Technical Report No. 3/07, October 2007

A BI-VARIATE KAPLAN-MEIER ESTIMATOR VIA AN INTEGRAL EQUATION

Arusharka Sen and Winfried Stute

TECHNICAL REPORT # 03/07, OCTOBER 2007
DEPARTMENT OF MATHEMATICS AND STATISTICS, CONCORDIA UNIVERSITY

A bi-variate Kaplan-Meier estimator via an integral equation

Arusharka Sen, *Department of Mathematics and Statistics, Concordia University, Montreal, Canada*

Winfried Stute, *Mathematisches Institut, Justus-Liebig University, Giessen, Germany*

1. Introduction

Let $(X_{i1}, X_{2i}),\ 1 \leq i \leq n$, be independent and identically distributed (iid) nonnegative random vectors, each having a bi-variate distribution function (d.f.) $F(\cdot,\cdot)$ and representing a bi-variate failure or survival time, such as those for 'twins' or two components of the same machine. Suppose further that these vectors are subject to *random censoring from the right* by another, independent set of iid random vectors $(Y_{i1}, Y_{2i}),\ 1 \leq i \leq n$, each having d.f. $G(\cdot,\cdot)$, so that we can only observe $(\delta_{1i}, \delta_{2i}, Z_{i1}, Z_{2i}),\ 1 \leq i \leq n$, where $\delta_{ji} = \mathbf{1}\{X_{ji} \leq Y_{ji}\},\ Z_{ji} = \min(X_{ji}, Y_{ji}),\ j = 1, 2,\ 1 \leq i \leq n$, and $\mathbf{1}(A)$ denotes the indicator function of the event A. Our goal is to estimate $F(\cdot,\cdot)$ or equivalently, its survivor function $\bar{F}(\cdot,\cdot) := P\{X_1 > \cdot, X_2 > \cdot\}$ based on the observed data, i.e., to obtain *the* bi-variate version of the celebrated Kaplan-Meier estimator.

This problem has a surprisingly long history; see Prentice *et al* (2004), Gill *et al* (1995) for more references. Notable among the estimators are those derived by Dabrowska (1988) and van der Laan (1995). Note that the former can assign negative values to some events whereas the latter is inexplicit, although asymptotically efficient under some strong conditions such as complete observation of the censoring variables. Gill *et al* (1995) show that three estimators proposed in the literature, including that of Dabrowska (1988), are efficient under complete independence (of all the X's and Y's) and continuity.

In Section 2 of this paper we show that a bi-variate (in fact, multivariate) survivor function is an *eigenfunction* corresponding to *eigenvalue* unity of the cumulative hazard function (looked upon as an integral operator). The estimator is then obtained in Section 3 as a solution to the empirical version of the eigenfunction equation, which is a matrix eigenvector problem. Section 4 gives the general solution to the eigenfunction problem. In Section 5, the estimator is linearized by the *functional* Δ*-method*, and its *influence function* is shown to be asymptotically efficient.

It must be mentioned that one version of the empirical matrix eigenvector problem (Eq.(2.1), Section 2) was also obtained by Prentice *et al* (2004). However, their solution was incorrect.

2. The integral equation

Let $\bar{F}(x_1, \ldots, x_m) = P\{X_1 > x_1, \ldots, X_m > x_m\}$ be the survival function of an m-dimensional random vector $X = (X_1, \ldots, X_m),\ m \geq 1$. Then $\bar{F}(\cdot, \ldots, \cdot)$ satisfies the integral equation

$$\bar{F}(x_1-, \ldots, x_m-) = \int_{[x_1,\infty)\times\cdots\times[x_m,\infty)} \bar{F}(t_1-, \ldots, t_m-) \frac{dF(t_1, \ldots, t_m)}{\bar{F}(t_1-, \ldots, t_m-)} \tag{2.1}$$

Let us look at $m = 2$ only for the sake of convenience. Now for censored data, we have

$$\frac{dF(t_1, t_2)}{\bar{F}(t_1-, t_2-)} = \frac{\bar{G}(t_1-, t_2-)dF(t_1, t_2)}{\bar{G}(t_1-, t_2-)\bar{F}(t_1-, t_2-)},$$

where $G(\cdot, \cdot)$ is the censoring distribution. Thus Eq.(2.1) becomes

$$\bar{F}(x_1-, x_2-) = \int_{[x_1,\infty)\times[x_2,\infty)} \bar{F}(t_1-, t_2-) \frac{dH^{11}(t_1, t_2)}{\bar{H}(t_1-, t_2-)}, \tag{2.2}$$

and $F(\cdot, \cdot)$ can be estimated as a *solution* to the empirical version of Eq.(2.2):

$$\bar{F}_n(x_1-, x_2-) = \int_{[x_1,\infty)\times[x_2,\infty)} \bar{F}_n(t_1-, t_2-) \frac{dH_n^{11}(t_1, t_2)}{\bar{H}_n(t_1-, t_2-)}, \tag{2.3}$$

where as usual, $H_n^{11}(t_1, t_2) = n^{-1} \sum_{i=1}^n \delta_{1i}\delta_{2i}\mathbf{1}\{Z_{1i} \leq t_1, Z_{2i} \leq t_2\}$, and $\bar{H}_n(t_1, t_2) = n^{-1} \sum_{i=1}^n \mathbf{1}\{Z_{1i} > t_1, Z_{2i} > t_2\}$.

Equations (2.1) and (2.3) obviously represent *eigenvalue* problems, i.e., $\bar{F}(x_1-, x_2-)$ and $\bar{F}_n(x_1-, x_2-)$ are *eigenvectors* corresponding to the eigenvalue 1 for the integral operators $\int_{[\cdot,\infty)\times[\cdot,\infty)}(dF(t_1, t_2)/\bar{F}(t_1-, t_2-))$ and $\int_{[\cdot,\infty)\times[\cdot,\infty)}(dH_n^{11}(t_1, t_2)/\bar{H}_n(t_1-, t_2-))$, respectively.

To solve Eq.(2.3), we may assume that the estimator gives mass $p_i \geq 0$ to the observation $(Z_{1i}, Z_{2i}),\ 1 \leq i \leq n$, so that

$$\bar{F}_i := \bar{F}_n(Z_{1i}-, Z_{2i}-) = \sum_{j=1}^n a_{ij}p_j,$$

where

$$a_{ij} = \begin{cases} 1 & \text{if } Z_{1j} \geq Z_{1i},\ Z_{2j} \geq Z_{2i} \\ 0 & \text{otherwise;} \end{cases}$$

Further, let $b_i := \triangle H_n^{11}(Z_{1i}, Z_{2i})/\bar{H}_n(Z_{1i}-, Z_{2i}-) = n^{-1}\delta_{1i}\delta_{2i}/\bar{H}_n(Z_{1i}-, Z_{2i}-)$. Then Eq.(2.3), with $x_1 = Z_{1i},\ x_2 = Z_{2i},\ 1 \leq i \leq n$, may be rewritten as

$$\begin{aligned} \sum_{j=1}^n a_{ij}p_j &= \sum_{k=1}^n \bar{F}_n(Z_{1k}-, Z_{2k}-)a_{ik}b_k \\ &= \sum_{k=1}^n a_{ik}b_k\left(\sum_{l=1}^n a_{kl}p_l\right) \\ &= \sum_{l=1}^n \left(\sum_{k=1}^n a_{ik}b_k a_{kl}\right) p_l. \qquad (2.4) \end{aligned}$$

