

Asymptotic properties of Kaplan–Meier estimator and hazard estimator for censored survival time with LENQD data

Yongming Li, Weicai Pang, Ziqing Feng & Naiyi Li

To cite this article: Yongming Li, Weicai Pang, Ziqing Feng & Naiyi Li (16 Jan 2024): Asymptotic properties of Kaplan–Meier estimator and hazard estimator for censored survival time with LENQD data, *Statistical Theory and Related Fields*, DOI: [10.1080/24754269.2024.2302754](https://doi.org/10.1080/24754269.2024.2302754)

To link to this article: <https://doi.org/10.1080/24754269.2024.2302754>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 16 Jan 2024.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

Asymptotic properties of Kaplan–Meier estimator and hazard estimator for censored survival time with LENQD data

Yongming Li ^a, Weicai Pang^b, Ziqing Feng^c and Naiyi Li^d

^aSchool of Mathematics and Computer Science, Shangrao Normal University, Shangrao, People's Republic of China; ^bSchool of Mathematics and Statistics, Nanning Normal University, Nanning, People's Republic of China; ^cCollege of Sciences, Nanchang University, Nanchang, People's Republic of China; ^dSchool of Mathematics and Computer, Guangdong Ocean University, Zhanjiang, People's Republic of China

ABSTRACT

In this paper, we consider the estimators of distribution function and hazard rate for censored survival time. First, some properties and inequalities are established for linearly extended negative quadrant-dependent sequence as auxiliary results. Then by applying the properties and inequalities, we investigate the strong consistency and strong representation for the Kaplan–Meier estimator and hazard rate estimator with censored linearly extended negative quadrant-dependent data. Under some mild conditions, we derive that the rates of strong consistency are near $O(n^{-1/2} \log^{1/2} n)$ and also obtain the strong representations with the remainder of order $O(n^{-1/2} \log^{1/2} n)$. The results established here extend and generalize the corresponding ones in recent literature.

ARTICLE HISTORY

Received 22 December 2021
Revised 25 November 2023
Accepted 25 December 2023

KEYWORDS

Linearly extended negative quadrant-dependent data; Kaplan–Meier estimator; strong consistency; strong representation

(2010) MATHEMATICS

SUBJECT CLASSIFICATIONS
62G20; 62G05

1. Introduction

The definitions of negatively quadrant dependence (NQD), negative dependence (ND), negative association (NA), linearly negative quadrant dependence (LNQD) and extended negative dependence (END) were given by Block et al. (1982), Joag-Dev and Proschan (1983), Lehmann (1966), Liu (2009) and Newman (1984), respectively. Their specific definitions are as follows.

Definition 1.1 (see Lehmann, 1966): The pair of random variables (X, Y) is said to be NQD, if for all real numbers x and y ,

$$P(X \leq x, Y \leq y) \leq P(X \leq x)P(Y \leq y).$$

Definition 1.2 (see Block et al., 1982): A finite family of random variables $\{X_1, \dots, X_n\}$ is said to be ND, if for all real numbers x_1, \dots, x_n such that both inequalities

$$P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) \leq M \prod_{i=1}^n P(X_i \leq x_i)$$

and



$$P(X_1 > x_1, X_2 > x_2, \dots, X_n > x_n) \leq M \prod_{i=1}^n P(X_i > x_i)$$

hold. An infinite sequence $\{X_n, n \geq 1\}$ is said to be ND if every finite subsequence is ND.

Definition 1.3 (see Joag-Dev & Proschan, 1983): A finite family of random variables $\{X_1, \dots, X_n\}$ is said to be NA, if for every pair of disjoint subsets A and B of $\{1, 2, \dots, n\}$,

$$\text{Cov}(f_1(X_i, i \in A), f_2(X_j, j \in B)) \leq 0,$$

whenever $f_1(\cdot)$ and $f_2(\cdot)$ are coordinatewise increasing (or decreasing) and the covariance exists. An infinite sequence $\{X_n, n \geq 1\}$ is said to be NA if every finite subsequence is NA.

CONTACT Yongming Li  ylm1019@163.com  School of Mathematics and Computer Science, Shangrao Normal University, Shangrao 334001, People's Republic of China

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Definition 1.4 (see Newman, 1984): A sequence of random variables $\{X_n, n \geq 1\}$ is said to be LNQD, if, for any disjoint subsets $A, B \subset \{1, 2, \dots\}$ and positive real numbers $\{r_1, r_2, \dots\}$ or negative real numbers $\{r_1, r_2, \dots\}$, the pair

$$\left(\sum_{i \in A} r_i X_i, \sum_{j \in B} r_j X_j \right) \text{ is NQD.}$$

Definition 1.5 (see Liu, 2009): A finite family of random variables $\{X_1, \dots, X_n\}$ is said to be END, if there exists a constant $M \geq 1$ such that both inequalities

$$P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) \leq M \prod_{i=1}^n P(X_i \leq x_i)$$

and

$$P(X_1 > x_1, X_2 > x_2, \dots, X_n > x_n) \leq M \prod_{i=1}^n P(X_i > x_i)$$

hold for all real numbers x_1, \dots, x_n . An infinite sequence $\{X_n, n \geq 1\}$ is said to be END if every finite subsequence is END.

