

RESEARCH

Open Access



# Nonparametric rank based estimation of bivariate densities given censored data conditional on marginal probabilities

Alan D. Hutson

Correspondence:  
alan.hutson@roswellpark.org  
Roswell Park Cancer Institute,  
Department of Biostatistics and  
Bioinformatics, Elm and Carlton  
Streets, Buffalo, NY 14263, USA

## Abstract

In this note we develop a new Kaplan-Meier product-limit type estimator for the bivariate survival function given right censored data in one or both dimensions. Our derivation is based on extending the constrained maximum likelihood density based approach that is utilized in the univariate setting as an alternative strategy to the approach originally developed by Kaplan and Meier (1958). The key feature of our bivariate survival function is that the marginal survival functions correspond exactly to the Kaplan-Meier product limit estimators. This provides a level of consistency between the joint bivariate estimator and the marginal quantities as compared to other approaches. The approach we outline in this note may be extended to higher dimensions and different censoring mechanisms using the same techniques.

**Keywords:** Product-limit estimator, Bivariate survival function, Maximum likelihood, Linear programming

**Mathematics Subject Classification (MSC):** 62N01; 62G07

## 1 Introduction

In this note we develop a new Kaplan-Meier product-limit type estimator for the bivariate survival function given right censored data in one or both dimensions. Our derivation is based on extending the constrained maximum likelihood density based approach (Satten and Datta 2001; Zhou 2005) that is utilized in the univariate setting as an alternative strategy to the classical discrete nonparametric hazard function approach (Kaplan and Meier 1958). There are several methods for estimating a bivariate survival function across different censoring patterns that have been proposed in the literature based on extending the univariate hazard function approach or creating various decompositions (Akritas and Van Keilegom 2003; Gill et al. 1995; Lin and Ying 1993; Prentice et al. 2004; Wang and Wells 1997). In general, they are somewhat complex to compute and may have deficiencies such as negative mass estimates at given points (Prentice et al. 2004). The large sample theory involving these estimators is quite technical (Gill et al. 1995). To the best of our knowledge one of the key limitations of all of the completely nonparametric bivariate survival function estimators developed to-date is that they yield marginal estimators that may not be equivalent to the product-limit estimator corresponding to each dimension. Our estimator, framed as a sparse multinomial estimation problem given simplex constraints, remedies this issue. In addition, in terms of future work our method may

be extended to higher dimensions and other censoring mechanisms (left and interval) using the techniques outlined in this note. We also can consider support over the entire real line.

In terms of background, we start by outlining the nonparametric maximum likelihood based density estimator in the univariate setting given right censored data. We can utilize this estimator to define a survival function estimator, which is equivalent to the product-limit estimator. We then move to the bivariate setting in the next section as a direct extension of this approach. As noted above there are two approaches towards arriving at the Kaplan-Meier product-limit estimator (Kaplan and Meier 1958). The well-known nonparametric textbook approach focuses on utilizing the discrete hazard function to define the parameters of interest. Towards this end let  $X_1, X_2, \dots, X_n$  denote i.i.d. failure times and let  $C_1, C_2, \dots, C_n$  denote the corresponding i.i.d. non-informative right censoring times,  $i = 1, 2, \dots, n$ . Given right censoring we only observe  $g \leq n$  of the  $X$ 's. Now let  $0 < x_{(1)} < x_{(2)} < \dots < x_{(g)}$  be the distinct ordered observed failure times. The classic maximum likelihood based derivation of the product-limit estimator starts by assuming the underlying distribution is discrete with probabilities  $\pi_j = P(X = x_{(j)}), j = 1, 2, \dots, g$  for  $g \leq n$ . Given a discrete hazard of  $h_j = P(X = x_{(j)} | X \geq x_{(j)})$  for  $0 \leq h_j \leq 1$  we have that  $\pi_1 = h_1, \pi_2 = (1 - h_1)h_2, \dots, \pi_g = (1 - h_1)(1 - h_2) \dots (1 - h_{g-1})h_g$ . Then the estimator of  $S(x) = P(X > x) = \prod_{x_{(j)} \leq x} (1 - h_j)$  in the discrete case is given as

$$\hat{S}(x) = \prod_{x_{(j)} \leq x} (1 - \hat{h}_j). \tag{1}$$

The estimates of the discrete hazard parameters are obtained from maximizing the log-likelihood

$$\log L = \sum_{j=1}^g d_j \log h_j + (r_j - d_j) \log(1 - h_j), \tag{2}$$

where  $d_j$  denotes the number of events and  $r_j$  denotes the number at risk at time  $x_{(j)}, j = 1, 2, \dots, g$ , e.g. See Cox and Oakes (1984) for details of this derivation. The maximization of (2) with respect to the parameters  $h_j, j = 1, 2, \dots, g$ , yields the oft utilized estimates  $\hat{h}_j = d_j/r_j$ , where  $r_j$  denotes the number of subjects at risk at time  $x_{(j)}$  and  $d_j$  denotes the number of subjects who fail at time  $x_{(j)}$ . For a technical treatment with respect to the behavior of the product-limit estimator and how it translates to continuous case see Chen and Lo (1997). It is well-known, but not immediately obvious, that the product-limit estimator reduces to the classic empirical estimator of the survival function  $\hat{S}(x) = 1 - \sum_{i=1}^n I(x_{(i)} \leq x)/n$  when there are no censored observations, where  $I(\cdot)$  denotes the indicator function. This is the starting point for most of the bivariate survival estimators found in the literature (Gill et al. 1995).

As an alternative approach to the Kaplan-Meier construction we start by estimating the density function first in a nonparametric fashion from which the survival function and distribution functions are then readily estimated. This approach mirrors the classic parametric maximum likelihood estimation given right censored data in terms of the likelihood containing a density function component and a survival function component whose relative contributions depends upon whether or not an observation is censored. In this framework denote the observed values as  $T_i = \min(X_i, C_i), i = 1, 2, \dots, n$ . Note that in our alternative derivation we allow the more general assumption that both  $X$  and  $C$  may have support over the entire real line as compared to the more common

restriction in survival modeling that  $X$  and  $C$  have support only on the positive real line. Furthermore, denote the censoring indicator variable as  $\delta_i = I_{(X_i \leq C_i)}$ , denote the ordered observed  $T_i$ 's as  $t_{(1)} < t_{(2)} < \dots < t_{(n)}$  and define the parameters of interest as  $\pi_i = P(X \leq t_{(i)} | \delta_i = 1) - P(X < t_{(i)} | \delta_i = 1)$ ,  $i = 1, 2, \dots, n$ . The parameter definition is the justification and linkage for this estimator towards underlying continuous data (Owen 1988). Note that similar to the traditional product-limit estimator  $\pi_i = 0$  if  $\delta_i = 0$  by definition. Now, given right-censoring we only observe  $j \leq n$  of the  $X$ 's, where  $j = \sum_{i=1}^n \delta_i$ .