Equation for $\mathbf{p} = (p_1, \ldots, p_n)$. In matrix notation, Eq.(2.4) becomes

$$\mathbf{Ap}{=}\mathbf{ABAp}, \ \sum_{i=1}^{n} p_i = 1, \tag{2.5}$$

where $\mathbf{A} = ((a_{ij})),\ \mathbf{p} = (p_1, \ldots, p_n),\ \mathbf{B} = \text{diag}\ (b_1, \ldots, b_n)$. Now order $(Z_{1i}, Z_{2i}),\ 1 \le i \le n$, in the increasing order of the first coordinate, i.e., as $(Z_{1i:n}, Z_{[2i:n]}),\ 1 \le i \le n$, where $Z_{11:n} \le \cdots \le Z_{1n:n}$ and $Z_{[2i:n]},\ 1 \le i \le n$, are the corresponding concommitants. Then, with any suitable convention for breaking ties, $\mathbf{A}$ becomes an *upper-triangular* matrix, i.e.,

$$a_{ij} = \begin{cases} 0 & \text{if } j < i \\ 1 & \text{if } j = i \\ 1 \text{ or } 0 & \text{if } j > i \end{cases}$$

Note that for univariate ordered data, $a_{ij} = 1$ for all $j \ge i$. Thus $\mathbf{A}$ now becomes invertible, and Eq.(2.5) becomes

$$\mathbf{p}{=}\mathbf{BAp}, \ \sum_{i=1}^{n} p_i = 1. \tag{2.6}$$

Remark 2.1. Equation (2.6) is also obtained, by a different heuristics and in a more complicated form, by Prentice et al (2004), as follows: they start with the motivation, *theoretical hazard* $\approx$ *empirical hazard*. This leads to the equation

$$\frac{p_i}{\sum_{j=1}^{n} a_{ij} p_j} = n^{-1}\delta_{1i}\delta_{2i}/\bar{H}_n(Z_{1i}-, Z_{2i}-) = b_i,$$

which is exactly Eq.(2.6). However, they put $a_{ij} = 1 - d_{ij}$, where $d_{ij} = \mathbf{1}\{Z_{1j} < Z_{1i} \text{ or } Z_{2j} < Z_{2i}\}$, and rewrite the equation as $\mathbf{(I{+}BD)p{=}b}$, where $\mathbf{b} = (b_1, \ldots, b_n)$. Further restriction to only the non-zero components of $\mathbf{b}$, say $\hat{\mathbf{b}} = (b_1, \ldots, b_s)$, and the corresponding $\hat{\mathbf{p}} = (p_1, \ldots, p_s)$ gives an equation of the form

$$\hat{\mathbf{A}}\hat{\mathbf{p}} = \mathbf{1},$$

where $\mathbf{1} = (1, \ldots, 1)$. From this they conclude that there always is a solution for $\mathbf{p}$ in Eq.(2.6), and the latter is unique. Section 2 below shows very clearly that *neither* conclusion is true.

Equation for $\bar{\mathbf{F}} = (\bar{F}_1, \ldots, \bar{F}_n)$. Since obviously $\bar{\mathbf{F}} = \mathbf{Ap}$, and also

$$1 = \bar{F}_n(0-, 0-) = \int_{[0,\infty)\times[0,\infty)} \bar{F}_n(t_1-, t_2-)\frac{dH_n^{11}(t_1, t_2)}{\bar{H}_n(t_1-, t_2-)},$$

we may write Eq.(2.5) alternatively as

$$\bar{\mathbf{F}} = \mathbf{AB}\bar{\mathbf{F}}, \ \sum_{i=1}^{n} b_i \bar{F}_i = 1, \tag{2.7}$$

3. Solutions to Equations (2.6) and (2.7)

Equation (2.6). Put $\delta_{1i}\delta_{2i} = \delta_i$. Under the ordering that led to Eq.(2.6), we then have $\mathbf{BA} = ((b_i a_{ij}))$, where

$$b_i a_{ij} = \frac{\delta_i a_{ij}}{n\bar{H}_n(Z_{1i}-, Z_{2i}-)} = \frac{\delta_i a_{ij}}{\sum_{k=1}^n \mathbf{1}\{Z_{1k} \geq Z_{1i}, Z_{2k} \geq Z_{2i}\}} = \frac{\delta_i a_{ij}}{\sum_{k=i}^n a_{ik}}$$

In particular, $\mathbf{BA}$ too is upper-triangular with diagonal elements ($b_i = \delta_i / \sum_{k=i}^n a_{ik},\ 1 \leq i \leq n$), which are of course the *eigenvalues* of $\mathbf{BA}$. Hence there are *three* possibilities with Eq.(2.6):

1) $b_{i_0} = 1$ FOR SOME i_0, $b_i < 1$ FOR $i \neq i_0$: UNIQUE SOLUTION.

   This is equivalent to $\delta_i = 1,\ a_{ik} = 0,\ k > i$, *uniquely* for $i = i_0$. In this case Eq.(2.6) has only one *linearly independent* solution $\mathbf{p}$ (i.e., the eigen-subspace of $\mathbf{BA}$ for eigenvalue 1 has dimension 1), hence a unique solution with $\sum_{i=1}^n p_i = 1$. This solution is obtained as follows: without loss of generality (WLOG), assume $i_0 = n$, i.e., $b_n = \delta_n = 1$. (If $i_0 < n$, interchange Row-i_0 and Row-n in $\mathbf{BA}$.)

   Then we have, for $1 \leq i \leq n-1$,

$$p_i = b_i \sum_{j=i}^n a_{ij} p_j,$$

   which gives

$$\begin{aligned} p_i &= \frac{b_i}{1-b_i} \sum_{j=i+1}^n a_{ij} p_j \\ &= c_i \Big[1 + \sum_{j>i} a_{ij} c_j + \sum_{j>i} \sum_{k>i} a_{ij} a_{jk} c_j c_k + \cdots \\ &\quad + a_{i,i+1} a_{i+1,i+2} \ldots a_{n-2,n-1} c_{i+1} \ldots c_{n-1} \Big] p_n, \qquad (3.1) \end{aligned}$$

   where $c_i = b_i/(1-b_i) = \delta_i/(n\bar{H}_n(Z_{1i}-, Z_{2i}-) - \delta_i)$, $1 \leq i \leq n-1$, and

$$\begin{aligned} p_n &= 1 \Big/ \Big[1 + \sum_{i=1}^{n-1} c_i + \sum_{i<j} c_i c_j a_{ij} + \sum_{i<j<k} c_i c_j c_k a_{ij} a_{jk} + \cdots \\ &\quad + c_1 \ldots c_{n-1} a_{12} a_{23} \ldots a_{n-2,n-1} \Big] \qquad (3.2) \end{aligned}$$

   Note that we got Eq.(3.2) using the condition $\sum_{i=1}^n p_i = 1$.

   In the univariate case, $a_{ij} = 1$ for $j \geq i$ so that $b_i = \delta_i/(n-i+1)$, and

$$\begin{aligned} p_n &= 1 / \textstyle\prod_{j=1}^{n-1} [1 + c_j] &= \textstyle\prod_{j=1}^{n-1} [1 - b_j], \\ p_i &= c_i p_n \textstyle\prod_{j=i+1}^{n-1} [1 + c_j] &= b_i \textstyle\prod_{j=1}^{i-1} [1 - b_j],\ 1 \leq i \leq n-1, \end{aligned}$$

   which is exactly the Kaplan-Meier estimator.

