Ann Inst Stat Math (2014) 66:913–930
DOI 10.1007/s10463-013-0433-x

Empirical likelihood bivariate nonparametric maximum likelihood estimator with right censored data

Jian-Jian Ren · Tonya Riddlesworth

Received: 2 January 2013 / Revised: 6 June 2013 / Published online: 26 October 2013

Abstract This article considers the estimation for bivariate distribution function (d.f.) $F_0(t, z)$ of survival time T and covariate variable Z based on bivariate data where T is subject to right censoring. We derive the empirical likelihood-based bivariate nonparametric maximum likelihood estimator $\hat{F}_n(t, z)$ for $F_0(t, z)$, which has an explicit expression and is *unique* in the sense of empirical likelihood. Other nice features of $\hat{F}_n(t, z)$ include that it has only nonnegative probability masses, thus it is monotone in bivariate sense. We show that under $\hat{F}_n(t, z)$, the conditional d.f. of T given Z is of the same form as the Kaplan–Meier estimator for the univariate case, and that the marginal d.f. $\hat{F}_n(\infty, z)$ coincides with the empirical d.f. of the covariate sample. We also show that when there is no censoring, $\hat{F}_n(t, z)$ coincides with the bivariate empirical d.f. For discrete covariate Z, the strong consistency and weak convergence of $\hat{F}_n(t, z)$ are established. Some simulation results are presented.

Keywords Bivariate data · Bivariate right censored data · Doubly censored data · Empirical likelihood · Maximum likelihood estimator · Right censored data

J.-J. Ren's research was partially supported by NSF grants DMS-0604488 and DMS-0905772/DMS-1232424.

J.-J. Ren (✉)
Department of Mathematics, University of Maryland, College Park, MD 20742, USA
e-mail: jjren@umd.edu

T. Riddlesworth
Department of Mathematics, Tennessee Technological University,
110 University Drive, Cookeville, TN 38505, USA
e-mail: triddlesworth@tntech.edu

1 Introduction

In the analysis of survival data, we often encounter situations where the response variable is the survival time T and is subject to right censoring, but the p-dimensional vector Z of covariates with components such as treatments, gender, etc., are completely observable. For simplicity of presentation, here we consider the case that covariate Z is a scalar rather than a vector, i.e., Z with dimension $p = 1$, noting that the generalization of our results in this article to multivariate case with $p > 1$ is straightforward. Specifically, suppose that

$$(T_1, Z_1), (T_2, Z_2), \ldots, (T_n, Z_n) \tag{1}$$

is a random sample of (T, Z), but the actually observed survival data are the bivariate data with one coordinate subject to random right censoring as follows:

$$(V_1, \delta_1, Z_1), (V_2, \delta_2, Z_2), \ldots, (V_n, \delta_n, Z_n), \tag{2}$$

where $V_i = \min\{T_i, C_i\}$, $\delta_i = I\{T_i \leq C_i\}$, and C_i is the right censoring variable with distribution function (d.f.) F_C and is independent of (T_i, Z_i). We call (2) the BD1RC data. In practice, if one wishes to use the nonparametric approach (i.e., without imposing any model assumptions) in the study of the relation between the right censored response variable T and the completely observable covariate variable Z, a natural thing to do is to estimate the bivariate distribution function (d.f.) $F_0(t, z)$ of (T, Z) based on observed survival data. In this article, we derive the *bivariate nonparametric maximum likelihood estimator* (BNPMLE) $\hat{F}_n(t, z)$ for $F_0(t, z)$ based on BD1RC data (2) using the empirical likelihood method (Owen 1988), and study its asymptotic properties under certain conditions.

To our best knowledge, there are currently no published works on the problem we consider here. In Akritas (1994), an estimator $\hat{F}_A$ for $F_0(t, z)$ with BD1RC data (2) was constructed using the conditional survival distribution and kernel estimator approach, thus estimator $\hat{F}_A$ is kernel and bandwidth dependent, and is not a maximum likelihood estimator in any sense. One related problem was studied by Lin and Ying (1993) who constructed the estimator for $F_0(t, z)$ based on the bivariate sample (1) that is subject to the same univariate right censoring on both components T_i and Z_i simultaneously. Another related and more complicated problem is the estimator for $F_0(t, z)$ based on the so-called bivariate right censored data, i.e., each component of (T_i, Z_i) in the bivariate sample (1) is subject to its own right censoring variable. Such a problem was considered by Dabrowska (1988, 1989), among others. The estimator by Dabrowska (1988) and those by others were constructed mainly based on or related to the representation of the bivariate survival function in terms of distribution functions of the data; see van der Laan (1996) for a nice review and discussion on this topic. A problem less related to our problem in this article is the bivariate right censored data considered by van der Laan (1996) under a different setting where the bivariate right censoring vector is discrete and is always observed. The work most closely related to ours is that by Ren and Gu (1997), where they constructed a bivariate distribution function estimator based on bivariate survival data which is subject to double censoring in one coordinate. Since right censoring is a special case of double censoring, our above BD1RC data (2) is a special case of that considered in Ren and Gu (1997). The estimator by Ren and Gu (1997) was given by the product of the empirical distribution estimator for $F_Z(z)$ and the conditional NPMLE for $P\{T \leq t \mid Z \leq z\}$, which is the univariate version of the NPMLE for doubly censored data given by Mykland and Ren (1996).

In comparison to the problems with the bivariate data subject to double censoring, univariate right censoring and bivariate right censoring studied by Ren and Gu (1997), Lin and Ying (1993), and Dabrowska (1988), respectively, our BD1RC data (2) problem is encountered far more frequently in practical situations, thus its solution is of important interest because the BNPMLE $\hat{F}_n(t, z)$ for $F_0(t, z)$ based on BD1RC data (2) provides tools for the studies of related nonparametric inference problems.

In terms of methodology, most estimators proposed for bivariate data subject to censoring are ad hoc, and either are kernel and bandwidth dependent (e.g., Akritas 1994) or contain negative probability masses; see discussions in van der Laan (1996). For instance, both bivariate distribution estimators by Dabrowska (1988) and Ren and Gu (1997) contain negative probability masses, thus not monotone in bivariate sense. In this context, the method of solving the self-consistency equation for censored data (Efron 1967) via the EM algorithm (Dempster et al. 1977; Turnbull 1976) is a possible procedure for finding the maximum likelihood distribution estimator. However, this approach as a general methodology has its limitation for censored bivariate data; see Dabrowska (1988), van der Laan (1996), among others. With all these in mind, we consider the empirical likelihood method (Owen 1988) for the problem studied in this article.

It is well known that the generality of the empirical likelihood approach in the contexts of censored data and nonparametric inferences is very attractive and useful. In particular, with the usual constraints or relevant estimating equations imposed, the resulting NPMLE based on the empirical likelihood does not contain any negative probability masses, thus it is monotone in bivariate sense for the problem under consideration here. However, despite these nice properties, it is also well known that the computation based on the empirical likelihood can be very challenging, often precisely due to these imposed constraints and/or the estimating equations. We show in this paper that when the computation issue is or can be resolved, the empirical likelihood method can indeed provide very nice solution for the censored bivariate data problem.

The main results of this article are organized as follows. In Sect. 2, we derive the empirical likelihood-based bivariate nonparametric maximum likelihood estimator (BNPMLE) $\hat{F}_n(t, z)$ for $F_0(t, z)$ with BD1RC data (2), which has an explicit expression and is *unique* in the sense of empirical likelihood. We show that under BNPMLE $\hat{F}_n(t, z)$, the conditional distribution function of T given Z is of the same form as the Kaplan–Meier estimator for the univariate case, while the marginal d.f. $\hat{F}_n(\infty, z)$ coincides with the empirical d.f. of the covariate sample $Z_1, \ldots, Z_n$ in (2). We also show that when there is no censoring, BNPMLE $\hat{F}_n(t, z)$ coincides with the bivariate empirical d.f. of sample (1). At the end of Sect. 2, we provide some discussions on the structure of the BNPMLE $\hat{F}_n(t, z)$ and its extension to the case of p-variate covariate $\boldsymbol{Z}$ with $p > 1$, and we also discuss the relation of our BNPMLE $\hat{F}_n(t, z)$ to the estimators by Akritas (1994) and by Ren and Gu (1997), respectively. In Sect. 3, we show that if the covariate variable Z in (2) is discrete, then our BNPMLE $\hat{F}_n(t, z)$ is uniformly strong consistent and converges to a centered Gaussian process as $n \to \infty$. Section 4 presents some simulation results which compare the performance of BNPMLE $\hat{F}_n(t, z)$ and the bivariate distribution estimator $\hat{F}_{RG}(t, z)$ by Ren and Gu (1997). All proofs are given in Sect. 5.

