma_{31}(\vec{\Lambda}) = \int_{[0,t]} \frac{\Lambda_{10}(du, v-)\Lambda_{01}(u-, dv)}{\beta(u,v)}$$

and

$$\Gamma_{32}(\vec{\Lambda}) = \int_{[0,t]} \frac{\Lambda_{11}(du, dv)}{\beta(u,v)}.$$

Assume that $\|(\Lambda_{11}^n, \Lambda_{11}^n\#)\|_v < M < \infty$ *and*

1. $\beta > \delta > 0$ *on* $[0, \tau]$ *for certain* $\delta > 0$.

2. *There exists a sequence of uniformly in* n *finite (signed) measures* μ_{2n} *so that* $\Lambda_{10}^n(u, dv) \ll \mu_{2n}(dv)$ *for all* u. *Similarly for* $\Lambda_{10}, \Lambda_{10}^n{}^\#, \Lambda_{01}, \Lambda_{01}^n, \Lambda_{01}^n{}^\#$.

3. *There exists a sequence of uniformly in* n *finite (signed) measures* μ_{1n} *so that* $\Lambda_{10}^n(du, v) \ll \mu_{2n}(du)$ *for all* u. *Similarly for* $\Lambda_{10}, \Lambda_{10}^n{}^\#, \Lambda_{01}, \Lambda_{01}^n, \Lambda_{01}^n{}^\#$.

4. $\left\| \frac{\Lambda_{10}^n(du, v)}{\mu_{1n}(du)} \right\|_\infty < M$ *and* $\left\| \frac{\Lambda_{10}^n(u, dv)}{\mu_{2n}(dv)} \right\|_\infty < M$ *for certain* $M < \infty$ *(uniform boundedness of the Radon-Nykodym derivatives). Similarly for* $\Lambda_{10}, \Lambda_{10}^n{}^\#, \Lambda_{01}, \Lambda_{01}^n, \Lambda_{01}^n{}^\#$

If $\vec{h}_n{}^\# \equiv \sqrt{n}(\vec{\Lambda}_n^\# - \vec{\Lambda}_n) \to \vec{h}$, *then we have:*

$$\sqrt{n}(\Gamma(\vec{\Lambda}_n{}^\#) - \Gamma(\vec{\Lambda}_n)) - d\Gamma(\vec{\Lambda})(\vec{h}_n^\#) \to 0, \tag{27}$$

for a certain continuous linear map $d\Gamma(\Lambda)$:

$$(D[0, \tau], \|\cdot\|_\infty)^3 \to (D[0, \tau], \|\cdot\|_\infty).$$

Proof. – We will give the proof of the characterization of ordinary compact differentiability, *i.e.* we replace Λ_n by Λ and $\Lambda_n^\#$ by Λ_n in (27). The reader can easily verify that the proof goes through when we do not do this. We have by telescoping:

$$\begin{aligned}
&\sqrt{n}(\Gamma_{31}(\Lambda_{01}^n, \Lambda_{10}^n) - \Gamma_{31}(\Lambda_{01}, \Lambda_{10}))\\
&= \sqrt{n}\left(\int\!\!\int \frac{\left\{\begin{array}{r} \beta(u,v)(\Lambda_{10}^n(du, v-)\Lambda_{01}^n(u-, dv) \\ -\Lambda_{10}(du, v-)\Lambda_{01}(u-, dv)) \end{array}\right\}}{\beta_n(u,v)\beta(u,v)}\right)\\
&\quad + \sqrt{n}\left(\int\!\!\int \frac{(\beta - \beta_n)(u,v)\Lambda_{10}(du, v-)\Lambda_{01}(u-, dv)}{\beta_n(u,v)\beta(u,v)}\right)\\
&= \int\!\!\int \frac{h_{10}^n(du, v-)\Lambda_{01}(u-, dv) + h_{01}^n(u-, dv)\Lambda_{10}^n(du, v-)}{\beta_n(u,v)}\\
&\quad + \int\!\!\int \sqrt{n}\frac{\beta - \beta_n}{\beta_n\beta}(u,v)\Lambda_{10}(du, v-)\Lambda_{01}(u-, dv).
\end{aligned}$$

It is easy to check that $\sqrt{n}\dfrac{\beta - \beta_n}{\beta_n\beta} \to H(u,v)$ for a fixed function $H(h_{10}, h_{01})$ linear in (h_{10}, h_{01}) which we will not write down. So the last term converges in $\|\cdot\|_\infty$ to

$$\int\!\!\int H(h_{10}, h_{01})(u,v)\Lambda_{10}(du, v-)\Lambda_{01}(u-, dv).$$

Notice that the supposed limit $d\Phi(\vec{\Lambda})$ is a continuous linear map because all terms can be defined by integration by parts with Lemma 2.2. We only consider the second integral. The first is dealt similarly. The difference between the second integral and its supposed limit can be rewritten by telescoping as the following sum of terms:

$$\begin{aligned}
&\int\!\!\int \frac{(h_{01}^n - h_{01})(u-, dv)\Lambda_{10}^n(du, v-)}{\beta_n(u,v)}\\
&\quad + \int\!\!\int \left(\frac{1}{\beta_n} - \frac{1}{\beta}\right)(u,v) + h_{01}(u-, dv)\Lambda_{10}(du, v-)\\
&\quad + \int\!\!\int \frac{(\Lambda_{10}^n - \Lambda_{10})(du, v-)h_{01}(u-, dv)}{\beta_n(u,v)}.
\end{aligned}$$

Term i. – Use corollary 2.1 with $H = h_{01}^n - h_{01}$, $\Lambda = \Lambda_{10}^n, \beta = \beta_n$. Then apply integration by parts (the second part of Lemma 2.2) and bound this term by the supnorm of $h_{10}^n - h_{10}$ times integrals like $\displaystyle\int \left|\frac{1}{\beta}\Lambda_{10}(du, v-)\Lambda_{01}(u-, dv)\right|$. For the rest we refer to the techniques in illustration 2 where we show by using the assumptions 1-4 that this variation is bounded.

Term ii. – Similar to our illustration II with $h(du, dv)$ replaced by $h_{01}(u-, dv)\Lambda_{10}(du, v-)$.

Term iii. – Substitute $h_{01} = (h_{01} - h_{01}^m) + h_{01}^m$. Now bound the term with $(h_{01} - h_{01}^m)$ in the supnorm of $(h_{01} - h_{01}^m)$ times a constant, and bound the term with h_{01}^m in the supnorm of $\Lambda_{10}^n - \Lambda_{10}$ times the variation of h_{01}^m, both in exactly the same way as we did in term i. Now, let $m \to \infty$ slowly enough (Helly-Bray Lemma 2.5).

The proof for Γ_{32} is similar, but easier. □

3.2.2. *The bivariate product-integral*

The essential ingredient for establishing differentiability results for the product-integral is the Duhamel equation. For the univariate product integral theory we refer to Gill and Johansen (1990). They also sketch how the proofs can be generalized to the multivariate product-integral. Here, we will present and prove the bivariate analogues of the Kolmogorov equations and Duhamel equation and finally state the differentiability result for the bivariate product-integral. For any signed measure L on $\mathbb{R}^2$ set

$$P((s,t], L) = \prod_{(s,t]} (1 + L(du, dv)), \tag{28}$$

where the bivariate product-integral

$$P_L(s,t] \equiv P((s,t], L) = \prod_{(s,t]} (1 + L(du, dv))$$

is defined as the limit of finite products of

$$\prod_{i,j=1}^{m} (1 + L((u_{i-1}, v_{j-1}), (u_i, v_j)]))$$

over partition-elements

$$J_{i,j} \equiv ((u_{i-1}, v_{j-1}), (u_i, v_j)] \text{ with } \max_{i,j}\{| (u_{i-1}, v_{j-1}) - (u_i, v_j) |\}$$

converging to zero. The ordering (specifying in what way we multiply over the elements of the partition) of this product is not relevant by the commutativity of multiplication in $\mathbb{R}$, but for our proofs we choose the video ordering (left under to right under then to left under one strip higher etc.). The proof that this product-integral is uniquely defined (that each sequence of partitions of rectangles with mesh converging to zero has the same limit) is exactly the same as the proof for the univariate product-integral as given in Gill and Johansen (1990), p. 1515.

Remark. – We use the same notation $P((s,t],L)$ for the bivariate Peano-series and the bivariate product integral. In one dimension these two are the same and in two dimensions the same properties (Kolmogorov equations, Duhamel equation) can be proved. By using the total ordering in $\mathbb{R}^2$ we can obtain all one dimensional results and we can go back and forth from total ordering to partial ordering.

PROPERTY 3.1. – $\prod_{(0,t]}(1+L(du,dv)) \leq \exp(\|L\|_v)$ *So if* L *is of bounded variation, then* $t \longrightarrow P((0,t],L)$ *is bounded in supnorm.*

This follows immediately from $1+ \mid L(J_{i,j}) \mid \leq \exp(\mid L(J_{i,j}) \mid)$. We will see that we can easily get generalizations of the Kolmogorov and Duhamel equation of the univariate case by replacing univariate intervals by rectangles with respect to the total (video) ordering. That is indeed what we will do. Then we will show that we can rewrite the obtained results in terms of rectangles with respect to the usual partial ordering.

LEMMA 3.1. – *Write* $(0,t] = \{x \in \mathbb{R}^2 : 0 < x \leq t\}$ *for an interval with respect to the partial ordering on* $\mathbb{R}^2$. *Denote* $]]0,t]]$ *for an interval with respect to the total (video) ordering on* $\mathbb{R}^2$: $(x,y) \in]]0,t]] \Leftrightarrow 0 < y < t_2$ *or* $y = t_2, 0 < x \leq t_1$. *Then for* $s \leq t$

$$(0,t] \cap]]0,s]] = (0,t_1] \times (0,s_2) \cup (0,s_1] \times \{s_2\}$$
$$(0,t] \cap]]s,\infty]] = (s_1,t_1] \times \{s_2\} \cup (0,t_1] \times (s_2,t_2].$$

The lemma says that we can describe these intersections as the union of one two dimensional rectangle and a one-dimensional line segment, both with respect to the partial ordering. The proof is trivial.

