Published in final edited form as:

Lifetime Data Anal. 2018 January ; 24(1): 3–27. doi:10.1007/s10985-016-9383-y.

Nonparametric Estimation of the Multivariate Survivor Function: The Multivariate Kaplan–Meier Estimator

Ross L. Prentice and

Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle,
Washington, USA

Shanshan Zhao

National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina,
USA, shanshan.zhao@nih.gov

Abstract

The Dabrowska (1988) product integral representation of the multivariate survivor function is extended, leading to a nonparametric survivor function estimator for an arbitrary number of failure time variates that has a simple recursive formula for its calculation. Empirical process methods are used to sketch proofs for this estimator's strong consistency and weak convergence properties. Summary measures of pairwise and higher-order dependencies are also defined and nonparametrically estimated. Simulation evaluation is given for the special case of three failure time variates.

Keywords

Censoring; Dabrowska estimator; Failure times; Kaplan–Meier estimator; Multivariate;
Nonparametric; Product integral; Survivor function; Trivariate dependency

1 Introduction

It is a great pleasure to contribute to this LIDA issue in honor of our esteemed colleague Jack Kalbfleisch. Jack has been a leader in the development and application of failure time data methods for well over four decades. We, along with many others, have learned much from Jack's innovative developments and careful writing. One of us (RP) has benefited also from many years of active collaboration with Jack on failure time regression and related methods and has benefited from many years of valued friendship with Jack and his wife Sharon and their family.

For this occasion we have chosen to present some new work on nonparametric estimation of the multivariate survivor function. While core methods for univariate failure time data analysis, including Kaplan and Meier (1958) curves and Cox (1972) regression, are well established, and are known to have good moderate and large sample size behavior, bivariate and higher-dimensional failure time data analysis methods are less developed. Here we focus on nonparametric estimation of the multivariate survivor function, building on the product integral representation of Dabrowska (1988). The status of nonparametric estimators of the bivariate survivor function, including the Dabrowska estimator, is reviewed in Section 2. The

trivariate generalization of this representation and a corresponding empirical estimator is given in Section 3. Summary measures of bivariate and trivariate dependency are also defined and nonparametrically estimated, and simulation evaluations are provided. Generalization to an arbitrary number of failure time variates are given in Section 4. Asymptotic distribution theory is sketched in Section 5, followed by discussion in Section 6.

2 Nonparametric estimation of the bivariate survivor functions

2.1 Background

Consider failure time variates $T_1 > 0$ and $T_2 > 0$ having a joint survivor function F , so that $P(t_1, t_2) = P(T_1 > t_1, T_2 > t_2)$. Suppose that an independent random sample of size n is drawn from a study population with (T_1, T_2) values that are subject to independent right censorship by variate (C_1, C_2) so that one observes $X_{1\ell} = T_{1\ell} \wedge C_{1\ell}$, $\delta_{1\ell} = I[X_{1\ell} = T_{1\ell}]$, $X_{2\ell} = T_{2\ell} \wedge C_{2\ell}$ and $\delta_{2\ell} = I[X_{2\ell} = T_{2\ell}]$ for $\ell = 1, \dots, n$, where ' $\wedge$ ' denotes minimum and $I[\cdot]$ denotes an indicator variable.

The nonparametric likelihood function for F can be written

$$L = \prod_{\ell=1}^n F(dX_{1\ell}, dX_{2\ell})^{\delta_{1\ell}\delta_{2\ell}} \{ -F(dX_{1\ell}, X_{2\ell}) \}^{\delta_{1\ell}(1-\delta_{2\ell})} \{ -F(X_{1\ell}, dX_{2\ell}) \}^{(1-\delta_{1\ell})\delta_{2\ell}} F(X_{1\ell}, X_{2\ell})^{(1-\delta_{1\ell})(1-\delta_{2\ell})}, \quad (1)$$

adopting the usual convention that failures precede censoring in the event of tied times. Denote by $t_{11} < t_{12} < \dots < t_{1J}$ the ordered uncensored T_1 values in the sample, and by $t_{21} < t_{22} < \dots < t_{2J}$ the ordered uncensored T_2 values, and set $t_{10} = t_{20} = 0$. L can be maximized by a discrete survivor function F having jumps at grid points (t_{1i}, t_{2j}) for $i = 0, \dots, I$ and $j = 0, \dots, J$ such that $(t_{1i}, t_{2j}) \in R = \{(t_1, t_2); X_{1\ell} \geq t_1, \text{ and } X_{2\ell} \geq t_2 \text{ for some } \ell\}$. However, the resulting partially maximized likelihood is overparameterized (e.g., Campbell, 1981; Hanley and Parnes, 1983; Tsai et al, 1986; van der Laan, 1996; Akritas and Van Keilegom, 2003; Prentice, 2014). This overparameterization occurs because the data at some grid points in the risk region (R) of the data are uninformative concerning local dependence between T_1 and T_2 . More specifically, consider the 2×2 table that distinguishes whether T_1 is equal to or greater than t_{1i} and whether T_2 is equal to or greater than t_{2j} among the $r_{ij} = \#\{\ell; X_{1\ell} \geq t_{1i}, X_{2\ell} \geq t_{2j}\}$ 'individuals' at risk at uncensored failure time grid point (t_{1i}, t_{2j}) . Denote the table entries by d_{ij}^{11} double failures, d_{ij}^{10} failures on T_1 but not T_2 , d_{ij}^{01} failures on T_2 but not T_1 , and d_{ij}^{00} non-failures on both T_1 and T_2 at (t_{1i}, t_{2j}) so that $r_{ij} = d_{ij}^{11} + d_{ij}^{10} + d_{ij}^{01} + d_{ij}^{00}$. The data at $\{(t_{1\ell}, t_{2m}); \ell \leq i, m \leq j, (\ell, m) \neq (i, j)\}$ determine the marginal totals for this table, and if either or both of the marginal totals $d_{ij}^{11} + d_{ij}^{10}$ or $d_{ij}^{11} + d_{ij}^{01}$ equals zero then $d_{ij}^{11} = 0$ necessarily, and there is no possibility of a double failure at (t_{1i}, t_{2j}) given the earlier data. Similarly, if any of the four marginal totals takes value zero, then all table entries are specified by the earlier data, and the data are uninformative about local

dependence at this gridpoint. The lack of a unique nonparametric likelihood-based estimator for F derives from the inability to nonparametrically estimate a distinct point mass failure probability at these grid points.

There have been two principal responses to this nonparametric maximum likelihood (NPML) overparameterization problem. One has been to repair the NPML procedure by imposing regularity conditions on F (e.g. Tsai et al, 1986), by data reduction (van der Laan, 1996), or by making a local independence specification for point mass assignments at uninformative grid points (Prentice, 2014). The other involves inserting empirical estimators of single and double failure rates into representations for F . We will consider the latter approach first.

2.2 Plug-in estimators of F

The discrete failure time special case of the Dabrowska (1988) product integral representation of F can be written

$$F(t_1, t_2) = \prod_0^{t_1} \frac{F(s_1, 0)}{F(s_1^-, 0)} \prod_0^{t_2} \frac{F(0, s_2)}{F(0, s_2^-)} \prod_0^{t_1} \prod_0^{t_2} \frac{F(s_1, s_2) F(s_1^-, s_2^-)}{F(s_1, s_2^-) F(s_1^-, s_2)}$$

The first two product integrals are the marginal survival probabilities $F(t_1, 0)$ and $F(0, t_2)$ respectively, while the double product integral, which reduces to $F(t_1, t_2) / \{F(t_1, 0)F(0, t_2)\}$ following massive cancellations, characterizes the pairwise dependency between T_1 and T_2 given the marginal survivor functions. Upon dividing each factor in this double product

argument by $F(s_1^-, s_2^-)$ this product integral factor can be recognized as the ratio of the conditional double survival probability at (s_1, s_2) to the product of corresponding single variable survival probability factors, a product that would obtain under local independence between T_1 and T_2 at (s_1, s_2) .

The Dabrowska estimator $\hat{F}$ is obtained by inserting empirical estimates for each of the conditional survival probabilities in F giving, in the above notation,

$$\hat{F}(t_1, t_2) = \prod_{t_{1i} \leq t_1} \frac{d_{i0}^{00}}{r_{i0}} \prod_{t_{2j} \leq t_2} \frac{d_{0j}^{00}}{r_{0j}} \prod_{t_{1i} \leq t_1} \prod_{t_{2j} \leq t_2} \frac{d_{ij}^{00}/r_{ij}}{\left\{ \frac{d_{ij}^{10} + d_{ij}^{00}}{r_{ij}} \frac{d_{ij}^{01} + d_{ij}^{00}}{r_{ij}} \right\}} \quad (2)$$

for all $(t_1, t_2) \in R$. Note that the final factor in (2) takes the local independence value of one whenever $d_{ij}^{11} = d_{ij}^{10} = 0$ and/or $d_{ij}^{11} = d_{ij}^{01} = 0$ and can be defined to take value zero if $d_{ij}^{00} = 0$. This estimator is readily calculated recursively at all grid points (t_{1k}, t_{2j}) in the risk region of the data, starting with Kaplan–Meier estimators for the marginal survivor functions, using

$$\hat{F}(t_{1i}, t_{2j}) = \frac{\hat{F}(t_{1i}, t_{2j}^-) \hat{F}(t_{1i}^-, t_{2j})}{\hat{F}(t_{1i}^-, t_{2j}^-)} \left\{ \frac{d_{ij}^{00} r_{ij}}{(d_{ij}^{10} + d_{ij}^{00})(d_{ij}^{01} + d_{ij}^{00})} \right\}$$

if $d_{ij}^{00} \neq 0$, along with $\hat{F}(t_{1i}, t_{2j}) = 0$ if $d_{ij}^{00} = 0$.