It should be noted that our empirical likelihood-based BNPMLE $\hat{F}_n(t, z)$ can be expressed as an integral of the conditional NPMLE, i.e., the Kaplan–Meier estimator in the univariate case; for details see Remark 1 at the end of Sect. 2. This is a quite interesting discovery in this article on the empirical likelihood-based distribution estimators. As pointed out in Remark 1, such a formulation is somewhat related to the conditional survival distribution approach used by Akritas (1994) for constructing his estimator $\hat{F}_A$, but it has nothing to do with the derivation of our BNPMLE $\hat{F}_n(t, z)$, because $\hat{F}_n(t, z)$ is an empirical likelihood-based maximum likelihood estimator.

2 Nonparametric maximum likelihood estimator

To derive the bivariate maximum likelihood estimator (BNPMLE) for bivariate d.f. $F_0(t, z)$ of (T, Z) based on the BD1RC data (2), we let

$$\begin{aligned} &U_1 < \cdots < U_m \text{ be all distinct values among } V_1, \ldots, V_n \\ &Y_1 < \cdots < Y_q \quad \text{be all distinct values among } Z_1, \ldots, Z_n \end{aligned} \tag{3}$$

and let F_C and f_C denote the d.f. and density function of censoring variable C_i, respectively. Since the essential idea of the empirical likelihood method is to consider those d.f.'s with support on observed data points (see Owen 1988) as the candidates for the NPMLE of $F_0(t, z)$, we treat $F_0(t, z)$ as a "*discrete*" bivariate d.f. in the following derivation of its empirical likelihood function. Note that with a "*discrete*" $F_0(t, z)$, the likelihood of BD1RC data (2) is given by

$$\begin{aligned} \prod_{i=1}^{n} P\{V = V_i, \delta = \delta_i, Z = Z_i\} &= \prod_{i=1}^{n} (P\{T_i = V_i, T_i \leq C_i, Z = Z_i\})^{\delta_i} (P\{C_i = V_i, T_i > C_i, Z = Z_i\})^{1-\delta_i} \\ &= \prod_{i=1}^{n} (P\{T = V_i, Z = Z_i\} \bar{F}_C(V_i))^{\delta_i} (P\{T > V_i, Z = Z_i\} f_C(V_i))^{1-\delta_i}, \end{aligned} \tag{4}$$

which is proportional to

$$\begin{aligned} \prod_{i=1}^{n} (P\{T = V_i, Z = Z_i\})^{\delta_i} (P\{T > V_i, Z = Z_i\})^{1-\delta_i} &= \prod_{i=1}^{n} (\mathrm{d}F_0(V_i, Z_i))^{\delta_i} (F_0(\infty, \mathrm{d}Z_i) - F_0(V_i, \mathrm{d}Z_i))^{1-\delta_i} \\ &= \prod_{i=1}^{m} \prod_{j=1}^{q} (\mathrm{d}F_0(U_i, Y_j))^{\delta_{ij}} (F_0(\infty, \mathrm{d}Y_j) - F_0(U_i, \mathrm{d}Y_j))^{n_{ij}-\delta_{ij}}, \end{aligned} \tag{5}$$

where $\mathrm{d}F_0(t,z) = P\{T = t, Z = z\}$, $F_0(t, \mathrm{d}z) = F_0(t,z) - F_0(t, z-)$, and

$$n_{ij} = \sum_{k=1}^{n} I\{V_k = U_i, Z_k = Y_j\}; \quad \delta_{ij} = \sum_{k=1}^{n} I\{V_k = U_i, \delta_k = 1, Z_k = Y_j\} \tag{6}$$

for $1 \le i \le m, 1 \le j \le q$. Thus, the empirical likelihood function for bivariate distribution function $F_0(t,z)$ of (T, Z) with BD1RC data (2) is given by

$$L(F) = \prod_{i=1}^{m} \prod_{j=1}^{q} \left(\mathrm{d}F(U_i, Y_j)\right)^{\delta_{ij}} \left(F(\infty, \mathrm{d}Y_j) - F(U_i, \mathrm{d}Y_j)\right)^{n_{ij}-\delta_{ij}}, \tag{7}$$

where F is any bivariate d.f., and by denoting P_F as the probability under F we have $\mathrm{d}F(t,z) = P_F\{T = t, Z = z\}$ and $F(t, \mathrm{d}z) = P_F\{T \le t, Z = z\} = F(t,z) - F(t, z-)$. Observe that (6) implies $n_{1j} + \cdots + n_{mj} \ge 1$ for any $1 \le j \le q$, and that $n_{ij} = 0$ implies $\delta_{ij} = 0$. Hence, letting

$$m_j = \max\{k \,|\, n_{kj} > 0\}, \qquad 1 \le j \le q \tag{8}$$

we have $n_{ij} = \delta_{ij} = 0$ for all $1 \le j \le q$, $m_j < i \le m$; in turn, empirical likelihood function (7) for F_0 is equivalently written as

$$L(F) = \prod_{j=1}^{q} \prod_{i=1}^{m_j} (\mathrm{d}F(U_i, Y_j))^{\delta_{ij}} (F(\infty, \mathrm{d}Y_j) - F(U_i, \mathrm{d}Y_j))^{n_{ij}-\delta_{ij}}. \tag{9}$$

To find the BNPMLE for F_0 based on empirical likelihood function (9), we restrict all possible candidates to those bivariate d.f.'s that assign all their probability masses to points (U_i, Y_j) and line segments $L_j = \{(t, Y_j) \in \mathbb{R}^2;\, t > U_m\}$ for $1 \le i \le m, 1 \le j \le q$, which writes likelihood function (9) as follows:

$$L(F) = \prod_{j=1}^{q} \prod_{i=1}^{m_j} (p_{ij})^{\delta_{ij}} \left(\sum_{k=i+1}^{m_j+1} p_{kj} \right)^{n_{ij}-\delta_{ij}} \equiv L(\boldsymbol{p}), \tag{10}$$

where

$$\boldsymbol{p} = (p_{11} \cdots p_{m_1+1,1} \cdots p_{1q} \cdots, p_{m_q+1,q}) \tag{11}$$

$$F(t,z) = \sum_{i=1}^{m} \sum_{j=1}^{q} q_{ij} I\{U_i \le t, Y_j \le z\}, \quad \text{for } t \le U_m,\ z \in \mathbb{R} \tag{12}$$

satisfy $q_{ij} = P_F\{T = U_i, Z = Y_j\}$ for $1 \le i \le m$, $1 \le j \le q$; and satisfy

$$\begin{cases} p_{ij} = q_{ij} = \mathrm{d}F(U_i, Y_j) = P_F\{\, T = U_i, Z = Y_j \,\}, \ \text{for } 1 \le j \le q, 1 \le i \le m_j \\ q_{m+1,\, j} = P_F\{\, (T, Z) \in L_j \,\} = P_F\{\, T > U_m, Z = Y_j \,\}, \ \text{for } 1 \le j \le q \\ p_{m_j+1,\, j} = P_F\{\, T > U_{m_j}, Z = Y_j \,\} = \sum_{i=m_j+1}^{m+1} q_{ij}, \ \text{for } 1 \le j \le q \\ \sum_{j=1}^{q} \sum_{i=1}^{m_j+1} p_{ij} = \sum_{i=1}^{m+1} \sum_{j=1}^{q} q_{ij} = 1. \end{cases} \tag{13}$$

Hence, the BNPMLE $\hat{F}_n(t, z)$ for $F_0(t, z)$ is the solution that maximizes above likelihood function $L(F) = L(\boldsymbol{p})$ in (10).