PROPOSITION 3.6 (Kolmogorov equations). – *Denote* $(0,t]$ *for an interval with respect to the partial ordering on* $\mathbb{R}^2$ *and denote* $]]0,t]]$ *for an interval with respect to the total ordering on* $\mathbb{R}^2$. *The bivariate product-integral* $P \equiv P_L$ *satisfies:*

$$\begin{aligned} P_L(s,t] &= 1 + \int_{(s,t]} P_L\{(s,t] \cap]]s,u[[\}L(du) \\ &= 1 + \int_{(s,t]} P_L\{(s,t] \cap]]u,t]]\}L(du). \end{aligned}$$

Proof. – We prove the first equality. Consider a finite partition π_m of $(s,t]$ of rectangles with diameter smaller than h_m. Replace the product-integrals by a finite product over this partition. Then the integral is an integral of a simple function with respect to the measure L. Because of the identity $\prod_{i=1}^{m}(1+a_i) = 1+\sum_{i=1}^{m}\prod_{j=1}^{i-1}(1+a_j)a_i$ it follows that the equality holds for this finite partition. By the convergence of this product to the product integral for $h_m \to 0$ (*see* definition of product integral) the left-hand side $P_L^m(s,t]$ and the integrand on the right-hand side $P_L^m\{(s,t]\cap]]s,u[[\}$ converge to $P_L(s,t]$ and $P_L\{(s,t]\cap]]s,u[[\}$, respectively. By using the dominated convergence theorem it is now straightforward to show that the right hand side converges for this sequence of partitions to $1+\int_{(s,t]} P_L\{(s,t]\cap]]s,u[[\}L(du)$. □

COROLLARY 3.1. – *If L is of bounded variation, then $t \to P((0,t],L)$ is of bounded variation.*

Proof. – This follows straightforwardly from property 3.1 and the Kolmogorov equations. For the precise argument see the proof of Theorem 3.4. □

LEMMA 3.2 (Duhamel equation with total ordering). – *We have:*

$$\begin{aligned} &P_\alpha(0,t] - P_\beta(0,t] \\ &\quad = \int_{(0,t]} P_\alpha\{(0,t]\cap]]0,s[[\}\, d(\alpha-\beta)(s)\, P_\beta\{(0,t]\cap]]s,\infty[[\}. \end{aligned} \tag{29}$$

Proof. – The proof is the same as the proof for the Kolmogorov equations except that we now have to use the telescoping-identity $\prod_{i=1}^{n} a_i - \prod_{i=1}^{n} b_i = \sum_{i=1}^{n}\prod_{j=1}^{i-1} a_j(a_i-b_i)\prod_{j=i+1}^{n} b_j$. □

Now, with these two lemmas we are able to write down a Duhamel equation which involves product-integrals over rectangles or lines with lower and upper corner chosen out of the corners of $(s,t]$. We can simplify this to only product-integrals over rectangles and lines with lower corner at $(0,0)$ as follows.

LEMMA 3.3.

$$P_\alpha(s,t] = \frac{P_\alpha(0,t]P_\alpha(0,s]}{P_\alpha(0,(s_1,t_2)]P_\alpha(0,(t_1,s_2)]},$$

which is the generalized ratio of the product-integral over a rectangle with lower corner at $(0,0)$ and upper corner at one of the four corners of $(s,t]$.

Proof. – The proof follows straightforwardly from the multiplicativity of the product-integral. □

PROPOSITION 3.7 (Duhamel equation). – *Define*

$$V(s,t) \equiv P_\alpha\{(0,t_1] \times (0,s_2)\} P_\alpha\{(0,s_1] \times \{s_2\}\} P_\beta\{(s_1,t_1] \times \{s_2\}\} P_\beta\{(0,t_1] \times (s_2,t_2]\},$$

where $P_\beta\{(s_1,t_1] \times \{s_2\}\}$ *and* $P_\beta\{(0,t_1] \times (s_2,t_2]\}$ *can be written as a generalized ratio of product-integrals over rectangles with lower corners at* $(0,0)$ *and upper corners with coordinates taken from* s *and* t (see *Lemma* 3.3). *Then*

$$P_\alpha(0,t] - P_\beta(0,t] = \int_{(0,t]} V(s,t) d(\alpha - \beta)(s). \tag{30}$$

All these product-integrals are of bounded (uniformly in t) variation in s by application of the Property 3.1. So by our repeated integration by parts formula Corollary 12 we can do integration by parts so that $\alpha - \beta$ appears as a function.

Proof (Duhamel equation). – Firstly, apply Lemma 3.2. Then by Lemma 3.1 and the multiplicativity of P_α we can write the product-integrals as a product over product-integrals over rectangles and hyperplanes with respect to the partial ordering. Finally apply Lemma 3.3. □

THEOREM 3.2. – *The bivariate product-integral* $P : (D[0,\tau], \|\cdot\|_\infty) \to (D[0,\tau], \|\cdot\|_\infty)$:

$$L \mapsto \prod_{[0,t]} (1 - L(du, dv))$$

satisfies the characterization of weak continuous differentiability, as stated in (27), *for sequences* $\|L_n\|_v < C$, $\|L_n^\#\|_v < C$ *converging to a signed measure* L.

It is already known that it holds for the univariate product-integral (Gill and Johansen, 1990).

Proof. – For this we refer to the differentiability proof of the bivariate Peano series in the preceding section: the same ingredients (Kolmogorov equations, Duhamel, repeated integration by parts) have to be used in the same way. □

3.3. Prentice-Cai representation

Recall the Prentice representation

$$\begin{aligned}\overline{F}(t) &= \Theta_1(\Lambda_{10}(\cdot,0), \Lambda_{01}(0,\cdot), R)\\ &= \Theta_1(\Lambda_{10}(\cdot,0), \Lambda_{01}(0,\cdot), \Theta_2(\widetilde{L}))\\ &= \Theta_1(\Lambda_{10}(\cdot,0), \Lambda_{01}(0,\cdot), \Theta_2(\Theta_3(\vec{\Lambda})\\ &= \Theta(\vec{\Lambda}),\end{aligned}$$

where Θ_1 is a product of two univariate product integrals w.r.t. $\Lambda_{10}(\cdot,0)$ and $\Lambda_{01}(0,\cdot)$, respectively, times R; $\Theta_2 = \Phi_2$ is the Volterra representation; Θ_3 is the $\widetilde{L}$ mapping which has the same structure (slightly easier) as the $\Gamma_3 = L$ mapping of Dabrowska's representation. So the weak continuous differentiability characterization has been proved for Θ_1 in Gill and Johansen (1990), Θ_2 is proved in Theorem 3.1, Θ_3 is proved by copying the proof of Proposition 3.5. The chain rule provides us now with the weak continuous differentiability characterization for Θ.

3.4. Differentiability theorem for Γ, Φ and Θ

THEOREM 3.3. – *All three representations are defined in section 1. Let Γ be the Dabrowska representation and $\vec{\Lambda}$ the vector of hazard measures corresponding with $\overline{F}$ as defined in section 1: $\overline{F} = \Gamma(\vec{\Lambda})$.*

DABROWSKA REPRESENTATION.

ASSUMPTIONS. – *Assume that $\|\Lambda_{11}^n\|_v < M < \infty$, $\|\Lambda_{11n}{}^{\#}\|_v < M < \infty$ and*

1. $\overline{F}(\tau) > 0$.

2. *There exists a sequence of uniformly in n finite (signed) measures μ_{2n} so that $\Lambda_{10}^n(u, dv) \ll \mu_{2n}(dv)$ for all u. Similarly for $\Lambda_{10}, \Lambda_{10}^n{}^{\#}, \Lambda_{01}, \Lambda_{01}^n, \Lambda_{01}^n{}^{\#}$.*

3. *There exists a sequence of uniformly in n finite (signed) measures μ_{1n} so that $\Lambda_{10}^n(du, v) \ll \mu_{2n}(du)$ for all u. Similarly for $\Lambda_{10}, \Lambda_{10}^n{}^{\#}, \Lambda_{01}, \Lambda_{01}^n, \Lambda_{01}^n{}^{\#}$.*

4. $\left\| \dfrac{\Lambda_{10}^n(du,v)}{\mu_{1n}(du)} \right\|_\infty < M$ *and* $\left\| \dfrac{\Lambda_{10}^n(u,dv)}{\mu_{2n}(dv)} \right\|_\infty < M$ *for certain $M < \infty$ (uniform boundedness of the Radon-Nykodym derivatives). Similarly for $\Lambda_{10}, \Lambda_{10}^n{}^{\#}, \Lambda_{01}, \Lambda_{01}^n, \Lambda_{01}^n{}^{\#}$.*

Let $Z_{\vec{\Lambda}}^{n\#} \equiv \sqrt{n}(\vec{\Lambda}_n^{\#} - \vec{\Lambda}_n) \rightarrow Z_{\vec{\Lambda}}$. *We have the following differentiability result for* Γ:

$$\sqrt{n}(\Gamma(\vec{\Lambda}_n^{\#}) - \Gamma(\vec{\Lambda}_n)) - d\Gamma(\vec{\Lambda})(Z_{\vec{\Lambda}}^{n\#}) \rightarrow 0,$$

for a continuous linear map $d\Gamma(\Lambda) : (D[0,\tau], \|\cdot\|_\infty)^3 \rightarrow (D[0,\tau], \|\cdot\|_\infty)$.