A rather similar estimator, that arises from a Peano series, rather than a product integral, representation for $F(t_1, t_2)/\{F(t_1, 0)F(0, t_2)\}$, is given by Prentice and Cai (1992).

These estimators have been shown to be strongly consistent for F , and $n^{1/2}(\hat{F} - F)$ has been shown to converge weakly to a mean zero Gaussian process, and bootstrap procedures have been shown to apply, over a follow-up region $[0, \tau_1] \times [0, \tau_2]$ where $P(X_1 > \tau_1, X_2 > \tau_2) > 0$. For example Gill et al (1995) show these properties using the continuity and compact differentiability of the product integral and Peano series transformations. They also show the estimators to be nonparametric efficient in the special case of complete independence of T_1 , T_2 , C_1 and C_2 . Simulation studies by various authors (e.g., Prentice and Cai, 1992) show these estimators to perform well with moderate sample sizes, though they are not expected to be nonparametric efficient away from the complete independence special case just mentioned. Not all plug-in estimators of F enjoy these favorable properties. For example a plug-in estimator attributed to Peter Bickel (Dabrowska, 1988), formed by plugging Kaplan–Meier marginal survivor function estimators and empirical double failure hazard into an inhomogenous Volterra integral equation representation for F is also strongly consistent and weakly convergent to a Gaussian process, but it has comparatively much poorer efficiency properties than do the Dabrowska and Prentice–Cai estimators.

The Dabrowska and Prentice–Cai estimators, as well as other available plug-in estimators, have the undesirable property that they can assign negative point mass at grid points in the risk region, or along uncensored failure time half-lines beyond the risk region. This problem can be linked to the presence of doubly censored data, for which the corresponding n^{-1} probability assigned by an NPML procedure is not confined to a single grid point or grid line. To address this issue Prentice (2014) considered an estimator that begins by setting aside all doubly censored observations that are interior to the risk region (those on the boundary were retained to preserve the full risk region). The NPML for the remaining data, among estimators that make a local independence specification whenever $d_{ij}^{11} + d_{ij}^{10} = 0$ and/or $d_{ij}^{11} + d_{ij}^{01} = 0$, is the Dabrowska estimator given above, which avoids negative mass after excluding the doubly censored observations. The essential equivalence between NPMLEs and self-consistent estimators (Tsai and Crowley, 1985) was then used to develop a ‘self-consistent NPMLE’ for F . Simulation studies (Prentice, 2014) revealed some modest moderate sample size efficiency improvements, compared to the Dabrowska estimator, for estimating the distribution function $\hat{F}(t_1, t_2) = P(T_1 > t_1, T_2 > t_2)$, but corresponding survivor function estimators, including those for the marginal survivor functions, were not usefully improved.

It seems reasonable to conclude that estimators based on maximization of (1), with a free parameter at each informative grid point are unlikely to offer much improvement over the Dabrowska or Prentice–Cai estimators. In Table 1 we report some results from further extensive simulation studies evaluating and comparing the moderate sample size performance of these two estimators. As in Prentice and Cai (1992), data are generated under the Clayton (1978) bivariate survival model, given by

$$F(t_1, t_2) = \left\{ F(t_1, 0)^{-\theta} + F(0, t_2)^{-\theta} - 1 \right\}^{-1/\theta} \vee 0$$

with unit exponential marginal survivor functions and with cross ratio $1 + \theta$ either equal to one (independence) or equal to 5 (strong positive dependence). The censoring times C_1 and C_2 were chosen to be independent and exponentially distributed with a mean of 0.5, so that there is a two-thirds probability of censoring for both time variables. Simulation sample mean and standard deviation estimate are shown at selected percentiles for the marginal distributions of T_1 and T_2 , based on 1000 generated samples at each θ value and at each of the sample sizes 30, 60, 120 and 240. Note the good practical performance of both estimators at each θ -value, even at a sample size of 30 at which there are an expected 10 uncensored failures for each of T_1 and T_2 , and an expected $30/9=3.33$ doubly uncensored failures under independence between T_1 and T_2 . Also, one can see somewhat smaller standard errors for the Prentice–Cai compared to the Dabrowska estimator at small sample sizes, but this difference essentially disappears by $n = 240$. Additional simulations under even stronger positive dependence ($\theta = 10$) yielded similar findings. Because the explicit product integral representation underlying the Dabrowska estimator carries over conveniently to trivariate and higher-dimensional failure time data, we will consider Dabrowska-type survivor function estimators in subsequent sections.

2.3 Additional NPMLE proposals

The other bivariate survivor function research pathway alluded to above also considers maximum likelihood estimation under (1) with a restricted set of parameters. The work of van der Laan (1996) is particularly noteworthy. He partitioned each failure time axis, imposed additional censorship to the lower endpoint of the interval in which each subject's potential censoring time resided, then distributed singly censored observations within strips formed by the respective partitions in an iterative fashion. By requiring the partition bandwidth to go to zero at small power of n he showed that a nonparametric efficient repaired NPMLE could be obtained. This impressive result is attended, however, by the need to specify partitions and partition bandwidths as $n \rightarrow \infty$, the need for potential censoring times for uncensored observations, and the need to arbitrarily accommodate singly censored observations that exceed the largest doubly uncensored observation in the pertinent strip. Furthermore, very large sample sizes may be needed for the estimator's efficiency to approach the nonparametric lower bound. This estimator does not seem to have received much use in applications.

The trivariate dependency component of (3) compares $F(t_1, t_2, t_3)$ to the corresponding value under independence of the three variates after allowing for marginal bivariate dependencies. Expression (3) decomposes these marginal probabilities and these pairwise and trivariate dependency probability ratios into corresponding local product integral factors. Importantly, each survival probability in this expression can be divided by the same survival probability with each s or s^- argument replaced by s^- (with $0^- = 0$) without changing the right side of (3). It follows that a nonparametric estimator $\hat{F}(t_1, t_2, t_3)$ can be obtained by plugging in empirical estimates for each of these ratios, for every $(t_1, t_2, t_3) \in R$. More specifically, denote the argument for a survival probability on the right side of (3) at (s_1, s_2, s_3) , including zero s -values, by $F(u_1, u_2, u_3)$ where $u_i = s_i$ or s_i^- for $i = 1, 2, 3$. $\hat{F}(t_1, t_2, t_3)$ is obtained by replacing each $F(u_1, u_2, u_3)$ on the right side of (3) by

$$\begin{aligned} & \tilde{F}(u_1, u_2, u_3) \\ &= \# \left\{ \ell; X_{i\ell} > s_i \text{ if } u_i = s_i \text{ or } X_{i\ell} \geq s_i \text{ if } u_i = s_i^- \text{ for each } i=1, 2, 3 \right\} \\ & \quad / \# \{ \ell; X_{i\ell} \geq s_i, \text{ each } i=1, 2, 3 \} \end{aligned}$$

Dabrowska (1988) referred to her nonparametric survivor function estimator as Kaplan–Meier estimator on the plane. Continuing this terminology one can refer to $\hat{F}$ as the trivariate Kaplan–Meier estimator.

The nonparametric estimator, $\hat{F}$, can be calculated by first calculating KM marginal survival function estimators, then calculating Dabrowska bivariate survivor function estimators of the form (2) for each of the three pairwise marginal survivor functions, then calculating $\hat{F}$ away from its coordinate axes, at uncensored failure time grid point $(t_{1i}, t_{2j}, t_{3k}) \in R$, using the recursive expression

$$\begin{aligned} & \hat{F}(t_{1i}, t_{2j}, t_{3k}) \\ &= \frac{\hat{F}(t_{1i}, t_{2j}, t_{3k}^-) \hat{F}(t_{1i}, t_{2j}^-, t_{3k}) \hat{F}(t_{1i}^-, t_{2j}, t_{3k}) \hat{F}(t_{1i}, t_{2j}^-, t_{3k}^-)}{\hat{F}(t_{1i}, t_{2j}^-, t_{3k}^-) \hat{F}(t_{1i}^-, t_{2j}, t_{3k}^-) \hat{F}(t_{1i}, t_{2j}^-, t_{3k})} \\ & \quad \times \left\{ \frac{\tilde{F}(t_{1i}, t_{2j}, t_{3k}) \tilde{F}(t_{1i}, t_{2j}^-, t_{3k}^-) \tilde{F}(t_{1i}^-, t_{2j}, t_{3k}) \tilde{F}(t_{1i}, t_{2j}^-, t_{3k}^-)}{\tilde{F}(t_{1i}, t_{2j}, t_{3k}^-) \tilde{F}(t_{1i}, t_{2j}^-, t_{3k}) \tilde{F}(t_{1i}^-, t_{2j}, t_{3k})} \right\}. \quad (4) \end{aligned}$$

This nonparametric estimator can be used for data display and for exploratory data analysis. For example three-dimensional survivor function estimators may be plotted overall and for various subsets of the data defined by treatment groups in randomized trials, or covariate subsets in cohort studies more generally. Plots of average cross ratios or concordance rates (e.g., Fan et al, 2000) can be used to explore the nature of dependency between pairs of the three failure time variates. For example an average cross ratio estimator $\hat{C}_{110}$ for T_1 and T_2 over $[0, t_1] \times [0, t_2]$ can be based on

$$\hat{C}_{110}(t_1, t_2, 0) = \left\{ \sum_{t_{1i} \leq t_1} \sum_{t_{2j} \leq t_2} \hat{F}(\Delta t_{1i}, t_{2j}^-, 0) \hat{F}(t_{1i}^-, \Delta t_{2j}, 0) \hat{F}(t_{1i}^-, t_{2j}^-, 0)^{-1} / \sum_{t_{1i} \leq t_1} \sum_{t_{2j} \leq t_2} \hat{F}(\Delta t_{1i}, \Delta t_{2j}, 0) \right\}^{-1}$$

$\hat{C}_{110}$ values close to one for (t_1, t_2) values throughout the risk region of the data are consistent with independence between T_1 and T_2 for $T_3 = 0$, while values close to a constant are consistent with a Clayton–Oakes model (Clayton, 1978; Oakes, 1982, 1986, 1989) with positive or negative dependence depending on whether this constant value exceeds or is less than one.