It should be noted that if $m_j < m$, the values of q_{ij}'s for $m_j < i \le m$ have no effects to the value of likelihood function (10). Thus, we can only derive the BNPMLE in terms of $\boldsymbol{p}$ (11) for $L(\boldsymbol{p})$. Let $\hat{\boldsymbol{p}}$ denote the solution of the following optimization problem:

$$\begin{cases} \max\ L(\boldsymbol{p}) = \prod_{j=1}^{q} \prod_{i=1}^{m_j} (p_{ij})^{\delta_{ij}} \left(\sum_{k=i+1}^{m_j+1} p_{kj} \right)^{n_{ij}-\delta_{ij}} \\ \text{subject to: } 0 \le p_{ij} \le 1, \ \text{for } 1 \le j \le q, 1 \le i \le m_j; \ \sum_{j=1}^{q} \sum_{i=1}^{m_j+1} p_{ij} = 1. \end{cases} \tag{14}$$

With proofs deferred to Sect. 5, the theorem below gives the solution and properties of $\hat{\boldsymbol{p}}$.

Theorem 1 *For any* $1 \le i \le m$, $1 \le j \le q$, *we denote*

$$N_{ij} = n_{ij} + \cdots + n_{mj} = \sum_{k=1}^{n} I\{V_k \ge U_i, Z_k = Y_j\}. \tag{15}$$

Then, the solution $\hat{\boldsymbol{p}}$ *of* (14) *is unique and satisfies the following:*

(i) *For any* $1 \le j \le q$, $1 \le i \le m_j$, *we have* $\hat{p}_{ij} > 0$ *if and only if* $\delta_{ij} > 0$*;*

(ii) *For any* $1 \le j \le q$, $1 \le i \le m_j$, *we have* $\sum_{k=i}^{m_j+1} \hat{p}_{kj} > 0$*;*

(iii) *For any* $1 \le j \le q$, *with notation* $\prod_{k=1}^{0} c_k \equiv 1$ *we have*

$$\begin{cases} \hat{p}_{ij} = \left(\frac{\delta_{ij}}{N_{ij}}\right) \left(\frac{N_{1j}}{n}\right) \prod_{k=1}^{i-1} \left(1 - \frac{\delta_{kj}}{N_{kj}}\right), \ \textit{for } 1 \le i \le m_j \\ \hat{p}_{m_j+1,\, j} = \frac{N_{1j}}{n} - \sum_{i=1}^{m_j} \hat{p}_{ij}. \end{cases} \tag{16}$$

Although Theorem 1 shows that the BNPMLE is unique in terms of $\boldsymbol{p}$ in (11), such uniqueness does not seem obvious in terms of F as given by (12)–(13), because if $m_j < m$ and $p_{m_j+1,\, j} > 0$ for some $1 \le j \le m$, it is not obvious how probability mass $p_{m_j+1,\, j}$ is distributed among $q_{m_j+1,\, j}, \ldots, q_{mj}, q_{m+1,\, j}$. To deal with this issue, we notice that (6), (8) and (15) imply that for any $1 \le j \le q$,

$$\begin{cases} n_{m_j,j} > 0 \ \Rightarrow \ N_{1j} \geq N_{2j} \geq \cdots \geq N_{m_j,j} > 0 \\ n_{ij} = \delta_{ij} = N_{ij} = 0, \quad \text{for } m_j < i \leq m \text{ when } m_j < m. \end{cases} \tag{17}$$

This means that points (U_i, Y_j) for $m_j < i \leq m$ are not observed among (V_k, Z_k)'s in data (2), thus by the usual empirical likelihood treatment these points (U_i, Y_j) are not assigned any probability masses, i.e., we have in (12)–(13) that for any $1 \leq j \leq q$,

$$q_{ij} = 0, \quad \text{for } m_j < i \leq m \tag{18}$$

$$p_{m_j+1,j} = P_F\{T > U_{m_j}, Z = Y_j\} = q_{m+1,j} = P_F\{T > U_m, Z = Y_j\}. \tag{19}$$

Hence, in sense of the empirical likelihood method the BNPMLE $\hat{F}_n(t, z)$ for $F_0(t, z)$ is *uniquely* given by the following same formula (16) of $\hat{p}_{ij}$'s due to (17)–(19):

$$\begin{cases} \hat{F}_n(t, z) = \sum_{i=1}^{m} \sum_{j=1}^{q} \hat{q}_{ij} I\{U_i \leq t, Y_j \leq z\}, \quad \text{for } t \leq U_m, \ z \in \mathbb{R} \\ \hat{q}_{ij} = \left(\frac{\delta_{ij}}{N_{ij}}\right)\left(\frac{N_{1j}}{n}\right) \prod_{k=1}^{i-1} \left(1 - \frac{\delta_{kj}}{N_{kj}}\right), \quad \text{for } 1 \leq i \leq m, \ 1 \leq j \leq q \\ \hat{q}_{m+1,j} = P_{\hat{F}_n}\{T > U_m, Z = Y_j\} = \left(\frac{N_{1j}}{n}\right) - \sum_{i=1}^{m} \hat{q}_{ij}, \quad \text{for } 1 \leq j \leq q \end{cases} \tag{20}$$

where $0/0$ is set as 0 whenever it occurs. With proofs given in Sect. 5, the following corollaries give further properties of the BNPMLE $\hat{F}_n(t, z)$.

Corollary 1 *The BNPMLE $\hat{F}_n(t, z)$ in (20) can be expressed by*

$$\hat{F}_n(t, z) = \sum_{j=1}^{q} P_{\hat{F}_n}\{Z = Y_j\} \hat{F}_n(t \mid Z = Y_j) I\{Y_j \leq z\}, \quad \textit{for } t \leq U_m, \ z \in \mathbb{R} \tag{21}$$

where under bivariate d.f. $\hat{F}_n$, we have $P_{\hat{F}_n}\{Z = Y_j\} = N_{1j}/n$, and $\hat{F}_n(t \mid Z = Y_j)$ is the conditional d.f. of T given $Z = Y_j$ satisfying

$$\bar{\hat{F}}_n(t \mid Z = Y_j) = P_{\hat{F}_n}\{T > t \mid Z = Y_j\} = \prod_{U_k \leq t} \left(1 - \frac{\delta_{kj}}{N_{kj}}\right), \quad \textit{for } t \leq U_m, \ 1 \leq j \leq q. \tag{22}$$

Corollary 2 *For BNPMLE $\hat{F}_n(t, z)$ in (20), the marginal d.f. $\hat{F}_n(\infty, z)$ of Z coincides with the empirical d.f. of sample $Z_1, \ldots, Z_n$ in (2).*

Corollary 3 *When there is no censoring in data (2), BNPMLE $\hat{F}_n(t, z)$ in (20) coincides with the bivariate empirical d.f. of the sample (1).*

Remark 1 Structure of BNPMLE $\hat{F}_n(t, z)$. Note that $\hat{F}_n(t \mid Z = Y_j)$ given in (22) is the Kaplan–Meier estimator in the univariate case with variable t, denoted as $\hat{F}_{KM}(t \mid Y_j)$. Thus, from Corollary 2, we know that $\hat{F}_n(t, z)$ shown in (21) is the sum of the products of the marginal NPMLE $\hat{F}_{n,Z}(z) \equiv \hat{F}_n(\infty, z)$ for $F_Z(z)$ and the conditional NPMLE $\hat{F}_{KM}(t \,|\, z)$ for $F_0(t \,|\, Z = z)$. This implies that (21) can be expressed as:

$$\hat{F}_n(t, z) = \int_{-\infty}^{z} P_{\hat{F}_n}\{T \le t, \,|\, Z = u\} d\hat{F}_{n,Z}(u) = \int_{-\infty}^{z} \hat{F}_{KM}(t \,|\, u) d\hat{F}_n(\infty, u) \tag{23}$$

for $t \le U_m$, $z \in \mathbb{R}$, where the first part of the equation is the idea that Akritas (1994) used to construct his kernel-based estimator for $F_0(t, z)$. In contrast, our BNPMLE $\hat{F}_n$ here is an empirical likelihood-based MLE, thus is not kernel and bandwidth dependent; and the second part of Eq. (23) holding for the resulting empirical likelihood-based BNPMLE $\hat{F}_n$ is an interesting new discovery in this article, but as shown above it is not the derivation tool used for obtaining the BNPMLE. Another matter worthy of mentioning is that with some algebraic work, it can be shown that the marginal d.f. of $\hat{F}_n(t, \infty)$ of T is not the Kaplan–Meier estimator in the univariate case; a special and easier case for this study is when there are no ties among V_i's and there are no ties among Z_i's. Due to a different method used, the marginal distribution of the bivariate distribution estimator with bivariate right censored data (a more difficult problem, of course) constructed by Dabrowska (1988) is the Kaplan–Meier estimator in the univariate case, but her estimator is not monotone and contains negative probability masses (see page 1485 of Dabrowska 1988). In contrast, our estimator $\hat{F}_n(t, z)$ in (20) is the unique maximum likelihood estimator in the sense of empirical likelihood, and is monotone with only nonnegative probability masses.