Maximum likelihood estimation is now carried forth under the constraint that  $\sum_{i=1}^n \pi_i = 1$  similar to classic maximum likelihood based empirical density estimation. The classic product-limit estimator is derived through a straightforward extension of the uncensored case and starts with the same assumption that the  $X_i$ 's are functionally discrete, i.e. the true distributions of interest are continuous and we are discretizing the time scale with respect to are definition of the  $\pi_i$ 's. In the case of observed ties in the data under the assumption of a truly continuous underlying distribution we can arbitrarily rank order those respective observations and combine the point masses corresponding to the respective given ties. In our alternative formulation the likelihood accounting for right-censoring now takes a form similar to the parametric setting and is given by

$$L = \prod_{i=1}^{n-1} \pi_i^{\delta_i} \left( 1 - \sum_{j=1}^i \pi_j \right)^{1-\delta_i} \times \left( 1 - \sum_{j=1}^{n-1} \pi_j \right) \tag{3}$$

The last term in the likelihood corresponds to the constraint that  $\sum_{i=1}^n \pi_i = 1$ . If the last observation is censored this implies as in the traditional approach that we still may have an improper distribution of the survival function. Hence, by definition we set this to be an uncensored observation per asymptotic consistency arguments (Chen and Lo 1997). Obviously the likelihood at (3) reduces to the likelihood for the classical empirical estimator given no censoring with  $\hat{\pi}_i = 1/n$ .

The form of the likelihood at (3) has been presented in other contexts such as empirical likelihood constrained maximum likelihood estimation and inference, e.g. see Zhou (2005) and the references there within. The constraint that  $\sum_{i=1}^n \pi_i = 1$  yields  $n - 1$  score equations given by  $s_j = \partial \log L / \partial \pi_j$ . Solving the system of equations that  $s_j = 0$  for  $j = 1, 2, \dots, n - 1$  yields the following nonparametric maximum likelihood estimates for the  $\pi_i$ 's given as

$$\hat{\pi}_i = \begin{cases} \frac{\delta_1}{n}, & i = 1, \\ \frac{\prod_{j=1}^{i-1} ((n-j+1) - \delta_j) \delta_i}{\prod_{j=1}^i (n+j-i)}, & i > 1, \end{cases} \tag{4}$$

where  $\hat{\pi}_n = 1 - \sum_{j=1}^{n-1} \hat{\pi}_j$ , see Satten and Datta (2001).

In addition, it follows straightforward that similar to the standard empirical distribution function estimator we have for right-censored data that

$$\hat{F}(x) = \sum_{i=1}^n \hat{\pi}_i I_{(t_{(i)} \leq x)} \tag{5}$$

and

$$\hat{S}(x) = 1 - \hat{F}(x) = 1 - \sum_{i=1}^n \hat{\pi}_i I_{(t_{(i)} \leq x)}, \tag{6}$$

where  $\hat{\pi}_i$ 's are given at (4) and  $\hat{\pi}_n = 1 - \sum_{j=1}^{n-1} \hat{\pi}_j$ . The estimator of the survival function given at (6) is equivalent to the product-limit form at (1) in terms of the actual estimated survival probabilities. This is the jumping off point of our new bivariate survival function estimator.

In Section 2 we outline the new constrained maximum likelihood procedure used to develop a bivariate estimator of the joint density  $f(x, y)$  from which estimators of the bivariate distribution function  $F(x, y)$  and bivariate survival function are readily calculated. In Section 3 we provide some illustrative toy examples followed by two real data examples. We finish with some basic conclusions.

### 2 Constrained maximum likelihood estimation

In this section we describe the process for estimating  $f(x, y)$  nonparametrically conditional on marginal constraints. This in turn will lead to an estimator for the bivariate distribution function  $F(x, y)$  and survival function  $S(x, y)$ . Towards this end let  $(X_i, Y_i)$ ,  $i = 1, 2, \dots, n$ , be independent and identically distributed pairs of bivariate failure times with joint probability density function  $f(x, y)$  and corresponding cumulative distribution function  $F(x, y)$  with  $S(x, y) = 1 - F(x, y)$ . Furthermore, let  $(C_i^x, C_i^y)$ ,  $i = 1, 2, \dots, n$ , be independent and identically bivariate distributed pairs of censoring variables. Under right censoring in each dimension we observe

$$(S_i, T_i) = ((X_i \wedge C_i^x), (Y_i \wedge C_i^y)) \text{ and } (\delta_i^x, \delta_i^y) = (I(X_i < C_i^x), I(Y_i < C_i^y)),$$

$$i = 1, 2, \dots, n,$$

where  $\wedge$  denotes the minimum between pairs of random variables and  $I(\cdot)$  denotes the indicator function. For this note we assume the  $(X_i, Y_i)$ 's and  $(C_i^x, C_i^y)$ 's are absolutely continuous and are also pairwise independent from each other. In the nonparametric setting the distribution and survival functions are defined as follows:

$$F(x, y) = \sum_{i=1}^n \sum_{j=1}^n \pi_{i,j} I_{(s_{(i)} \leq x, t_{(j)} \leq y)} \tag{7}$$

$$S(x, y) = 1 - F(x, y), \tag{8}$$

where the parameters, i.e. the  $\pi_{i,j}$ 's,  $i = 1, 2, \dots, n$ ,  $j = 1, 2, \dots, n$ , are in essence weights between 0 and 1 and are defined in detail below at (17).

Now similar to the univariate case, outlined in the introduction, denote the parameters corresponding to the nonparametric estimators of the marginal densities  $f_x$  and  $f_y$  as

$$\pi_{r_{s_i}, \cdot} = \Delta F_s(S_{r_{s_i}}) = F_s(S_{r_{s_i}}) - F_s(S_{r_{s_i}-}) = P(S_{r_{s_i}} \leq s_{r_{s_i}}) - P(S_{r_{s_i}} < s_{r_{s_i}}) \text{ and}$$

$$\pi_{\cdot, r_{t_j}} = \Delta F_t(T_{r_{t_j}}) = F_t(T_{r_{t_j}}) - F_t(T_{r_{t_j}}-) = P(T_{r_{t_j}} \leq t_{r_{t_j}}) - P(T_{r_{t_j}} < t_{r_{t_j}}), \tag{9}$$

where we denote the ranks of the observed failure or censoring times per each margin as  $r_{s_i} = \text{rank}(S_i)$  and  $r_{t_j} = \text{rank}(T_j)$ , respectively, corresponding to the order statistics  $S_{(1)} < S_{(2)} < \dots < S_{(n)}$  and  $T_{(1)} < T_{(2)} < \dots < T_{(n)}$ . The parameters at (9) are instrumental with respect to defining the simplex constraints used in our maximization procedure described below. The inter-relationships between the cell probabilities, the  $\pi_{i,j}$ 's, and the marginal probabilities, the  $\pi_{r_{s_i}, \cdot}$ 's and  $\pi_{\cdot, r_{t_j}}$ 's, are defined in detail below at (20) and (21), respectively.