2) $b_i = 1$ FOR $i \in I$, WHERE $I = \{i_1, \ldots, i_r\}$, $r \geq 2$, AND $b_i < 1$ FOR $i \notin I$: MULTIPLE SOLUTIONS.

In this case we do not have a unique solution to Eq.(2.6), satisfying $\sum_{i=1}^n p_i = 1$, but $r \geq 2$ linearly independent solutions. We can, however, enforce a unique solution by imposing the restriction

$$p_{i_1} = \cdots = p_{i_r}. \tag{3.3}$$

Assume as in Case-1 that $i_r = n$, WLOG. Note that $c_i = b_i/(1-b_i)$ is not defined for $i \in I$, but by Eq.(3.1)–(3.2) it is clear that Eq.(3.3) is ensured if we let $c_i = 1,\ i \in I$. Thus the modified solution is again given by Eq.(3.1)–(3.2), putting $c_i = 1,\ a_{ik} = 0,\ k > i$, for $i \in I$.

3) $b_i < 1$ FOR $1 \leq i \leq n$: NO SOLUTION.

In this case 1 is not an eigenvalue of $\mathbf{BA}$, and Eq.(2.6) has *no* non-zero solution. We can, however, produce a pseudo-solution that leads to a *defective* estimator, i.e., one satisfying $\sum_{i=1}^n p_i < 1$. Augment the vector $\mathbf{p} = (p_1, \ldots, p_n)$ to $\tilde{\mathbf{p}} = (p_1, \ldots, p_n, p_{n+1})$, the matrix $\mathbf{B}$ to $\tilde{\mathbf{B}} = \text{diag}\ (b_1, \ldots, b_n, b_{n+1})$ where $b_{n+1} = 1$, and the matrix $\mathbf{A}$ to $\tilde{\mathbf{A}} = ((a_{ij}))_{(n+1)\times(n+1)}$ where $a_{i,n+1} = 1$ for $1 \leq i \leq n+1$, $a_{n+1,j} = 0$ for $1 \leq j \leq n$. Then the equation

$$\tilde{\mathbf{p}} = \tilde{\mathbf{B}}\tilde{\mathbf{A}}\tilde{\mathbf{p}}$$

obviously satisfies the condition of Case-1 with n replaced by $(n+1)$. Hence its unique solution is obtained from Eq.(3.1)–(3.2) with n replaced by $(n+1)$. Note that in this solution p_{n+1} represents the excess mass and $(p_1, \ldots, p_n)$ the defective estimator.

In the univariate case defective estimator is obtained when $\delta_n = 0$, and the excess mass is given by (following Case 1)

$$\begin{aligned} p_{n+1} &= 1/\textstyle\prod_{j=1}^n [1 + c_j] \\ &= \textstyle\prod_{j=1}^n [1 - b_j] = 1 - \sum_{i=1}^n p_i, \end{aligned}$$

Equation (2.7). To avoid unnecessary repetitions, let us consider only Case-1, namely, $b_n = 1$, $b_i < 1$ for $1 \leq i \leq n-1$, which gives a unique solution. Eq.(2.7) can be written as

$$\bar{F}_i = \sum_{j=i}^n a_{ij} b_j \bar{F}_j,\ 1 \leq i \leq n,$$

which in this case leads to the solution

$$\begin{aligned} \bar{F}_i &= (1-b_i)^{-1} \sum_{j=i+1}^n a_{ij} b_j \bar{F}_j \\ &= (1-b_i)^{-1} \Bigg[1 + \sum_{j>i} a_{ij} c_j + \sum_{j>i} \sum_{k>i} a_{ij} a_{jk} c_j c_k + \cdots \\ &\quad + a_{i,i+1} a_{i+1,i+2} \ldots a_{n-2,n-1} c_{i+1} \ldots c_{n-1} \Bigg] \bar{F}_n,\ 1 \leq i \leq n-1, \end{aligned} \tag{3.4}$$

where, as before, $c_i = b_i/(1-b_i) = \delta_i/(n\bar{H}_n(Z_{1i}-, Z_{2i}-) - \delta_i)$, $1 \le i \le n-1$, and

$$\begin{aligned}\bar{F}_n &= 1/\left[1 + \sum_{i=1}^{n-1} c_i + \sum_{i<j} c_i c_j a_{ij} + \sum_{i<j<k} c_i c_j c_k a_{ij} a_{jk} + \cdots\right. \\ &\quad \left. + c_1 \ldots c_{n-1} a_{12} a_{23} \ldots a_{n-2,n-1}\right] \qquad (3.5)\end{aligned}$$

Note that we got Eq.(3.5) using the condition $\sum_{i=1}^{n} b_i \bar{F}_i = 1$.

4. General Solution to the Eigenvalue Equation (2.1)

Let us consider the general eigenvalue problem in Eq.(2.1) above:

$$q(x) = \int \mathbf{1}\{t \ge x\} q(t) \frac{dF(t)}{\bar{F}(t-)}, \tag{4.1}$$

with the initial condition $q(0) = 1$. Here $t = (t_1, t_2)$ and $x = (x_1, x_2)$ represent vector variables, and the inequality is defined in the coordinate-wise sense.

Uniqueness of the solution. Since $\bar{F}(x-)$ is always a solution to Eq.(4.1), it is clear that we have a unique solution, under $q(0) = 1$, if every solution is of the form $c\bar{F}(x-)$, i.e., the eigen-subspace corresponding to the eigenvalue 1 is of dimension 1.

However, it is not hard to see that uniqueness fails if there are more than one point x satisfying $0 < P\{X = x\} = P\{X \ge x\}$, because for every such point Eq.(4.1) is of the form: $q(x) = q(x)$. Note that in 1-dimension there *cannot be* more than one such point. Hence we have the following result:

THEOREM 4.1. Denote $\Delta_F(x) = P\{X = x\}$, and suppose that the set,

$$N(F) := \{x | 0 < \Delta_F(x) = P\{X \ge x\}\},$$

contains *at most one* point. Then every solution to Eq.(4.1) is of the form $q(x) = c\bar{F}(x-)$.