Remark 2 Relation to the Estimator by Ren and Gu (1997). In Ren and Gu (1997), a bivariate distribution function estimator $\hat{F}_{RG}(t, z)$ was constructed based on bivariate survival data which is subject to double censoring in one coordinate. Since right censoring is a special case of double censoring, our BD1RC data (2) is a special case of that considered in Ren and Gu (1997). Their estimator $\hat{F}_{RG}(t, z)$ is given by the product of the empirical distribution estimator for $F_Z(z)$ and the conditional NPMLE for $P\{T \le t \,|\, Z \le z\}$, thus its marginal d.f. $\hat{F}_{RG}(t, \infty)$ coincides with the NPMLE in the univariate case for doubly censored data given by Mykland and Ren (1996). From Chang and Yang (1987), we know that when there is no left censoring, the NPMLE by Mykland and Ren (1996) is the same as the Kaplan–Meier estimator. Thus, with our BD1RC data (2) the marginal distribution $\hat{F}_{RG}(t, \infty)$ of the estimator by Ren and Gu (1997) is the Kaplan–Meier estimator in the univariate case. Hence, estimator $\hat{F}_{RG}(t, z)$ with our BD1RC data (2) is different from our BNPMLE $\hat{F}_n(t, z)$ in (20), and is not the maximum likelihood estimator in the sense of empirical likelihood. Moreover, estimator $\hat{F}_{RG}(t, z)$ with our BD1RC data (2) is not monotone and contains negative probability masses.

Remark 3 Extension to p-variate covariate $\boldsymbol{Z}$ *with* $p > 1$. If $\boldsymbol{Z}_i$'s in (2) is p-variate with $p > 1$, then we let $\boldsymbol{Y}_j \in \mathbb{R}^p$ in (3), now without ordering, represent all distinct vectors of $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$. Following the proofs of Theorem 1, we can show that the empirical likelihood NPMLE $\hat{F}_n(t, \boldsymbol{z})$ for $F_0(t, \boldsymbol{z})$ is given by (20) with $Y_j \le z$ and $Z = Y_j$ replaced by $\boldsymbol{Y}_j \le \boldsymbol{z}$ and $\boldsymbol{Z} = \boldsymbol{Y}_j$, respectively, where $(\boldsymbol{Y}_j \le \boldsymbol{z}) \equiv (Y_{1j} \le z_1, \ldots, Y_{pj} \le z_p)$.

3 Asymptotic properties for discrete covariate

In the analysis of survival data, we often encounter the situation where the covariates Z are discrete or categorical. For instance, in practice covariate variable Z may represent treatment levels or methods. With the proofs deferred to Sect. 5, the following theorems establish some asymptotic results of our BNPMLE $\hat{F}_n(t, z)$ given by (20) with BD1RC data (2) where the covariate variable Z is discrete.

Theorem 2 *For BD1RC data* (2)*, assume that covariate variable Z is discrete with q possible values given by the second line of* (3)*, and assume that those conditions required in Corollary 1.2 of* Stute and Wang (1993) *hold for distributions of lifetime variable T and continuous censoring variable C. Then,*

$$\sup_{0\leq t\leq \tau_V,\, z\in\mathbb{R}} |\hat{F}_n(t, z) - F_0(t, z)| \stackrel{a.s.}{\to} 0, \quad as\ n \to \infty \tag{24}$$

where τ_V is given as in Stute and Wang (1993)*.*

Theorem 3 *For BD1RC data* (2)*, assume that covariate variable Z is discrete with q possible values given by the second line of* (3)*, and assume that those conditions in* Gill (1983) *hold for distributions of lifetime variable T and censoring variable C. Then, $\sqrt{n}(\hat{F}_n - F_0)$ weakly converges to a centered Gaussian process as $n \to \infty$.*

Remark 4 Above Theorem 2 and Theorem 3 give asymptotic results when the covariate variable Z in BD1RC data (2) is discrete, which, of course, is of special importance in the analysis of survival data; in fact, based on our BNPMLE $\hat{F}_n(t, z)$ in this paper Ren and Riddlesworth (2012) constructed the empirical likelihood ratio confidence interval for conditional survival probabilities of T given discrete covariate $Z = z_0$. However, for the case when Z is continuous, the methods used in our proofs for these theorems do not apply, because for continuous Z, quantity q in Eq. (21) is not a fixed constant when $n \to \infty$. Thus, the proofs for continuous covariate variable Z are much more involved technically and need further studies. For now, some simulation results on $\hat{F}_n$ with continuous covariate Z are presented and discussed in the next section.

4 Simulations

This section presents some simulation results on our BNPMLE $\hat{F}_n(t, z)$ given by (20) with right censored data (2), where both cases with discrete and continuous covariate variable Z are under consideration. Since our BD1RC data (2) is a special case of that considered by Ren and Gu (1997) (see above Remark 2 in Sect. 2), here we also make comparison between BNPMLE $\hat{F}_n(t, z)$ and the bivariate distribution estimator $\hat{F}_{RG}(t, z)$ by Ren and Gu (1997) for right censored data (2).

One practical issue in the actual computation of BNPMLE $\hat{F}_n(t, z)$ that needs to be noted is that in (20) we have $\hat{q}_{m+1,j} > 0$ for some js, which is the same issue as that with the Kaplan–Meier estimator in the univariate case. Due to equation (19), a natural thing to do is to evenly distribute the probability mass $\hat{q}_{m+1,j}$, whenever positive, to points $(U_{m_j+1}, Y_j), \ldots, (U_m, Y_j)$. This is done in all our simulation computations for $\hat{F}_n(t, z)$; noting that something similar is routinely done for the univariate Kaplan–Meier estimator.

Let Exp(μ) represent the exponential distribution with mean μ. Our simulation studies consider right censored data (2) with $F_T = \text{Exp}(1)$ as the d.f. of T, $F_C = \text{Exp}(3)$ the d.f. of right censoring variable C, and $F_Z = \text{U}\{1, 2, 3, 4, 5\}$ the d.f. of discrete covariate variable Z, where (T, C) is independent of Z and U$\{1, 2, 3, 4, 5\}$ represents the uniform distribution on points 1, 2, 3, 4, 5. To compare the performance of $\hat{F}_n(t, z)$ and $\hat{F}_{RG}(t, z)$ with the d.f. $F_0(t, z)$ of (T, Z), we generate 1000 such samples (2) with $n = 50, 100, 200, 500, 1000$, respectively. For each n, Table 1 includes the right censoring percentage of the generated samples, and includes the simulation average of $\|\hat{F}_n - F_0\|$ and $\|\hat{F}_{RG} - F_0\|$ with the simulation standard deviation (s.d.) given in the parenthesis, where the uniform norm $\|\cdot\|$ is taken over all sample points $(V_i, Z_i), i = 1, \ldots, n$.

To compare the performance of $\hat{F}_n(t, z)$ and $\hat{F}_{RG}(t, z)$ with continuous covariate variable Z, we conduct the simulation studies in Table 1 with $F_Z = \text{U}(0, 1)$ as the d.f. of Z, and include the results in Table 2, where U(0, 1) represents the uniform distribution on interval (0, 1). Moreover, the simulation studies in Table 1 are repeated with $F_C = \text{Exp}(3)$ and $F_{T|Z} = \text{Exp}(Z)$ as the conditional d.f. of T given Z, and the results are included in Table 3, where Z is a continuous r.v. with p.d.f. $f_Z(z) = 2/z^2$ if $1 < z < 2$; 0, elsewhere.