PRENTICE-CAI REPRESENTATION. – *The same statement holds for the Prentice-Cai representation* $\overline{F} = \Theta(\vec{\Lambda})$.

VOLTERRA REPRESENTATION. – *The same differentiability result holds for* $\overline{F} = \Phi(\vec{\Lambda})$ *with the assumptions 2, 3 and 4 replaced by:*

$$\|(\Lambda_{10}^n, \Lambda_{01}^n)\|_v < M < \infty \qquad \textit{and} \qquad \|(\Lambda_{10}^{n\#}, \Lambda_{01}^{n\#})\|_v < M < \infty.$$

Proof. – This differentiability property has been proved for the univariate product integral in Gill and Johansen (1990) (so this gives it for Γ_1, Θ_1, Φ_2), for the bivariate product integral in Theorem 3.2 (so this gives it for Γ_2), for the bivariate Volterra representation (bivariate Peano series) in Theorem 3.1 (so this gives it for Θ_2, Φ_1), for the L mapping in Theorem 3.5 where we need the denominator assumptions (so this gives it for Γ_3, Θ_3), where one has to notice that Assumption 1 tells us that $\beta > 0$ (denominator in L and $\widetilde{L}$). Now, the theorem follows from the chain rule. □

4. THE ESTIMATORS

Let Φ, Γ and Θ denote the Volterra, Dabrowska and Prentice-Cai representation, respectively, which were defined and studied in sections 2 and 3. We now construct the estimators in models 1-3 which are based on these representations. From now everything indexed by n is random.

ESTIMATORS FOR MODEL 1. – *Estimation of a bivariate distribution function with known marginals.*

In this model the marginal distributions are known. Recall the Definition 1 (or call it representation in terms of F) of the integrated hazard $\vec{\Lambda} = \vec{\Lambda}(F) = (\Lambda_{10}(F), \Lambda_{01}(F), \Lambda_{11}(F))$. We estimate the hazards with their natural empirical estimators. So let

$$\overline{F}_n(t) \equiv \frac{1}{n} \sum_{i=1}^{n} I[(X,Y)_i > (t_1, t_2)],$$

be the empirical survival function. Then we take $\vec{\Lambda}_n = \vec{\Lambda}(F_n)$. So $\vec{\Lambda}_n$ has the following coordinates:

$$\Lambda_{11}^n \equiv \int_{(0,t]} \frac{I[\overline{F}_n(s-) > 0]}{\overline{F}_n(s-)} dF_n(s),$$
$$\Lambda_{10}^n \equiv \int_{(0,t_1]} \frac{I[\overline{F}_n(u-, t_2) > 0]}{\overline{F}_n(u-, t_2)} F_n(du, t_2),$$
$$\Lambda_{01}^n \equiv \int_{(0,t_2]} \frac{I[\overline{F}_n(t_1, v-) > 0]}{\overline{F}_n(t_1, v-)} F_n(t_1, dv).$$

DABROWSKA ESTIMATOR. – We only have to estimate the L operator which captures the dependence structure, and this L operator was a nice functional of $\vec{\Lambda}$. Recall the representation (2):

$$\overline{F} = \Gamma_1(\Lambda_{10}(\cdot, 0), \Lambda_{01}(0, \cdot), \Gamma_2\Gamma_3(\vec{\Lambda})).$$

Because $\Lambda_{10}(\cdot, 0) = \Lambda_{10}(\cdot, 0)(F_1^0)$, $\Lambda_{01}(0, \cdot) = \Lambda_{01}(0, \cdot)(F_2^0)$ are known we only have to plug in $\vec{\Lambda}_n$ for $\vec{\Lambda}$ in order to get the estimator based on the Dabrowska representation. So

$$\begin{aligned} \overline{F}_n^D(t) &= \Gamma_1(\Lambda_{10}(\cdot, 0), \Lambda_{01}(0, \cdot), \Gamma_2\Gamma_3(\vec{\Lambda}_n))(t) \\ &\text{for} \quad t \in E_n^+ \equiv \{s : \overline{F}_n(s-) > 0\} \end{aligned} \tag{31}$$

which is just a product: $\overline{F}_1^0(t_1)\overline{F}_2^0(t_2) \prod_{(0,t]}(1 - L(\vec{\Lambda}_n))$. Using the Dabrowska representation tells us that $\overline{F}_n = \overline{F}_{1n}\overline{F}_{2n} \prod(1 - L(\vec{\Lambda}_n))$. So the estimator simplifies to

$$\overline{F}_n^D = \frac{\overline{F}_1^0 \overline{F}_2^0}{\overline{F}_{1n}\overline{F}_{2n}} \overline{F}_n.$$

VOLTERRA ESTIMATOR. – Recall the Volterra representation (8):

$$\overline{F} = \Phi_1(\Phi_2(\Lambda_{10}(\cdot, 0), \Lambda_{01}(0, \cdot)), \Lambda_{11}),$$

where $\Psi = 1 - F_1^0 - F_2^0 = \Phi_2(\Lambda_{10}(\cdot, 0), \Lambda_{01}(0, \cdot))$ is completely known and thereby need not to be estimated. So the Volterra estimator is now given by:

$$\overline{F}_n^V(t) = \Phi_1(\Phi_2(\Lambda_{10}(\cdot, 0), \Lambda_{01}(0, \cdot)), \Lambda_{11}^n)(t) \qquad \text{for} \quad t \in E_n^+. \tag{32}$$

PRENTICE-CAI ESTIMATOR. – Recall the Prentice and Cai representation (11):

$$\overline{F}^{PC} = \Theta_1(\Lambda_{10}(\cdot,0), \Lambda_{01}(0,\cdot), \Theta_2(\Theta_3(\Lambda_{10}(\cdot,0), \Lambda_{01}(0,\cdot), \vec{\Lambda}))).$$

Also here we only have to estimate the $\widetilde{L} = \Theta_3$-mapping which is a functional in $(\Lambda_{10}(\cdot,0), \Lambda_{01}(0,\cdot), \vec{\Lambda})$. Define

$$K_n \equiv \Theta_3(\Lambda_{10}(\cdot,0), \Lambda_{01}(0,\cdot), \vec{\Lambda}_n)).$$

So then the Prentice-Cai estimator is given by:

$$\left.\begin{aligned} &\overline{F}_n^{PC}(t)\\ &\quad = \Theta_1(\Lambda_{10}(\cdot,0), \Lambda_{01}(0,\cdot), \Theta_2(\Theta_3(\Lambda_{10}(\cdot,0), \Lambda_{01}(0,\cdot), \vec{\Lambda}_n)))(t)\\ &\quad\ \ \text{for} \quad t \in E_n^+\\ &\quad = \overline{F}_1^0(t_1)\overline{F}_2^0(t_2)\Theta_2(K_n)(t). \end{aligned}\right\} \quad (33)$$

ESTIMATORS FOR MODEL 2. – *The "three-sample" model.*

As above we consider all three representations as mappings from $(\Lambda_{10}(\cdot,0), \Lambda_{01}(0,\cdot)) = (\Lambda_{10}(\cdot,0)(F_1), \Lambda_{01}(0,\cdot)(F_2))$ and $\vec{\Lambda} = \vec{\Lambda}(F)$ to itself. We obtain the estimators for model 2 by replacing these by: $(\Lambda_{10}^n(\cdot,0), \Lambda_{01}^n(0,\cdot)) = (\Lambda_{10}(\cdot,0)(F_1^{n_1+n_2}), \Lambda_{01}(0,\cdot)(F_2^{n_1+n_3}))$ and $\vec{\Lambda}_n = \vec{\Lambda}(F^{n_1})$, respectively, where F^{n_1} is the (joint) empirical distribution of F using the n_1 observations of sample 1, $F_1^{n_1+n_2}$ is the marginal empirical distribution of F_1 based on the $n_1 + n_2$ observations from sample 1 and 2, and $F_2^{n_1+n_3}$ is the marginal empirical distribution of F_2 based on the $n_1 + n_3$ observations from sample 1 and 3. In other words we use all the appropriate marginal data to estimate the marginals F_1, F_2. We are now just back in model 1 if we set $F_1^0 = F_1^{n_1+n_2}$ and $F_2^0 = F_2^{n_1+n_3}$ and forget the second and third sample. In other words, we just use the auxiliary samples to estimate the marginals and then we just take the estimators proposed in model 1 (with known marginals replaced by these estimated marginals) using the first sample. Therefore, we obtain the following estimators for model 2:

DABROWSKA ESTIMATOR. – Define $V_n^+ \equiv \{s : \overline{F}^{n_1}(s-) > 0\}$. Then

$$\overline{F}_n^D(t) = \frac{\overline{F}_1^{n_1+n_2}(t_1)\overline{F}_2^{n_1+n_3}(t_2)}{\overline{F}_1^{n_1}(t_1)\overline{F}_2^{n_1}(t_2)}\overline{F}^{n_1}(t) \qquad \text{for} \quad t \in V_n^+.$$

VOLTERRA ESTIMATOR.

$$\begin{aligned}\overline{F}_n^V(t) =& \Phi_1(\Phi_2(\Lambda_{10}(\cdot,0)(F_1^{n_1+n_2}),\\ &\Lambda_{01}(0,\cdot)(F_2^{n_1+n_3})), \Lambda_{11}(F^{n_1}))(t) \qquad \text{for} \quad t \in V_n^+, \quad (34)\end{aligned}$$

where $\Phi_2(\Lambda_{10}(\cdot,0)(F_1^{n_1+n_2}),\Lambda_{01}(0,\cdot)(F_2^{n_1+n_3})) = 1 - F_1^{n_1+n_2} - F_2^{n_1+n_3}$.