PROOF: Put $r(x) = q(x)/\bar{F}(x-)$, then Eq.(4.1) can be written as

$$r(x) = \frac{\int \mathbf{1}\{t \ge x\} r(t) dF(t)}{\bar{F}(x-)}.$$

First assume $r(x)$ is a simple function, of the form $r(x) = \sum_{j=1}^{k} c_j \mathbf{1}_{A_j}(x)$, where $\cup_j A_j = \mathbb{R}^2$, $A_j \cap A_{j'} = \emptyset$ for $j \ne j'$. Then the equation further reduces to: $c_l = \sum_{j=1}^{k} c_j p_j(x)$ for $x \in A_l$, where $p_j(x) = P\{X \in A_j \cap [x, \infty)\}/\bar{F}(x-)$, so that $\sum_{j=1}^{k} p_j(x) = 1$. Further, for all $x \notin N(F)$, $p_j(x) > 0$ for at least two indices $1 \le j \le k$. Now we have $0 = \sum_{j=1}^{k} (c_j - c_l) p_j(x)$, hence taking $c_l = \min_j c_j$ we get $c_1 = \cdots = c_k$. The case of a general $r(x)$ follows.(??)□

Solution in 1 dimension. In 1-dimension we may write Eq.(4.1), for any $M > x$ such that $\bar{F}(M-) > 0$, as

$$\begin{aligned} \bar{F}(x-) &= \int \mathbf{1}\{M > t \geq x\} \bar{F}(t-) \frac{dF(t)}{\bar{F}(t-)} + \bar{F}(M-), \\ &= \bar{F}(M-) + \sum_{n=1}^{\infty} \int \cdots \int \mathbf{1}\{M > t_n \geq \cdots \geq t_1 \geq x\} \bar{F}(M-) \frac{dF(t_n)}{\bar{F}(t_n-)} \cdots \frac{dF(t_1)}{\bar{F}(t_1-)}, \\ &\quad \text{iterating the previous equality;} \qquad (4.2) \end{aligned}$$

now using the initial condition $\bar{F}(0-) = 1$ we get

$$\bar{F}(M-) = \left[1 + \sum_{n=1}^{\infty} \int \cdots \int \mathbf{1}\{M > t_n \geq \cdots \geq t_1 \geq 0\} \frac{dF(t_n)}{\bar{F}(t_n-)} \cdots \frac{dF(t_1)}{\bar{F}(t_1-)}\right]^{-1},$$

so that

$$\bar{F}(x-) = \frac{1 + \sum_{n=1}^{\infty} \int \cdots \int \mathbf{1}\{M > t_n \geq \cdots \geq t_1 \geq x\} \frac{dF(t_n)}{\bar{F}(t_n-)} \cdots \frac{dF(t_1)}{\bar{F}(t_1-)}}{1 + \sum_{n=1}^{\infty} \int \cdots \int \mathbf{1}\{M > t_n \geq \cdots \geq t_1 \geq 0\} \frac{dF(t_n)}{\bar{F}(t_n-)} \cdots \frac{dF(t_1)}{\bar{F}(t_1-)}} \qquad (4.3)$$

Example 4.1. If $F(x) = 1 - e^{-\lambda x},\ x \geq 0,\ \lambda > 0$, then $\frac{dF(t)}{\bar{F}(t-)} = \lambda dt$, and Eq.(4.3) reduces to, for any $M > x$,

$$\bar{F}(x) = \frac{\exp(\lambda(M-x))}{\exp(\lambda M)} = \exp(-\lambda x).$$

Solution in 2 or more dimensions. In 2 or higher dimensions Eq.(4.2) no longer holds because of the loss of the linear order in these dimensions. However, denote $M > x$ if $M_j \geq x_j$ with strict inequality for at least one $j = 1, 2$. Then for any $M > x$ with $\bar{F}(M-) > 0$,

$$\bar{F}_M(x) := P\{M > X > x\},\ \bar{F}_M^+(x) := \bar{F}_M(x) + \bar{F}(M-).$$

Then obviously $\lim_{M\to\infty} \bar{F}_M(x) = \lim_{M\to\infty} \bar{F}_M^+(x) = \bar{F}(x)$, and for $M > x$,

$$\begin{aligned} \bar{F}_M^+(x-) &= \int \mathbf{1}\{M > t \geq x\} \bar{F}_M^+(t-) \frac{dF(t)}{\bar{F}_M^+(t-)} + \bar{F}(M-) \\ &= \bar{F}(M-) \left[1 + \sum_{n=1}^{\infty} \int \cdots \int \mathbf{1}\{M > t_n \geq \cdots \geq t_1 \geq x\} \frac{dF(t_n)}{\bar{F}_M^+(t_n-)} \cdots \frac{dF(t_1)}{\bar{F}_M^+(t_1-)}\right], \\ &\quad (4.4) \end{aligned}$$

iterating the previous equality. Next, using the initial condition $\bar{F}_M^+(0-) = F(M-) + \bar{F}(M-)$ we get

$$\bar{F}_M^+(x-) = [F(M-) + \bar{F}(M-)] \frac{[1 + \sum_{n=1}^{\infty} \int \cdots \int \mathbf{1}\{M > t_n \geq \cdots \geq t_1 \geq x\} \frac{dF(t_n)}{\bar{F}_M^+(t_n-)} \cdots \frac{dF(t_1)}{\bar{F}_M^+(t_1-)}]}{1 + \sum_{n=1}^{\infty} \int \cdots \int \mathbf{1}\{M > t_n \geq \cdots \geq t_1 \geq 0\} \frac{dF(t_n)}{\bar{F}_M^+(t_n-)} \cdots \frac{dF(t_1)}{\bar{F}_M^+(t_1-)}}.$$

Hence finally,

$$\bar{F}(x-) = \lim_{M\to\infty} \bar{F}_M^+(x-) = \lim_{M\to\infty} \frac{[1+\sum_{n=1}^{\infty}\int\cdots\int \mathbf{1}\{M > t_n \geq \cdots \geq t_1 \geq x\}\frac{dF(t_n)}{\bar{F}(t_n-)}\cdots\frac{dF(t_1)}{\bar{F}(t_1-)}]}{1+\sum_{n=1}^{\infty}\int\cdots\int \mathbf{1}\{M > t_n \geq \cdots \geq t_1 \geq 0\}\frac{dF(t_n)}{\bar{F}(t_n-)}\cdots\frac{dF(t_1)}{\bar{F}(t_1-)}}. \tag{4.5}$$

EXAMPLE 4.2. Let $F(x_1, x_2) = F_1(x_1)F_2(x_2)$, where $F_j(x) = 1-e^{-\lambda_j x}$, $j = 1, 2$. Then the right-hand side of Eq.(4.5) reduces to

$$\begin{aligned}
& \lim_{M\to\infty}\left[\sum_{r=0}^{\infty}\frac{(\lambda_1(M_1-x_1))^r}{r!}\frac{(\lambda_2(M_2-x_2))^r}{r!}\right] \Big/ \left[\sum_{r=0}^{\infty}\frac{(\lambda_1 M_1)^r}{r!}\frac{(\lambda_2 M_2)^r}{r!}\right] \\
= \; & e^{-\lambda_1 x_1}e^{-\lambda_2 x_2}\lim_{M\to\infty}\left(\left[\sum_{r=0}^{\infty}e^{-\lambda_1(M_1-x_1)}e^{-\lambda_2(M_2-x_2)}\frac{(\lambda_1(M_1-x_1))^r}{r!}\frac{(\lambda_2(M_2-x_2))^r}{r!}\right] \Big/ \right. \\
& \left.\left[\sum_{r=0}^{\infty}e^{-\lambda_1 M_1}e^{-\lambda_2 M_2}\frac{(\lambda_1 M_1)^r}{r!}\frac{(\lambda_2 M_2)^r}{r!}\right]\right) \\
= \; & e^{-\lambda_1 x_1}e^{-\lambda_2 x_2} = \bar{F}(x_1, x_2),
\end{aligned}$$

since the limit in the first equality above equals one. (??)