From the above-mentioned literatures and definitions, we know that NQD, ND, NA, LNQD and END sequences are widely used in multivariate statistical analysis and reliability theory and survival analysis, and have been received more and more attention recently. For example, Wang and Zhang (2006) provided a Berry–Esseen theorem for LNQD sequence; Ko et al. (2007) discussed the strong convergence and central limit theorem for weighted sums of LNQD sequence; Wang et al. (2010) established some exponential inequalities and by which obtained complete convergence based on LNQD sequence; Li et al. (2012) provided some inequalities for LNQD sequence and by which investigated the asymptotic normality of weighted function estimator for regression function; Shen and Zhu (2015) studied the complete convergence for weighted sums of LNQD sequence and presented some complete convergence for arrays of rowwise LNQD sequence; Hu and Jiang (2018) discussed the uniformly asymptotic normality of sample quantile estimator for LNQE sequence; Hu and Wang (2020) investigated the Berry–Esseen bound of wavelet estimator for nonparametric regression model based on LNQD sequence.

Furthermore, we note that END sequence is more general than ND sequence in that it can reflect not only a negative dependence structure but also a positive one to some extent. Motivated by END and LNQD sequences, Li et al. (2023) put forth the definition of linearly extended negative quadrant dependence (LENQD) as follows.

Definition 1.6 (see Li et al., 2023): The pair (X, Y) is said to ENQD, if there exists a dominating constant $M \geq 1$, such that both inequalities

$$P(X \leq x, Y \leq y) \leq M P(X \leq x)P(Y \leq y)$$

and

$$P(X > x, Y > y) \leq M P(X > x)P(Y > y)$$

hold for every real x, y . A sequence of random variables $\{X_n, n \geq 1\}$ is said to be EPNQD, if for all $i, j \geq 1, i \neq j$, the pair (X_i, X_j) is ENQD. A sequence of random variables $\{X_n, n \geq 1\}$ is said to be LENQD with dominating constant $M \geq 1$, if for any disjoint subsets $A, B \subset \{1, 2, \dots\}$ and positive real numbers $\{l_1, l_2, \dots\}$ or negative real numbers $\{l_1, l_2, \dots\}$, the pair

$$\left(\sum_{i \in A} l_i X_i, \sum_{j \in B} l_j X_j \right) \text{ is ENQD with the same dominating constant } M.$$

From Examples 1.1 and 1.2 and Remark 1.1 in Li et al. (2023), it is easily seen that independent random variables, NA, END and LNQD random variables are LENQD random variables. Therefore, LENQD sequence has also wide applications in multivariate statistical analysis and reliability theory and survival analysis.

We know that Kaplan–Meier estimate is one of the commonly used methods in survival analysis and has received considerable attention in the literature. For example, Liebscher (2002) derived the rates of uniform strong convergence for kernel density estimators and hazard rate estimators for right censoring based on a stationary strong-mixing sequence; Liang and Una-Alvarez (2009) studied the convergence of Kaplan–Meier estimator

based on the stationary strong-mixing data; Wu and Chen (2013) established the strong representation results of Kaplan–Meier estimator for censored NA data. More recently, Shen and Wang (2016) investigated the strong convergence properties for Kaplan–Meier estimator and hazard estimator based on censored NSD data; Anevski (2017) derived the limit distribution results of Kaplan–Meier estimator for distribution function based on right censored observations with a stationary time series; Li and Zhou (2019) and Li and Zhou (2020) respectively investigated the Kaplan–Meier estimator and hazard estimator for censored WOD and END data; Ahmed and Flandre (2020) proposed a weighted Kaplan–Meier estimators to estimate the HIV-1 RNA reduction; Nematolahi et al. (2020) derived the asymptotic distribution of Kaplan–Meier estimator under PROS sampling design; Wu et al. (2022) investigated the rates of strong consistency and the strong representations for the Kaplan–Meier estimator and hazard estimator with censored WOD data.

However, the work mentioned above focuses basically on strong mixing dependent data, NA, NSD, END and WOD data. Noting that NA, END and LNQD imply LENQD, and we naturally ask whether those results can still hold for more comprehensive LENQD data in theory. From this, it is of interest to discuss the Kaplan–Meier estimator and hazard rate estimator of survival time based on censored LENQD data. Therefore, in this paper, we will investigate the strong consistency and the strong representation for Kaplan–Meier estimator and hazard rate estimator with censored LENQD data.

The remaining parts of this paper are structured as follows. Section 2 briefly provides some notations, properties and inequalities, while Section 3 is devoted to setting up some auxiliary results. Finally, we discuss and derive the strong consistency and strong representation for Kaplan–Meier estimator and hazard estimator based on censored LENQD data in Section 4.

2. Some notations, properties and inequalities

To facilitate the notation reference below, we begin with a brief review of Kaplan–Meier estimator and hazard rate estimator as follows.