PRENTICE-CAI ESTIMATOR. – Define

$$K_n = \Theta_3(\Lambda_{10}(\cdot,0)(F_1^{n_1+n_2}),\Lambda_{01}(0,\cdot)(F_2^{n_1+n_3}),\vec{\Lambda}(F^{n_1})).$$

Then

$$\overline{F}_n^{PC}(t) = \overline{F}_1^{n_1+n_2}(t_1)\overline{F}_2^{n_1+n_3}(t_2)\Theta_2(K_n)(t) \qquad \text{for} \quad t \in V_n^+.$$

Remark. – If we only use the bivariate sample for the K_n in the Prentice-Cai estimator, then we get the same simplification as we had with the Dabrowska estimator in model 1:

$$\overline{F}_n^{PC} = \frac{\overline{F}_1^{n_1+n_2}\overline{F}_2^{n_1+n_3}}{\overline{F}_1^{n_1}\overline{F}_2^{n_1}}\overline{F}^{n_1}.$$

ESTIMATORS FOR MODEL 3. – *Bivariate random censoring.*

Here, one sees the advantage of choosing the hazards as parameters of the model: they are perfectly suited for censored data. Define the following subdistributions of the data corresponding with the four kinds of censoring which can occur.

$$H_{ij}(t) \equiv P(X \le t_1, Y \le t_2, \delta = i, \epsilon = j) \qquad \text{for} \quad i,j \in \{0,1\}, \ \ t \in \mathbb{R}_+^2.$$

and

$$H(t) \equiv P(X \le t_1, Y \le t_2) = \sum_{i,j} H_{ij}(t).$$

Then, on $[0,\tau]$ with $\overline{H}(\tau-) = \overline{FG}(\tau-) > 0$,

$$\begin{aligned}
\Lambda_{11}(t) &= \int_{[0,t]} \frac{\overline{G}(s-)}{\overline{F}(s-)\overline{G}(s-)} dF(s) \\
&= \int_{[0,t]} \frac{1}{\overline{H}(s-)} dH_{11}(s) \\
\Lambda_{10}(t) &= \int_{[0,t_1]} \frac{\overline{G}(u-,t_2)}{\overline{F}(u-,t_2)\overline{G}(u-,t_2)} F(du,t_2) \\
&= \int_{[0,t_1]} \frac{1}{\overline{H}(u-,t_2)} (H_{11}+H_{10})(du,t_2) \\
\Lambda_{01}(t) &= \int_{[0,t_2]} \frac{\overline{G}(t_1,v-)}{\overline{F}(t_1,v-)\overline{G}(t_1,v-)} F(t_1,dv) \\
&= \int_{[0,t_2]} \frac{1}{\overline{H}(t_1,v-)} (H_{11}+H_{01})(t_1,dv).
\end{aligned}$$

If we define $\vec{H} = (H_{10}, H_{01}, H_{00}, H_{11})$, then $\vec{\Lambda} = \vec{\Lambda}(\vec{H})$. Let

$$H_{nij} \equiv \frac{1}{n} \sum_{k=1}^{n} I(X_k \leq t_1, Y_k \leq t_2, \delta_k = i, \epsilon_k = j)$$

be the empirical distribution of H_{ij} for $i, j \in \{0, 1\}$ and $\vec{H}_n = (H_{n10}, H_{n01}, H_{n00}, H_{n11})$. We estimate $\vec{\Lambda}$ with

$$\vec{\Lambda}_n = \vec{\Lambda}(\vec{H}_n) \qquad \text{for} \quad t \in W_n^+ \equiv \{t : \overline{H}(t) > 0\}.$$

In other words $\vec{\Lambda}_n$ is given by the formulas above with H_{ij} replaced by H_{nij}.

DABROWSKA ESTIMATOR. – Recall the representation $S = \Gamma(\vec{\Lambda})$. We have

$$\overline{F}_n^D(t) = \Gamma(\vec{\Lambda}_n)(t) \qquad \text{for} \quad t \in W_n^+, \tag{35}$$

which equals the product $\overline{F}_{1n}\overline{F}_{2n} \prod(1 - L(\vec{\Lambda}_n))$ where $\overline{F}_{1n}$, $\overline{F}_{2n}$ are the univariate Kaplan-Meier estimators of $\overline{F}_1, \overline{F}_2$, respectively.

VOLTERRA ESTIMATOR. – Recall the representation $S = \Phi(\vec{\Lambda})$. We have

$$\overline{F}_n^V = \Phi(\vec{\Lambda}_n) \qquad \text{for} \quad t \in W_n^+, \tag{36}$$

where $\Phi_2(\vec{\Lambda}_n) = 1 - F_{1n} - F_{2n}$.

PRENTICE-CAI ESTIMATOR. – Recall the representation $S = \Theta(\vec{\Lambda})$. So

$$\overline{F}_n^{PC} = \Theta(\vec{\Lambda}_n) \qquad \text{for} \quad t \in W_n^+, \tag{37}$$

which is equal to the product $F_{1n}F_{2n}\Theta_2(\Theta_3(\vec{\Lambda}_n))$.

5. ASYMPTOTIC PROPERTIES OF THE ESTIMATORS

We will use the results of section 3 to establish functional central limit theorems for the estimators defined in section 4 for models 1-3. As outlined in section 1, we do this by applying the functional delta-method theorems in Wellner (1993) to the representations Φ, Γ and Θ. Since the necessary characterization of weak continuous Hadamard differentiability has been established in section 3, the remaining hypothesis of the delta-method which we need to verify here is weak convergence of the normalized arguments of

Φ, Γ and Θ. We use the modern weak convergence theory, due to Hoffmann-Jørgensen (1984) and Dudley (1985) following an evolution from Dudley (1966), which makes measurability an irrelevant issue, but measurability in the limit is required. With this theory one does not have to give up the Borel-sigma algebra but (in non-separable metric spaces D) it gives up the goal of inducing distributions on D equipped with some sigma-algebra of subsets. Weak convergence will now be equivalent to convergence of inner and outer probabilities of Borel P-continuity sets. We consider random processes as elements of the cadlag function space endowed with the supnorm and the Borel sigma algebra. We will denote this space with $(D[0,\tau], \|\cdot\|_\infty, \beta)$. If X_n converges weakly to a random X in this sense we will denote it with $X_n \stackrel{D}{\Longrightarrow} X$. For an extensive weak convergence theory we refer to van der Vaart and Wellner (1993).

Let P_n as usual be the empirical distribution using the i.i.d. data $X_i, i = 1, \ldots, n$, $X_i \sim P_0$. Now, we consider a bootstrap sample of n i.i.d. observations $X_i^\#, i = 1, \ldots, n$ with $X_i^\# \sim P_n$, We denote the empirical distribution based on this bootstrap sample with $P_n^\#$ and estimators of (say) θ based on this bootstrap sample with $\theta_n^\#$. The question of interest is: does $\sqrt{n}(\theta_n^\# - \theta_n) \stackrel{D}{\Longrightarrow} Z^\#$ a.s. (*i.e.* given the data $X_i, i = 1, \ldots$) as $n \to \infty$, where $Z^\# \stackrel{d}{=} Z$ and $\sqrt{n}(\theta_n - \theta) \stackrel{D}{\Longrightarrow} Z$?

Recall section 4 where we defined the estimators.

MODEL 1. – For all three representations we used the representation $F \to \vec{\Lambda}(F)$ for estimating $\vec{\Lambda}$: for each representation we plugged in the same $\vec{\Lambda}_n = \vec{\Lambda}(F_n)$ for $\vec{\Lambda} = \vec{\Lambda}(F)$. Let $V_1 = \vec{\Lambda}(F), V_{1n} = \vec{\Lambda}(F_n)$ and $V_{1n}^\# = \vec{\Lambda}(F_n^\#)$. So we need to prove that $\sqrt{n}(V_{1n} - V_1) \stackrel{D}{\Longrightarrow} Z_1$ and

$$\sqrt{n}(V_{1n}^\# - V_{1n}) \stackrel{D}{\Longrightarrow} Z_1 \text{a.s.} \quad \text{in} \quad (\mathrm{D}[0,\tau])^3$$

for certain Gaussian process Z_1.

MODEL 2. – For all three representations we used the representation $F \to \vec{\Lambda}(F)$, $F_1 \to \Lambda_{10}(\cdot, 0)(F_1)$ and $F_2 \to \Lambda_{01}(0, \cdot)(F_2)$ for estimating $\vec{\Lambda}$ and $\Lambda_{10}(\cdot, 0), \Lambda_{01}(0, \cdot)$, respectively: for each representation we plugged in the same $\vec{\Lambda}_n = \vec{\Lambda}(F^{n_1})$ for $\vec{\Lambda}(F)$ and $(\Lambda_{10}(\cdot,0)(F_1^{n_1+n_2}), \Lambda_{01}(0,\cdot)(F_2^{n_1+n_3})$ for $(\Lambda_{10}(\cdot,0)(F_1), \Lambda_{01}(0,\cdot)(F_2))$. Let $V_2 = (\Lambda_{10}(\cdot,0)(F_1), \Lambda_{01}(0,\cdot)(F_2), \vec{\Lambda}(F))$ and let V_{2n} and $V_{2n}^\#$ be defined as above by substituting $(F_1^{n_1+n_2}, F_2^{n_1+n_3}, F^{n_1})$ and $(F_1^{n_1+n_2\#}, F_2^{n_1+n_3\#}, F^{n_1}\#)$, respectively. So we need to prove that $\sqrt{n}(V_{2n} - V_2) \stackrel{D}{\Longrightarrow} Z_2$ and

$$\sqrt{n}(V_{2n}^\# - V_{2n}) \stackrel{D}{\Longrightarrow} Z_2 \text{a.s.} \quad \text{in} \quad (\mathrm{D}[0,\tau])^5$$

for certain Gaussian process Z_2. We will do asymptotics as $n_1 \to \infty$, irrespative of what n_2 and n_3 do.