Remark 5 It should be noticed that in Table 1, the covariate variable Z is discrete, while in Tables 2, 3, Z is continuous. It also should be noticed that in Tables 1, 2,

Table 1 Comparison of $\hat{F}_n$, $\hat{F}_{RG}$, F_0 with right censored samples

Sample size	Average $\|\hat{F}_n - F_0\|$ (s.d.)	Average $\|\hat{F}_{RG} - F_0\|$ (s.d.)	Censoring %
$n = 50$	0.1345 (.0399)	0.1337 (.0391)	25.1
$n = 100$	0.1007 (.0273)	0.0996 (.0271)	25.1
$n = 200$	0.0739 (.0200)	0.0735 (.0198)	25.2
$n = 500$	0.0484 (.0122)	0.0482 (.0122)	25.1
$n = 1000$	0.0349 (.0089)	0.0348 (.0089)	25.1
Distributions	$F_T = \text{Exp}(1),\ F_C = \text{Exp}(3),\ F_Z = \text{U}\{1, 2, 3, 4, 5\}$		

Table 2 Comparison of $\hat{F}_n$, $\hat{F}_{RG}$, F_0 with right censored samples

Sample size	Average $\|\hat{F}_n - F_0\|$ (s.d.)	Average $\|\hat{F}_{RG} - F_0\|$ (s.d.)	Censoring %
$n = 50$	0.1286 (.0387)	0.1332 (.0381)	25.1
$n = 100$	0.0971 (.0282)	0.0999 (.0279)	25.1
$n = 200$	0.0741 (.0204)	0.0747 (.0193)	25.2
$n = 500$	0.0506 (.0131)	0.0499 (.0121)	25.1
$n = 1000$	0.0395 (.0107)	0.0366 (.0088)	24.9
Distributions	$F_T = \text{Exp}(1),\ F_C = \text{Exp}(3),\ F_Z = \text{U}(0, 1)$		

Table 3 Comparison of $\hat{F}_n$, $\hat{F}_{RG}$, F_0 with right censored samples

Sample size	Average $\|\hat{F}_n - F_0\|$ (s.d.)	Average $\|\hat{F}_{RG} - F_0\|$ (s.d.)	Censoring %
$n = 50$	0.1286 (.0386)	0.1366 (.0400)	31.6
$n = 100$	0.0988 (.0284)	0.1033 (.0284)	31.6
$n = 200$	0.0764 (.0222)	0.0771 (.0210)	31.4
$n = 500$	0.0553 (.0142)	0.0511 (.0119)	31.4
$n = 1000$	0.0451 (.0106)	0.0366 (.0092)	31.6
Distributions	$F_{T\|Z} = \text{Exp}(Z),\ F_C = \text{Exp}(3),\ F'_Z(z) = 2z^{-2}I\{1 < z < 2\}$		

lifetime variable T and covariate variable Z are independent, while in Table 3, T and Z are dependent. Nonetheless, clearly all Tables 1, 2, 3 show that our BNPMLE $\hat{F}_n$ and the bivariate distribution estimator $\hat{F}_{RG}$ by Ren and Gu (1997) perform similarly, and that with rather high censoring percentages, the estimation errors decrease as the sample size n increases. As already mentioned in Remark 2 of Sect. 2, $\hat{F}_{RG}$ has negative probability masses, while our BNPMLE $\hat{F}_n$ has only nonnegative probability masses, which is a desirable property for a bivariate distribution estimator in practice. Moreover, Ren and Gu (1997) showed that under mild conditions with discrete or continuous Z, we have $\|\hat{F}_{RG} - F_0\| \stackrel{a.s.}{\to} 0$ and $\hat{F}_{RG}$ converges to a centered Gaussian process as $n \to \infty$. Evidently, our results in Table 1 support our Theorems 2-3 in Sect. 3 on the asymptotic properties of BNPMLE $\hat{F}_n$ with discrete covariate variable Z, and our results in Tables 2, 3 indicate that our Theorems 2-3 may very well hold for $\hat{F}_n$ with continuous covariate variable Z.

5 Proofs

Proof of Theorem 1 (i)–(ii) The proofs for Parts (i)–(ii) are given in three separate cases as follows. If $\delta_{ij} > 0$, then we have $\hat{p}_{ij} > 0$ in order to have $L(\hat{\boldsymbol{p}}) > 0$ in (10); in turn, we have $\sum_{k=i}^{m_j+1} \hat{p}_{kj} \geq \hat{p}_{ij} > 0$.

If $\delta_{ij} = 0$ and $n_{ij} > 0$, then in (10) we have for $\hat{\boldsymbol{p}}$:

$$\left(\hat{p}_{ij}\right)^{\delta_{ij}} \left(\sum_{k=i+1}^{m_j+1} \hat{p}_{kj}\right)^{n_{ij}-\delta_{ij}} = \left(\sum_{k=i+1}^{m_j+1} \hat{p}_{kj}\right)^{n_{ij}} > 0, \tag{25}$$

which implies $\sum_{k=i}^{m_j+1} \hat{p}_{kj} \geq \sum_{k=i+1}^{m_j+1} \hat{p}_{kj} > 0$. Assume $\hat{p}_{ij} > 0$. Then, if letting $\tilde{p}_{ij} = 0$, $\tilde{p}_{m_j+1,\,j} = (\hat{p}_{ij} + \hat{p}_{m_j+1,\,j})$ and $\tilde{p}_{kl} = \hat{p}_{kl}$ for the rest of $1 \leq l \leq q, 1 \leq k \leq (m_l + 1)$, we have $L(\tilde{\boldsymbol{p}}) > L(\hat{\boldsymbol{p}})$ because in (10) we have

$$\prod_{k=1}^{m_j} (\hat{p}_{kj})^{\delta_{kj}} \left(\sum_{l=k+1}^{m_j+1} \hat{p}_{lj}\right)^{n_{kj}-\delta_{kj}}$$

$$
\begin{aligned}
&= \left\{ \prod_{k=1}^{i-1} (\hat{p}_{kj})^{\delta_{kj}} (\hat{p}_{k+1,\, j} + \cdots + \hat{p}_{ij} + \cdots + \hat{p}_{m_j+1,\, j})^{n_{kj}-\delta_{kj}} \right\} \\
&\quad \times \left(\sum_{k=i+1}^{m_j+1} \hat{p}_{kj} \right)^{n_{ij}} \left\{ \prod_{k=i+1}^{m_j} (\hat{p}_{kj})^{\delta_{kj}} (\hat{p}_{k+1,\, j} + \cdots + \hat{p}_{m_j+1,\, j})^{n_{kj}-\delta_{kj}} \right\} \\
&< \left\{ \prod_{k=1}^{i-1} (\tilde{p}_{kj})^{\delta_{kj}} (\tilde{p}_{k+1,\, j} + \cdots + \tilde{p}_{ij} + \cdots + \tilde{p}_{m_j+1,\, j})^{n_{kj}-\delta_{kj}} \right\} \\
&\quad \times \left(\sum_{k=i+1}^{m_j+1} \tilde{p}_{kj} \right)^{n_{ij}} \left\{ \prod_{k=i+1}^{m_j} (\tilde{p}_{kj})^{\delta_{kj}} (\tilde{p}_{k+1,\, j} + \cdots + \tilde{p}_{m_j+1,\, j})^{n_{kj}-\delta_{kj}} \right\} \qquad (26)
\end{aligned}
$$

due to (25) and the facts: $(\hat{p}_{ij} + \hat{p}_{m_j+1,\, j}) = (\tilde{p}_{ij} + \tilde{p}_{m_j+1,\, j})$, $\tilde{p}_{m_j+1,\, j} > \hat{p}_{m_j+1,\, j}$. But $L(\tilde{\boldsymbol{p}}) > L(\hat{\boldsymbol{p}})$ is a contradiction. Hence, we have $\hat{p}_{ij} = 0$.