Let $\{X_1, \dots, X_n\}$ be survival times with an unknown continuous distribution function $F(x) = P(X_i \leq x)$ satisfying $F(0) = 0$, and $\{Y_1, \dots, Y_n\}$ be random censoring times with an unknown distribution function $G(y) = P(Y_i \leq y)$ satisfying $G(0) = 0$. Suppose that $\{X_1, \dots, X_n\}$ and $\{Y_1, \dots, Y_n\}$ are mutually independent. Let survival times X_i be censored on the right by the censoring times Y_i , so that one can only observe (T_i, δ_i) , where

$$T_i = X_i \wedge Y_i \quad \text{and} \quad \delta_i = I(X_i \leq Y_i).$$

Here and thereafter, $a \wedge b$ means the minimum of a and b , and $I(A)$ means the indicator function of an event A .

Denote the number of uncensored observations less than or equal to t by

$$N_n(t) = \sum_{i=1}^n I(T_i \leq t, \delta_i = 1) = \sum_{i=1}^n I(X_i \leq t \wedge Y_i),$$

and the number of censored or uncensored observations greater than or equal to t by

$$M_n(t) = \sum_{i=1}^n I(T_i \geq t).$$

The Kaplan–Meier estimator of $F(x)$ was proposed by Kaplan and Meier (1958) as follows:

$$\widehat{F}_n(x) = 1 - \prod_{s \leq x} \left[1 - \frac{dN_n(s)}{M_n(s)} \right], \quad \text{where } dN_n(s) = N_n(s) - N_n(s-).$$

It is further assumed that $F(x)$ has a density $f(x)$, and then the hazard rate $\lambda(x)$ can be written as

$$\lambda(x) = \frac{d}{dx} (-\log \bar{F}(x)) = \frac{f(x)}{\bar{F}(x)}, \quad \text{where } F(x) < 1 \text{ and } \bar{F}(x) = 1 - F(x).$$

The distribution function and empirical distribution function of $\{T_n, n \geq 1\}$ are defined as

$$L(t) = 1 - \bar{F}(t)\bar{G}(t) = 1 - [1 - F(t)][1 - G(t)], \quad \bar{L}(t) = 1 - L(t)$$

and

$$L_n(t) = \frac{1}{n} \sum_{i=1}^n I(T_i < t) = 1 - \frac{M_n(t)}{n}, \quad \bar{L}_n(t) = 1 - L_n(t).$$

And the cumulative and empirical cumulative hazard rate functions are defined by

$$\Lambda(x) = -\log \bar{F}(x) = \int_0^x \frac{dF(t)}{\bar{F}(t)} = \int_0^x \frac{dF_*(t)}{\bar{L}(t)}$$

and

$$\widehat{\Lambda}_n(x) = \int_0^x \frac{dN_n(t)}{M_n(t)} = \int_0^x \frac{dF_{*n}(t)}{\bar{L}_n(t)},$$

where $F_*(t)$ and its empirical distribution function $F_{*n}(t)$ are defined as

$$F_*(t) = P(T_1 \leq t, \delta_1 = 1) = \int_0^\infty F(t \wedge y) dG(y) = \int_0^t \bar{G}(y) dF(y)$$

and

$$F_{*n}(t) = \frac{1}{n} \sum_{i=1}^n I(T_i \leq t, \delta_i = 1) = \frac{N_n(t)}{n}.$$

Noting that $N(t)$ is a step function, and $dN(T_{(k)}) = \delta_{(k)}$, $k = 1, 2, \dots, n$, we can easily rewrite the above estimators of $\widehat{\Lambda}_n(x)$ and $\widehat{F}_n(x)$ as follows:

$$\widehat{\Lambda}_n(x) = \sum_{k=1}^n \frac{I(T_{(k)} \leq x, \delta_{(k)} = 1)}{n - k + 1} \quad (1)$$

and

$$\widehat{F}_n(x) = 1 - \prod_{i=1}^n \left(1 - \frac{\delta_{(k)}}{n - k + 1} \right)^{I(T_{(k)} \leq x, \delta_{(k)} = 1)}, \quad (2)$$

where $T_{(1)} \leq T_{(2)} \leq \dots \leq T_{(n)}$ denotes the order statistics of T_1, T_2, \dots, T_n , and $\delta_{(i)}$ is the concomitant of $T_{(i)}$.

Next, we give a basic property and an inequality of LENQD sequence, which are helpful in proving the theorems in Sections 3 and 4.

- Lemma 2.1** (see Li et al., 2023): (i) Let (X, Y) be ENQD random variables with dominating coefficient $M \geq 1$. If f and g are nondecreasing (or nonincreasing) functions, then $(f(X), g(Y))$ is ENQD random variables, and its dominating constant M remains unchanged.
- (ii) Let $\{X_n, n \geq 1\}$ be LENQD sequence with dominating coefficient $M \geq 1$, and f_n be nondecreasing (or non-increasing) functions. Then $\{f_n(X_n), n \geq 1\}$ is LENQD sequence, and the dominating constant M remains unchanged.