MODEL 3. – For all three representations we used the representation $\vec{H} \to \vec{\Lambda}(\vec{H})$ for estimating $\vec{\Lambda}$: for each representation we plugged in the same $\vec{\Lambda}_n = \vec{\Lambda}(\vec{H}_n)$ for $\vec{\Lambda}(\vec{H})$. Let $V_3 = (\vec{\Lambda}(\vec{H}))$ and let V_{3n} and $V_{3n}^{\#}$ be defined as above by substituting $\vec{H}_n$ and $\vec{H}_n^{\#}$, respectively. So we need to prove that $\sqrt{n}(V_{3n} - V_3) \stackrel{D}{\Longrightarrow} Z_3$ and

$$\sqrt{n}(V_{3n}^{\#} - V_{3n}) \stackrel{D}{\Longrightarrow} Z_3 \text{a.s.} \quad \text{in} \quad (\mathrm{D}[0,\tau])^3$$

for certain Gaussian process Z_3.

By the functional delta-method (Wellner, 1992), applied to the representations V_i, $i = 1, 2, 3$, it suffices [for showing that $\sqrt{n}(V_{in}^{\#} - V_{in}) \stackrel{D}{\Longrightarrow} Z_i$ a.s., $i = 1, 2, 3$] to show that the representations $V_1(F), V_2(F_1, F_2, F)$ and $V_3(\vec{H})$ satisfy the characterization of weak continuous differentiability and that the bootstrap works for the plugged in empirical processes:

$$\begin{aligned}
X_n^{\#} &\equiv \sqrt{n}(F_n^{\#} - F_n) \stackrel{D}{\Longrightarrow} X^1 \text{a.s.} \\
Y_n^{\#} &\equiv \sqrt{n}(F^{n_1 \#} - F^{n_1}, F_1^{n_1+n_2 \#} - F_1^{n_1+n_2}, F_2^{\# n_1+n_3} - F_2^{n_1+n_3}) \\
&\quad \stackrel{D}{\Longrightarrow} X^2 \text{a.s.} \\
Z_n^{\#} &\equiv \sqrt{n}(\vec{H}_n^{\#} - \vec{H}_n) \stackrel{D}{\Longrightarrow} X^3 \text{a.s.}
\end{aligned}$$

where the Gaussian processes X^i, $i = 1, 2, 3$ are the limit processes of the non-bootstrapped processes X_n, Y_n, Z_n, respectively.

Firstly, we will state a lemma which will easily provide us with the weak continuous differentiability of the three representations V_i, $i = 1, 2, 3$.

LEMMA 5.1. – *The functional*

$$A : (F, G) \mapsto \int F(s) dG(s)$$

satisfies the characterization of weak continuous differentiability at any point (F, G) *where* F *and* G *are of bounded variation for sequences* $(F_n, G_n), (F_n^{\#}, G_n^{\#})$ *of bounded variation uniformly in* n [see (27)].

This lemma has been proved in illustration I and this mapping is also contained in the mapping L and $\tilde{L}$: the integration by parts Lemma 2.2 and the Helly-Bray Lemma 2.5 are the only ingredients we need in order to carry out the univariate proof in Gill (1989).

Recall the representations of V_i, $i = 1,2,3$: they are all compositions of $Y \to \frac{1}{Y}$ and A. So the weak continuous differentiability of V_i, $i = 1,2,3$ follows directly by the weak continuous differentiability of $Y \to \frac{1}{Y}$ at a $Y > \delta > 0$ on $[0,\tau]$ for a $\delta > 0$ and application of lemma 5.1 and the chain rule. In models 1 and 2 $\overline{F}$ plays the role of Y and in model 3 $\overline{FG}$ plays the role of Y. So for models 1 and 2 we need the assumption that $[0,\tau]$ is chosen such that $\overline{F}(\tau) > 0$ and in model 3 we need that $\overline{F}(\tau)\overline{G}(\tau) > 0$.

It remains to verify that the bootstrap works for the empirical processes we plugged into V_i, $i = 1,2,3$ which follows straightforwardly from well known empirical process theory:

MODEL 1. – F_n is just the usual empirical process indexed by the indicators. So the weak convergence of X_n and the bootstrapped $X_n^{\#}$ (a.s.) are well known.

MODEL 2. – We need to show weak convergence of Y_n and $Y_n^{\#}$ to the same limit process X^2.

Let $n = n_1 + n_2 + n_3$ and suppose that $\lambda_{n1} \equiv n_1/n \to \lambda_1 > 0$ and $\lambda_{ni} \equiv n_i/n \to \lambda_i \geq 0$, $i = 2,3$. It is well known (just empirical processes indexed by the indicators) that

$$\begin{gathered}(\sqrt{n_1}(F^{n_1\#} - F^{n_1}), \sqrt{n_1+n_2}(F_1^{n_1+n_2\#} - F_1^{n_1+n_2}),\\ \sqrt{n_1+n_3}(F_2^{n_1+n_2\#} - F_2^{n_1+n_3})) \stackrel{D}{\Longrightarrow} W\end{gathered}$$

a.s. in $(D[0,\tau])^3$ for $n \to \infty$ and for a certain Brownian bridge $W = (W_1, W_2, W_3)$. Because $\sqrt{n/n_1} \to 1/\sqrt{\lambda_1}$, $\sqrt{n/n_1+n_2} \to \sqrt{1/(\lambda_1+\lambda_2)}$ and $\sqrt{n/(n_1+n_3)} \to \sqrt{1/(\lambda_1+\lambda_3)}$ this implies (just write for the first coordinate $\sqrt{n} = \sqrt{n/n_1}\sqrt{n_1}$ and similarly for the second and third coordinate):

$$\sqrt{n}((F^{n_1\#} - F^{n_1}), (F_1^{n_1+n_2\#} - F_1^{n_1+n_2}), (F_2^{n_1+n_2\#} - F_2^{n_1+n_3})) \stackrel{D}{\Longrightarrow} X^2$$

a.s. in $(D[0,\tau])^3$ for $n \to \infty$ and where

$$X^2 = \left(\frac{1}{\sqrt{\lambda_1}}W_1, \sqrt{\frac{1}{\lambda_1+\lambda_2}}W_2, \sqrt{\frac{1}{\lambda_1+\lambda_3}}W_3\right).$$

MODEL 3. – Again, the weak convergence of $Z_n^{\#}$ is well known from empirical process theory.

5.1. Final theorem

THEOREM 5.1 (Functional central limit theorems for the estimators $\overline{F}_n^D, \overline{F}_n^{PC}$ and $\overline{F}_n^V$ in the models 1, 2 and 3). – *Suppose model i, $i \in \{1,2,3\}$, holds and assume that*

$$\begin{aligned} \overline{F}(\tau) > 0 \quad & if \quad i = 1 \ \ or \ \ i = 2 \\ \overline{FG}(\tau) > 0 \quad & if \quad i = 3. \end{aligned}$$

Recall the definitions of $X^i, V_1 = V_1(F)$, $V_2 = V_2(F_1, F_2, F)$, $V_3 = V_3(\vec{H})$ made in this section and the representations $V_i \rightarrow \Gamma(V_i)$, $V_i \rightarrow \Phi(V_i)$, $V_i \rightarrow \Theta(V_i)$ given in section 1, $i = 1,2,3$. We denote their derivatives with $d\Gamma, d\Phi$ and $d\Theta$.

DABROWSKA'S ESTIMATOR. – *For models $i = 1,2,3$.*

$$\overline{F}_n^D \rightarrow \overline{F} \ \ a.s.$$

and

$$\sqrt{n}(\overline{F}_n^D - \overline{F}) \xrightarrow{D} (d\Gamma(V_i) \circ dV_i(F))(X^i) \quad in \quad (D[0,\tau], \|\cdot\|_\infty, \beta)$$

for a continuous linear map

$$d\Gamma(V_i) \circ dV_i(F) : (D[0,\tau], \|\cdot\|_\infty)^3 \rightarrow (D[0,\tau], \|\cdot\|_\infty)$$

if $i = 1$ or $i = 3$ and

$$d\Gamma(V_2) : (D[0,\tau], \|\cdot\|_\infty)^5 \rightarrow (D[0,\tau], \|\cdot\|_\infty).$$

Moreover,

$$\sqrt{n}(\overline{F}_n^{\#D} - \overline{F}_n^D) \xrightarrow{D} d\Gamma(V_i)(X) a.s. \quad in \quad (D[0,\tau], \|\cdot\|_\infty).$$

So this estimator is consistent, its normalized version converges weakly to a Gaussian process and the bootstrap is asymptotically valid.

PRENTICE AND CAI'S ESTIMATOR. – *The same statement holds: just replace $d\Gamma(V_i)$ by $d\Theta(V_i)$.*

VOLTERRA ESTIMATOR. – *The same statement holds: just replace $d\Gamma(V_i)$ by $d\Phi(V_i)$.*

Proof. – We have to verify the conditions of the (bootstrap) functional delta-method for Hadamard differentiable functionals (Theorem 3.4 and

Corollary 5.2 in Wellner 1992) and apply it to Γ, Θ and Φ all three considered as functionals in V_1 for model 1, V_2 for model 2 and V_3 for model 3. Then we have the weak convergence result by application of Theorem 3.2 and the bootstrap result by application of Corollary 5.2. The weak convergence of $\sqrt{n}(V_{in}^{\#} - V_{in})$ (a.s.) and of $\sqrt{n}(V_{in} - V_i)$ $i = 1, 2, 3$, have been shown above. For the differentiability condition we only need to verify the conditions of Theorem 3.3. Assumption 1 in Theorem 3.3 is $\overline{F}(\tau) > 0$. For the other assumptions it suffices to show that $\vec{\Lambda}$, $\vec{\Lambda}_n$ and $\vec{\Lambda}_n^{\#}$ satisfy the Assumptions 2-4 stated in Theorem 3.3 (the bounded Radon-Nykodym derivatives assumptions). Here, one has to notice that assumption 2-4 for $\Lambda_{11}^n, \Lambda_{11}^{n\#}$ are stronger than the requirement that these functions are of bounded variation uniformly in n.