Alternative form of solution. Consider only the 2-dimensional case. Take $M > 0$ and a partition $0 = s_{10} < s_{11} < \cdots < s_{1N_1} = M_1$, $0 = s_{20} < s_{21} < \cdots < s_{2N_2} = M_2$. Then a discretized version of Eq.(4.4) for this partition would be

$$\bar{F}_i = \sum_{j\geq i}\bar{F}_j\Delta_j,\ \sum_j \bar{F}_j\Delta_j = 1 \tag{4.6}$$

where $i = (i_1, i_2)$, $j = (j_1, j_2)$, $1 \leq i_l, j_l \leq N_l$, $l = 1, 2$, $\bar{F}_i = \bar{F}(s_{1i_1}, s_{2i_2})$, and

$$\Delta_j = \int_{[s_{1j_1}, s_{1,j_1+1})\times[s_{2j_2}, s_{2,j_2+1})} dF(t)/\bar{F}(t-).$$

Now Eq.(4.6) can be solved as in Eq.(3.4)–(3.5) to get

$$\bar{F}_i = \frac{(1-\Delta_i)^{-1}\left[1+\sum_{r\geq 1}\mathbf{1}\{N > j_r > \cdots > j_1 > i\}\frac{\Delta_{j_r}}{1-\Delta_{j_r}}\cdots\frac{\Delta_{j_1}}{1-\Delta_{j_1}}\right]}{1+\sum_{r\geq 1}\mathbf{1}\{N > j_r > \cdots > j_1\}\frac{\Delta_{j_r}}{1-\Delta_{j_r}}\cdots\frac{\Delta_{j_1}}{1-\Delta_{j_1}}}; \tag{4.7}$$

Letting the partition-size go to ∞ we get the following expression from Eq.(4.7) (remember that $a > b$ if $a_j \geq b_j$ with *strict* inequality for at least one $j = 1, 2$):

$$\bar{F}(x-) = \lim_{M\to\infty}\frac{(1-\Delta_F(x))^{-1}\left[1+\sum_{n=1}^{\infty}\int\cdots\int\mathbf{1}\{M > t_n > \cdots > t_1 > x\}\frac{dF(t_n)}{1-\Delta_\Lambda(t_n)}\cdots\frac{dF(t_1)}{1-\Delta_\Lambda(t_1)}\right]}{1+\sum_{n=1}^{\infty}\int\cdots\int\mathbf{1}\{M > t_n > \cdots > t_1 \geq 0\}\frac{dF(t_n)}{1-\Delta_\Lambda(t_n)}\cdots\frac{dF(t_1)}{1-\Delta_\Lambda(t_1)}}, \tag{4.8}$$

where $\Delta_\Lambda(x) = \int_{\{x\}} dF(t)/\bar{F}(t-)$. The equivalence of Eq.(4.5) and (4.8) can be seen via the well-known expansion: $(1-x)^{-1} = 1 + x + x^2 + \cdots$ for $|x| < 1$. Eq.(4.8) was also obtained by Prentice et al (2004) in an incorrect form.

5. Influence function of the estimator

Note that Equations (2.3)–(2.6) and their solutions are completely *dimension-free*, i.e., is valid for $\Delta_i := (\delta_{1i}, \ldots, \delta_{mi})$, $Z_i = (Z_{1i}, \ldots, Z_{mi})$ for $m \geq 1$, with the definitions $\delta_i = \prod_{j=1}^m \delta_{ji}$ and $a_{ik} = \mathbf{1}\{Z_k \geq Z_i\}$ where the inequality is defined in the coordinate-wise sense. Hence in this section we shall use scalar notation also for vector variables, with the above interpretation.

Now to derive the *influence functions* for the estimators $\bar{F}_n(x)$ and $\int \varphi dF_n$ for a given $\varphi(\cdot)$, let P denote the distribution of (δ, Z) and P_n the empirical distribution of (δ_i, Z_i), $1 \leq i \leq n$. Also, let $T_x(P) := \bar{F}(x-)$, $T_\varphi(P) := \int \varphi dF$, and let $T_x(P_n)$, $T_\varphi(P_n)$ be their estimators, respectively, obtained via Eq.(2.6). Thus we rewrite Eq. (2.2) and (2.3) as the eigenvalue problems

$$\begin{aligned} T_x(P) &= \int \mathbf{1}\{t \geq x\} T_t(P) \frac{dH^{11}(t)}{\bar{H}(t-)}, \\ T_x(P_n) &= \int \mathbf{1}\{t \geq x\} T_t(P_n) \frac{dH_n^{11}(t)}{\bar{H}_n(t-)}, \end{aligned} \tag{1}$$

with the initial conditions $T_0(P) = 1$, $T_0(P_n) = 1$.

Note also that, for a function $\varphi(\cdot)$ satisfying $\varphi(x) = 0$ if $x \notin [0, \tau]$ for some τ with $\bar{H}(\tau) > 0$,

$$\begin{aligned} T_\varphi(P) &= \int \varphi(t) T_t(P) \frac{dH^{11}(t)}{\bar{H}(t-)}, \\ T_\varphi(P_n) &= \int \varphi(t) T_t(P_n) \frac{dH_n^{11}(t)}{\bar{H}_n(t-)}. \end{aligned} \tag{2}$$

Section 4 shows that the functional $T_x(P) = T_x(H^{11}, H)$ is *Hadamard differentiable*, i.e.,

$$\lim_{\varepsilon \to 0} \left[T_x(H^{11} + \varepsilon h_\varepsilon^{11}, H + \varepsilon h_\varepsilon) - T_x(H^{11}, H) \right] / \varepsilon$$

exists if $h_\varepsilon^{11} \to h^{11}$, $h_\varepsilon \to h$, in the domain of sub-probability measures on $\mathbb{R}_+^m$.

The influence functions for $T_x(P_n)$, $T_\varphi(P_n)$ can now be derived as

$$L_x(P_n) = \sum_{i=1}^n l_x(\delta_i, Z_i)/n = \lim_{\varepsilon \to 0} [T_x(P + \varepsilon(P_n - P)) - T_x(P)]/\varepsilon, \tag{3}$$

and $L_\varphi(P_n) = \sum_{i=1}^n l_\varphi(\delta_i, Z_i)/n$ similarly.

Now let $P_{\varepsilon,n} = P + \varepsilon(P_n - P) = (1 - \varepsilon)P + \varepsilon P_n$, and note that

$$\begin{aligned} H_{\varepsilon,n}^{11}(t) &:= H^{11}(t) \text{ corresponding to } P_{\varepsilon,n} = (1 - \varepsilon) H^{11}(t) + \varepsilon H_n^{11}(t), \\ \bar{H}_{\varepsilon,n}(t) &:= \bar{H}(t) \text{ corresponding to } P_{\varepsilon,n} = (1 - \varepsilon)\bar{H}(t) + \varepsilon \bar{H}_n(t). \end{aligned}$$

Then from Eq.(10),

$$\lim_{\varepsilon \to 0} [T_x(P_{\varepsilon,n}) - T_x(P)]/\varepsilon$$

$$
\begin{aligned}
&= \lim_{\varepsilon\to 0}\int \mathbf{1}\{t\ge x\}\left[T_t(P_{\varepsilon,n})\frac{dH^{11}_{\varepsilon,n}(t)}{\bar H_{\varepsilon,n}(t-)} - T_t(P)\frac{dH^{11}(t)}{\bar H(t-)}\right]/\varepsilon \\
&= \lim_{\varepsilon\to 0}\int \mathbf{1}\{t\ge x\}\frac{[T_t(P_{\varepsilon,n})-T_t(P)]}{\varepsilon}\frac{dH^{11}_{\varepsilon,n}(t)}{\bar H_{\varepsilon,n}(t-)} \\
&\quad +\lim_{\varepsilon\to 0}\int \mathbf{1}\{t\ge x\}T_t(P)\left[\frac{dH^{11}_{\varepsilon,n}(t)}{\bar H_{\varepsilon,n}(t-)} - \frac{dH^{11}(t)}{\bar H(t-)}\right]/\varepsilon. \qquad (4)
\end{aligned}
$$