If $\delta_{ij} > 0$ and $n_{ij} = 0$, then we have $i < m_j$ due to (8) and (18), and we have $\sum_{k=i}^{m_j+1} \hat{p}_{kj} \geq (\hat{p}_{m_j,\, j} + \hat{p}_{m_j+1,\, j}) > 0$, because with $n_{m_j,j} > 0$, from the proofs above we know that $\delta_{m_j,j} > 0$ implies $\hat{p}_{m_j,\, j} > 0$, and $\delta_{m_j,j} = 0$ implies $\hat{p}_{m_j,\, j} = 0$, $\hat{p}_{m_j+1,\, j} > 0$. Assume $\hat{p}_{ij} > 0$. Then, if letting $\tilde{p}_{ij} = 0$, $\tilde{p}_{m_j,\, j} = (\frac{1}{2}\, \hat{p}_{ij} + \hat{p}_{m_j,\, j})$, $\tilde{p}_{m_j+1,\, j} = (\frac{1}{2}\, \hat{p}_{ij} + \hat{p}_{m_j+1,\, j})$ and $\tilde{p}_{kl} = \hat{p}_{kl}$ for the rest of $1 \leq l \leq q$, $1 \leq k \leq (m_l + 1)$, we have

$$
(\hat{p}_{m_j,\, j})^{\delta_{m_j,j}} (\hat{p}_{m_j+1,\, j})^{n_{m_j,j}-\delta_{m_j,j}} < (\tilde{p}_{m_j,\, j})^{\delta_{m_j,j}} (\tilde{p}_{m_j+1,\, j})^{n_{m_j,j}-\delta_{m_j,j}}, \qquad (27)
$$

in turn, we have $L(\tilde{\boldsymbol{p}}) > L(\hat{\boldsymbol{p}})$ because in (10) we have

$$
\begin{aligned}
&\prod_{k=1}^{m_j} (\hat{p}_{kj})^{\delta_{kj}} \left(\sum_{l=k+1}^{m_j+1} \hat{p}_{lj} \right)^{n_{kj}-\delta_{kj}} \\
&\quad = \left\{ \prod_{k=1}^{i-1} (\hat{p}_{kj})^{\delta_{kj}} (\hat{p}_{k+1,\, j} + \cdots + \hat{p}_{ij} + \cdots + \hat{p}_{m_j,\, j} + \hat{p}_{m_j+1,\, j})^{n_{kj}-\delta_{kj}} \right\} \\
&\qquad \times \left\{ \prod_{k=i+1}^{m_j-1} (\hat{p}_{kj})^{\delta_{kj}} (\hat{p}_{k+1,\, j} + \cdots + \hat{p}_{m_j,\, j} + \hat{p}_{m_j+1,\, j})^{n_{kj}-\delta_{kj}} \right\} \\
&\qquad \times \left\{ (\hat{p}_{m_j,\, j})^{\delta_{m_j,j}} (\hat{p}_{m_j+1,\, j})^{n_{m_j,j}-\delta_{m_j,j}} \right\} \\
&\quad < \left\{ \prod_{k=1}^{i-1} (\tilde{p}_{kj})^{\delta_{kj}} (\tilde{p}_{k+1,\, j} + \cdots + \tilde{p}_{ij} + \cdots + \tilde{p}_{m_j,\, j} + \tilde{p}_{m_j+1,\, j})^{n_{kj}-\delta_{kj}} \right\} \\
&\qquad \times \left\{ \prod_{k=i+1}^{m_j-1} (\tilde{p}_{kj})^{\delta_{kj}} (\tilde{p}_{k+1,\, j} + \cdots + \tilde{p}_{m_j,\, j} + \tilde{p}_{m_j+1,\, j})^{n_{kj}-\delta_{kj}} \right\}
\end{aligned}
$$

$$\times\left\{(\tilde{p}_{m_j,j})^{\delta_{m_j,j}}(\tilde{p}_{m_j+1,j})^{n_{m_j,j}-\delta_{m_j,j}}\right\} \tag{28}$$

due to (27) and the facts: $(\hat{p}_{ij}+\hat{p}_{m_j,j}+\hat{p}_{m_j+1,j}) = (\tilde{p}_{ij}+\tilde{p}_{m_j,j}+\tilde{p}_{m_j+1,j})$, $\tilde{p}_{m_j,j} > \hat{p}_{m_j,j}$, $\tilde{p}_{m_j+1,j} > \hat{p}_{m_j+1,j}$. But $L(\tilde{\boldsymbol{p}}) > L(\hat{\boldsymbol{p}})$ is a contradiction. Hence, we have $\hat{p}_{ij} = 0$. □

Proof of Theorem 1 (iii) and Uniqueness From Theorem 1 (ii), we consider the following substitution for $1 \le j \le q$, $1 \le i \le m_j$:

$$a_{ij} = \frac{p_{ij}}{b_{ij}} \quad \text{and} \quad b_{ij} = \sum_{k=i}^{m_j+1} p_{kj} > 0, \tag{29}$$

which imply

$$b_{i+1,j} = \sum_{k=i+1}^{m_j+1} p_{kj} = b_{ij} - p_{ij} \quad \text{and} \quad 1 - a_{ij} = \frac{b_{i+1,j}}{b_{ij}}. \tag{30}$$

Since

$$\begin{aligned}
\prod_{j=1}^{q}\prod_{i=1}^{m_j}(1-a_{ij})^{N_{i+1,j}} &= \prod_{j=1}^{q}\prod_{i=1}^{m_j}\left(\frac{b_{i+1,j}}{b_{ij}}\right) \\
&= \prod_{j=1}^{q}\frac{(b_{2j})^{n_{2j}+\cdots+n_{m_j,j}}}{(b_{1j})^{n_{2j}+\cdots+n_{m_j,j}}} \times \frac{(b_{3j})^{n_{3j}+\cdots+n_{m_j,j}}}{(b_{2j})^{n_{3j}+\cdots+n_{m_j,j}}} \\
&\quad \times \cdots \times \frac{(b_{m_j+1,j})^{n_{m_j+1,j}}}{(b_{m_j,j})^{n_{m_j+1,j}}} \\
&= \prod_{j=1}^{q}\frac{(b_{2j})^{n_{2j}}(b_{3j})^{n_{3j}}\cdots(b_{m_j,j})^{n_{m_j,j}}}{(b_{1j})^{n_{2j}+\cdots+n_{m_j,j}}}(b_{m_j+1,j})^{n_{m_j+1,j}} \\
&= \left(\prod_{j=1}^{q}\prod_{i=1}^{m_j}(b_{ij})^{n_{ij}}\right)\left(\prod_{j=1}^{q}\frac{1}{(b_{1j})^{n_{1j}+\cdots+n_{m_j,j}}}\right) \\
&= \left(\prod_{j=1}^{q}\prod_{i=1}^{m_j}(b_{ij})^{n_{ij}}\right)\Big/\left(\prod_{j=1}^{q}(b_{1j})^{N_{1j}}\right),
\end{aligned}$$

from (29)–(30) we can write (10) as

$$
\begin{aligned}
L(\boldsymbol{p}) &= \prod_{j=1}^{q}\prod_{i=1}^{m_j}(a_{ij}b_{ij})^{\delta_{ij}}(b_{i+1,j})^{n_{ij}-\delta_{ij}} \\
&= \prod_{j=1}^{q}\prod_{i=1}^{m_j}(a_{ij}b_{ij})^{\delta_{ij}}((1-a_{ij})b_{ij})^{n_{ij}-\delta_{ij}} \\
&= \left(\prod_{j=1}^{q}\prod_{i=1}^{m_j}(b_{ij})^{n_{ij}}\right)\left(\prod_{j=1}^{q}\prod_{i=1}^{m_j}(a_{ij})^{\delta_{ij}}(1-a_{ij})^{n_{ij}-\delta_{ij}}\right) \\
&= \left(\prod_{j=1}^{q}\prod_{i=1}^{m_j}(1-a_{ij})^{N_{i+1,j}}\right)\left(\prod_{j=1}^{q}(b_{1j})^{N_{1j}}\right)\left(\prod_{j=1}^{q}\prod_{i=1}^{m_j}(a_{ij})^{\delta_{ij}}(1-a_{ij})^{n_{ij}-\delta_{ij}}\right) \\
&= \left(\prod_{j=1}^{q}(b_{1j})^{N_{1j}}\right)\left(\prod_{j=1}^{q}\prod_{i=1}^{m_j}(a_{ij})^{\delta_{ij}}(1-a_{ij})^{N_{ij}-\delta_{ij}}\right) \equiv G(\boldsymbol{a},\boldsymbol{b}). \qquad (31)
\end{aligned}
$$