VERIFICATION OF ASSUMPTIONS 2-4 OF THEOREM 3.3. – We will prove these conditions for $\Lambda_{10}(t) = -\int \overline{F}(du, t_2)/\overline{F}(u-, t_2)$. It will be clear that the proof for Λ_{01} and Λ_{11} is similar. We have

$$\Lambda_{10}(du, v) = -\frac{\overline{F}(du, t_2)}{\overline{F}(u-, t_2)} \leq \frac{1}{\overline{F}(\tau)} \overline{F}(du, 0).$$

Therefore we have $\Lambda_{10}(du, v) \ll \overline{F}(du, 0)$ and $\Lambda_{10}(du, v)/\overline{F}(du, 0) \leq 1/\overline{F}(\tau)$ (*i.e.* Radon-Nykodym derivative is bounded). Furthermore we have:

$$\begin{aligned} \Lambda_{10}(u, dv) &= -\int_{(0,u]} \frac{\overline{F}(ds, dv)}{\overline{F}(s-, v)} + \int_{(0,u]} \frac{\overline{F}(ds, v)}{\overline{F}(s-, v)^2} \overline{F}(s-, dv) \\ &\leq \frac{1}{\overline{F}(\tau)} \overline{F}(0, dv) + \frac{1}{\overline{F}(\tau)^2} \overline{F}(0, dv). \end{aligned}$$

Therefore we also have $\Lambda_{10}(du, v) \ll \overline{F}(0, dv)$ and $\Lambda_{10}(u, dv)/\overline{F}(du, 0) \leq 1/\overline{F}(\tau) + 1/\overline{F}(\tau)^2$. This proves conditions 2-4 for Λ_{10} by setting $\mu_1 = \overline{F}_1$ and $\mu_2 = \overline{F}_2$ (the marginals of $\overline{F}$). Notice now that $\Lambda_{10}^n(t) = -\int \overline{F}_n(du, t_2)/\overline{F}_n(u-, t_2)$ for certain random survivor function $\overline{F}_n$ which converges a.s. in model 1 and 2 to $\overline{F}$ and in model 3 to $\overline{FG}$. So in the proof of Assumptions 2-4 we just replace F by F_n in order to obtain bounds $1/\overline{F_n}(\tau)$ and $1/\overline{F_n}(\tau) + 1/\overline{F_n}(\tau)^2$ for the Radon-Nykodym derivatives. By the almost sure convergence of F_n, these bounds are bounded uniformly in n (if n large enough) a.s. Finally, we have $\Lambda_{10}^{n\#}(t) = -\int \overline{F}_n^{\#}(du, t_2)/\overline{F}_n^{\#}(u-, t_2)$ for certain random survivor function $\overline{F}_n^{\#}$ which converges a.s. in model 1 and 2 to $\overline{F}$ and in model 3 to $\overline{FG}$. Therefore the same proof works. This completes the verification of the Assumption 2-4.

We can now apply Theorem 3.3 and thereby we can apply the functional delta-method Theorem 3.4 and Corollary 5.2 in Wellner (1993). This proves the weak convergence and bootstrap results of the theorem.

The consistency follows from the continuity of the representations Γ, Θ, Φ in V_i and the almost sure consistency of V_{in} to V_i in supnorm and on its turn the consistency of V_{in} follows from the continuity of the representations V_i in the empirical processes plugged in and the almost sure consistency of these empirical processes (Glivenko-Cantelli). □

Remark. – So far we did not write down the influence curves (derivatives) $d\Gamma(V_i) \circ dV_i(F)(X^i)$, $d\Theta(V_i) \circ dV_i(F)(X^i)$ and $d\Phi(V_i) \circ dV_i(F)(X^i)$ of the estimators because these formulas are large and not necessary for this work. The variance of these influences curves equal the variance of the limiting distributions of the estimators. Therefore, the influence curves become useful if one wants to estimate the variance of the limiting distribution or in any other efficiency analysis. Below we will write down the proof of efficiency of the Dabrowska and Prentice-Cai estimator in case of independence, and thereby also give an illustration of how an influence curve can be fairly easily obtained.

6. INFLUENCE CURVES

If an estimator is a compactly differentiable function of the empirical distribution of an i.i.d. sample, then it is asymptotically linear by application of the functional delta-method (Gill, 1989); one can write

$$\Theta_n = \Theta + \frac{1}{n}\sum_{i=1}^{n} IC_\Theta(X_i) + o_P(n^{-\frac{1}{2}}),$$

where $IC_\Theta(X_i)$, called the influence curve at the point X_i is the derivative of the function in question applied to the centred empirical process, at sample size 1, based on the single observation X_i. This follows from linearity of the derivative and the fact that an empirical distribution function is a sample average. One has $E_\Theta(IC_\Theta(X_i)) = 0$, while $\mathrm{Var}(IC_\Theta(X_i))$ is the asymptotic variance of $\sqrt{n}(\Theta_n - \Theta)$. So it is not surprising that the influence curve plays an important role in efficiency and robustness studies.

Restrict attention to model 3 (the bivariate random censorship model). We discuss here computation of the influence curves of our three estimators

$\overline{F}_n^D(t)$, $\overline{F}_n^V(t)$, $\overline{F}_n^{PC}(t)$, for given t, as function of a bivariate censored observation (X, Y, δ, ϵ). The form of the influence curve also depends on the point at which we make the calculations, *i.e.* on the assumed "true" values of F and G.

In principle, using the chain rule, one can write down formulas by applying the derivative of each composing mapping in turn. The resulting formulas are very large and not very illuminating. The procedure can be speeded up by noting the following algorithm for computing the derivative of our mappings, applied to any function: consider integrals and product-integrals as ordinary sums and products, consider differentials dF, dh etc. as ordinary variables indexed by (e.g.) t; apply the usual rules of algebra, and then convert back to a proper mathematical expresion by replacing sums and products involving differentials by the 'obvious' integrals or product integrals. This also applies to the Peano series since it is an infinite sum of multiple integrals.

The above statement is trivially true if the distributions involved are discrete. By approximating the continuous distributions by discrete distributions and using that the algorithm is correct for discrete distributions, the result for continuous distributions follows straightforwardly from appropriate continuity of the compact derivative in the sense that $IC_{\Theta_n}(X) \to IC_{\Theta}(X)$ for sequences $\Theta_n \to \Theta$. So the whole idea which makes this algorithm work is that by appropriate continuity of the derivative one can determine the derivative at a general point from the derivative at a discrete approximation and the derivative at a discrete approximation is obtained by applying the usual rules of algebra (*i.e.* the algorithm is then trivially correct). It is proved for the Dabrowska representation in van der Laan, 1990. We will not prove it here.

We will compute the influence curve by direct formal algebraic manipulation of the representations of the estimators. We will use the chain rule in the sense that we will decompose the calculation in two steps: from the empirical distributions to the empirical hazards, and from the empirical hazards to the survivor functions.

Also we will only compute the influence curve at a special point where much significance occurs: namely F is continuous, $F = F_1 F_2$, and $G = G_1 G_2$. We call this "complete independence" (of all survival and all censoring variables), and continuity of survival. The simplification caused by independence of the survival variables is obvious. Continuity of survival means that all unpleasant terms like $1/(1 - \Delta\Lambda)$, both arising as derivatives and as part of the representations themselves (the β function in the Dabrowska and Prentice-Cai representations) disappear completely. Also terms arising from the derivatives of β disappear, a more subtle point (this is shown by using the $d - \Delta$-interchange lemma and that by continuity the underlying hazards have no jumps), but fortunately true. Finally independence of censoring makes the probabilistic structure of the influence curves easier still and also allows optimality calculations (computation of the efficient influence curve) to be done explicitly.

The finding will be: at complete independence the Dabrowska and the Prentice-Cai estimators are efficient (*see* e.g. Bickel *et al.*, 1993, for efficiency theory). We prove this "at continuity" and conjecture it is also true without this restriction. The Volterra estimator is not efficient at this point. The result means that the Dabrowska and Prentice-Cai are almost equivalent and close to efficient while the Volterra estimator is much inferior. This finding has been supported by extensive simulations (Bakker, Prentice-Cai, Pruitt).