As described in Section 1, and with a simple modification of the notation, the parameter estimates derived from the form of the likelihood at (3) corresponding to the marginal density  $f_x$  were derived by Satten and Datta (2001) as

$$\hat{\pi}_{i.} = \begin{cases} \frac{\delta_{(1)}^x}{n}, & i = 1, \\ \frac{\prod_{j=1}^{i-1} ((n-j+1) - \delta_{(j)}^x) \delta_{(i)}^x}{\prod_{j=1}^i (n+j-i)}, & i > 1, \end{cases} \tag{10}$$

where we set  $\delta_{(n)}^x = 1$ . In the case of no censoring all  $\hat{\pi}_{i.}$ 's are equal to  $1/n$ .

It follows straightforward that similar to the standard empirical distribution function estimator we have for right-censored data that

$$\hat{F}_x(x) = \sum_{i=1}^n \hat{\pi}_{i.} I_{(s(i) \leq x)} \tag{11}$$

and

$$\hat{S}_x(x) = 1 - \hat{F}_x(x) = 1 - \sum_{i=1}^n \hat{\pi}_{i.} I_{(s(i) \leq x)}, \tag{12}$$

where  $\hat{\pi}_{i.}$ 's are given at (10) and  $\hat{\pi}_{n.} = 1 - \sum_{j=1}^{n-1} \hat{\pi}_{j.}$ . It should be obvious that  $\hat{\pi}_{i.} = 0$  from (10) when  $\delta_{(i)}^x = 0$  in the case of a censored observation.

Similarly, we have the estimated parameters corresponding to the marginal density  $f_y$ ,  $\delta_{(n)}^y = 1$ , given as

$$\hat{\pi}_{.j} = \begin{cases} \frac{\delta_{(1)}^y}{n}, & j = 1, \\ \frac{\prod_{i=1}^{j-1} ((n-i+1) - \delta_{(i)}^y) \delta_{(j)}^y}{\prod_{i=1}^j (n+i-j)}, & j > 1, \end{cases} \tag{13}$$

which yields the marginal estimators of the distribution and survival functions as

$$\hat{F}_y(y) = \sum_{j=1}^n \hat{\pi}_{.j} I_{(t(j) \leq y)} \tag{14}$$

and

$$\hat{S}_y(y) = 1 - \hat{F}_y(y) = 1 - \sum_{j=1}^n \hat{\pi}_{.j} I_{(t(j) \leq y)}, \tag{15}$$

where  $\hat{\pi}_{.j}$ 's are given at (13) and  $\hat{\pi}_{.n} = 1 - \sum_{i=1}^{n-1} \hat{\pi}_{i.}$ .

Let us now define the parameters associated with the nonparametric likelihood corresponding to the joint density  $f(x, y)$ . In terms of determining the relevant parameters for use in the nonparametric likelihood model we need to define an indicator function as a type of bookkeeping feature given censoring information from both marginal distributions. Towards this end let

$$\delta_{i,j} = \begin{cases} 1, & \text{if } \delta_i^x \delta_i^y = 1, \forall i = 1, 2, \dots, n \\ 1, & \text{if } (1 - \delta_i^x) \delta_i^y = 1, \forall i = r_{s_j+1}, \dots, n, j = 1, 2, \dots, n, \\ 1, & \text{if } \delta_i^x (1 - \delta_i^y) = 1, \forall i = 1, 2, \dots, n, j = r_{t_i+1}, \dots, n, \\ 1, & \text{if } (1 - \delta_i^x) (1 - \delta_i^y) = 1, \forall i = r_{s_j+1}, \dots, n, j = r_{t_i+1}, \dots, n, \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

Then  $\forall (r_{s_i}, r_{t_j})$  combinations for which  $\delta_{r_{s_i}, r_{t_j}} \neq 0, \delta_{r_{s_i}}^x \neq 0$  and  $\delta_{r_{t_j}}^y \neq 0, i = 1, 2, \dots, n$  and  $j = 1, 2, \dots, n$  the parameters of interest in our nonparametric model corresponding to the joint density  $f(x, y)$  are given as

$$\begin{aligned} \pi_{r_{s_i}, r_{t_j}} &= \Delta F \left( S_{r_{s_i}}, T_{r_{t_j}} \right) \\ &= F \left( S_{r_{s_i}}, T_{r_{t_j}} \right) - F \left( S_{r_{s_i}}, T_{r_{t_j}} - \right) - F \left( S_{r_{s_i}} - , T_{r_{t_j}} \right) + F \left( S_{r_{s_i}} - , T_{r_{t_j}} - \right) \\ &= P \left( S_{r_{s_i}} \leq s_{r_{s_i}}, T_{r_{t_j}} \leq t_{r_{t_j}} \right) - P \left( S_{r_{s_i}} \leq s_{r_{s_i}}, T_{r_{t_j}} < t_{r_{t_j}} \right) \\ &\quad - P \left( S_{r_{s_i}} < s_{r_{s_i}}, T_{r_{t_j}} \leq t_{r_{t_j}} \right) + P \left( S_{r_{s_i}} < s_{r_{s_i}}, T_{r_{t_j}} < t_{r_{t_j}} \right), \end{aligned} \tag{17}$$

else we define  $\pi_{r_{s_i}, r_{t_j}} = 0$ , i.e.  $\pi_{r_{s_i}, r_{t_j}} = 0$  if  $\delta_{r_{s_i}, r_{t_j}} = 0$  or  $\delta_{r_{s_i}}^x = 0$  or  $\delta_{r_{t_j}}^y = 0$ . In the specific case where there is no censoring for both  $X$  and  $Y$  the number of parameters is of size  $n$  and the corresponding maximum likelihood estimator for  $\pi_{r_{s_i}, r_{t_j}}$  is  $1/n$  as per the standard empirical density estimator, i.e. there is a point mass of  $1/n$  per each set of paired observations.