A corresponding ratio measure can also be defined for trivariate dependency assessment. One can express the conditional survival probabilities from (3) in terms of hazard rates in an obvious notation, for example, by

$$\begin{aligned} F(s_1, s_2^-, s_3^-) / F(s_1^-, s_2^-, s_3^-) &= 1 - \Lambda_{100}(ds_1, s_2^-, s_3^-), \\ F(s_1, s_2, s_3^-) / F(s_1^-, s_2^-, s_3^-) &= 1 - \Lambda_{100}(ds_1, s_2^-, s_3^-) - \Lambda_{010}(s_1^-, ds_2, s_3^-) + \Lambda_{110}(ds_1, ds_2, s_3^-), \quad \text{and} \\ F(s_1, s_2, s_3) / F(s_1^-, s_2^-, s_3^-) &= 1 - \Lambda_{100}(ds_1, s_2^-, s_3^-) - \Lambda_{010}(s_1^-, ds_2, s_3^-) - \Lambda_{001}(s_1^-, s_2^-, ds_3) + \Lambda_{110}(ds_1, ds_2, s_3^-) + \Lambda_{101}(ds_1, s_2^-, ds_3) + \Lambda_{011}(s_1^-, ds_2, ds_3) + \Lambda_{111}(ds_1, ds_2, ds_3). \end{aligned}$$

(5)

Substituting these expressions, along with corresponding expressions interchanging the role of s_1, s_2 and s_3 into (3) gives an expression for $F(t_1, t_2, t_3)$ as a function of single, double and triple hazard rates at $s_i = t_i, i = 1, 2, 3$. Also, from the triple product integral in (3) one sees that there will be no trivariate dependency at (s_1, s_2, s_3) if

$$\begin{aligned} &F(s_1, s_2, s_3) \\ &\quad / F(s_1^-, s_2^-, s_3^-) \\ &= F(s_1, s_2, s_3^-) F(s_1, s_2^-, s_3) F(s_1^-, s_2, s_3) \\ &\quad / \{F(s_1, s_2^-, s_3^-) F(s_1^-, s_2, s_3^-) F(s_1^-, s_2^-, s_3)\}. \end{aligned}$$

Setting the left side of the last expression in (5) equal to this quantity uniquely determines a triple failure hazard rate

$$\begin{aligned}
& \Lambda_{111}^0(ds_1, ds_2, ds_3) \\
&= 1 - \Lambda_{100}(ds_1, s_2^-, s_3^-) \\
&\quad - \Lambda_{010}(s_1^-, ds_2, s_3^-) \\
&\quad - \Lambda_{001}(s_1^-, s_2^-, ds_3) \\
&\quad + \Lambda_{110}(ds_1, ds_2, s_3^-) \\
&\quad + \Lambda_{101}(ds_1, s_2^-, ds_3) \\
&\quad + \Lambda_{011}(s_1^-, ds_2, ds_3) \\
&\quad - \frac{F(s_1, s_2, s_3^-) F(s_1, s_2^-, s_3) F(s_1^-, s_2, s_3)}{F(s_1, s_2^-, s_3^-) F(s_1^-, s_2, s_3^-) F(s_1^-, s_2^-, s_3)}, \quad (6)
\end{aligned}$$

for which the trivariate dependency factor in (3) takes value one at (s_1, s_2, s_3) . One can thereby define a local trivariate dependency measure at (s_1, s_2, s_3) by

$\Lambda_{111}(ds_1, ds_2, ds_3) / \Lambda_{111}^0(ds_1, ds_2, ds_3)$, which compares the actual triple failure hazard rate to that under no trivariate dependence, on a relative scale. An average trivariate dependency ratio $C_{111}(t_1, t_2, t_3)$ over a region $[0, t_1] \times [0, t_2] \times [0, t_3]$, where (t_1, t_2, t_3) is in the support of the (X_1, X_2, X_3) , can be defined by

$$C_{111}(t_1, t_2, t_3)^{-1} = \frac{\int_0^{t_1} \int_0^{t_2} \int_0^{t_3} \{ \Lambda_{111}^0(ds_1, ds_2, ds_3) / \Lambda_{111}(ds_1, ds_2, ds_3) \} F(ds_1, ds_2, ds_3)}{\int_0^{t_1} \int_0^{t_2} \int_0^{t_3} F(ds_1, ds_2, ds_3)}.$$

This measure can be nonparametrically estimated by $\hat{C}_{111}$, where

$$\hat{C}_{111}(t_1, t_2, t_3) = \int_0^{t_1} \int_0^{t_2} \int_0^{t_3} \hat{F}(\Delta s_1, \Delta s_2, \Delta s_3) / \int_0^{t_1} \int_0^{t_2} \int_0^{t_3} \hat{F}(s_1^-, s_2^-, s_3^-) \hat{\Lambda}_{111}^0(\Delta s_1, \Delta s_2, \Delta s_3),$$

where $\hat{\Lambda}_{111}^0$ is obtained by everywhere inserting $\hat{F}$ for F in (6).

3.2 Clayton–Oakes model simulation evaluation

Absolutely continuous failure times were generated to study properties of the trivariate Kaplan–Meier estimator $\hat{F}$. With absolutely continuous failure times the survivor function decomposition can be written in exponential form as

$$\begin{aligned}
& F(t_1, t_2, t_3) \\
&= \exp \left\{ - \int_0^{t_1} \Lambda_{100}(ds_1, 0, 0) - \int_0^{t_2} \Lambda_{010}(0, ds_2, 0) - \int_0^{t_3} \Lambda_{001}(0, 0, ds_3) + \int_0^{t_1} \int_0^{t_2} \Lambda_{110}(ds_1, ds_2, 0) + \int_0^{t_1} \int_0^{t_3} \Lambda_{101}(ds_1, 0, ds_3) \right. \\
&\quad \left. + \int_0^{t_2} \int_0^{t_3} \Lambda_{011}(0, ds_2, ds_3) - \int_0^{t_1} \int_0^{t_2} \int_0^{t_3} \Lambda_{111}(ds_1, ds_2, ds_3) \right\} \quad (7)
\end{aligned}$$

where, for example

$$A_{110}(ds_1, ds_2, s_3) = \Lambda_{110}(ds_1, ds_2, s_3) - \Lambda_{100}(ds_1, s_2, s_3) \Lambda_{010}(s_1, ds_2, s_3),$$

and where

$$\begin{aligned} A_{111}(ds_1, ds_2, ds_3) &= \Lambda_{111}(ds_1, ds_2, ds_3) \\ &\quad - \Lambda_{100}(ds_1, s_2, s_3) \Lambda_{010}(s_1, ds_2, s_3) \Lambda_{001}(s_1, s_2, ds_3) \\ &\quad - A_{110}(ds_1, ds_2, s_3) \Lambda_{001}(s_1, s_2, ds_3) \\ &\quad - A_{101}(ds_1, s_2, ds_3) \Lambda_{010}(s_1, ds_2, s_3) \\ &\quad - A_{011}(s_1, ds_2, ds_3) \Lambda_{100}(ds_1, s_2, s_3). \end{aligned}$$

Failure times (T_1, T_2, T_3) were generated from the specialized Clayton–Oakes model

$$F(t_1, t_2, t_3) = \left\{ F(t_1, 0, 0)^{-\theta} + F(0, t_2, 0)^{-\theta} + F(0, 0, t_3)^{-\theta} - 2 \right\}^{-1/\theta} \vee 0 \quad (8)$$

which has bivariate and trivariate dependencies controlled by the single parameter $\theta \in [-1, \infty)$. See Prentice (2016) for a broader class of Clayton–Oakes models of dimension 2. Failure times from (8) are everywhere absolutely continuous for $\theta \geq -1/3$. Straightforward calculations give, at any continuity point (s_1, s_2, s_3) ,

$$\Lambda_{110}(ds_1, ds_2, s_3) = (1 + \theta) \Lambda_{100}(ds_1, s_2, s_3) \Lambda_{010}(s_1, ds_2, s_3)$$

so that the cross ratio function at $(T_1, T_2) = (s_1, s_2)$ for $T_3 > s_3$ takes value $1 + \theta$, independently of (s_1, s_2, s_3) , with the identical cross ratio function for the other pairs of failure time variates. One can also calculate

$$\Lambda_{111}(ds_1, ds_2, ds_3) = (1 + \theta)(1 + 2\theta) \Lambda_{100}(ds_1, s_2, s_3) \Lambda_{010}(s_1, ds_2, s_3) \Lambda_{001}(s_1, s_2, ds_3),$$

from which the trivariate dependency function A_{111} is given by

$$A_{111}(ds_1, ds_2, ds_3) = 2\theta^2 \Lambda_{100}(ds_1, s_2, s_3) \Lambda_{010}(s_1, ds_2, s_3) \Lambda_{001}(s_1, s_2, ds_3).$$