6.1. Computation of the influence curves

We do not go through the computation in detail but suffice with the remark that each step is made rigorous by application of our differentiability results for all mappings which occur. Since we are going to suppress s, t etc. a different notation is more convenient. We replace n by $\widehat{\ }$ and use $1, 2$ to indicate functions only depending on the first or second variable. In particular we use:

$$\Lambda, \quad \Lambda_1, \quad \Lambda_2, \quad \Lambda_{1\backslash 2} \quad \text{and} \quad \Lambda_{2\backslash 1}$$

instead of

$$\Lambda_{11}(\cdot,\cdot), \quad \Lambda_{10}(\cdot,0), \quad \Lambda_{01}(0,\cdot), \quad \Lambda_{10}(\cdot,\cdot) \quad \text{and} \quad \Lambda_{01}(\cdot,\cdot).$$

The influence curves for $\widehat{\Lambda}, \widehat{\Lambda}_1$ etc. are very simple and we denote them as follows:

$$\begin{aligned} d\widehat{\Lambda} - d\Lambda &\approx \frac{dM}{y} \\ d\widehat{\Lambda}_i - d\Lambda_i &\approx \frac{dM_i}{y_i} \quad i = 1 \text{ or } 2 \\ d\widehat{\Lambda}_{i\backslash j} - d\Lambda_{i\backslash j} &\approx \frac{dM_{i\backslash j}}{y} \quad i, j = 1, 2 \text{ or } 2, 1. \end{aligned}$$

Here for one bivariate censored observation (X, Y, δ, ϵ),

$$\begin{aligned}
M(s,t) &= I\{X \le s, Y \le t, \delta = 1, \epsilon = 1\} \\
&\quad - \int_0^s \int_0^t I\{X \ge u, Y \ge v\}\Lambda(du, dv) \\
M_1(s) &= I\{X \le s, \delta = 1\} - \int_0^s I\{X \ge u\}\Lambda_1(du) \\
M_2(t) &= I\{Y \le t, \epsilon = 1\} - \int_0^t I\{Y \ge v\}\Lambda_2(dv) \\
M_{1\backslash 2}(s,t) &= I\{X \le s, Y \ge t, \delta = 1\} - \int_0^s I\{X \ge u, Y \ge t\}\Lambda_{1\backslash 2}(ds, t) \\
y(s,t) &= P(X \ge s, Y \ge t) \\
y_1(s) &= P(X \ge s).
\end{aligned}$$

Using $\prod(1 + dL)$ for the product integral of L and $P_{[0,t]}(L)$ for the Peano series $P([0,t] : L)$ of L, we note that

$$\prod(1 + d\widehat{L}) - \prod(1 + dL) \approx \prod(1 + dL) \int \frac{d\widehat{L} - dL}{(1 + \Delta L)} \tag{38}$$

$$P_{[0,t]}(\widehat{L}) - P_{[0,t]}(L) \approx \int P_{[0,s)}(L)(\widehat{L} - L)(ds) P_{(s,t]}(L). \tag{39}$$

Our three representations are:

$$D: \prod_{[0,t_1]}(1 - d\Lambda_1) \prod_{[0,t_2]}(1 - d\Lambda_2) \prod_{[0,t]}\left(1 + \frac{d\Lambda - d\Lambda_{1\backslash 2} d\Lambda_{2\backslash 1}}{(1 - \Delta\Lambda_{1\backslash 2})(1 - \Delta\Lambda_{2\backslash 1})}\right)$$

$$\begin{aligned}
PC:\ & P_{[0,t_1]}(\Lambda_1) P_{[0,t_2]}(\Lambda_2) P_{[0,t]} \\
& \times \left(\int\!\!\int \frac{d\Lambda - d\Lambda_{1\backslash 2} d\Lambda_2 - d\Lambda_{2\backslash 1} d\Lambda_1 + d\Lambda_1 d\Lambda_2}{(1 - \Delta\Lambda_1)(1 - \Delta\Lambda_2)}\right) \\
V:\ & (P_{[0,t_1]}(\Lambda_1) + P_{[0,t_2]}(\Lambda_2) - 1) \\
& - \int \big(P_{[0,s_1]}(\Lambda_1) + P_{[0,s_2]}(\Lambda_2) - 1\big) P_{(s,t]}\Lambda(ds),
\end{aligned}$$

where we set $-d_2 P_{(s,t]}(\Lambda) = P_{(s,t]}\Lambda(ds)$. This last equality follows trivially from the Kolmogorov equations. In the one-dimensional case $P_{[0,t_i]}(\Lambda_i) = \prod_{[0,t_i]}(1 + d\Lambda_i)$, $i = 1, 2$.

This gives us then, by inspection [just notice that the denominator of L and $\widetilde{L}$ do not contribute to the influence curve by the $d - \triangle$ interchange lemma and noting that $f(\triangle s) = 0$ if f is continuous]

$$D: \overline{F}\left\{-\int \frac{dM_1}{y_1} - \int \frac{dM_2}{y_2} + \int\int \left(\frac{dM - dM_{1\backslash 2} d\Lambda_{2\backslash 1} - dM_{2\backslash 1} d\Lambda_{1\backslash 2}}{y}\right)\right\}$$

and for Prentice-Cai we have

$$PC: \overline{F}\left\{-\int \frac{dM_1}{y_1} - \int \frac{dM_2}{y_2}\right\} + \overline{F}_1\overline{F}_2\left\{\int P_{[0,s)}(L)(d\widehat{L} - d\widetilde{L})(ds)P_{(s,t]}(L)\right\}$$

where

$$(d\widehat{L} - d\widetilde{L}) \approx \left(\frac{dM}{y} - d\Lambda_{1\backslash 2}\frac{dM_2}{y_2} - dM_{1\backslash 2}\frac{d\Lambda_2}{y} - d\Lambda_{2\backslash 1}\frac{dM_1}{y_1} - dM_{2\backslash 1}\frac{d\Lambda_1}{y} + \frac{dM_1}{y_1}d\Lambda_2 + \frac{dM_2}{y_2}d\Lambda_1\right).$$

Finally, we have by using (38)

$$\begin{aligned} V: &\left(-\overline{F}_1\int \frac{dM_1}{y_1} - \overline{F}_2\int \frac{dM_2}{y_2}\right) \\ &+ \int \left(\overline{F}_1\int \frac{dM_1}{y_1} + \overline{F}_2\int \frac{dM_2}{y_2}\right)_- P(\Lambda)d\Lambda \\ &- \int_0^t \Psi(s)\frac{dM(s)}{y(s)}P_{(s,t]}(\Lambda) \\ &+ \int_0^t \Psi(s)\Lambda(ds)\int_s^t P_{(s,u)}(\Lambda)\frac{dM(u)}{y(u)}P_{(u,t]}(\Lambda). \end{aligned}$$

For convenience we denote $P_{(u,t]}(\Lambda)$ with $P(u,t]$. Let's first simplify the last two terms of the Volterra influence curve. We have by Fubini and the Kolmogorov equation $\overline{F}(t) = \Psi(t) + \int_0^t \Psi(s)\Lambda(ds)P(s,t)$:

$$\begin{aligned} &\int_0^t \Psi(s)\Lambda(ds)\int_s^t P_{(s,u)}(\Lambda)\frac{dM(u)}{y(u)}P_{(u,t]}(\Lambda) \\ &= \int_0^t \left(\int_0^u \Psi(s)\Lambda(ds)P(s,u)\right)\frac{dM(u)}{y(u)}P(u,t] \\ &= \int_0^t (\overline{F}(u) - \Psi(u))\frac{dM(u)}{y(u)}P(u,t]. \end{aligned}$$

Therefore we have the following simple form of the influence curve of the Volterra estimator:

$$\begin{aligned} V: & \left(-\overline{F}_1 \int \frac{dM_1}{y_1} - \overline{F}_2 \int \frac{dM_2}{y_2}\right) \\ & + \int \left(\overline{F}_1 \int \frac{dM_1}{y_1} + \overline{F}_2 \int \frac{dM_2}{y_2}\right)_- P(\Lambda) d\Lambda \\ & - \int_0^t \overline{F}(s) \frac{dM(s)}{y(s)} P(s,t]. \end{aligned}$$

Next, simplification arises on assuming independence in F and G. Then $\Lambda_{1\backslash 2} = \Lambda_1$, $\Lambda = \Lambda_1 \Lambda_2$, $y = y_1 y_2$, $P(\widetilde{L}) = 1$ ($\widetilde{L} = 0$) and $L = 0$. The influence curve of Volterra does not simplify much under independence and therefore we will not proceed with writing out the Volterra influence curve under independence.

$$\begin{aligned} D, PC: \ & \overline{F}\Big\{ - \int \frac{dM_1}{y_1} - \int \frac{dM_2}{y_2} \\ & + \int\int \left(\frac{dM - dM_{1\backslash 2} d\Lambda_2 - dM_{2\backslash 1} d\Lambda_1}{y_1 y_2} \right) \Big\} \end{aligned}$$

and notice that by cancellation of terms PC simplifies to exactly the same influence curve as Dabrowska's! Now, let dN_1, dN_2, Y_1, Y_2 be defined by $dM = dN - Y d\Lambda$, $dM_{1\backslash 2} = dN_1 Y_2 - Y_1 Y_2 d\Lambda_1$ etc. Then we obtain for Dabrowska and Prentice-Cai

$$\begin{aligned} & \overline{F}\Big\{ - \int \frac{dM_1}{y_1} - \int \frac{dM_2}{y_2} \\ & + \int\int \left(\frac{\left\{ \begin{array}{l} dN_1 dN_2 - Y_1 Y_2 d\Lambda_1 d\Lambda_2 - dN_1 Y_2 d\Lambda_2 \\ \quad + Y_1 Y_2 d\Lambda_1 d\Lambda_2 - dN_2 Y_1 d\Lambda_1 + Y_1 Y_2 d\Lambda_1 d\Lambda_2 \end{array} \right\}}{y_1 y_2} \right) \Big\} \\ & = \overline{F}\Big\{ - \int \frac{dM_1}{y_1} - \int \frac{dM_2}{y_2} + \int \frac{dM_1}{y_1} \int \frac{dM_2}{y_2} \Big\}. \end{aligned}$$

We will now show that this is also the optimal influence curve.