Lemma 2.2 ((see Li et al., 2023) Bernstein-type inequality): Let $\{X_n, n \geq 1\}$ be a sequence of LENQD random variables with a dominating coefficient $M \geq 1$, and $EX_n = 0$, $|X_n| \leq d_n$ a.s. for each $n \geq 1$, where $\{d_n, n \geq 1\}$ is a sequence of positive constants. Assume that $t > 0$ such that $t \cdot \max_{1 \leq i \leq n} d_i \leq 1$. Then for any $\varepsilon > 0$,

$$P \left(\left| \sum_{i=1}^n X_i \right| > \varepsilon \right) \leq 2M \exp \left\{ -t\varepsilon + t^2 \sum_{i=1}^n EX_i^2 \right\}.$$

3. Some auxiliary results

We will establish two auxiliary results, which play an important role in the proof of the main results in Section 4.

Theorem 3.1: The union of independent sets of ENQD random variables is also ENQD.

Proof: Let $\mathbf{X} = \{X_1, X_2\}$ and $\mathbf{Y} = \{Y_1, Y_2\}$ be two independent random vectors, the pairs (X_1, X_2) and (Y_1, Y_2) be ENQD. We shall show that the pair of random vector $(\mathbf{X}, \mathbf{Y}) = \{(X_1, Y_1), (X_2, Y_2)\}$ is ENQD.

Noting that (X_1, X_2) is ENQD, and (Y_1, Y_2) is also ENQD, then by the independence of \mathbf{X} and \mathbf{Y} , we know there exist constants $M_1 \geq 1$ and $M_2 \geq 1$, and $M = M_1 M_2$, such that

$$\begin{aligned}
 & P(X_1 > x_1, Y_1 > y_1; X_2 > x_2, Y_2 > y_2) \\
 &= P(X_1 > x_1, X_2 > x_2) P(Y_1 > y_1, Y_2 > y_2) \\
 &\leq M_1 \prod_{i=1}^2 P(X_i > x_i) \cdot M_2 \prod_{i=1}^2 P(Y_i > y_i) \\
 &= MP(X_1 > x_1, Y_1 > y_1) P(X_2 > x_2, Y_2 > y_2).
 \end{aligned} \tag{3}$$

Similarly, we can obtain that

$$\begin{aligned}
 & P(X_1 \leq x_1, Y_1 \leq y_1; X_2 \leq x_2, Y_2 \leq y_2) \\
 &= P(X_1 \leq x_1, X_2 \leq x_2) P(Y_1 \leq y_1, Y_2 \leq y_2) \\
 &\leq M_1 \prod_{i=1}^2 P(X_i \leq x_i) \cdot M_2 \prod_{i=1}^2 P(Y_i \leq y_i) \\
 &= MP(X_1 \leq x_1, Y_1 \leq y_1) P(X_2 \leq x_2, Y_2 \leq y_2).
 \end{aligned} \tag{4}$$

By relations of (3) and (4), and Definition 1.6, we know that (X_1, Y_1) and (X_2, Y_2) are ENQD random variables. This completes the proof. \blacksquare

Theorem 3.2: Let $\{X_n, n \geq 1\}$ be a sequence of LENQD random variables with a dominating coefficient $M \geq 1$, and having unknown distribution function $F(x)$ with bounded probability density function $f(x)$. Setting $F_n(x) = n^{-1} \sum_{i=1}^n I(X_i \leq x)$ as the empirical distribution function, and taking $\tau_n = n^{-1/2} \log^{1/2} n$, then

$$\sup_x |F_n(x) - F(x)| = O(\tau_n), \text{ a.s.}$$

Proof: By Lemma 2 of Yang (2003), let $\{x_{n,k}\}$ satisfy $F(x_{n,k}) = k/n$ for $n \geq 3$ and $k = 1, \dots, n-1$, and then we obtain that

$$\sup_{-\infty < x < +\infty} |F_n(x) - F(x)| \leq \max_{1 \leq k \leq n-1} |F_n(x_{n,k}) - F(x_{n,k})| + \frac{2}{n}.$$

It is easy to see that $n\tau_n \rightarrow \infty$. Therefore, for all n large enough and $\varepsilon \geq 4\sqrt{3}$, we have $2/n < \varepsilon\tau_n/2$. Hence, we can get that

$$\begin{aligned}
 P\left(\sup_x |F_n(x) - F(x)| > \varepsilon\tau_n\right) &\leq P\left(\max_{1 \leq k \leq n-1} |F_n(x_{n,k}) - F(x_{n,k})| > \varepsilon\tau_n/2\right) \\
 &\leq \sum_{k=1}^{n-1} P(|F_n(x_{n,k}) - F(x_{n,k})| > \varepsilon\tau_n/2).
 \end{aligned} \tag{5}$$

Set $\xi_{j,k} = I(X_j < x_{n,k}) - EI(X_j < x_{n,k})$, $j = 1, \dots, n$. Then it follows from Lemma 2.1 that $\{\xi_{j,k}\}$ is still LENQD random variables with $\xi_{j,k} = 0$, $|\xi_{j,k}| \leq 2$. Taking $t = \varepsilon\tau_n/4$, and applying Lemma 2.2, we can obtain that

$$\begin{aligned}
 P(|F_n(x_{n,k}) - F(x_{n,k})| > \varepsilon\tau_n/2) &= P\left(\left|\sum_{j=1}^n \xi_{j,k}\right| > \varepsilon\tau_n/2\right) \\
 &\leq 2M \exp\left\{-t\varepsilon n\tau_n/2 + t^2 \sum_{j=1}^n E\xi_{j,k}^2\right\} \\
 &\leq 2M \exp\{-t\varepsilon n\tau_n/2 + nt^2\} \\
 &\leq 2M \exp\left\{-\frac{\varepsilon^2 n\tau_n^2}{16}\right\} \leq 2Mn^{-3}.
 \end{aligned} \tag{6}$$

Thus, with relations (5) and (6), we obtain

$$P\left(\sup_x |F_n(x) - F(x)| > \varepsilon \tau_n\right) \leq 2Mn^{-2}.$$

Therefore, Theorem 3.2 holds. ■

4. Main results

We can obtain the strong consistency and strong representation for Kaplan–Meier estimator and hazard estimator based on censored LENQD data as follows.