Since

$$
\left[\frac{dH^{11}_{\varepsilon,n}(t)}{\bar H_{\varepsilon,n}(t-)} - \frac{dH^{11}(t)}{\bar H(t-)}\right]/\varepsilon = \left[\frac{(1-\varepsilon)dH^{11}(t)+\varepsilon dH^{11}_n(t)}{(1-\varepsilon)\bar H(t-)+\varepsilon \bar H_n(t-)} - \frac{dH^{11}(t)}{\bar H(t-)}\right]/\varepsilon,
$$

it follows from Eq.(13) that the influence function $L_x(P_n)$ of $T_x(P_n) = \bar F_n(x-)$ satisfies the linear equation

$$
\begin{aligned}
L_x(P_n) &= \int \mathbf{1}\{t\ge x\}L_t(P_n)\frac{dH^{11}(t)}{\bar H(t-)} \\
&\quad + \int \mathbf{1}\{t\ge x\}T_t(P)\frac{\bar H(t-)dH^{11}_n(t) - \bar H_n(t-)dH^{11}(t)}{\bar H^2(t-)};
\end{aligned}
$$

since $T_t(P) = \bar F(t-)$, this simplifies to

$$
L_x(P_n) - \int \mathbf{1}\{t\ge x\}L_t(P_n)\frac{dF(t)}{\bar F(t-)} = \int \mathbf{1}\{t\ge x\}\left[\frac{dH^{11}_n(t)}{\bar G(t-)} - \bar H_n(t-)\frac{dF(t)}{\bar H(t-)}\right] \qquad (5)
$$

Further, using Eq.(2) the influence function $L_\varphi(P_n)$ of $T_\varphi(P_n) = \int \varphi dF_n$ is given by

$$
\begin{aligned}
L_\varphi(P_n) &= \lim_{\varepsilon\to 0}[T_\varphi(P_{\varepsilon,n}) - T_\varphi(P)]/\varepsilon \\
&= \lim_{\varepsilon\to 0}\int \varphi(t)\left[T_t(P_{\varepsilon,n})\frac{dH^{11}_{\varepsilon,n}(t)}{\bar H_{\varepsilon,n}(t-)} - T_t(P)\frac{dH^{11}(t)}{\bar H(t-)}\right]/\varepsilon \\
&= \int \varphi(t)L_t(P_n)\frac{dF(t)}{\bar F(t-)} \\
&\quad + \int \varphi(t)\left[\frac{dH^{11}_n(t)}{\bar G(t-)} - \bar H_n(t-)\frac{dF(t)}{\bar H(t-)}\right] \qquad (6)
\end{aligned}
$$

Now from Theorem 4.1, Eq.(5) does not have a unique solution in general, because the corresponding homogeneous equation, $l(x) - \int \mathbf{1}\{t\ge x\}l(t)\frac{dF(t)}{\bar F(t-)} = 0$, has the general solution $c\bar F(x-)$ for any scalar c (i.e., $\bar F(x-)$ is the eigenvector spanning the 1-dimensional eigenspace).

However, a unique solution, which gives the *unique influence function*, is obtained if we impose the initial condition $L_0(P_n) = 0$; this is the natural condition to impose in view of the uncensored case, where we know that

$$
L_x(P_n) = n^{-1}\sum_{i=1}^{n}\mathbf{1}\{X_i\ge x\} - \bar F(x-).
$$

Thus in view of Theorem 4.1, and focusing only on the special case where $F(\cdot)$, $G(\cdot)$ have densities $f(\cdot)$, $g(\cdot)$, we rewrite Eq.(5) as a *Volterra integral equation*:

$$
\begin{aligned}
L_x(P_n) - \int L_t(P_n)K(x,dt) &= z_n(x) \qquad (7) \\
L_0(P_n) &= 0, \qquad (8)
\end{aligned}
$$

where

$$K(x,dt) = \mathbf{1}\{t \geq x\} dF(t)/\bar{F}(t), \;\; K(0,dt) = dF(t)/\bar{F}(t)$$

and

$$z_n(x) = \int \mathbf{1}\{t \geq x\} \left[\frac{dH_n^{11}(t)}{\bar{G}(t)} - \bar{H}_n(t) \frac{dF(t)}{\bar{H}(t)} \right].$$

THEOREM 2. Under continuity of $F(\cdot)$, $G(\cdot)$, the influence function for $T_x(P_n)$ is *asymptotically efficient* and is given by the unique solution to Eq.(7)–(8):

$$L_x(P_n) = a_x(P_n) - \bar{F}(x) a_0(P_n)$$

where

$$a_x(P_n) = z_n(x) + \sum_{r=1}^{\infty} \int \cdots \int z_n(y_r) K(x, dy_1) \cdots K(y_r, dy_r). \tag{9}$$

PROOF: It is easy to see that the infinite *Neumann* series in $a_x(P_n)$ is convergent, so that $a_x(P_n)$ is finite, and it is a solution to Eq.(17). In view of Lemma 1, a general solution to Eq.(7) is therefore of the form

$$L_x^c(P_n) = a_x(P_n) + c\bar{F}(x).$$

Setting $L_0^c(P_n) = 0$ gives $c = -a_0(P_n)$, hence the unique solution to Eq.(7)–(8) is given by $L_x(P_n)$.

As for asymptotic efficiency, $L_x(P_n)$ obviously belongs to the *closed linear span* of the empirical process $\{z_n(x), \; x \geq 0\}$, which obvioulsy belongs to the *tangent space* of the model at P, hence so does $L_x(P_n)$.

To complete the proof, we need to verify Eq.(3.7), Theorem 3.1, of van der Vaart (1991), which in this case boils down to

$$\mathbf{A}^* L_x(P_n) = n^{-1} \sum_{i=1}^{n} \mathbf{1}\{X_i \geq x\} - \bar{F}(x), \tag{10}$$

where $\mathbf{A}$ is the *score* operator:

$$\mathbf{A}h = E(h(X_i, 1 \leq i \leq n) | \delta_i, Z_i, 1 \leq i \leq n)$$

and

$$\mathbf{A}^* L = E(L(\delta_i, Z_i, 1 \leq i \leq n) | X_i, 1 \leq i \leq n)$$

is the *dual* of $\mathbf{A}$.