It now follows similar to that of the univariate case at (3) that the likelihood function for the joint density  $f(x, y)$  defined through the parameters at (17) is given as

$$\begin{aligned} L &= \prod_{i=1}^n \pi_{r_{s_i}, r_{t_i}}^{\delta_i^x \delta_i^y} \left( \sum_{j=r_{s_i}+1}^n \delta_j^x \pi_{j, r_{t_i}} \right)^{(1-\delta_i^x) \delta_i^y} \left( \sum_{k=t_{s_i}+1}^n \delta_k^y \pi_{r_{s_i}, k} \right)^{\delta_i^x (1-\delta_i^y)} \\ &\quad \times \left( \sum_{j=r_{s_i}+1}^n \sum_{k=t_{s_i}+1}^n \delta_j^x \delta_k^y \pi_{j, k} \right)^{(1-\delta_i^x)(1-\delta_i^y)}, \end{aligned} \tag{18}$$

where the components of the likelihood correspond to the four possible right-censoring combinations for the  $(\delta_i^x, \delta_i^y)$  pairs, i.e. there is a simple point mass given no censoring else probability is shifted to the right similar to the classic Kaplan-Meier estimator given the various censoring patterns. The objective is to maximize  $L$  at (18) subject to the simplex constraints

$$\sum_{i=1}^n \sum_{j=1}^n \delta_{i,j} \delta_i^x \delta_j^y \pi_{i,j} = 1, \tag{19}$$

$$\sum_{j=1}^n \delta_{i,j} \delta_j^y \pi_{i,j} = \pi_{i,\cdot}, \tag{20}$$

$$\sum_{i=1}^n \delta_{i,j} \delta_i^x \pi_{i,j} = \pi_{\cdot,j}, \tag{21}$$

$$\text{if } \delta_{i,j} = \delta_i^x = \delta_j^y = 1 \text{ then } 0 \leq \pi_{i,j} \leq 1, \forall i = 1, \dots, n, j = 1, \dots, n. \tag{22}$$

The constraints at (20) and (21) pertain to the marginal constraints where  $\pi_{i,\cdot}$  and  $\pi_{\cdot,j}$  are defined at (9). Our approach is similar to the problems described for multinomial distribution parameter estimation given sparse data and a class of linear simplex constraints (Liu 2000). The argument for replacing  $\pi_{i,\cdot}$  and  $\pi_{\cdot,j}$  at (9) with their corresponding estimators  $\hat{\pi}_{i,\cdot}$  and  $\hat{\pi}_{\cdot,j}$  at (10) and (13), respectively, follows similar to the classic  $R \times C$  contingency table exact inference case. The contribution of the  $\hat{\pi}_{i,\cdot}$ 's and  $\hat{\pi}_{\cdot,j}$ 's to the multinomial distribution given sparse data and linear simplex constraints corresponding to  $\hat{\pi}_{r_{s_i}, r_{t_j}}$ 's of interest depend on the data through the censoring values for the  $\delta^x$ 's and  $\delta^y$ 's. The joint distribution of the  $\hat{\pi}_{i,\cdot}$ 's and  $\hat{\pi}_{r_{s_i}, r_{t_j}}$ 's is identical to the distribution of the

$\hat{\pi}_{r_{s_i}, r_{t_j}}$ 's since by definition the  $\hat{\pi}_{i,\cdot}$ 's are determined by the  $\hat{\pi}_{r_{s_i}, r_{t_j}}$ 's. The same holds in the other dimension relative to the  $\hat{\pi}_{\cdot, j}$ 's, i.e. the  $\hat{\pi}_{r_{s_i}, r_{t_j}}$ 's are sufficient statistics in terms of determining the parameters that define the marginal densities.

**Steps in the parameter estimation:**

1. Given the observed censoring pattern utilize the constraints (19), (20) and (21) to define the parameter space.
2. Obtain the estimates of marginal probabilities  $\pi_{i,\cdot}$  and  $\pi_{\cdot, j}$  given by  $\hat{\pi}_{i,\cdot}$  and  $\hat{\pi}_{\cdot, j}$  at (10) and (13), respectively. Substitute the estimates into (20) and (21) after first processing step 1 above.
3. Utilize standard maximum likelihood technique on the likelihood defined at (18) to solve for the remaining unknown parameters given the constraints defined at (19)–(22).

It then follows that the estimators of the bivariate distribution function and survival function are given as:

$$\hat{F}(x, y) = \sum_{i=1}^n \sum_{j=1}^n \hat{\pi}_{i,j} I(s_{(i)} \leq x, t_{(j)} \leq y) \tag{23}$$

$$\hat{S}(x, y) = 1 - \hat{F}(x, y), \tag{24}$$

respectively. Some small sample toy examples will be provided in the next section in order to illustrate the process followed by some real data examples.

Note that if censoring occurs solely in either the  $x$  or  $y$  dimension the likelihood at (18) reduces substantially in complexity and has a form very similar to that of the univariate setting at (3).

**Comment** Large sample variance estimates for  $\hat{F}(x, y)$  and  $\hat{S}(x, y)$  are conceptually straightforward in that they follow standard methods based on obtaining the co-information matrix with dimensions that vary as a function of the proportion of censored observations. For small samples this is straightforward. However for moderate to large samples and from a programming point of view, this becomes a rather complex computational problem such that we would recommend either bootstrap or jackknife methodologies for the purpose of variance estimation.

**3 Examples**

In this section we provide a few straightforward small sample scenarios in order to illustrate the maximization process for the estimator of  $f(x, y)$  and  $S(x, y)$  given various censoring patterns. This is followed by two real data examples used by previous researchers in the past to illustrate this type of estimator. It is important to note that as we present the results we set  $\delta_{(n)}^x = \delta_{(n)}^y = 1$  by definition. The rationale for this was described above.

**3.1 Toy data examples**

**Example 1.** For  $n = 6$  we have the following data:

$$\begin{aligned} \mathbf{s} &= (6.1, 4.5, 6.2, 4.8, 5.9, 3.3), r_s = (5, 2, 6, 3, 4, 1), \delta^x = (1, 1, 0, 1, 0, 1), \\ \mathbf{t} &= (9.6, 2.6, 7.2, 4.1, 7.7, 5.0), r_t = (6, 1, 4, 2, 5, 3), \delta^y = (0, 1, 1, 1, 0, 1). \end{aligned}$$

The vector of parameters of interest defined by the censoring patterns as per (16) is given as  $\boldsymbol{\pi} = (\pi_{1,3}, \pi_{2,1}, \pi_{3,2}, \pi_{5,6}, \pi_{6,4}, \pi_{6,6})$ . The goal is to maximize the likelihood

$$L = \pi_{1,3}\pi_{2,1}\pi_{3,2}\pi_{5,6}\pi_{6,4} (\pi_{5,6} + \pi_{6,6})$$

subject to the simplex constraints from (19), (20) and (21), respectively, and given as:

1.  $\pi_{1,3} + \pi_{2,1} + \pi_{3,2} + \pi_{5,6} + \pi_{6,4} + \pi_{6,6} = 1,$
2.  $\pi_{1,3} = 1/6, \pi_{2,1} = 1/6, \pi_{3,2} = 1/6, \pi_{5,6} = 1/4, \pi_{6,4} + \pi_{6,6} = 1/4,$
3.  $\pi_{2,1} = 1/6, \pi_{3,2} = 1/6, \pi_{1,3} = 1/6, \pi_{6,4} = 1/6, \pi_{5,6} + \pi_{6,6} = 1/3.$

We can see that in this small sample setting that estimates for  $\boldsymbol{\pi} = (\pi_{1,3}, \pi_{2,1}, \pi_{3,2}, \pi_{5,6}, \pi_{6,4}, \pi_{6,6})$  in this case are determined solely by the marginal constraints. In general, for moderate to large sample sizes this will not be the case.