From Theorem 1 (i)–(ii) and (29)–(30), we know that a solution that maximizes $G(\boldsymbol{a}, \boldsymbol{b})$ satisfies: $b_{1j} > 0$; $a_{ij} > 0$ if $\delta_{ij} > 0$; $0 \le a_{ij} < 1$ if $(N_{ij} - \delta_{ij}) > 0$. Thus, the optimization problem (14) is equivalent to:

$$
\begin{cases}
\max\ \log G(\boldsymbol{a},\boldsymbol{b}) = \left(\sum_{j=1}^{q} N_{1j}\log b_{1j}\right) + G_1(\boldsymbol{a}) \\
\text{subject to: } 0 \le a_{ij} \le 1,\ 0 < b_{1j} \le 1,\ \text{for } 1 \le j \le q, 1 \le i \le m_j;\ \sum_{j=1}^{q} b_{1j} = 1,
\end{cases}
\qquad (32)
$$

where setting $0\log 0 = 0$ whenever it occurs, we have

$$
G_1(\boldsymbol{a}) = \sum_{j=1}^{q}\sum_{i=1}^{m_j}(\delta_{ij}\log a_{ij} + (N_{ij}-\delta_{ij})\log(1-a_{ij})). \qquad (33)
$$

Since at least one of δ_{ij} and $N_{ij} - \delta_{ij}$ is positive for any $1 \le j \le q, 1 \le i \le m_j$ due to (17), thus for $G_1(\boldsymbol{a})$ in (33) we have

$$
\frac{\partial G_1}{\partial a_{ij}} = \left(\frac{\delta_{ij}}{a_{ij}} - \frac{N_{ij}-\delta_{ij}}{1-a_{ij}}\right), \quad \frac{\partial^2 G_1}{\partial a_{ij}^2} = -\left(\frac{\delta_{ij}}{a_{ij}^2} + \frac{N_{ij}-\delta_{ij}}{(1-a_{ij})^2}\right) < 0, \quad \frac{\partial^2 G_1}{\partial a_{ij}\,\partial a_{kl}} = 0, \qquad (34)
$$

where $(i, j) \ne (k, l)$, and $0/0$ is set as 0 whenever it occurs. Hence, $G_1(\boldsymbol{a})$ is concave and is uniquely maximized by

$$
\hat{a}_{ij} = \frac{\delta_{ij}}{N_{1j}}, \quad 1 \le j \le q, 1 \le i \le m_j. \qquad (35)
$$

Using Lagrange multipliers, it can be shown by Theorem 4.3.8 of Bazaraa et al. (1993) that $\sum_{j=1}^{q} N_{1j} \log b_{1j}$ subject to $0 < b_{1j} \leq 1$, $1 \leq j \leq q$ and $\sum_{j=1}^{q} b_{1j} = 1$ is uniquely maximized by

$$\hat{b}_{1j} = \frac{N_{1j}}{n}, \quad 1 \leq j \leq q. \tag{36}$$

The uniqueness of $\hat{\boldsymbol{p}}$ and the proof of Theorem 1 (iii) follow from the uniqueness of (35)–(36) and from noticing that

$$\hat{b}_{1j} = \sum_{i=1}^{m_j+1} \hat{p}_{ij}, \quad 1 \leq j \leq q \tag{37}$$

and that (29)–(30) and (35)–(36) imply for $1 \leq j \leq q$, $1 \leq i \leq m_j$:

$$\hat{p}_{ij} = \hat{a}_{ij}\hat{b}_{ij}(1-\hat{a}_{1j})\cdots(1-\hat{a}_{i-1,j}) = \left(\frac{\delta_{ij}}{N_{ij}}\right)\left(\frac{N_{1j}}{n}\right)\prod_{k=1}^{i-1}\left(1-\frac{\delta_{kj}}{N_{kj}}\right). \tag{38}$$

□

Proof of Corollary 1 For any $t \leq U_m$, $z \in \mathbb{R}$, from (20) we have

$$\begin{aligned}
\hat{F}_n(t,z) &= \sum_{i=1}^{m}\sum_{j=1}^{q}\left\{\left(\frac{\delta_{ij}}{N_{ij}}\right)\left(\frac{N_{1j}}{n}\right)\prod_{k=1}^{i-1}\left(1-\frac{\delta_{kj}}{N_{kj}}\right) I\{U_i \leq t, Y_j \leq z\}\right\} \\
&= \sum_{j=1}^{q}\sum_{i=1}^{m}\left(\frac{N_{1j}}{n}\right)\left\{\prod_{k=1}^{i-1}\left(1-\frac{\delta_{kj}}{N_{kj}}\right) - \prod_{k=1}^{i}\left(1-\frac{\delta_{kj}}{N_{kj}}\right)\right\} I\{U_i \leq t, Y_j \leq z\} \\
&= \sum_{j=1}^{q}\left(\frac{N_{1j}}{n}\right)\left\{1-\prod_{U_k \leq t}\left(1-\frac{\delta_{kj}}{N_{kj}}\right)\right\} I\{Y_j \leq z\}.
\end{aligned} \tag{39}$$

Since (20) implies

$$P_{\hat{F}_n}\{Z = Y_j\} = P_{\hat{F}_n}\{T \leq U_m, Z = Y_j\} + P_{\hat{F}_n}\{T > U_m, Z = Y_j\} = \left(\frac{N_{1j}}{n}\right), \tag{40}$$

then under $\hat{F}_n$ in (39) the conditional d.f. of T given $Z = Y_j$ is given by

$$\begin{aligned}
\hat{F}_n(t \mid Z = Y_j) &= P_{\hat{F}_n}\{T \leq t \mid Z = Y_j\} = \frac{P_{\hat{F}_n}\{T \leq t, Z = Y_j\}}{P_{\hat{F}_n}\{Z = Y_j\}} \\
&= \frac{\hat{F}_n(t, Y_j) - \hat{F}_n(t, Y_j-)}{N_{1j}/n} = \frac{(N_{1j}/n)\left\{1-\prod_{U_k \leq t}\left(1-\frac{\delta_{kj}}{N_{kj}}\right)\right\}}{N_{1j}/n} \\
&= 1-\prod_{U_k \leq t}\left(1-\frac{\delta_{kj}}{N_{kj}}\right),
\end{aligned} \tag{41}$$

where $t \leq U_m$. Hence, (21)–(22) follow from (39)–(41). □

Proof of Corollary 2 From (20), we have that for any $(t, z) \in \mathbb{R}^2$,

$$\hat{F}_n(t, z) = \sum_{i=1}^{m} \sum_{j=1}^{q} \hat{q}_{ij} I\{U_i \le t, Y_j \le z\} + \sum_{j=1}^{q} \hat{q}_{m+1,\,j} I\{t > U_m, z = Y_j\}, \tag{42}$$

which, by (15), gives the marginal d.f. of Z as follows:

$$\begin{aligned}
\hat{F}_n(\infty, z) &= P_{\hat{F}_n}\{T \le U_m, Z \le z\} + P_{\hat{F}_n}\{T > U_m, Z \le z\} \\
&= \hat{F}_n(U_m, z) + \sum_{j=1}^{q} P_{\hat{F}_n}\{T > U_m, Z = Y_j\} I\{Y_j \le z\} \\
&= \sum_{j=1}^{q} \sum_{i=1}^{m} \hat{q}_{ij} I\{Y_j \le z\} + \sum_{j=1}^{q} \hat{q}_{m+1,\,j} I\{Y_j \le z\} \\
&= \sum_{j=1}^{q} \left\{ \left(\sum_{i=1}^{m} \hat{q}_{ij} \right) + \hat{q}_{m+1,\,j} \right\} I\{Y_j \le z\} = \sum_{j=1}^{q} \left(\frac{N_{1j}}{n} \right) I\{Y_j \le z\} \\
&= n^{-1} \sum_{j=1}^{q} \sum_{k=1}^{n} I\{Z_k = Y_j\} I\{Y_j \le z\} = n^{-1} \sum_{k=1}^{n} I\{Z_k \le z\}.
\end{aligned} \tag{43}$$