The trivariate dependency ratio measure at (s_1, s_2, s_3) is simply

$$\Lambda_{111}(ds_1, ds_2, ds_3) / \Lambda_{111}^0(ds_1, ds_2, ds_3) = (1 + \theta)(1 + 2\theta) / (1 + 3\theta)$$

at any (s_1, s_2, s_3) and the average trivariate dependency ratio over $[0, t_1] \times [0, t_2] \times [0, t_3]$ is also $C_{111}(t_1, t_2, t_3) = (1 + \theta)(1 + 2\theta) / (1 + 3\theta)$ for any (t_1, t_2, t_3) .

Without loss of generality we can assume T_i to have unit exponential marginal distribution for $i = 1, 2, 3$. Failure times (t_1, t_2, t_3) from (8) with $\theta = -1/3$ are then readily generated from independent uniform $(0, 1)$ variate values (u_1, u_2, u_3) by considering probability integral transformations of the marginal T_1 density, the conditional density for T_2 given $T_1 = t_1$, and the conditional density for T_3 given $(T_1, T_2) = (t_1, t_2)$. Specifically one can write

$$\begin{aligned} t_1 &= -\log u_1, \\ t_2 &= \theta^{-1} \log \left\{ 1 - u_1^{-\theta} \left(1 - u_2^{-\theta(1+\theta)^{-1}} \right) \right\} \quad \text{if } \theta \neq 0 \quad \text{and} \quad t_2 = -\log u_2 \quad \text{if } \theta = 0, \quad \text{and} \\ t_3 &= \theta^{-1} \log \left\{ 1 - u_1^{-\theta} u_2^{-\theta(1+\theta)^{-1}} \left(1 - u_3^{-\theta(1+2\theta)^{-1}} \right) \right\} \quad \text{if } \theta \neq 0 \quad \text{and} \quad t_3 = -\log u_3 \quad \text{if } \theta = 0. \end{aligned}$$

Failure times were generated for samples of size $n = 30, 60$ and 120 at θ values of zero (independence), 2 (moderate positive dependence), and 5 (strong positive dependence). Censoring times c_1, c_2 and c_3 were generated from independent exponential distributions each with expectation 0.5 , so that there is a two-thirds probability of censoring for each $x_i = t_i \wedge c_i, i = 1, 2, 3$.

Table 2 presents summary statistics for the trivariate Kaplan–Meier estimator at values $0.163, 0.357$, or 0.598 for each of t_1, t_2 , and t_3 based on 1000 simulations. Note that $\exp(-0.163) = 0.85$, $\exp(-0.357) = 0.70$ and $\exp(-0.598) = 0.55$. The estimator can be seen to be rather accurate at each θ -value, even at sample sizes as small as $n = 30$. Sample standard deviations decline with increasing sample size in the expected manner.

The pairwise cross ratio function estimator $\hat{C}_{110}$ was also examined in these simulations. This estimator depends substantially on the number of (T_1, T_2) double failures compared to that under independence between the two variates. The expected number of such double failures is small under some simulation scenarios. For example, at $\theta = 0$ and $n = 30$ this expected number is only $30/9 = 3.33$ overall and the expected number, for example, over $[0, 0.163] \times [0, 0.163] \times [0, \infty)$ is only $3.33(0.15)^2 = 0.075$. Even at $n = 120$ this expectation is only 0.30 . Nevertheless, corresponding simulation means (sample standard deviations) at $n = 120$ were $1.023 (0.445)$ at $\theta = 0$; $3.163 (1.604)$ at $\theta = 2$; and $6.268 (3.010)$ at $\theta = 5$. These can be observed to somewhat overestimate the targeted $1 + \theta$ values and to be imprecise, as is to be expected. The average trivariate dependency ratio C_{111} depends substantially on the number of triple failures in the sample, which is very small under any of these simulation scenarios. An additional set of simulations were run under the same configurations, but dropping all censoring, to examine properties of C_{110} and C_{111} in scenarios with a larger number of double and triple failures. Table 3 provides summary statistics from these simulations in the special case of $\theta = 2$ and $n = 120$. One can see that C_{110} incorporates only a slight overestimation, while C_{111} overestimates a little more, and that both estimators are fairly precise at time points away from the origin.

3.3 Normal transformation model simulation evaluation

A reviewer suggested that we include a simulation under another trivariate failure time model. One could readily simulate data under a frailty model, but the specialized Clayton–Oakes model (8) has a frailty model interpretation. In fact for $\theta > 0$ (8) arises from a gamma

frailty model that has been transformed so that the marginal survivor functions are free of the dependency parameter θ .

We carried out, instead, a further simulation under a trivariate normal transformation model. See Li et al (2008) for a detailed development of semi-parametric likelihood estimation under a bivariate normal transformation model. Briefly, trivariate normal observations (w_1, w_2, w_3) having means of zero, variances of one, and common correlation ρ were generated from independent standard normal variates (v_1, v_2, v_3) via $w_1 = v_1$, $w_2 = \rho v_1 + (1 - \rho^2)^{1/2} v_2$, and $w_3 = \rho v_1 + \rho \left(\frac{1-\rho}{1-\rho}\right)^{1/2} v_2 + \left\{\frac{(1-\rho)(1+2\rho)}{1+\rho}\right\}^{1/2} v_3$. Failure times having unit exponential marginals and a trivariate normal transformation model were then obtained as $t_i = -\log\{\Phi(w_i)\}$, $i = 1, 2, 3$, where Φ is the standard normal distribution function. The trivariate survivor function F is given by

$$F(t_1, t_2, t_3) = \int_{w_1}^{\infty} \int_{w_2}^{\infty} \int_{w_3}^{\infty} \phi_{111}(u_1, u_2, u_3) du_1 du_2 du_3, \quad (9)$$

where $w_i = \Phi^{-1}(e^{-t_i})$, $i = 1, 2, 3$, and

$$\begin{aligned} \phi_{111}(u_1, u_2, u_3) &= (2\pi)^{-3/2} \{(1-\rho)(1+2\rho)\}^{-1/2} \exp \left[-\{2(1-\rho)\}^{-1} (u_1^2 + u_2^2 + u_3^2) + \rho \{2(1-\rho)(1+2\rho)\}^{-1} (u_1 + u_2 + u_3)^2 \right] \\ &\quad . \end{aligned}$$

From this expression one has, for example, single, double and triple failure hazard rates at (t_1, t_2, t_3) of

$$\begin{aligned} \lambda_{100}(t_1, t_2, t_3) &= -\left(\frac{dw_1}{dt_1}\right) \int_{w_2}^{\infty} \int_{w_3}^{\infty} \phi_{111}(w_1, w_2, w_3) du_2 du_3 / F(t_1, t_2, t_3), \\ \lambda_{110}(t_1, t_2, t_3) &= \left(\frac{dw_1}{dt_1}\right) \left(\frac{dw_2}{dt_2}\right) \int_{w_3}^{\infty} \phi_{111}(w_1, w_2, w_3) du_3 / F(t_1, t_2, t_3), \quad \text{and} \\ \lambda_{111}(t_1, t_2, t_3) &= -\left(\frac{dw_1}{dt_1}\right) \left(\frac{dw_2}{dt_2}\right) \left(\frac{dw_3}{dt_3}\right) \phi_{111}(w_1, w_2, w_3) / F(t_1, t_2, t_3), \end{aligned}$$

where $dw_i/dt_i = -\Phi(w_i)\phi(w_i)^{-1}$, $i = 1, 2, 3$, and ϕ is the standard normal density. Tables 4 and 5 present simulation summary statistics under the same configurations as Tables 2 and 3, except that the normal transformations model, at ρ values of 0.0, 0.2, and 0.5 is substituted for the Clayton–Oakes model. The target values for the average bivariate and trivariate dependency measures C_{110} and C_{111} do not have a simple form as a function of ρ , and were calculated by numerical integration.

The trivariate Kaplan–Meier estimator performs well (Table 4) under the normal transformation model, with some overestimation at the smallest sample size ($n = 30$). Simulation summary statistics for the average bivariate and trivariate dependency ratio estimators are shown in Table 5, based on 1000 simulations at $n = 120$ with no censoring. The $\hat{C}_{110}$ estimator incorporates some modest upward bias under these simulation conditions. The targets for $\hat{C}_{111}$ require very complex numerical integrations under this simulation configuration, and are omitted from Table 5.