6.2. Optimal influence curve under complete independence

Denote the bivariate censored data with V: so $V = (X, Y, \delta, \epsilon)$. The score operator for $\overline{F}$ is given by:

$$\dot{l}: L^2(F) \to L^2(P_{F,G}) : \dot{l}(h)(V) = E_F(h(S,T) \mid V).$$

This follows from the general formula for the score operator in missing data models (*see* Bickel *et al.*, 1993, section 6.4). Then the information operator is given by

$$\dot{l}^\top \dot{l} : L^2(F) \rightarrow L^2(F) : \dot{l}^\top \dot{l}(h)(S,T) = E_{P_{F,G}}(E_F(h(S,T) \mid V) \mid (S,T)).$$

Define $\kappa_t \equiv I_{(t,\infty)} - \overline{F}(t) \in L^2(F)$. Then the efficient influence curve for estimating $\overline{F}(t)$ is given by

$$\widetilde{l}(F,t,\cdot) = \dot{l}(\dot{l}^\top \dot{l})^{-1}(\kappa_t) \in L^2(P_{F,G}).$$

For this general formula for the efficient influence curve we refer to Bickel *et al.*, 1993, section 6.4. Assume now complete independence. Let $t = (t_1, t_2)$, $\kappa_{t_1} \equiv I_{(t_1,\infty)} - \overline{F_1}(t_1)$, $\kappa_{t_2} \equiv I_{(t_2,\infty)} - \overline{F_2}(t_2)$. Define h_1 (univariate function of S) by $\dot{l}^\top \dot{l}(h_1) = \kappa_{t_1}$ and h_2 (univariate function of T) by $\dot{l}^\top \dot{l}(h_2) = \kappa_{t_2}$. Then by complete independence [notice that $\dot{l}^\top \dot{l}(h_1 h_2) = \dot{l}^\top \dot{l}(h_1) \dot{l}^\top \dot{l}(h_2)$] we have

$$\begin{aligned} \dot{l}^\top \dot{l}(h_1 h_2 + h_1 \overline{F}_2 + h_2 \overline{F}_1) &= \dot{l}^\top \dot{l}(h_1) \dot{l}^\top \dot{l}(h_2) + \overline{F}_2 \dot{l}^\top \dot{l}(h_1) + \overline{F}_1 \dot{l}^\top \dot{l}(h_2) \\ &= \kappa_{t_1} \kappa_{t_2} + \overline{F}_2 \kappa_{t_1} + \overline{F}_1 \kappa_{t_2} \\ &= \kappa_t. \end{aligned}$$

So under complete independence we have:

$$\widetilde{l}(F,t,\cdot) = \dot{l}(h_1 h_2 + h_1 \overline{F}_2 + h_2 \overline{F}_1).$$

Again, by complete independence we have $\dot{l}(h_1 h_2) = \dot{l}_1(h_1) \dot{l}_2(h_2)$ where $\dot{l}_1(h_1) = E(h_1(S) \mid (X, \delta))$ and $\dot{l}_2(h_2) = E(h_2(T) \mid (Y, \epsilon))$. $\dot{l}_1(h_1)$ is the efficient influence curve for estimating $\overline{F_1}(t)$ for the univariate censoring model where we only observe (X, δ) and we have a same statement for $\dot{l}_2(h_2)$. So $\dot{l}_i(h_i)$, $i = 1, 2$, equals the influence curve of the Kaplan-Meier estimator for estimating F_i which is given by: $IC_i \equiv -\overline{F}_i \int \frac{dM_i}{y_i}$, $i = 1, 2$. So under complete independence we have

$$\begin{aligned} \widetilde{l}(F,t,\cdot) &= IC_1 IC_2 + IC_1 \overline{F}_2 + IC_2 \overline{F}_1 \\ &= \overline{F} \left\{ \int \frac{dM_1}{y_1} \int \frac{dM_2}{y_2} - \int \frac{dM_1}{y_1} - \int \frac{dM_2}{y_2} \right\} \end{aligned} \tag{40}$$

and this is exactly the influence curve of the Dabrowska and Prentice-Cai estimator under complete independence. This proves that the Dabrowska and

Prentice-Cai estimator are efficient under complete independence. Finally notice that (40) provides us with a nice and simple formula for the variance of the efficient influence curve:

$$\mathrm{Var}(\widetilde{l}(F,t,\cdot)) = \mathrm{Var}(IC_1)\mathrm{Var}(IC_2) + \overline{F}_1^2\mathrm{Var}(IC_2) + \overline{F}_2^2\mathrm{Var}(IC_1).$$

For example, in the case that T_1, T_2, C_1, C_2 are all four independent and uniform $(0,1)$, the reader can easily verify that this variance equals:

$$\frac{1}{4}(1 + (1-t_1)^2 + (1-t_2)^2 - 3(1-t_1)^2(1-t_2)^2).$$

Computer simulations for the Prentice-Cai and Dabrowska estimator show that this limiting variance is already closely approximated for $n = 100$ (*see* Bakker, 1990, Prentice-Cai, 1992*a*, 1992*b*).

ACKNOWLEDGMENTS

This paper began with Peter Bickel's discovery of the estimator based on the Volterra representation in Seattle in August, 1986, and was originally conceived as a project involving Bickel, Gill, and Wellner. All the present authors owe thanks to Peter for his interest in this research over an extended period, and for a number of valuable suggestions.

REFERENCES

P. K. Andersen, Ø. Borgan, R. D. Gill and N. Keiding, *Statistical Models Based on Counting Processes*, Springer-Verlag, New York, 1992.

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

P. J. Bickel, Y. Ritov and J. A. Wellner, Efficient estimation of linear functionals of a probability measure P with known marginal distributions, *Ann. Statist.*, Vol. **19**, 1991, pp. 1316-1346.

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

M. D. Burke, Estimation of a bivariate survival function under random censorship, *Biometrika*, Vol. **75**, 1988, pp. 379-382.

D. M. Dabrowska, Kaplan Meier Estimate on the Plane, *Ann. Statist.*, Vol. **16**, 1988, pp. 1475-1489.

D. M. Dabrowska, Kaplan Meier Estimate on the Plane: Weak Convergence, LIL, and the Bootstrap, *J. Multivar. Anal.*, Vol. **29**, 1989, pp. 308-325.

R. M. Dudley, Weak convergence of probabilities on nonseparable metric spaces, and empirical measures on Euclidean spaces, *Illinois J. Math.*, Vol. **10**, 1966, pp. 109-126.

R. M. Dudley, Fréchet differentiability, p-variation and universal Donsker classes. *Ann. Prob.* Vol. **20**, 1991, pp. 1968-1983.

R. M. DUDLEY, *Empirical processes, p-variation for* $p \leq 2$ *and the quantile and* $\int FdG$ *operator*, unpublished manuscript 1992.

R. D. GILL, Non- and Semi-parametric Maximum Likelihood Estimators and the von Mises Method (Part 1), *Scand. J. Statist.*, Vol. **16**, 1989, pp. 97-128.

R. D. GILL and S. JOHANSEN, A survey of product integration with a view towards application in survival analysis, *Ann. Statist.*, Vol. **18**, 1990, pp. 1501-1555.

R. D. GILL, Multivariate survival analysis, *Theory Prob. Appl.*, Vol. **37**, 1992, pp. 18-31 and 307-328 (in Russian; English translation to appear).

E. GINÉ and J. ZINN, Bootstrapping general empirical measures, *Ann. Probability*, Vol. **18**, 1990, pp. 284-301.

J. HOFFMANN-JØRGENSEN, *Stochastic Processes on Polish Spaces*, unpublished manuscript, 1984.

L. V. KANTOROVICH and G. P. AKILOV, *Functional Analysis*, translated by Howard L. Silcock, New York, Permagon Press, 1982.

M. J. van der LAAN, *Dabrowska's multivariate product limit estimator and the delta-method*, Master's Thesis, Dept. of Math., Utrecht Univ., 1990.

M. J. van der LAAN, *Analysis of Pruitt's estimator of the bivariate survival function*, preprint No. 648, Dept of Math., Utrecht Univ., 1991.

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

M. LOÈVE, *Probability Theory*, van Nostrand, New York, 1955.

G. NEUHAUS, On weak convergence of stochastic processes with multidimensional time parameter, *Ann. Math. Statist.*, Vol. **42**, 1971, pp. 1285-1295.

D. POLLARD, *Convergence of Stochastic Processes*, Springer-Verlag, New York, 1984.

R. L. PRENTICE and J. CAI, Covariance and survivor function estimation using censored multivariate failure time data, *Biometrika*, Vol. **79**, 1992*a*, pp. 495-512.

R. L. PRENTICE and J. CAI, *Marginal and conditional models for the analysis of multivariate failure time data*. Klein J. P. and Goel P. K., Eds., *Survival Analysis State of the Art*, Kluwer, Dordrecht, 1992*b*.

R. C. PRUITT, On negative mass assigned by the bivariate Kaplan-Meier estimator, *Ann. Stat.*, Vol. **19**, 1991*a*, pp. 443-453.

R. C. PRUITT, *Small sample comparisons of five bivariate survival curve estimators*, Technical Report No. 559, University of Minnesota, 1991*a*.

R. C. PRUITT, *Strong consistency of self-consistent estimators: general theory and an application to bivariate survival analysis*, Technical Report No. 543, University of Minnesota, 1991*b*.

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

A. SHEEHY and J. A. WELLNER, *Uniformity in P of some limit theorems for empirical measures and processes*, Technical Report No. 134, Dept. of Statistics, Univ. of Washington, 1988.

A. SHEEHY and J. A. WELLNER, Uniform Donsker Classes of Functions, *Ann. Probability*, Vol. **20**, 1992, pp. 1983-2030.

G. R. SHORACK and J. A. WELLNER, *Empirical Processes*. Wiley, New York, 1986.

A. W. VAN der VAART and J. A. WELLNER, *Prohorov and continuous mapping theorems in the Hoffmann-Jørgensen weak convergence theory, with application to convolution and asymptotic minimax theorems*, Tech. Report No. 157. Dept. of Statistics, University of Washington, Seattle, 1989.

A. W. VAN der VAART and J. A. WELLNER, *Weak Convergence and Empirical Processes*, Springer Verlag, New York, to appear 1995.

J. A. WELLNER, Covariance formulas via marginal martingales, *Statistica Neerlandica*, Vol. **48**, 1994.

(Manuscript received February 22, 1993;
revised version received February 28, 1994.)