□

Proof of Corollary 3 When there is no censoring, in (6) we have $n_{ij} = \delta_{ij}$ for all $1 \le i \le m,\ 1 \le j \le q$. Thus, from (20) and (15) we have for $1 \le i \le m$, $1 \le j \le q$:

$$\begin{aligned}
\hat{q}_{ij} &= \left(\frac{n_{ij}}{N_{ij}} \right) \left(\frac{N_{1j}}{n} \right) \prod_{k=1}^{i-1} \left(1 - \frac{n_{kj}}{N_{kj}} \right) = \left(\frac{n_{ij}}{N_{ij}} \right) \left(\frac{N_{1j}}{n} \right) \prod_{k=1}^{i-1} \left(\frac{N_{k+1,\,j}}{N_{kj}} \right) \\
&= \left(\frac{n_{ij}}{N_{ij}} \right) \left(\frac{N_{1j}}{n} \right) \left\{ \frac{N_{2j}}{N_{1j}} \cdot \frac{N_{3j}}{N_{2j}} \times \cdots \times \frac{N_{i,\,j}}{N_{i-1,\,j}} \right\} = \left(\frac{n_{ij}}{n} \right).
\end{aligned} \tag{44}$$

Thus, from (6) we have in (20):

$$\begin{aligned}
\hat{F}_n(t, z) &= n^{-1} \sum_{i=1}^{m} \sum_{j=1}^{q} n_{ij} I\{U_i \le t, Y_j \le z\} \\
&= n^{-1} \sum_{i=1}^{m} \sum_{j=1}^{q} \sum_{k=1}^{n} I\{V_k = U_i, Z_k = Y_j\} I\{U_i \le t, Y_j \le z\} \\
&= n^{-1} \sum_{k=1}^{n} \sum_{i=1}^{m} \sum_{j=1}^{q} I\{V_k = U_i \le t, Z_k = Y_j \le z\} \\
&= n^{-1} \sum_{k=1}^{n} I\{V_k \le t, Z_k \le z\},
\end{aligned} \tag{45}$$

which is the bivariate empirical d.f. of sample (1). □

Proof of Theorem 2 In (3), for all possible values of Z we denote the following:

$$z_k = Y_k,\;\; P_k = P\{Z = z_k\},\;\; F_{0k}(t) = P\{T \le t \mid Z = z_k\}, \tag{46}$$

where $k = 1, \ldots, q$. Then, the bivariate d.f. $F_0(t, z)$ of (T, Z) can be expressed as

$$F_0(t, z) = P\{T \le t, Z \le z\} = \sum_{z_k \le z} P\{T \le t, Z = z_k\} = \sum_{z_k \le z} P_k\, F_{0k}(t). \tag{47}$$

From (3), (15) and (21), we have

$$\hat{P}_k \equiv P_{\hat{F}_n}\{Z = z_k\} = n^{-1} \sum_{i=1}^{n} I\{V_i \ge U_1,\; Z_i = z_k\} = n^{-1} \sum_{i=1}^{n} I\{Z_i = z_k\}. \tag{48}$$

Since $\hat{F}_{nk}(t) \equiv \hat{F}_n(t \mid Z = z_k)$ in (22) is the Kaplan–Meier estimator for $F_{0k}(t)$ in (46) as in the univariate case, then from (21), (47)–(48) and Corollary 1.2 of Stute and Wang (1993) we know that (24) follows from

$$\begin{aligned}\sup_{0 \le t \le \tau_V,\, z \in \mathbb{R}} |\hat{F}_n(t, z) - F_0(t, z)| &= \sup_{0 \le t \le \tau_V,\, z \in \mathbb{R}} \Big|\sum_{z_k \le z} \hat{P}_k\, \hat{F}_{nk}(t) - F_0(t, z)\Big| \\ &\le \sum_{k=1}^{q} \Big(|\hat{P}_k - P_k| + \sup_{0 \le t \le \tau_V} |\hat{F}_{nk}(t) - F_{0k}(t)|\Big),\end{aligned} \tag{49}$$

because for each $1 \le k \le q$, $\hat{P}_k$ is a strong consistent estimator for P_k and $\hat{F}_{nk}(t)$ is a uniform strong consistent estimator for $F_{0k}(t)$. $\square$

Proof of Theorem 3 Using the notations in the proof of Theorem 5, since $\hat{F}_{nk}(t)$ is the Kaplan–Meier estimator for $F_{0k}(t)$ in the univariate case, then from (21), (47)–(48) and Gill (1983) we know that the proof follows from

$$\begin{aligned}\sqrt{n}[\hat{F}_n(t, z) - F_0(t, z)] &= \sqrt{n} \sum_{k=1}^{q} [\hat{P}_k\, \hat{F}_{nk}(t) - P_k\, F_{0k}(t)] I\{z_k \le z\} \\ &= \sum_{k=1}^{q} \Big(\hat{P}_k\, \sqrt{n}[\hat{F}_{nk}(t) - F_{0k}(t)] \\ &\quad + \sqrt{n}(\hat{P}_k - P_k) F_{0k}(t)\Big)\, I\{z_k \le z\} \\ &\overset{w}{\Rightarrow} \sum_{k=1}^{q} (P_k \mathbb{G}_k + \mathbb{Z}_k F_{0k}(t)) I\{z_k \le z\}, \quad \text{as } n \to \infty\end{aligned} \tag{50}$$

because for each $1 \le k \le q$, $\sqrt{n}(\hat{P}_k - P_k)$ weakly converges to zero-mean normal r.v. $\mathbb{Z}_k$ and $\sqrt{n}[\hat{F}_n(t, z) - F_0(t, z)]$ weakly converges to centered Gaussian process $\mathbb{G}_k$ as $n \to \infty$. $\square$

Acknowledgments The authors thank the Editor/Associate Editor and two referees for their comments and suggestions on the earlier draft of this article.

References

Akritas, M. A. (1994). Nearest neighbor estimation of a bivariate distribution under random censoring. *The Annals of Statistics*, *22*, 1299–1327.

Bazaraa, M. S., Sherali, H. D., Shetty, C. M. (1993). *Nonlinear programming, theory and algorithms* (2nd ed.). USA: Wiley.

Chang, M. N., Yang, G. L. (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. *The Annals of Statistics*, *15*, 1536–1547.

Dabrowska, D. M. (1988). Kaplan-Meier estimate on the plane. *The Annals of Statistics*, *16*, 1475–1489.

Dabrowska, D. M. (1989). Kaplan-Meier estimate on the plane: weak convergence, LIL, and the bootstrap. *Journal of Multivariate Analysis*, *29*, 308–325.

Dempster, A. P., Laird, N. M., Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM-algorithm. *Journal of the Royal Statistical Society, Series B*, *39*, 1–38.

Efron, B. (1967). The two sample problem with censored data. In *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability* (pp. 831-853). Berkeley: University of California Press

Gill, R. (1983). Large sample behavior of the product-limit estimator on the whole line. *The Annals of Statistics*, *11*, 49–58.

Lin, D. Y., Ying, Z. L. (1993). A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika*, *80*, 573–581.

Mykland, P. A., Ren, J. (1996). Self-consistent and maximum likelihood estimation for doubly censored data. *The Annals of Statistics*, *24*, 1740–1764.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, *75*, 237–249.

Ren, J., Gu, M. G. (1997). Regression M-estimators for doubly censored data. *The Annals of Statistics*, *25*, 2638–2664.

Ren, J., Riddlesworth, T. (2012). Empirical likelihood ratio confidence intervals for conditional survival probabilities with right censored data. *Statistics and Its Interface*, *5*, 339–343.

Stute, W., Wang, J.-L. (1993). The strong law under random censorship. *The Annals of Statistics*, *21*, 1591–1607.

Turnbull, B. W. (1976). The empirical distribution with arbitrarily grouped censored and truncated data. *Journal of the Royal Statistical Society, Series B*, *38*, 290–295.

van der Laan, M. J. (1996). Efficient estimation in the bivariate censoring model and repairing NPMLE. *The Annals of Statistics*, *24*, 596–627.