Theorem 4.1: Let $\{X_n, n \geq 1\}$ and $\{Y_n, n \geq 1\}$ be two LENQD sequences with a dominating coefficient $M \geq 1$. Suppose $\{X_n, n \geq 1\}$ and $\{Y_n, n \geq 1\}$ are independent. Then, for any $0 < \tau < \tau_L$,

$$\sup_{0 \leq x \leq \tau} |\widehat{\Lambda}_n(x) - \Lambda(x)| = O(\tau_n), \text{ a.s.} \quad (7)$$

and

$$\sup_{0 \leq x \leq \tau} |\widehat{F}_n(x) - F(x)| = O(\tau_n), \text{ a.s.} \quad (8)$$

Theorem 4.2: Under the assumptions of Theorem 4.1, we have

$$\widehat{\Lambda}_n(x) - \Lambda(x) = -\frac{1}{n} \sum_{i=1}^n \eta(T_i, \delta_i, x) + R_{1n}(x) \quad (9)$$

and

$$\widehat{F}_n(x) - F(x) = \frac{\bar{F}(x)}{n} \sum_{i=1}^n \eta(T_i, \delta_i, x) + R_{2n}(x), \quad (10)$$

where, for $j = 1, 2$,

$$\sup_{0 \leq x \leq \tau} |R_{jn}(x)| = O(\tau_n), \text{ a.s.}, \quad \eta(t, \delta, x) = g(t \wedge x) - \frac{I(t \leq x, \delta = 1)}{\bar{L}(t)}, \quad g(x) = \int_0^x \frac{dF_*(t)}{\bar{L}^2(t)}.$$

Remark 4.1: By the auxiliary results in the previous section, we can prove Theorems 4.1 and 4.2 which are the extensions of those results in Wu and Chen (2013) from NA to LENQD, Li and Zhou (2020) from END to LENQD, Wu et al. (2022) from WOD to LENQD, and so on.

Now to prove Theorems 4.1 and 4.2 as follows.

Proof of Theorem 4.1: By Lemma 2.1 and Theorem 3.1, $\{T_n, n \geq 1\}$ and $\{(T_n, \delta_n), n \geq 1\}$ are also two LENQD sequences. Then, by Theorem 3.2 we obtain

$$\sup_{t \geq 0} |L_n(t) - L(t)| = O(\tau_n), \text{ a.s.} \quad (11)$$

and

$$\sup_{t \geq 0} |F_{*n}(t) - F_*(t)| = O(\tau_n), \text{ a.s.} \quad (12)$$

Following the same lines of the proof of Equations (3.1) and (3.2) in Li and Zhou (2020), by straightforward calculations, we have

$$\begin{aligned} & \widehat{\Lambda}_n(x) - \Lambda(x) \\ &= \int_0^x \frac{\bar{L}(t) - \bar{L}_n(t)}{\bar{L}_n(t)\bar{L}(t)} dF_*(t) + \frac{F_{*n}(t) - F_*(t)}{\bar{L}_n(t)} - \int_0^x (F_{*n}(t) - F_*(t)) d\left(\frac{1}{\bar{L}_n(t)}\right), \end{aligned}$$

and then, by the equations of (11) and (12), for $0 < \tau < \tau_L$, we have that

$$\begin{aligned} & \sup_{0 \leq x < \tau} |\widehat{\Lambda}_n(x) - \Lambda(x)| \\ & \leq \frac{\sup_{x \geq 0} |\bar{L}_n(x) - \bar{L}(x)|}{\bar{L}_n(\tau)\bar{L}(\tau)} + \frac{2 \sup_{x \geq 0} |F_{*n}(x) - F_*(x)|}{\bar{L}_n(\tau)} = O(\tau_n). \end{aligned}$$

Thus Equation (7) follows immediately.