3.4 Further simulations

To check further on the estimation behavior of $\hat{C}_{111}$ additional simulations like those reported in Table 3, but at $n = 180$, were conducted under the Clayton–Oakes model (8). As shown in Table 6 the $\hat{C}_{111}$ estimator is more precise, as expected at $n = 180$ compared to $n = 120$, and the upward bias is only partially eliminated, suggesting some overestimation unless the number of triple failures is quite large.

A reviewer has made the interesting suggestion that one could consider a recursive version of the procedure (4), iteratively replacing each $\tilde{F}(s_1, s_2, s_3)$ value after the first iteration by the current value of $\hat{F}(s_1, s_2, s_3) / \hat{F}(s_1^-, s_2^-, s_3^-)$, continuing until convergence is achieved. The resulting estimator of F would then have a type of self-consistency property away from the coordinate axes. We evaluated this idea under the simulation conditions leading to Table 2. Convergence was achieved quickly, usually within five iterations. The nonparametric estimator that emerged tended to have slightly smaller sample standard deviations compared to the trivariate Kaplan–Meier estimator, but also somewhat larger upward biases.

4 Higher-dimensional nonparametric survivor function estimation

Now consider an arbitrary number of failure time variates $T_i > 0, i = 1, \dots, m$. A nonparametric likelihood expression for the joint survivor function F , where $R(t_1, \dots, t_m) = P(T_1 > t_1, \dots, T_m > t_m)$, based on censored failure time data $(X_{i\ell}, \delta_{i\ell})$ for $i = 1, \dots, m, \ell = 1, \dots, n$, is readily specified extending that given in Section 3. It can be seen to be maximized by a discrete survivor function over the risk region of the data $R = \{(t_1, \dots, t_m); X_{i\ell} \geq t_i, \text{ each } i = 1, \dots, m, \text{ for some } \ell \in (1, \dots, n)\}$. A discrete representation for F was not given explicitly in Dabrowska (1988) but can be shown, through massive cancellations, to be given by

$$F(t_1, \dots, t_m) = \prod_{i=1}^m Q_i(t_1, \dots, t_m), \quad (10)$$

where $Q_1(t_1, \dots, t_m)$ is the product of the m single product integrals of the form

$$\prod_0^{t_1} F(s_1, 0, \dots, 0) / F(s_1^-, 0, \dots, 0)$$

for the marginal survival probabilities, $Q_2(t_1, \dots, t_m)$ is the product of the $\binom{m}{2}$ double product integrals of the form

$$\prod_0^{t_1} \prod_0^{t_2} F(s_1, s_2, 0, \dots, 0) / \{F(s_1^-, s_2^-, 0, \dots, 0) / \{F(s_1^-, s_2^-, 0, \dots, 0) F(s_1^-, s_2, 0, \dots, 0)\}\}$$

for the marginal pairwise dependencies, and for $i = 3, \dots, m$ $Q_i(t_1, \dots, t_m)$ is the product of

the $\binom{m}{i}$ i -dimensional product integrals of the form

$$\prod_0^{t_1} \prod_0^{t_2} \dots \prod_0^{t_i} \frac{A(s_1, \dots, s_i, 0, \dots, 0)}{B(s_1, \dots, s_i, 0, \dots, 0)}, \quad (11)$$

where $A(s_1, \dots, s_i, 0, \dots, 0)$ is the product of the 2^{i-1} factors $F(u_1, \dots, u_i, 0, \dots, 0)$ having $u_j = s_j^-$ for an even number of s_j 's and $u_j = s_j$ for the remaining arguments for $j = 1, \dots, i$, and $B(s_1, \dots, s_i, 0, \dots, 0)$ is the product of the complementary 2^{i-1} factors $F(u_1, \dots, u_i, 0, \dots, 0)$ having $u_j = s_j^-$ for an odd number of s_j 's and $u_j = s_j$ for the remaining arguments, $j = 1, \dots, i$. $Q_i(t_1, \dots, t_m)$ characterizes the i -variate dependencies among the m failure time variates, for $i = 2, \dots, m$.

Every survival probability of the form $F(u_1, \dots, u_i, 0, \dots, 0)$ that comprise $A(s_1, \dots, s_i, 0, \dots, 0)$ and $B(s_1, \dots, s_i, 0, \dots, 0)$ in these i -dimensional product integrals, for $i = 1, \dots, m$, can be replaced by $F(u_1, \dots, u_i, 0, \dots, 0) / F(s_1^-, \dots, s_i^-, 0, \dots, 0)$ without changing the right side of (10). A nonparametric estimator $\hat{F}(t_1, \dots, t_m)$ can then be obtained by replacing each such conditional survival probability by its empirical estimator

$$\begin{aligned} \tilde{F}(u_1, \dots, u_i, 0, \dots, 0) \\ = \# \left\{ \ell; X_{j\ell} > s_j \text{ if } u_j = s_j, \text{ or } X_{j\ell} \geq s_j \text{ if } u_j = s_j^-, \text{ all } j = 1, \dots, i \right\} \\ / \# \{ \ell; X_{j\ell} \geq s_j, j = 1, \dots, i \}. \end{aligned}$$

$\hat{F}$ can be calculated by first calculating the m marginal KM estimators, then recursively

calculating the $\binom{m}{2}$ Dabrowska marginal pairwise survival function estimators, continuing through the recursive calculation of the $m-1$ dimensional marginal survival probabilities, and finally calculating $\hat{F}$ away from its coordinate axes recursively by equating the m -variate dependency factor at uncensored data grid point $(t_1, \dots, t_m)$ to its corresponding empirical estimator giving

$$\hat{F}(t_1, \dots, t_m) = \frac{\hat{B}(t_1, \dots, t_m)}{\hat{A}_-(t_1, \dots, t_m)} \frac{\tilde{A}(t_1, \dots, t_m)}{\tilde{B}(t_1, \dots, t_m)}$$

where $\hat{B}(t_1, \dots, t_m)$ is $B(t_1, \dots, t_m)$ with all survival probabilities replaced by $\hat{F}$ value from preceding calculations, $\hat{A}_-(t_1, \dots, t_m)$ drops $F(t_1, \dots, t_m)$ from $A(t_1, \dots, t_m)$ and replaces each remaining survival probability by $\hat{F}$ from earlier calculations, and $\tilde{A}(t_1, \dots, t_m)$ and $\tilde{B}(t_1, \dots, t_m)$ are simply $A(t_1, \dots, t_m)$ and $B(t_1, \dots, t_m)$ with each survival probability replaced by its empirical conditional survival probability $\tilde{F}$ estimator. It follows that calculation of the nonparametric estimator, which can be referred to as the m -dimensional Kaplan–Meier estimator, is practical even for fairly large data sets, at least for a moderate number of failure time variates.

The nonparametric estimator $\hat{F}$ can be used for many exploratory purposes including estimated survival curves for a particular failure time variable under constraints on some or all of the other failure time variates; exploration of the nature and magnitude of any pairwise or higher order dependencies; plots of cumulative single, double, . . . , m -variate hazard functions; and comparisons of the estimators just mentioned among subsets of the study

sample, to name a few. Applications of these types may benefit from corresponding confidence intervals or bounds, requiring distributional approximations for $\hat{F}$.

5 Asymptotic distribution theory

Asymptotic distribution theory for the m -dimensional Kaplan–Meier estimator can be developed using empirical process theory. This approach is described in detail in Gill and Johansen (1990) for the Kaplan–Meier estimator ($m = 1$) and for the Dabrowska estimator ($m = 2$). Briefly for $m = 1$ one can define the counting process N , where

$$N(t) = n^{-1} \# \{k; X_k \leq t, \delta_k = 1\}$$

and the ‘at risk’ process Y , where

$$Y(t) = n^{-1} \# \{k; X_k \geq t\}.$$

The Glivenko–Cantelli and Donsker theorems apply to the empirical process (N, Y) to show it to be supremum norm consistent for, and weakly convergent in a Gaussian manner to its target, over a follow-up interval $[0, \tau]$ where $P(X > \tau) > 0$ as $n \rightarrow \infty$. The continuity and compact differentiability of the transformation from (N, Y) to the hazard function Λ over this region then shows $\hat{\Lambda}$, where

$$\hat{\Lambda}(t) = \sum_{t_i \leq t} N(\Delta t_i) / Y(t_i)$$

to be strongly consistent for its target Λ , and $n^{1/2}(\hat{\Lambda} - \Lambda)$ to be weakly convergent to a mean zero Gaussian process over $[0, \tau]$ as $n \rightarrow \infty$. Finally, the continuity and compact differentiability of the product integral transformation from Λ to F shows the Kaplan–Meier estimator to be strongly consistent for F , and $n^{1/2}(\hat{F} - F)$ to converge weakly to a mean zero Gaussian process. Empirical process theory also shows that bootstrap procedures apply to $\hat{F}$ for confidence interval and confidence band calculations, among other purposes.