The estimates given the constraints are provided in Table 1. In this specific case no maximization of the likelihood was needed. Note that the estimators of the marginal survivor functions  $\hat{S}_x(x)$  and  $\hat{S}_y(y)$  from (12) and (15), respectively, and based on the parameter estimates in Table 1 are exactly those corresponding to the product-limit estimator. The plot of the bivariate survival function  $\hat{S}(x, y)$  from (24) is given in Fig. 1, which is clearly monotone decreasing in both dimensions for increasing values of data with positive support.

**Example 2.** For  $n = 6$  we have the following data:

$$\begin{aligned} \mathbf{s} &= (8.9, 2.6, 3.7, 5.9, 7.9, 1.2), r_s = (6, 2, 3, 4, 5, 1), \delta^x = (0, 0, 1, 1, 1, 1), \\ \mathbf{t} &= (6.3, 9.6, 6.8, 0.1, 0.7, 4.2), r_t = (4, 6, 5, 1, 3, 2), \delta^y = (0, 0, 1, 1, 1, 1). \end{aligned}$$

The vector of parameters of interest defined by the censoring patterns as per (16) is given as  $\boldsymbol{\pi} = (\pi_{1,3}, \pi_{3,5}, \pi_{3,6}, \pi_{4,1}, \pi_{4,6}, \pi_{5,2}, \pi_{5,6}, \pi_{6,5}, \pi_{6,6})$ , which is of a slightly higher dimension as compared to Example 1. The goal is to maximize the likelihood

$$L = \pi_{1,3}\pi_{3,5}\pi_{4,1}\pi_{5,2} (\pi_{3,6} + \pi_{4,6} + \pi_{5,6} + \pi_{6,6}) (\pi_{6,5} + \pi_{6,6})$$

subject to the simplex constraints from (19), (20) and (21), respectively, and given as:

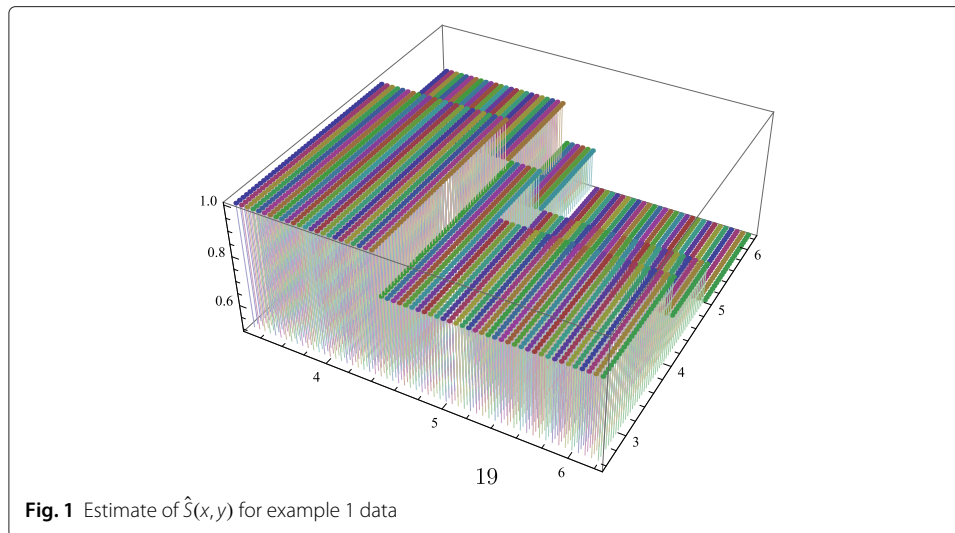
1.  $\pi_{1,3} + \pi_{3,5} + \pi_{3,6} + \pi_{4,1} + \pi_{4,6} + \pi_{5,2} + \pi_{5,6} + \pi_{6,5} + \pi_{6,6} = 1$
2.  $\pi_{1,3} = 1/6, \pi_{3,5} + \pi_{3,6} = 5/24, \pi_{4,1} + \pi_{4,6} = 5/24, \pi_{5,2} + \pi_{5,6} = 5/24, \pi_{6,5} + \pi_{6,6} = 5/24$
3.  $\pi_{4,1} = 1/6, \pi_{5,2} = 1/6, \pi_{1,3} = 1/6, \pi_{3,5} + \pi_{6,5} = 1/4, \pi_{3,6} + \pi_{4,6} + \pi_{5,6} + \pi_{6,6} = 1/4$

The results of the maximum likelihood estimates derived from (18) are provided in Table 2. In this case note that  $\hat{\pi}_{3,6}$  was at the boundary of the feasible region and set equal

**Table 1** Bivariate estimates for  $\hat{\pi}_{ij}, i = 1, 2, \dots, 6, j = 1, 2, \dots, 6$ , corresponding to example 1

$i/j$	$r_t = 1$	$r_t = 2$	$r_t = 3$	$r_t = 4$	$r_t = 5$	$r_t = 6$	$\hat{\pi}_{i.}$	$\delta_{r_{s_i}}^x$
$r_s = 1$	0	0	1/6	0	0	0	1/6	1
$r_s = 2$	1/6	0	0	0	0	0	1/6	1
$r_s = 3$	0	1/6	0	0	0	0	1/6	1
$r_s = 4$	0	0	0	0	0	0	0	0
$r_s = 5$	0	0	0	0	0	1/4	1/4	1
$r_s = 6$	0	0	0	1/6	0	1/12	1/4	1
$\hat{\pi}_{.j}$	1/6	1/6	1/6	1/6	0	1/3	1	
$\delta_{r_{t_j}}^y$	1	1	1	1	0	1		





to 0. The plot of the bivariate survival function  $\hat{S}(x, y)$  from (24) is given in Fig. 2, which as before is clearly monotone decreasing in both dimensions for increasing values of the data with positive support.