We continue with the proof of Equation (8). By $e^x = 1 + x + o(x)$, we can get

$$\begin{aligned} & \widehat{F}_n(x) - F(x) \\ & = e^{-\Lambda(x)} [\widehat{\Lambda}_n(x) - \Lambda(x) + o(\widehat{\Lambda}_n(x) - \Lambda(x))] + [1 - \widehat{F}_n(x)] \{-\widehat{\Lambda}_n(x) \\ & \quad - \log[1 - \widehat{F}_n(x)] + o(-\widehat{\Lambda}_n(x) - \log[1 - \widehat{F}_n(x)])\}. \end{aligned} \quad (13)$$

By the notations of $\widehat{\Lambda}_n(x)$ and $\widehat{F}_n(x)$ in (1) and (2), for $0 < \tau < \tau_L$, $0 \leq x \leq \tau$, we can compute that

$$\begin{aligned} & 0 \leq -\widehat{\Lambda}_n(x) - \log[1 - \widehat{F}_n(x)] \\ & = -\sum_{i=1}^n \frac{I(\delta_{(i)} = 1, T_{(i)} \leq x)}{n-i+1} - \sum_{i=1}^n I(\delta_{(i)} = 1, T_{(i)} \leq x) \log \frac{n-i}{n-i+1} \\ & \leq \sum_{i=1}^n I(\delta_{(i)} = 1, T_{(i)} \leq x) \frac{1}{(n-i)(n-i+1)} \\ & = \sum_{i=1}^{n-M_n(x)} \left(\frac{1}{n-i} - \frac{1}{n-i+1} \right) = \frac{1}{M_n(x)} - \frac{1}{n} \\ & \leq \frac{1}{n} \frac{1}{\bar{L}_n(x)} = O\left(\frac{1}{n}\right). \end{aligned} \quad (14)$$

Hence, from (13) and (14), we conclude that

$$\widehat{F}_n(x) - F(x) = \bar{F}(x) [\widehat{\Lambda}_n(x) - \Lambda(x) + o(\widehat{\Lambda}_n(x) - \Lambda(x))] + O\left(\frac{1}{n}\right). \quad (15)$$

Therefore, from (15), and by (7) again, (8) holds. The proof is completed. ■

Proof of Theorem 4.2: It is easily seen that

$$\begin{aligned} & \widehat{\Lambda}_n(x) - \Lambda(x) \\ & = \left(\int_0^x \frac{dF_{*n}(t)}{\bar{L}(t)} - \int_0^x \frac{\bar{L}_n(t)}{\bar{L}^2(t)} dF_*(t) \right) \\ & \quad + \int_0^x \left(\frac{1}{\bar{L}_n(t)} - \frac{1}{\bar{L}(t)} \right) d(F_{*n}(t) - F_*(t)) + \int_0^x \frac{(\bar{L}_n(t) - \bar{L}(t))^2}{\bar{L}^2(t)\bar{L}_n(t)} dF_*(t) \\ & =: J_1(x) + J_2(x) + J_3(x). \end{aligned} \quad (16)$$

Firstly, for $J_1(x)$, noting that $F_{*n}(t) = N_n(t)/n$ and $N_n(t)$ is a step function, we derive that

$$J_1(x) = \frac{1}{n} \sum_{i=1}^n \frac{I(T_i \leq x, \delta_i = 1)}{\bar{L}(T_i)} - \frac{1}{n} \sum_{i=1}^n g(x \wedge T_i) = -\frac{1}{n} \sum_{i=1}^n \eta(T_i, \delta_i, x). \quad (17)$$

Secondly, we analyse $J_2(x)$. By dividing the interval $[0, \tau]$ into $[x_1, x_2], \dots, [x_i, x_{i+1}], \dots, [x_{k_n}, x_{k_n+1}]$, such that $\Lambda(x_{i+1}) - \Lambda(x_i) = O(\tau_n)$, where $k_n = O(\tau_n^{-1})$, $0 = x_1 < x_2 < \dots < x_{k_n+1} = \tau$, and $\tau < \tau_\Lambda$, we have the following

decomposition:

$$\begin{aligned}
J_2(x) &= \int_0^x \left(\frac{1}{\bar{L}_n(t)} - \frac{1}{\bar{L}(t)} \right) d(F_{*n}(t) - F_*(t)) \\
&\leq c \max_{1 \leq i \leq k_n} \sup_{y \in [x_i, x_{i+1}]} |\bar{L}_n(y) - \bar{L}_n(x_i) - \bar{L}(y) + \bar{L}(x_i)| \\
&\quad + c \max_{1 \leq i \leq k_n} |F_{*n}(x_{i+1}) - F_{*n}(x_i) - F_*(x_{i+1}) + F_*(x_i)| + O(\tau_n^2) \\
&=: J_{21} + J_{22} + O(\tau_n^2).
\end{aligned} \tag{18}$$

We first evaluate J_{21} in (18). Let us divide each $[x_i, x_{i+1}]$ into subintervals $[x_{ij}, x_{i(j+1)}]$, $j = 1, \dots, b_n$, such that $|\bar{L}_n(x_{i(j+1)}) - \bar{L}(x_{ij})| = O(\tau_n^{3/2})$ uniformly in i, j , where $b_n = O(k_n^{1/2}) = O(\tau_n^{-1/2})$. By Theorem 4.1 and $|\bar{L}_n(y) - \bar{L}_n(x_{ij})| \leq \frac{1}{n} \leq O(\tau_n^{3/2})$ for $y \in [x_{ij}, x_{i(j+1)}]$, we conclude that