To verify Eq.(10), note that $L_x(P_n)$ is of the form $L_x(P_n) = \sum_{i=1}^n L_x(\delta_i, Z_i)/n$. We thus need to look only at

$$E[L_x(\delta_i, Z_i)|X_i] = E[a_x(\delta_i, Z_i)|X_i] - \bar{F}(x) E[a_0(\delta_i, Z_i)|X_i]$$

for each fixed i, $1 \leq i \leq n$. Let us consider only the bi-variate case, i.e., $m = 2$ and $X_i = (X_{1i}, X_{2i})$. Now

$$E[a_x(\delta_i, Z_i)|X_i] = E[z_i(x)|X_i] + \sum_{r=1}^{\infty} \int \cdots \int E[z_i(y_r)|X_i] K(x, dy_1) \cdots K(y_r, dy_r),$$

where

$$z_i(x) = \frac{\delta_i \mathbf{1}\{Z_i \geq x\}}{\bar{G}(Z_i)} - \int \mathbf{1}\{Z_i \geq t\}\mathbf{1}\{t \geq x\}\frac{dF(t)}{\bar{H}(t)};$$

further, for an arbitrary $h(\delta_i, Z_i)$ we have

$$\begin{aligned}
&E[h(\delta_{1i}, \delta_{2i}, Z_{1i}, Z_{2i})|X_i] \\
= &\int [h(1,1,X_{1i},X_{2i})\mathbf{1}\{X_{1i} \leq y_1, X_{2i} \leq y_2\} + h(1,0,X_{1i},y_2)\mathbf{1}\{X_{1i} \leq y_1, X_{2i} > y_2\} \\
&+h(0,1,y_1,X_{2i})\mathbf{1}\{X_{1i} > y_1, X_{2i} \leq y_2\} + h(0,0,y_1,y_2)\mathbf{1}\{X_{1i} > y_1, X_{2i} > y_2\}]\, dG(y_1,y_2),
\end{aligned}$$

so that

$$E[z_i(x)|X_i] = \mathbf{1}\{X_i \geq x\} - \int \mathbf{1}\{X_i \geq t\}\mathbf{1}\{t \geq x\}\frac{dF(t)}{\bar{F}(t)} = \mathbf{1}\{X_i \geq x\} - \int \mathbf{1}\{X_i \geq t\}K(x, dt).$$

Plugging the expression for $E[z_i(x)|X_i]$ back into that for $E[a_x(\delta_i, Z_i)|X_i]$ we see that the successive terms in the infinite series cancel each other out, so that we are left with

$$E[a_x(\delta_i, Z_i)|X_i] = \mathbf{1}\{X_i \geq x\} \text{ and } E[a_0(\delta_i, Z_i)|X_i] = \mathbf{1}\{X_i \geq 0\} = 1.$$

Hence finally,

$$E[L_x(\delta_i, Z_i)|X_i] = E[a_x(\delta_i, Z_i)|X_i] - \bar{F}(x)E[a_0(\delta_i, Z_i)|X_i] = \mathbf{1}\{X_i \geq x\} - \bar{F}(x),$$

and Eq.(10) is verified.$\square$

We now calculate the influence function in a few special cases.

Example 1: Uncensored data. In this case $H_n^{11}(\cdot) = H_n(\cdot) = F_n(\cdot)$, the usual empirical d.f., $\bar{G} \equiv 1$, $H^{11}(\cdot) = H(\cdot) = F(\cdot)$, $\bar{F} = \bar{H}$, so that Eq.(17) reduces to

$$L_x(P_n) - \int \mathbf{1}\{t \geq x\}L_t(P_n)\frac{dF(t)}{\bar{F}(t-)} = \bar{F}_n(x-) - \int \mathbf{1}\{t \geq x\}\bar{F}_n(t-)\frac{dF(t)}{\bar{F}(t-)},$$

which gives the general solution $L_x^c(P_n) = \bar{F}_n(x-) + c\bar{F}(x-)$, and under $L_0(P_n) = 0$, the unique solution

$$L_x(P_n) = \bar{F}_n(x-) - \bar{F}(x-) = n^{-1}\sum_{i=1}^{n}\mathbf{1}\{X_i \geq x\} - \bar{F}(x-).$$

The same also follows from Theorem 2 as $a_x(P_n)$ simplifies to $\bar{F}_n(x-)$.

Example 2: Influence function for $T_\varphi(P_n) = \int \varphi dF_n$.
From Eq.(6) and Theorem 2, we get for a $\varphi(\cdot)$ with compact support,

$$L_\varphi(P_n) = \int \varphi(t)a_t(P_n)\frac{dF(t)}{\bar{F}(t)} - a_0(P_n)\int \varphi(t)dF(t) + \int \varphi(t)\left[\frac{dH_n^{11}(t)}{\bar{G}(t)} - \bar{H}_n(t)\frac{dF(t)}{\bar{H}(t)}\right]$$

Example 3: Univariate censored data. Here for $r \geq 1$,

$$\begin{aligned}
& \int \cdots \int K(x, dy_1) \cdots K(y_r, dy_r) z_n(y_r) \\
= & \int \cdots \int \mathbf{1}\{x < y_1 < \cdots < y_r\} \frac{f(y_1)}{1 - F(y_1)} \cdots \frac{f(y_r)}{1 - F(y_r)} \\
& \left(\int \mathbf{1}\{t \geq y_r\} \left[\frac{dH_n^{11}(t)}{\bar{G}(t)} - \bar{H}_n(t) \frac{dF(t)}{\bar{H}(t)} \right] \right) dy_1 \ldots dy_r \\
= & \int \frac{(-\log \bar{F}(t) + \log \bar{F}(x))^r}{r!} \mathbf{1}\{t \geq x\} \left[\frac{dH_n^{11}(t)}{\bar{G}(t)} - \bar{H}_n(t) \frac{dF(t)}{\bar{H}(t)} \right],
\end{aligned}$$

so that

$$a_x(P_n) = \int \frac{\bar{F}(x)}{\bar{F}(t)} \mathbf{1}\{t \geq x\} \left[\frac{dH_n^{11}(t)}{\bar{G}(t)} - \bar{H}_n(t) \frac{dF(t)}{\bar{H}(t)} \right],$$

and hence

$$L_x(P_n) = a_x(P_n) - \bar{F}(x) a_0(P_n) = -\int \bar{F}(x) \mathbf{1}\{t < x\} \left[\frac{dH_n^{11}(t)}{\bar{H}(t)} - \bar{H}_n(t) \frac{dF(t)}{\bar{H}(t)\bar{F}(t)} \right] \tag{11}$$

In order to derive the general expression obtained by Stute (1995) (Eq.(1.7), Theorem 1.1), consider a general $\varphi(\cdot)$ with compact support. Then from Eq.(6) and Eq.(11),

$$\begin{aligned}
& L_\varphi(P_n) \\
= & -\int \varphi(t) \left(\int \mathbf{1}\{s < t\} \left[\frac{dH_n^{11}(s)}{\bar{H}(s)} - \bar{H}_n(s) \frac{dF(s)}{\bar{H}(s)\bar{F}(s)} \right] \right) dF(t) \\
& + \int \varphi(t) \left[\frac{dH_n^{11}(t)}{\bar{G}(t)} - \bar{H}_n(t) \frac{dF(t)}{\bar{H}(t)} \right]
\end{aligned} \tag{12}$$