The asymptotic theory development is similar for $m > 1$. For example for $m = 2$ one can define single and double failure counting processes N_{10} , N_{01} and N_{11} and at risk process Y by

$$\begin{aligned} N_{10}(t_1, t_2) &= n^{-1} \# \{k; X_{1k} \leq t_1, X_{2k} > t_2, \delta_{1k} = 1\}, \\ N_{01}(t_1, t_2) &= n^{-1} \# \{k; X_{1k} > t_1, X_{2k} \leq t_2, \delta_{2k} = 1\} \\ N_{11}(t_1, t_2) &= n^{-1} \# \{k; X_{1k} \leq t_1, X_{2k} \leq t_2, \delta_{1k} = 1, \delta_{2k} = 1\}, \\ Y(t_1, t_2) &= n^{-1} \# \{k; X_{1k} \geq t_1, X_{2k} \geq t_2\}. \end{aligned}$$

The Glivenko–Cantelli and Donsker theorems show $(N_{10}, N_{01}, N_{11}, Y)$ to be strongly consistent for, and weakly convergent in a Gaussian manner to its target over a region $[0, \tau_1] \times [0, \tau_2]$ where $P(X_1 > \tau_1, X_2 > \tau_2) > 0$ as $n \rightarrow \infty$. The continuity and compact

differentiability of the transformation to single and double failure hazard functions Λ_{10} , Λ_{01} and Λ_{11} shows $(\hat{\Lambda}_{10}, \hat{\Lambda}_{01}, \hat{\Lambda}_{11})$ to also have these asymptotic properties in relation to its target, where for example

$$\begin{aligned}\hat{\Lambda}_{10}(t_1, t_2) &= \sum_{t_{1i} \leq t_1} Y(t_{1i}, t_2)^{-1} N_{10}(\Delta t_{1i}, t_2) \quad \text{and} \\ \hat{\Lambda}_{11}(t_1, t_2) &= \sum_{t_{1i} \leq t_1} \sum_{t_{2j} \leq t_2} Y(t_{1i}, t_{2j})^{-1} N_{11}(\Delta t_{1i}, \Delta t_{2j}).\end{aligned}$$

Then the continuity and compact differentiability of the transformation from $(\Lambda_{10}, \Lambda_{01}, \Lambda_{11})$ to F via the Dabrowska representation

$$F(t_1, t_2) = \prod_0^{t_1} \{1 - \Lambda_{10}(ds_1, 0)\} \prod_0^{t_2} \{1 - \Lambda_{01}(0, ds_2)\} \prod_0^{t_1} \prod_0^{t_2} \left[1 + \frac{\Lambda_{11}(ds_1, ds_2) - \Lambda_{10}(ds_1, s_2^-) \Lambda_{01}(s_1^-, ds_2)}{\{1 - \Lambda_{10}(\Delta s_1, s_2^-)\} \{1 - \Lambda_{01}(s_1^-, \Delta s_2)\}} \right]$$

in conjunction with

$$1 + \frac{\hat{\Lambda}_{11}(\Delta t_1, \Delta t_2) - \hat{\Lambda}_{10}(\Delta t_1, t_2^-) \hat{\Lambda}_{01}(t_1^-, \Delta t_2)}{\{1 - \hat{\Lambda}_{10}(\Delta t_1, t_2^-)\} \{1 - \hat{\Lambda}_{01}(t_1^-, \Delta t_2)\}} = \frac{\hat{F}(t_1, t_2) \hat{F}(t_1^-, t_2^-)}{\hat{F}(t_1, t_2^-) \hat{F}(t_1^-, t_2)}$$

shows $\hat{F}$ to be strongly consistent for F , $n^{1/2}(\hat{F} - F)$ to be weakly convergent to a mean zero Gaussian process, and shows bootstrap procedures to apply to $\hat{F}$, over $[0, \tau_1] \times [0, \tau_2]$.

Asymptotic theory development is unchanged conceptually for $m > 2$. The joint survivor function F can be expressed in terms of single, double, . . . , m -variate hazard functions, each of which has an empirical estimator with asymptotic properties determined by empirical process theory.

More specifically, consider the representation (10) for discrete failure time variates. The survivor function at $(t_1, \dots, t_m)$ is given by (10) with its product integral factors of the form (11). Write

$$\begin{aligned}
& \frac{F(s_1, \dots, s_i, 0, \dots, 0)}{F(s_1^-, \dots, s_i^-, 0, \dots, 0)} \\
&= 1 - \Lambda_{10\dots 0}(ds_1, s_2^-, \dots, s_1^-, 0, \dots, 0) \\
&\quad - \dots - \Lambda_{0\dots 010\dots 0}(s_1^-, \dots, s_{i-1}^-, ds_i, 0, \dots, 0) \\
&\quad + \Lambda_{110\dots 0}(ds_1, ds_2, s_3^-, \dots, s_i^-, 0, \dots, 0) \\
&\quad + \dots + \Lambda_{0\dots 0110\dots 0}(s_1^-, \dots, s_{i-2}^-, ds_{i-1}, ds_i, 0, \dots, 0) \\
&\quad - \Lambda_{1110\dots 0}(ds_1, ds_2, ds_3, s_4^-, \dots, s_i^-, 0, \dots, 0) \\
&\quad - \dots - \Lambda_{0\dots 01110\dots 0}(s_1^-, \dots, s_{i-3}^-, ds_{i-2}, ds_{i-1}, ds_i, 0, \dots, 0) \\
&\quad + \dots + (-1)^i \Lambda_{1\dots 10\dots 0}(ds_1, \dots, ds_i, 0, \dots, 0). \tag{12}
\end{aligned}$$

Denote by $\Lambda_{1\dots 10\dots 0}^0(ds_1, \dots, ds_i, 0, \dots, 0)$ the i -variate hazard rate in (12) which gives a value $B(s_1, \dots, s_i, 0, \dots, 0) / \{A_-(s_1, \dots, s_i, 0, \dots, 0) F(s_1^-, \dots, s_i^-, 0, \dots, 0)\}$ for the left side of (12), given corresponding single, double, $\dots$, $i-1$ variate hazard rates at $(s_1, \dots, s_i, 0, \dots, 0)$, for all $(s_1, \dots, s_i, 0, \dots, 0)$. Then one can write

$$\begin{aligned}
& \prod_0^{t_1} \dots \prod_0^{t_i} \frac{A(s_1, \dots, s_i, 0, \dots, 0)}{B(s_1, \dots, s_i, 0, \dots, 0)} \\
&= \prod_0^{t_1} \dots \prod_0^{t_i} \left[1 + \left\{ \Lambda_{1\dots 10\dots 0}(ds_1, \dots, ds_i, 0, \dots, 0) - \Lambda_{1\dots 10\dots 0}^0(ds_1, \dots, ds_1, 0, \dots, 0) \right\} F(t_1^-, \dots, t_i^-, 0, \dots, 0) \right] \frac{A_-(t_1, \dots, t_i, 0, \dots, 0)}{B(t_1, \dots, t_i, 0, \dots, 0)} \tag{13}
\end{aligned}$$

The right side of this expression is in standard product integral form with all elements except $\Lambda_{1\dots 10\dots 0}(ds_1, \dots, ds_i, 0, \dots, 0)$ determined by hazard rates at preceding times $\{(v_1, \dots, v_i, 0, \dots, 0), v_j - s_j, j=1, \dots, i \text{ and } (v_1, \dots, v_i, 0, \dots, 0) - (s_1, \dots, s_i, 0, \dots, 0)\}$. The product integral expressions (10) and (11), in the form of the right side of (13) applies to discrete, continuous or mixed failure time variates. The Glivenko–Cantelli and Donsker theorems apply to the $2^m - 1$ counting processes $(N_{10\dots 0}, \dots, N_{0\dots 01}, N_{110\dots 0}, \dots, N_{0\dots 011}, N_{1110\dots 0}, \dots, N_{1\dots 1})$ and the at risk process Y over a region $[0, \tau_1] \times \dots \times [0, \tau_m]$ where $P(X_1 > \tau_1, \dots, X_m > \tau_m) > 0$. The strong consistency, weak Gaussian convergence and bootstrap applicability for these 2^m processes then imply the same properties for the corresponding $2^m - 1$ empirical hazard processes as estimators of corresponding hazard processes and the product integral representation for F in terms of (10), (11) and the hazard rate expressions (13) then imply $\hat{F}$ to be strongly consistent for F , $n^{1/2}(\hat{F} - F)$ to converge in distribution to a mean zero Gaussian process, and bootstrap procedures to apply to $\hat{F}$ over $[0, \tau_1] \times \dots \times [0, \tau_m]$.

Strong consistency, weak convergence to mean zero Gaussian processes and bootstrap applicability for average pairwise and higher order dependency estimators over $[0, \tau_1] \times \dots [0, \tau_m]$ follow readily from these properties for $\hat{F}$ in view of the continuity and compact differentiability of the transformations from $\hat{F}$ to these dependency estimators.

6 Discussion

This paper summarizes a substantial literature on nonparametric estimation of the bivariate survivor function under right censoring, and concludes that simply calculated plug-in estimations by Dabrowska (1988) and Prentice and Cai (1992) are good choices for most practical purposes, even though not nonparametric efficient away from the independence special case. The Dabrowska estimator arises from an attractive product integral decomposition of the bivariate survivor function.

Here, the Dabrowska representation is extended to the joint survivor function for an arbitrary number, m , of failure time variates. A decomposition of F is presented as a product of its m

marginal survivor functions, multiplied by its $\binom{m}{2}$ pairwise dependency functions,

multiplied by its $\binom{m}{3}$ trivariate dependency functions, . . . , and finally multiplied by its m -variate dependency function. These factors can be further expressed in terms of single, double, . . . , m -variate product integrals over functions of F that are readily estimated by inserting empirical estimate of conditional survival rates at grid points determined by the uncensored failure times in the sample. A development is sketched showing the resulting estimator $\hat{F}$ to be strongly consistent for F , showing $n^{1/2}(\hat{F} - F)$ to converge weakly to a mean zero Gaussian process, and showing bootstrap applicability. In spite of some negative probability assignments the estimator performs well in simulation studies with moderate sized samples.