$$\begin{aligned}
J_{21} &\leq \max_{1 \leq i \leq k_n} \max_{1 \leq j \leq b_n} |\bar{L}_n(x_{ij}) - \bar{L}_n(x_i) - \bar{L}(x_{ij}) + \bar{L}(x_i)| + O(\tau_n^{3/2}) \\
&= \max_{1 \leq i \leq k_n} \max_{1 \leq j \leq b_n} \left| \frac{1}{n} \sum_{k=1}^n (\eta_{ik} + \zeta_{ijk}) \right| + O(\tau_n^{3/2}),
\end{aligned} \tag{19}$$

where

$$\eta_{ik} = I(T_k \geq x_i) - EI(T_k \geq x_i), \quad \zeta_{ijk} = I(T_k \geq x_{ij}) - EI(T_k \geq x_{ij}), \quad k = 1, \dots, n,$$

and

$$\bar{L}_n(x_{ij}) - \bar{L}_n(x_i) - \bar{L}(x_{ij}) + \bar{L}(x_i) = \frac{1}{n} \sum_{k=1}^n (\eta_{ik} + \zeta_{ijk}).$$

According to Lemma 2.1, we know that $\{\eta_{ik}\}$ and $\{\zeta_{ijk}\}$ are both LENQD with the same dominating constant M , and

$$|\eta_{ik}| \leq 1, \quad |\zeta_{ijk}| \leq 1, \quad E\eta_{ik} = E\zeta_{ijk} = 0, \quad E\eta_{ik}^2 \leq 1, \quad E\zeta_{ijk}^2 \leq 1.$$

Then, applying Lemma 2.2, and taking $t = \tau_n$, we derive that

$$\begin{aligned}
&\mathbb{P} \left(\max_{1 \leq i \leq k_n} \max_{1 \leq j \leq b_n} \left| \frac{1}{n} \sum_{k=1}^n (\eta_{ik} + \zeta_{ijk}) \right| \geq 8\tau_n \right) \\
&\leq \sum_{i=1}^{k_n} \sum_{j=1}^{b_n} \mathbb{P} \left(\left| \sum_{k=1}^n (\eta_{ik} + \zeta_{ijk}) \right| \geq 8n\tau_n \right) \\
&\leq \sum_{i=1}^{k_n} \sum_{j=1}^{b_n} \mathbb{P} \left(\left| \sum_{k=1}^n \eta_{ik} \right| \geq 4n\tau_n \right) + \sum_{i=1}^{k_n} \sum_{j=1}^{b_n} \mathbb{P} \left(\left| \sum_{k=1}^n \zeta_{ijk} \right| \geq 4n\tau_n \right) \\
&\leq \sum_{i=1}^{k_n} \sum_{j=1}^{b_n} 4M \exp(-4n\tau_n^2 + n\tau_n^2) \\
&= 4Mk_nb_n \exp(-3 \log n) \leq CMn^{-2}.
\end{aligned} \tag{20}$$

By the Borel–Cantelli lemma, from (19) and (20), we obtain that

$$J_{21} = O(\tau_n), \text{ a.s.} \tag{21}$$

Next, we will deal with J_{22} in (18). Since $|F_*(x) - F_*(y)| \leq |\bar{L}(x) - \bar{L}(y)|$ for all x and y , similarly to the conclusion of J_{21} , we obtain that

$$J_{22} = O(\tau_n). \tag{22}$$

Therefore, from (18), (21) and (22), it follows that

$$J_2(x) = O(\tau_n). \tag{23}$$

Finally, for $J_3(x)$, by applying Theorem 2.2, we derive that

$$J_3(x) = \int_0^x \frac{[\bar{L}_n(t) - \bar{L}(t)]^2}{\bar{L}^2(t)\bar{L}_n(t)} dF_*(t) = O(\tau_n^2). \quad (24)$$

Therefore, combining with (16)–(18) and (23)–(24), we immediately obtain (9). And from (13) and (9), we obtain (10). This completes the proof of the theorem. ■

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Natural Science Foundation of China (12161075), Jiangxi Provincial Natural Science Foundation (20212ACB201006) and Natural Science Foundation of Guangdong Province (2022A1515010978).