**Example 3.** For  $n = 6$  we have the following data:

$$\mathbf{s} = (6.4, 7.0, 7.2, 8.1, 6.2, 7.4), r_s = (2, 3, 4, 6, 1, 5), \delta^x = (0, 1, 0, 0, 1, 0),$$

$$\mathbf{t} = (8.6, 7.5, 8.0, 4.8, 0.8, 4.0), r_t = (6, 4, 5, 3, 1, 2), \delta^y = (0, 1, 0, 0, 1, 0).$$

The vector of parameters of interest defined by the censoring patterns as per (16) is given as  $\boldsymbol{\pi} = (\pi_{1,1}, \pi_{3,4}, \pi_{3,6}, \pi_{6,4}, \pi_{6,6})$ . The goal is to maximize the likelihood

$$L = \pi_{1,1} \pi_{3,4} \pi_{6,6} (\pi_{3,6} + \pi_{6,6}) (\pi_{6,4} + \pi_{6,6})^2$$

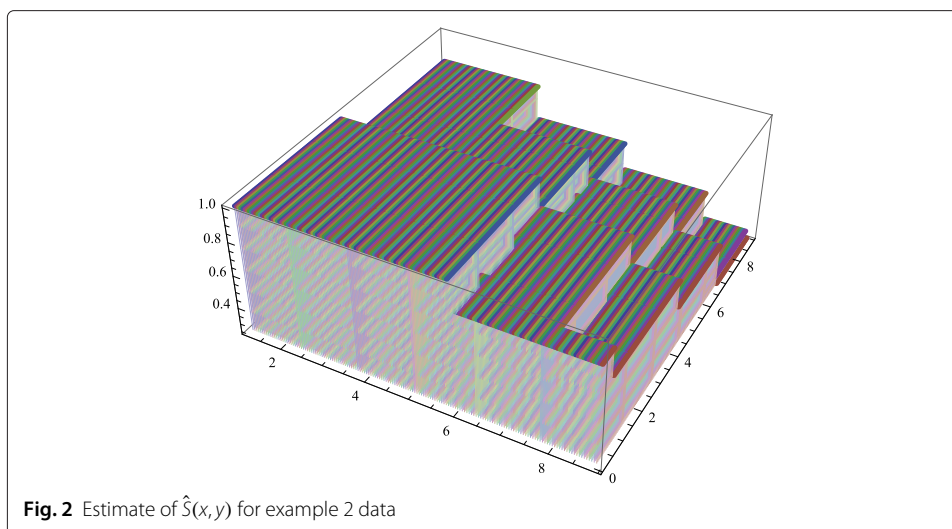
subject to the simplex constraints from (19), (20) and (21), respectively, and given as:

1.  $\pi_{1,1} + \pi_{3,4} + \pi_{3,6} + \pi_{6,4} + \pi_{6,6} = 1$
2.  $\pi_{1,1} = 1/6, \pi_{3,4} + \pi_{3,6} = 5/24, \pi_{6,4} + \pi_{6,6} = 5/8$
3.  $\pi_{1,1} = 1/6, \pi_{3,4} + \pi_{6,4} = 5/18, \pi_{3,6} + \pi_{6,6} = 5/9$

In this example case we have heavy censoring relative to the total number of observations. Similar to example 2,  $\hat{\pi}_{3,6}$  was at the boundary of the feasible region and set equal to 0. The results of the maximum likelihood estimates derived from (18) are provided in Table 3. The plot of the bivariate survival function  $\hat{S}(x, y)$  from (24) is given in Fig. 3,

**Table 2** Bivariate estimates for  $\hat{\pi}_{ij}, i = 1, 2, \dots, 6, j = 1, 2, \dots, 6$ , corresponding to example 2

$i/j$	$r_t = 1$	$r_t = 2$	$r_t = 3$	$r_t = 4$	$r_t = 5$	$r_t = 6$	$\hat{\pi}_{i.}$	$\delta_{r_{ij}}^x$
$r_s = 1$	0	0	1/6	0	0	0	1/6	1
$r_s = 2$	0	0	0	0	0	0	0	0
$r_s = 3$	0	0	0	0	5/24	0	5/24	1
$r_s = 4$	1/6	0	0	0	0	1/24	5/24	1
$r_s = 5$	0	1/6	0	0	0	1/24	5/24	1
$r_s = 6$	0	0	0	0	1/24	1/6	5/24	1
$\hat{\pi}_{.j}$	1/6	1/6	1/6	0	1/4	1/4	1	
$\delta_{r_{ij}}^y$	1	1	1	0	1	1		



which as before is clearly monotone decreasing in both dimensions for increasing values of data with positive support.

Note that if there was no censoring within examples 1–3 then the respective values for  $\hat{\pi}_{r_{s_i}, r_{t_i}}$ ,  $i = 1, 2, \dots, n$ , would be simply  $1/n$ , which corresponds to the maximum likelihood estimates for the classic empirical joint density function.

### 3.2 Real data examples

In this section we re-analyze two sets of real data utilized to demonstrate other approaches to bivariate survival function estimation (Akritas and Van Keilegom 2003; Wang and Wells 1997), see the references contained within relative to the source of the original data.

**Survival days of skin grafts in burn patients (Wang and Wells 1997).** For this data set we have  $n = 11$  paired survival times and censoring indicators for skin grafts in burn patients given as:

$$\mathbf{s} = (37, 19, 57, 93, 16, 22, 20, 18, 63, 29, 60),$$

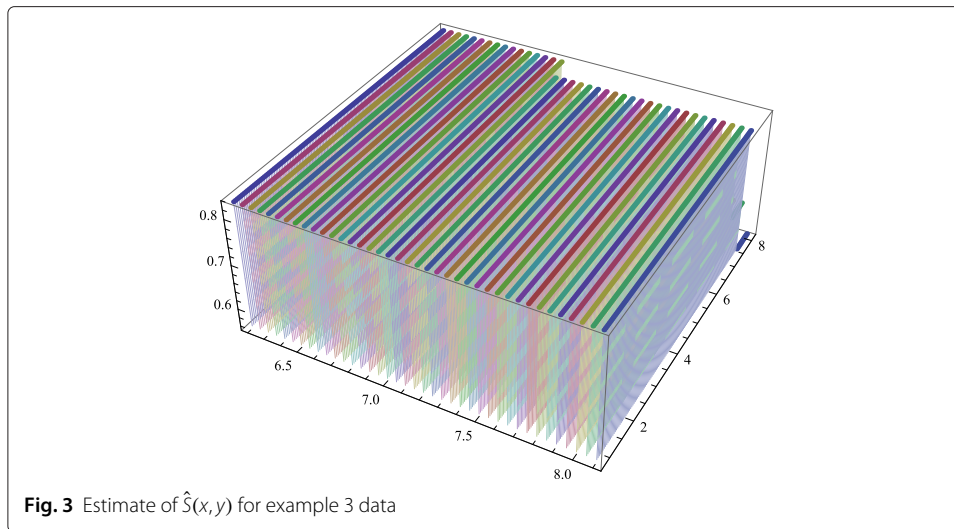
$$\mathbf{t} = (29, 13, 15, 26, 11, 17, 26, 21, 43, 15, 40),$$

$$\delta^x = (1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0),$$

$$\delta^y = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1).$$

**Table 3** Bivariate estimates for  $\hat{\pi}_{ij}$ ,  $i = 1, 2, \dots, 6, j = 1, 2, \dots, 6$ , corresponding to example 3