Consider the *second inner integral* in the *first term* of Eq.(12):

$$\begin{aligned}
& -\int \mathbf{1}\{s < t\} \bar{H}_n(s) \frac{dF(s)}{\bar{H}(s)\bar{F}(s)} \\
= & -n^{-1} \sum_{i=1}^n \int \mathbf{1}\{s < t \wedge Z_i\} \frac{dF(s)}{\bar{G}(s)\bar{F}^2(s)} \\
= & -n^{-1} \sum_{i=1}^n (\bar{G}(t \wedge Z_i) \bar{F}(t \wedge Z_i))^{-1} + 1 \\
& + \int \mathbf{1}\{s < t\} \bar{H}_n(s) \frac{dG(s)}{\bar{F}(s)\bar{G}^2(s)} \\
= & -(\bar{H}_n(t)/\bar{H}(t)) - n^{-1} \sum_{i=1}^n (\mathbf{1}\{Z_i \leq t\}/\bar{H}(Z_i)) + 1 \\
& + \int \mathbf{1}\{s < t\} \frac{\bar{H}_n(s) dG(s)}{\bar{H}(s)\bar{G}(s)},
\end{aligned} \tag{13}$$

where we have used integration-by-parts in the third line; now put Eq.(13) into Eq.(12) to get

$$L_\varphi(P_n) = n^{-1} \sum_{i=1}^n \left[\delta_i \varphi(Z_i)/\bar{G}(Z_i) + (1 - \delta_i) \gamma_\varphi(Z_i) - \Gamma_\varphi(Z_i) \right] - E_F(\varphi),$$

where $\gamma_\varphi(Z_i) = S_\varphi(Z_i)/\bar{H}(Z_i)$, $\Gamma_\varphi(Z_i) = \int \mathbf{1}\{Z_i > s\}[S_\varphi(s)/(\bar{H}(s)\bar{G}(s))]dG(s)$ and $S_\varphi(s) = \int \mathbf{1}\{s < t\}\varphi(t)dF(t)$. This is exactly the expression obtained by Stute (1995).

Example 4: Bivariate censored data with independent components.
Let us consider the bi-variate situation where $F(x_1, x_2) = F_1(x_1)F_2(x_2)$, $G(y_1, y_2) = G_1(y_1)G_2(y_2)$, where $F_j(\cdot),\ G_j(\cdot),\ j = 1, 2$, are the marginal distribution functions. In this case,

$$K(x, dt) = \mathbf{1}\{t_1 \geq x_1, t_2 \geq x_2\}\frac{f_1(t_1)f_2(t_2)}{(1 - F_1(t_1))(1 - F_2(t_2))}dt_1dt_2 = K(x_1, dt_1)K(x_2, dt_2).$$

Hence

$$\begin{aligned} & a_{x_1,x_2}(P_n) \\ = \ & \sum_{r=0}^{\infty}(r!)^{-2}\int\left(\prod_{j=1}^{2}\mathbf{1}\{t_j \geq x_j\}(-\log\bar{F}_j(t_j) + \log\bar{F}_j(x_j))^r\right) \\ & \left[\frac{dH_n^{11}(t_1, t_2)}{\bar{G}_1(t_1)\bar{G}_2(t_2)} - \bar{H}_n(t)\frac{dF_1(t_1)dF_2(t_2)}{\bar{H}_1(t_1)\bar{H}_2(t_2)}\right]. \end{aligned}$$

This, however, does not seem to simplify any further.

References.

1. Dabrowska, D.M. (1988). Kaplan-Meier estimate on the plane. Ann. Statist. **16**, 1475–1489.

2. Gill, R.D., van der Laan, M.J. and Wellner, J.A. (1995). Inefficient estimators of the bivariate survival function for three models. Ann. Inst. H. Poincar Probab. Statist. **31**, 545–597.

3. Prentice, R.L., Moodie, F. and Wu, J. (2004). Hazard-based nonparametric survivor function estimation. J. R. Stat. Soc. Ser. B Stat. Methodol. **66**, 305–319.

4. Stute, W. (1995). The central limit theorem under random censorship. Ann. Statist. **23**, 422–439.

5. van der Laan, M.J. (1996). Efficient estimation in the bivariate censoring model and repairing NPMLE. Ann. Statist. **24**, 596–627.

6. van der Vaart, A. (1991). On differentiable functionals. Ann. Statist. **19**, 178–204.

List of Recent Technical Reports

74. Qihe Tang, *Asymptotic Ruin Probabilities of the Renewal Model with Constant Interest Force and Regular Variation*, November 2004

75. Xiaowen Zhou, *On a Classical Risk Model with a Constant Dividend Barrier*, November 2004

76. K. Balasubramanian and M.I. Beg, *Three Isomorphic Vector Spaces–II*, December 2004

77. Michael A. Kouritzin and Wei Sun, *Rates for Branching Particle Approximations of Continuous–Discrete Filters*, December 2004

78. Rob Kaas and Qihe Tang, *Introducing a Dependence Structure to the Occurences in Studying Precise Large Deviations for the Total Claim Amount*, December 2004

79. Qihe Tang and Gurami Tsitsiashvili, *Finite and Infinite Time Ruin Probabilities in the Presence of Stochastic Returns on Investments*, December 2004

80. Alexander Melnikov and Victoria Skornyakova, *Efficient Hedging Methodology Applied to Equity–Linked Life Insurance*, February 2005

81. Qihe Tang, *The Finite Time Ruin Probability of the Compound Poisson Model with Constant Interest Force*, June 2005

82. Marc J. Goovaerts, Rob Kaas, Roger J.A. Laeven, Qihe Tang and Raluca Vernic, *The Tail Probability of Discounted Sums of Pareto–Like Losses in Insurance*, August 2005

83. Yogendra P. Chaubey and Haipeng Xu, *Smooth Estimation of Survival Functions under Mean Residual Life Ordering*, August 2005

84. Xiaowen Zhou, *Stepping–Stone Model with Circular Brownian Migration*, August 2005

85. José Garrido and Manuel Morales, *On the Expected Discounted Penalty Function for Lévy Risk Processes*, November 2005

86. Ze–Chun Hu, Zhi–Ming Ma and Wei Sun, *Extensions of Lévy–Khintchine Formula and Beurling–Deny Formula in Semi–Dirichlet Forms Setting*, February 2006

87. Ze–Chun Hu, Zhi–Ming Ma and Wei Sun, *Formulae of Beurling–Deny and Lejan for Non-Symmetric Dirichlet Forms*, February 2006

88. Ze–Chun Hu and Wei Sun, *A Note on Exponential Stability of the Non-Linear Filter for Denumerable Markov Chains*, February 2006

89. H. Brito–Santana, R. Rodríguez–Ramos, R. Guinovart–Díaz, J. Bravo–Castillero and F.J. Sabina, *Variational Bounds for Multiphase Transversely Isotropic Composites*, August 2006

90. José Garrido and Jun Zhou, *Credibility Theory for Generalized Linear and Mixed Models*, December 2006

91. Daniel Dufresne, José Garrido and Manuel Morales, *Fourier Inversion Formulas in Option Pricing and Insurance*, December 2006

92. Xiaowen Zhou, *A Superprocess Involving Both Branching and Coalescing*, December 2006

93. Yogendra P. Chaubey, Arusharka Sen and Pranab K. Sen, *A New Smooth Density Estimator for Non–Negative Random Variables*, January 2007

94. Md. Sharif Mozumder and José Garrido, *On the relation between the Lévy measure and the jump function of a Lévy process*, October 2007

95. Arusharka Sen and Winfried Stute, *A bi-variate Kaplan-Meier estimator via an integral equation*, October 2007

Copies of technical reports can be requested from:

Dr. Wei Sun
Department of Mathematics and Statistics
Concordia University
1455 de Maisonneuve Blvd. West,
Montreal, QC, H3G 1M8 CANADA