ORCID

Yongming Li  <http://orcid.org/0000-0003-3456-9230>

References

- Ahmed, I., & Flandre, P. (2020). Weighted Kaplan–Meier estimators motivating to estimate HIV-1 RNA reduction censored by a limit of detection. *Statistics in Medicine*, 39(7), 968–983. <https://doi.org/10.1002/sim.8455>
- Anevski, D. (2017). Functional central limit theorems for the Nelson–Aalen and Kaplan–Meier estimators for dependent stationary data. *Statistics and Probability Letter*, 124(1), 83–91. <https://doi.org/10.1016/j.spl.2017.01.005>
- Block, H. W., Savits, T. H., & Shaked, M. (1982). Some concepts of negative dependence. *The Annals of Probability*, 10(3), 765–772. <https://doi.org/10.1214/aop/1176993784>
- Hu, X. P., & Jiang, R. (2018). Uniformly asymptotic normality of sample quantiles estimator for linearly negative quadrant dependent samples. *Journal of Inequalities and Applications*, 2018(196), 1–12. <https://doi.org/10.1186/s13660-018-1788-6>
- Hu, X.P., & Wang, J. Y. (2020). A Berry–Esseen bound of wavelet estimation for a nonparametric regression model under linear process errors based on LNQD sequence. *Aims Mathematics*, 5(6), 6985–6995. <https://doi.org/10.3934/math.2020448>
- Joag-Dev, K., & Proschan, F. (1983). Negative association of random variables with applications. *Annals of Statistics*, 11(1), 286–295. <https://doi.org/10.1214/aos/1176346079>
- Kaplan, E. M., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(2), 457–481. <https://doi.org/10.1080/01621459.1958.10501452>
- Ko, M. H., Ryu, D. H., & Kim, T. S. (2007). Limiting behaviors of weighted sums for linearly negative quadrant dependent random variables. *Journal of the American Statistical Association*, 11(2), 511–522. <https://doi.org/10.11650/twjm/1500404705>
- Lehmann, E. (1966). Some concepts of dependence. *Annals of Mathematical Statistics*, 37(5), 1137–1153. <https://doi.org/10.1214/aoms/1177699260>
- Li, Y. M., Guo, J. H., & Li, N. Y. (2012). Some inequalities for a LNQD sequence with applications. *Journal of Inequalities and Applications*, 2012(216), 1–9. <https://doi.org/10.1186/1029-242X-2012-216>
- Li, Y. M., Pang, W. C., Feng, Z. Q., & Li, N. Y. (2023). On the linearly extended negative quadrant dependent random variables and its inequalities. *Communication Statistics-Theory and Method*, 52(24), 8696–8711. <https://doi.org/10.1080/03610926.2022.2068600>
- Li, Y. M., & Zhou, Y. (2019). Asymptotic properties of the Kaplan–Meier estimator and hazard rate estimator for right censored and widely orthant dependent data. *Acta Mathematicae Applicatae Sinica*, 42(1), 71–84. <https://doi.org/10.12387/C2019006>
- Li, Y. M., & Zhou, Y. (2020). The Kaplan–Meier estimator and hazard estimator for censored END survival time observations. *Communication Statistics-Theory and Method*, 49(11), 2690–2702. <https://doi.org/10.1080/03610926.2019.1580737>
- Liang, H. Y., & Una-Alvarez, J. (2009). A Berry–Esseen type bound in kernel density estimation for strong mixing censored samples. *Journal of Multivariate Analysis*, 100(6), 1219–1231. <https://doi.org/10.1016/j.jmva.2008.11.001>
- Liebscher, E. (2002). Kernel density and hazard rate estimation for censored data under α -mixing condition. *Annals of the Institute of Statistical Mathematics*, 54(1), 19–28. <https://doi.org/10.1023/A:1016157519826>
- Liu, L. (2009). Precise large deviations for dependent random variables with heavy tails. *Statistics and Probability Letter*, 79(9), 1290–1298. <https://doi.org/10.1016/j.spl.2009.02.001>
- Nematollahi, S., Nazari, S., Shayan, Z., Ayatollahi, S. M. T., & Amanati, A. (2020). Improved Kaplan–Meier estimator in survival analysis based on partially rank-ordered set samples. *Computational and Mathematical Methods in Medicine*, 2020(4), 1–11. <https://doi.org/10.1155/2020/7827434>
- Newman, C. M. (1984). Asymptotic independent and limit theorems for positively and negatively dependent random variables. In *Inequalities in statistics and probability* (Lincoln, Neb., 1982), IMS Lecture Notes–Monograph Series (Vol. 5, pp. 127–140). Inst Math Statist.
- Shen, A. T., & Wang, X. J. (2016). Kaplan–Meier estimator and hazard estimator for censored negatively superadditive dependent data. *Statistics*, 50(2), 377–388. <https://doi.org/10.1080/02331888.2015.1038269>
- Shen, A. T., & Zhu, H. Y. (2015). Complete convergence for weighted sums of LNQD random variables. *Stochastics*, 87(1), 160–169. <https://doi.org/10.1080/17442508.2014.931959>

- Wang, J. F., & Zhang, L. X. (2006). A Berry–Esseen theorem for weakly negatively dependent random variables and its applications. *Acta Mathematica Hungarica*, 110(4), 293–308. <https://doi.org/10.1007/s10474-006-0024-x>
- Wang, X. J., Hu, S. H., Yang, W. Z., & Li, X. Q. (2010). Exponential inequalities and complete convergence for a LNQD sequence. *Journal of the Korean Statistical Society*, 39(4), 555–564. <https://doi.org/10.1016/j.jkss.2010.01.002>
- Wu, Q. Y., & Chen, P. Y. (2013). Strong representation results of the Kaplan–Meier estimator for censored negatively associated data. *Journal of Inequalities and Applications*, 2013(1), 1–9. <https://doi.org/10.1186/1029-242X-2013-1>
- Wu, Y., Yu, W., & Wang, X. J. (2022). Strong representations of the Kaplan–Meier estimator and hazard estimator with censored widely orthant dependent. *Computational Statistics*, 37(1), 383–402. <https://doi.org/10.1007/s00180-021-01125-z>
- Yang, S. C. (2003). Consistency of nearest neighbor estimator of density function for negative associated samples. *Acta Mathematicae Applicatae Sinica*, 26(3), 385–395. https://doi.org/10.1300/J025v20n01_07