$i/j$	$r_t = 1$	$r_t = 2$	$r_t = 3$	$r_t = 4$	$r_t = 5$	$r_t = 6$	$\hat{\pi}_{i\cdot}$	$\delta_{r_{s_i}}^x$
$r_s = 1$	1/6	0	0	0	0	0	1/6	1
$r_s = 2$	0	0	0	0	0	0	0	0
$r_s = 3$	0	0	0	5/24	0	0	5/24	1
$r_s = 4$	0	0	0	0	0	0	0	0
$r_s = 5$	0	0	0	0	0	0	0	0
$r_s = 6$	0	0	0	5/72	0	5/9	5/8	1
$\hat{\pi}_{\cdot j}$	1/6	0	0	5/18	0	5/9	1	
$\delta_{r_{t_j}}^y$	1	0	0	1	0	1		



As you can see there is no censoring in the  $y$  component and only a moderate amount of censoring in the  $x$  component. Hence in most instances the  $\hat{\pi}_{r_{s_i}, r_{t_j}}$ 's will be equal to  $1/n$ . In this case the vector of parameters with the corresponding maximum likelihood estimates are given as:  $(\hat{\pi}_{1,1} = \frac{1}{11}, \hat{\pi}_{2,8} = \frac{1}{11}, \hat{\pi}_{3,2} = \frac{1}{11}, \hat{\pi}_{4,6} = \frac{1}{11}, \hat{\pi}_{5,5} = \frac{1}{11}, \hat{\pi}_{6,4} = \frac{1}{11}, \hat{\pi}_{7,9} = \frac{1}{11}, \hat{\pi}_{10,3} = \frac{3}{64}, \hat{\pi}_{10,10} = \frac{31}{704}, \hat{\pi}_{10,11} = \frac{1}{11}, \hat{\pi}_{11,3} = \frac{31}{704}, \hat{\pi}_{11,7} = \frac{1}{11}, \hat{\pi}_{11,10} = \frac{3}{64})$ . The likelihood for this example has the form

$$L = \pi_{1,1}\pi_{2,8}\pi_{3,2}\pi_{4,6}\pi_{5,5}\pi_{6,4}\pi_{7,9}\pi_{10,11} (\pi_{10,3} + \pi_{11,3}) \pi_{11,7} (\pi_{10,10} + \pi_{11,10}) \quad (25)$$

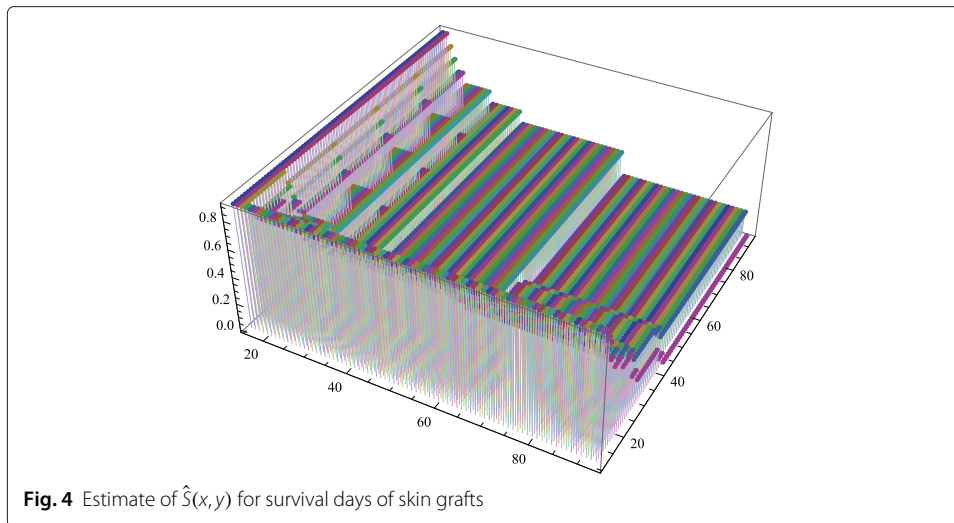
with simplex constraints from (19), (20) and (21), respectively, given as:

1.  $\pi_{1,1} + \pi_{2,8} + \pi_{3,2} + \pi_{4,6} + \pi_{5,5} + \pi_{6,4} + \pi_{7,9} + \pi_{10,3} + \pi_{10,10} + \pi_{10,11} + \pi_{11,3} + \pi_{11,7} + \pi_{11,10} = 1,$
2.  $\pi_{1,1} = 1/11, \pi_{2,8} = 1/11, \pi_{3,2} = 1/11, \pi_{4,6} = 1/11, \pi_{5,5} = 1/11, \pi_{6,4} = 1/11, \pi_{7,9} = 1/11, \pi_{10,3} + \pi_{10,10} + \pi_{10,11} = 2/11, \pi_{11,3} + \pi_{11,7} + \pi_{11,10} = 2/11$
3.  $\pi_{1,1} = 1/11, \pi_{3,2} = 1/11, \pi_{10,3} + \pi_{11,3} = 1/11, \pi_{6,4} = 1/11, \pi_{5,5} = 1/11, \pi_{4,6} = 1/11, \pi_{11,7} = 1/11, \pi_{2,8} = 1/11, \pi_{7,9} = 1/11, \pi_{10,10} + \pi_{11,10} = 1/11, \pi_{10,11} = 1/11.$

Again, we see the most of the parameters in the likelihood are determined via the marginal constraints, which should be obvious given the low percentage of censored

**Table 4** Estimated bivariate survival probabilities for skin graft data evaluated at the marginal quartiles

$u_x$	$u_y$	$\hat{Q}_x(u_x)$	$\hat{Q}_y(u_y)$	$\hat{S}(\hat{Q}_x(u_x), \hat{Q}_y(u_y))$
0.25	0.25	19.25	15.0	0.82
0.25	0.5	19.25	21.0	0.82
0.25	0.75	19.25	28.25	0.73
0.5	0.25	29.	15.0	0.73
0.5	0.5	29.	21.0	0.55
0.5	0.75	29.	28.25	0.45
0.75	0.25	59.25	15.0	0.73
0.75	0.5	59.25	21.0	0.55
0.75	0.75	59.25	28.25	0.45