Many failure time data analyses include covariates that may be evolving over the follow-up period of a study. The representation (10) can be generalized to include regression components for marginal hazard rates, and for bivariate and higher dimensional covariance rates. The absolutely continuous version of the representation arising from (10), (11) and (13) is particularly promising for this purpose. This topic will be developed by the authors for a separate communication.

Acknowledgements

Ross Prentice's work was partially supported by a grant (P01 CA-53996-38) from the National Cancer Institute. Shanshan Zhao's work was supported in part by the International Research Program of the NIH, National Institute of Environmental Health Sciences.

References

- Akritis MG, Van Keilegom I. Estimation of the bivariate and marginal distributions with censored data. *Journal of the Royal Statistical Society, Series B.* 2003; 65:457–471.
- Campbell G. Nonparametric bivariate estimation with randomly censored data. *Biometrika.* 1981; 68(2):417–422.

- Clayton DG. Model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*. 1978; 65:141–151.
- Cox DR. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society Series B (Methodological)*. 1972; 34(2):187–220.
- Dabrowska DM. Kaplan–Meier estimate on the plane. *Annals of Statistics*. 1988; 16:1475–1489.
- Fan J, Prentice RL, Hsu L. A class of weighted dependence measures for bivariate failure time data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2000; 62(1):181–190.
- Gill RD, Johansen S. A survey of product integration with a view towards application in survival analysis. *Annals of Statistics*. 1990; 18:1501–1555.
- Gill RD, van der Laan MJ, Wellner JA. Inefficient estimators of the bivariate survival function for three models. *Ann Inst H Poincaré Probab Statist*. 1995; 31(3):545–597.
- Hanley JA, Parnes MN. Nonparametric estimation of a multivariate distribution in the presence of censoring. *Biometrics*. 1983; 39(1):129–139. [PubMed: 6871341]
- Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*. 1958; 53(282):457–481.
- Li Y, Prentice RL, Lin X. Semiparametric maximum likelihood estimation in normal transformation models for bivariate survival data. *Biometrika*. 2008; 95:947–960. [PubMed: 19079778]
- Oakes D. A model for association in bivariate survival data. *Journal of the Royal Statistical Society Series B (Methodological)*. 1982:414–422.
- Oakes D. Semiparametric inference in a model for association in bivariate survival data. *Biometrika*. 1986; 73(2):353–361.
- Oakes D. Bivariate survival models induced by frailties. *Journal of the American Statistical Association*. 1989; 84(406):487–493.
- Prentice RL. Self-consistent nonparametric maximum likelihood estimator of the bivariate survivor function. *Biometrika*. 2014; 101(3):505–518. [PubMed: 25632162]
- Prentice RL. Higher dimensional Clayton–Oakes models for multivariate failure time data. *Biometrika*. 2016; 103:231–236. [PubMed: 27738350]
- Prentice RL, Cai J. Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika*. 1992; 79(3):495–512.
- Tsai WY, Crowley J. A large sample study of generalized maximum likelihood estimators from incomplete data via self-consistency. *The Annals of Statistics*. 1985:1317–1334.
- Tsai WY, Leurgans S, Crowley J. Nonparametric estimation of a bivariate survival function in the presence of censoring. *Ann Statist*. 1986; 14(4):1351–1365.
- van der Laan MJ. Efficient estimation in the bivariate censoring model and repairing NPMLE. *Annals of Statistics*. 1996; 24:596–627.

Table 1

Simulation summary statistics* for Dabrowska ($\hat{F}_D$) and Prentice–Cai ($\hat{F}_{PC}$) bivariate survivor function estimators under the Clayton (1978) model with cross ratios $1 + \theta$

Marginal Survival Rates F	$\theta = 0$ (independence)											
	(0.85, 0.85)				(0.70, 0.85)				(0.55, 0.85)			
	0.723				0.595				0.468			
n	30	60	120	240	30	60	120	240	30	60	120	240
$\hat{F}_D$	0.725	0.723	0.724	0.724	0.604	0.596	0.596	0.596	0.472	0.470	0.471	0.470
(SD $\hat{F}_D$)	(0.088)	(0.063)	(0.045)	(0.031)	(0.110)	(0.073)	(0.055)	(0.038)	(0.131)	(0.090)	(0.064)	(0.045)
$\hat{F}_{PC}$	0.725	0.723	0.724	0.724	0.604	0.596	0.596	0.596	0.472	0.469	0.471	0.470
(SD $\hat{F}_{PC}$)	(0.088)	(0.063)	(0.045)	(0.031)	(0.108)	(0.073)	(0.055)	(0.038)	(0.125)	(0.090)	(0.064)	(0.045)
	(0.70, 0.70)				(0.55, 0.70)				(0.55, 0.55)			
	0.490				0.389				0.303			
n	30	60	120	240	30	60	120	240	30	60	120	240
$\hat{F}_D$	0.498	0.491	0.492	0.492	0.391	0.391	0.391	0.389	0.307	0.305	0.309	0.307
(SD $\hat{F}_D$)	(0.116)	(0.078)	(0.057)	(0.404)	(0.140)	(0.093)	(0.063)	(0.045)	(0.162)	(0.113)	(0.069)	(0.047)
$\hat{F}_{PC}$	0.499	0.491	0.492	0.492	0.395	0.390	0.390	0.389	0.322	0.311	0.308	0.307
(SD $\hat{F}_{PC}$)	(0.113)	(0.077)	(0.056)	(0.404)	(0.126)	(0.090)	(0.063)	(0.045)	(0.133)	(0.094)	(0.066)	(0.046)
$\theta = 4$ (strong positive dependence)												
Marginal Survival Rates	(0.85, 0.85)				(0.70, 0.85)				(0.55, 0.85)			
F	0.771				0.666				0.539			
n	30	60	120	240	30	60	120	240	30	60	120	240
$\hat{F}_D$	0.774	0.773	0.773	0.772	0.674	0.668	0.668	0.667	0.541	0.540	0.542	0.541
(SD $\hat{F}_D$)	(0.080)	(0.570)	(0.042)	(0.030)	(0.102)	(0.069)	(0.052)	(0.037)	(0.134)	(0.089)	(0.061)	(0.046)
$\hat{F}_{PC}$	0.773	0.772	0.773	0.772	0.673	0.667	0.667	0.667	0.539	0.539	0.542	0.541
(SD $\hat{F}_{PC}$)	(0.080)	(0.570)	(0.042)	(0.030)	(0.102)	(0.070)	(0.052)	(0.037)	(0.131)	(0.089)	(0.061)	(0.046)
	(0.70, 0.70)				(0.55, 0.70)				(0.55, 0.55)			

F	0.608				0.516				0.468			
	30	60	120	240	30	60	120	240	30	60	120	240
n												
$\hat{F}_D$	0.613	0.607	0.609	0.609	0.524	0.519	0.520	0.519	0.471	0.469	0.474	0.472
(SD $\hat{F}_D$)	(0.112)	(0.074)	(0.056)	(0.034)	(0.139)	(0.092)	(0.061)	(0.044)	(0.177)	(0.110)	(0.069)	(0.049)
$\hat{F}_{PC}$	0.610	0.605	0.607	0.608	0.517	0.515	0.518	0.518	0.466	0.462	0.468	0.469
(SD $\hat{F}_{PC}$)	(0.109)	(0.073)	(0.055)	(0.039)	(0.125)	(0.087)	(0.060)	(0.044)	(0.134)	(0.093)	(0.064)	(0.047)

* Tabular entries are the sample mean (sample standard deviation) based on all simulations at which the point in question is in the risk region of the data. Numbers of contributing samples range from a low of 866 at $\theta = 0$ and marginal survival rates of (0.55,0.55) to 1000, with counts of 990 or more at $n = 120$, and counts of 1000 at $n = 240$.

Table 2

Simulation summary statistics^{*} for the trivariate Kaplan–Meier estimator $\hat{F}$, under the Clayton–Oakes survivor function (8) with unit exponential margins. Censoring variates were independent exponentials with mean 0.5.