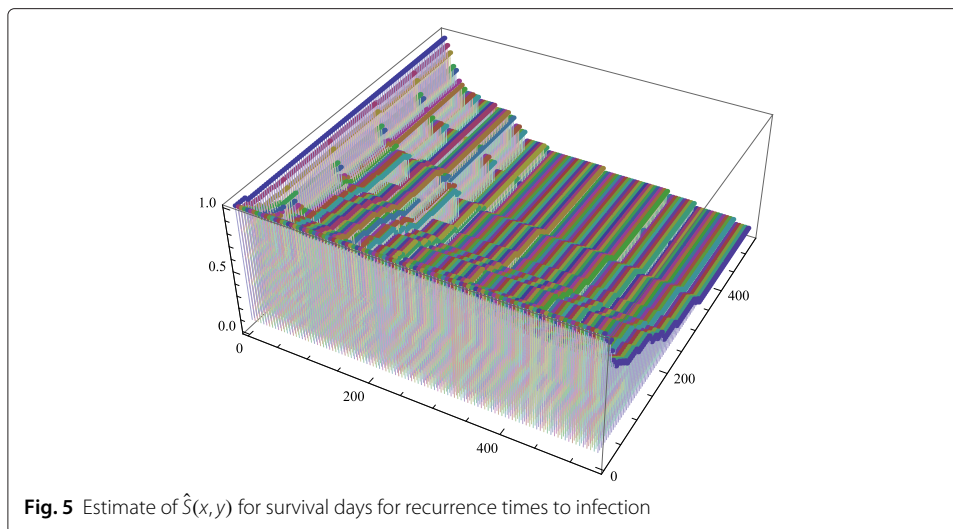
observations. The estimated bivariate survival probabilities  $\hat{S}(\hat{Q}_x(u_x), \hat{Q}_y(u_y))$  evaluated at the marginal quartiles are provided in Table 4. The joint bivariate survival function is plotted in Fig. 4. We see that our estimates provide a valid estimator of the joint survival function that is monotone decreasing in both dimensions.

**Recurrence times to infection at the point of insertion of a catheter for kidney patients using portable dialysis equipment (Akritas and Van Keilegom 2003).** For this data set we have  $n = 38$  paired survival times corresponding to infections times at two points along with paired censoring indicators. The data given below yields 487  $\pi_{i,j}$  parameters to be estimated. The constraints from (19), (20) and (21) and likelihood are not presented for this problem. Essentially the problem is a basic symbolic linear programming problem, which in our case was readily handled within Mathematica(Mathematica 8.0 for Linux, Wolfram Research Inc.,Champaign, IL). The number of independent free parameters to be estimated after considering the constraints was 248.

$$\mathbf{s} = (8, 23, 22, 447, 30, 24, 7, 511, 53, 15, 7, 141, 96, 149, 536, 17, 185, 292, 22, 15, 152, 402, 13, 39, 12, 113, 132, 34, 2, 130, 27, 5, 152, 190, 119, 54, 6, 63),$$

**Table 5** Estimated bivariate survival probabilities for recurrence times to infection at the point of insertion of a catheter for kidney patients at the marginal quantiles

$u_x$	$u_y$	$\hat{Q}_x(u_x)$	$\hat{Q}_y(u_y)$	$\hat{S}(\hat{Q}_x(u_x), \hat{Q}_y(u_y))$
0.25	0.25	15	16	0.95
0.25	0.5	15	39	0.92
0.25	0.75	15	154	0.81
0.5	0.25	46	16	0.91
0.5	0.5	46	39	0.85
0.5	0.75	46	154	0.68
0.75	0.25	149	16	0.90
0.75	0.5	149	39	0.77
0.75	0.75	149	154	0.58



$$\mathbf{t} = (16, 13, 28, 318, 12, 245, 9, 30, 196, 154, 333, 8, 38, 70, 25, 4, 177, 114, 159, 108, 562, 24, 66, 46, 40, 201, 156, 30, 25, 26, 58, 43, 30, 5, 8, 16, 78, 8),$$

$$\delta^x = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1),$$

$$\delta^y = (1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 0).$$

Due to the size of the problem we only present the estimated bivariate survival probabilities  $\hat{S}(\hat{Q}_x(u_x), \hat{Q}_y(u_y))$  evaluated at the marginal quartiles as provided in Table 5. The joint bivariate survival function is plotted in Fig. 5. We see that our estimates provide a valid estimator of the joint survival function that is monotone decreasing in both dimensions. It is straightforward to evaluate all potential survival probabilities and any estimators one wishes to derive from these given the sophistication of current software packages.

### 4 Conclusions

In this note we provide the only method for estimating a bivariate survival function such that the marginal estimators correspond exactly to the Kaplan-Meier product limit estimators such that there is a relative consistency between the marginal estimates derived via univariate or bivariate methods. Unlike other methods developed in the literature, our approach is generalizable to higher dimensions and different censoring mechanisms (interval and left censoring). Our methodology also opens up an alternative path for kernel smoothing of both the bivariate density, distribution function and survival function via use of the estimated  $\pi_{i,j}$ 's over the multinomial grid of non-zero mass. Using real data we illustrated the computational approach and feasibility of this new method of simplex constraint based maximum likelihood estimation as applied to right censored data.

Received: 19 August 2015 Accepted: 22 March 2016  
 Published online: 14 April 2016

### References

Akritis, MG, Van Keilegom, I: Estimation of bivariate and marginal distributions with censored data. *J. R. Stat. Soc. Ser. B.* **65**, 457–471 (2003)  
 Chen, K, Lo, S-H: On the rate of uniform convergence of the product-limit estimator: Strong and weak laws. *Ann. Stat.* **25**, 1050–1087 (1997)  
 Cox, DR, Oakes, D: *Analysis of Survival Data*. Chapman & Hall/CRC, New York (1984)

- Gill, RD, van der Lann, MJ, Wellner, JA: Inefficient estimators of the bivariate survival function for three models. *Annales de l'Institut Henri Poincaré*. **31**, 545–597 (1995)
- Kaplan, EL, Meier, P: Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **53**, 457–481 (1958)
- Lin, DY, Ying, Z: A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika*. **80**, 573–581 (1993)
- Liu, C: Estimation of discrete distributions with a class of simplex constraints. *J. Am. Stat. Assoc.* **95**, 109–120 (2000)
- Owen, AB: Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*. **75**, 237–249 (1988)
- Prentice, RL, Zoe Moodie, F, Wu, J: Hazard-based nonparametric survivor function estimation. *J. R. Stat. Soc. Ser. B*. **66**, 305–319 (2004)
- Satten, GA, Datta, S: The Kaplan-Meier Estimator as an Inverse-Probability-of-Censoring Weighted Average. *Am. Stat.* **55**, 207–210 (2001)
- Wang, W, Wells, MT: Nonparametric estimators of the bivariate survival function under simplified censoring condition. *Biometrika*. **84**, 863–880 (1997)
- Zhou, M: Empirical likelihood ratio with arbitrarily censored/truncated data by EM algorithm. *J. Comput. Graph. Stat.* **14**, 643–656 (2005)

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](http://springeropen.com)

---