				$\theta = 0$			$\theta = 2$			$\theta = 5$					
				F	$n = 30$	$\hat{F}$	$n = 120$	F	$n = 30$	$\hat{F}$	$n = 120$	F	$n = 30$	$\hat{F}$	$n = 120$
	t_1	t_2	t_3												
	0.163	0.163	0.163	0.614	0.614(0.099)	0.614(0.070)	0.616(0.048)	0.682	0.688(0.095)	0.683(0.066)	0.681(0.046)	0.732	0.737(0.089)	0.735(0.062)	0.730(0.043)
	0.163	0.163	0.357	0.506	0.506(0.114)	0.503(0.077)	0.506(0.055)	0.597	0.601(0.111)	0.598(0.074)	0.598(0.051)	0.652	0.654(0.101)	0.654(0.071)	0.653(0.050)
	0.163	0.163	0.598	0.397	0.394(0.135)	0.395(0.089)	0.396(0.064)	0.495	0.495(0.135)	0.495(0.090)	0.496(0.060)	0.537	0.539(0.131)	0.539(0.089)	0.538(0.060)
	0.163	0.357	0.357	0.417	0.412(0.129)	0.416(0.081)	0.420(0.058)	0.537	0.541(0.116)	0.539(0.079)	0.539(0.054)	0.607	0.610(0.110)	0.611(0.075)	0.608(0.054)
	0.163	0.357	0.598	0.327	0.325(0.139)	0.323(0.095)	0.327(0.069)	0.640	0.459(1.148)	0.458(0.099)	0.461(0.062)	0.521	0.519(0.147)	0.522(0.096)	0.522(0.062)
	0.163	0.598	0.598	0.257	0.265(0.147)	0.255(0.110)	0.259(0.076)	0.408	0.403(0.166)	0.407(0.116)	0.412(0.078)	0.478	0.474(0.163)	0.481(0.121)	0.480(0.074)
	0.357	0.357	0.357	0.343	0.339(0.140)	0.337(0.095)	0.345(0.062)	0.493	0.492(0.132)	0.492(0.085)	0.495(0.057)	0.575	0.576(0.127)	0.579(0.079)	0.575(0.056)
	0.357	0.357	0.598	0.270	0.266(0.145)	0.261(0.108)	0.265(0.074)	0.431	0.425(0.158)	0.431(0.102)	0.432(0.065)	0.507	0.506(0.157)	0.508(0.105)	0.508(0.065)
	0.357	0.598	0.598	0.212	0.216(0.143)	0.206(0.115)	0.214(0.077)	0.388	0.383(0.170)	0.390(0.123)	0.392(0.082)	0.470	0.468(0.164)	0.471(0.128)	0.470(0.082)
	0.598	0.598	0.598	0.166	0.178(0.139)	0.168(0.103)	0.172(0.087)	0.355	0.354(0.177)	0.359(0.136)	0.360(0.093)	0.444	0.434(0.190)	0.438(0.152)	0.445(0.098)

^{*} Tabular entries are sample mean (sample standard deviation) based on all simulations at which the point (t_1, t_2, t_3) is in the risk region of the data.

Table 3

Simulation summary statistics^{*} for the average pairwise marginal dependency estimator $\hat{C}_{110}$ and for the average trivariate dependency estimator $\hat{C}_{111}$ under the survivor function (8) with $\theta = 2$ and $n = 120$ with no censoring

t_1	t_2	C_{110}	$\hat{C}_{110}(\text{SD})$	t_1	t_2	t_3	C_{111}	$\hat{C}_{111}(\text{SD})$
0.163	0.163	3	3.103(1.309)	0.163	0.163	0.163	2.143	2.267(1.210)
0.163	0.357	3	3.098(0.991)	0.163	0.163	0.357	2.143	2.250(0.869)
0.163	0.598	3	3.081(0.896)	0.163	0.163	0.598	2.143	2.237(0.732)
0.357	0.357	3	3.060(0.719)	0.163	0.357	0.357	2.143	2.243(0.656)
0.357	0.598	3	3.050(0.650)	0.163	0.357	0.598	2.143	2.230(0.568)
0.598	0.598	3	3.034(0.575)	0.163	0.598	0.598	2.143	2.232(0.513)
				0.357	0.357	0.357	2.143	2.248(0.503)
				0.357	0.357	0.598	2.143	2.230(0.434)
				0.357	0.598	0.598	2.143	2.231(0.390)
				0.598	0.598	0.598	2.143	2.236(0.352)

* Tabular entries are sample means (sample standard deviations) based on all simulations for which the point $(t_1, t_2, 0)$ for C_{110} , and the point (t_1, t_2, t_3) for C_{111} , is in the risk region for the data.

Table 4

Simulation summary statistics* for the trivariate Kaplan–Meier estimator $\hat{F}$, under the trivariate normal transformation survivor function (9) with unit exponential margins. Censoring variates were independent exponentials with mean 0.5.

			$\rho = 0$			$\rho = 0.2$			$\rho = 0.5$		
t_1	t_2	t_3	F		$\hat{F}$	F		$\hat{F}$	F		$\hat{F}$
			$n = 30$	$n = 60$	$n = 120$	$n = 30$	$n = 60$	$n = 120$	$n = 30$	$n = 60$	$n = 120$
0.163	0.163	0.163	0.614	0.616(0.100)	0.614(0.069)	0.613(0.049)	0.642	0.645(0.098)	0.642(0.066)	0.642(0.048)	0.693(0.046)
0.163	0.163	0.357	0.506	0.513 (0.109)	0.504(0.075)	0.505(0.055)	0.541	0.546(0.111)	0.539(0.075)	0.541(0.055)	0.598(0.052)
0.163	0.163	0.598	0.397	0.405(0.130)	0.392(0.089)	0.395(0.063)	0.434	0.442(0.129)	0.429(0.090)	0.433(0.061)	0.489(0.062)
0.163	0.357	0.357	0.417	0.422(0.113)	0.415(0.081)	0.414(0.058)	0.459	0.464(0.116)	0.458(0.081)	0.457(0.059)	0.525(0.057)
0.163	0.357	0.598	0.327	0.339(0.126)	0.326(0.092)	0.325(0.063)	0.371	0.387(0.131)	0.368(0.092)	0.369(0.065)	0.438(0.063)
0.163	0.598	0.598	0.257	0.288 (0.133)	0.267 (0.097)	0.261 (0.070)	0.303	0.333(0.137)	0.310(0.097)	0.306(0.069)	0.371(0.070)
0.357	0.357	0.357	0.343	0.353 (0.118)	0.343 (0.086)	0.341 (0.060)	0.393	0.402(0.120)	0.394(0.086)	0.392(0.059)	0.471(0.059)
0.357	0.357	0.598	0.270	0.289(0.126)	0.274(0.095)	0.271(0.065)	0.320	0.338(0.133)	0.324(0.095)	0.323(0.065)	0.400(0.064)
0.357	0.598	0.598	0.212	0.247(0.131)	0.228(0.095)	0.221(0.068)	0.263	0.296(0.142)	0.278(0.096)	0.270(0.067)	0.345(0.072)
0.598	0.598	0.598	0.166	0.210(0.130)	0.187(0.094)	0.179(0.067)	0.218	0.255(0.143)	0.238(0.101)	0.227(0.072)	0.306(0.075)

* Tabular entries are sample mean (sample standard deviation) based on all simulations at which the point (t_1, t_2, t_3) is in the risk region of the data.

Table 5

Simulation summary statistics^{*} for the average pairwise marginal dependency estimator $\hat{C}_{110}$ and for the average trivariate dependency estimator $\hat{C}_{111}$ under the survivor function (9) with $\rho = 0.2$ and $n = 120$ with no censoring

t_1	t_2	C_{110}	$\hat{C}_{110}(\text{SD})$	t_1	t_2	t_3	$\hat{C}_{111}(\text{SD})$
0.163	0.163	1.721	1.823(0.956)	0.163	0.163	0.163	1.427(1.575)
0.163	0.357	1.592	1.686(0.666)	0.163	0.163	0.357	1.422(1.280)
0.163	0.598	1.521	1.591(0.560)	0.163	0.163	0.598	1.516(1.294)
0.357	0.357	1.487	1.543(0.429)	0.163	0.357	0.357	1.348(0.805)
0.357	0.598	1.428	1.473(0.343)	0.163	0.357	0.598	1.282(0.618)
0.598	0.598	1.375	1.403(0.265)	0.163	0.598	0.598	1.262(0.506)
				0.357	0.357	0.357	1.279(0.489)
				0.357	0.357	0.598	1.239(0.407)
				0.357	0.598	0.598	1.207(0.324)
				0.598	0.598	0.598	1.183(0.261)

* Tabular entries are sample means (sample standard deviations) based on all simulations for which the point $(t_1, t_2, 0)$ for C_{110} , and the point (t_1, t_2, t_3) for C_{111} , is in the risk region for the data.

Table 6

Additional simulation summary statistics^{*} for the average trivariate dependency estimate $\hat{C}_{111}$ under the Clayton–Oakes survivor function model (8) with unit exponential marginals and no censoring.

t_1	t_2	t_3	$C_{111}(\text{SD})$	$\hat{C}_{111}(\text{SD}), n = 20$	$\hat{C}_{111}(\text{SD}), n = 180$
0.163	0.163	0.163	2.143	2.267(1.210)	2.256(0.983)
0.163	0.163	0.357	2.143	2.250(0.869)	2.213(0.694)
0.163	0.163	0.598	2.143	2.237(0.732)	2.219(0.599)
0.163	0.357	0.357	2.143	2.243(0.656)	2.202(0.527)
0.163	0.357	0.598	2.143	2.230(0.568)	2.203(0.452)
0.163	0.598	0.598	2.143	2.232(0.513)	2.198(0.389)
0.357	0.357	0.357	2.143	2.248(0.503)	2.197(0.407)
0.357	0.357	0.598	2.143	2.230(0.434)	2.197(0.347)
0.357	0.598	0.598	2.143	2.231(0.390)	2.195(0.301)
0.598	0.598	0.598	2.143	2.236(0.352)	2.189(0.262)

^{*} Tabular entries are sample means (sample standard deviations) based on all simulations for which the point $(t_1, t_2, 0)$ is in the risk region for the data.