Journal of Inequalities and Applications
a SpringerOpen Journal

RESEARCH **Open Access**

Strong representation results of the Kaplan-Meier estimator for censored negatively associated data

Qunying Wu[1*] and Pingyan Chen[2]

*Correspondence: wqy666@glut.edu.cn
[1]College of Science, Guilin University of Technology, Guilin, 541004, P.R. China
Full list of author information is available at the end of the article

Abstract
In this paper, we discuss the strong convergence rates and strong representation of the Kaplan-Meier estimator and the hazard estimator based on censored data when the survival and the censoring times form negatively associated (NA) sequences. Under certain regularity conditions, strong convergence rates are established for the Kaplan-Meier estimator and the hazard estimator, and the Kaplan-Meier estimator and the hazard estimator can be expressed as the mean of random variables, with the remainder of order $n^{-1/2}\ln^{1/2} n$ a.s.

MSC: Primary 60F15; secondary 60F05

Keywords: NA sequence; random censorship model; Kaplan-Meier estimator; strong representation; strong convergence rate

1 Introduction and main results

Let $\{T_i; i \geq 1\}$ be a sequence of true survival times. Random variables (r.v.s) are not assumed to be mutually independent; it is assumed, however, that they have a common unknown continuous marginal distribution function (d.f.) $F(x) = P(T_i \leq x)$ such that $F(0) = 0$. Let the r.v.s T_i be censored on the right by the censoring r.v.s Y_i, so that one observes only (Z_i, δ_i), where

$$Z_i = \min(T_i, Y_i) := T_i \wedge Y_i \quad \text{and} \quad \delta_i = I(T_i \leq Y_i), \quad i = 1, \ldots, n.$$

Here and in the sequel, $I(A)$ is the indicator random variable of the event A. In this random censorship model, the censoring times Y_i, $i = 1, \ldots, n$, are assumed to have the common distribution function $G(y) = P(Y_i \leq y)$ such that $G(0) = 0$; they are also assumed to be independent of the r.v.s T_i's. The problem at hand is that of drawing nonparametric inference about F based on the censored observations (Z_i, δ_i), $i = 1, \ldots, n$. For this purpose, define two stochastic processes on $[0, \infty)$ as follows:

$$N_n(t) = \sum_{k=1}^{n} I(Z_k \leq t, \delta_k = 1) = \sum_{k=1}^{n} I(T_k \leq t \wedge Y_k),$$

the number of uncensored observations less than or equal to t, and

$$Y_n(t) = \sum_{k=1}^{n} I(Z_k \geq t),$$

the number of censored or uncensored observations greater than or equal to t. The following nonparametric estimation $\hat{F}_n$ of F due to Kaplan and Meier [1] is widely used to estimate F on the basis of the data (Z_i, δ_i):

$$\hat{F}_n(x) = 1 - \prod_{s \leq x} \left(1 - \frac{\mathrm{d}N_n(s)}{Y_n(s)}\right),$$

where $\mathrm{d}N_n(s) = N_n(s) - N_n(s-)$.

Let L be the distribution of the Z_i's, $\bar{L} := 1 - L$. Since the sequences $\{T_n; n \geq 1\}$ and $\{Y_n; n \geq 1\}$ are independent, it follows that $L = 1 - \bar{F}\bar{G} = 1 - (1 - F)(1 - G)$. The empirical d.f. $L_n(t)$ of L is defined by

$$L_n(t) := \frac{1}{n} \sum_{k=1}^{n} I(Z_k < t) = 1 - \frac{Y_n(t)}{n} := \frac{\bar{Y}_n(t)}{n},$$

where $\bar{Y}_n(t) = \sum_{k=1}^{n} I(Z_k < t)$.

Define (possibly infinite) times τ_F, τ_G and τ_L by

$$\tau_F = \inf\{y; F(y) = 1\}, \qquad \tau_G = \inf\{y; G(y) = 1\}, \qquad \tau_L = \inf\{y; L(y) = 1\}.$$

Then $\tau_L = \tau_F \wedge \tau_G$. By setting

$$F_*(t) = P(Z_1 \leq t, \delta_1 = 1) = P(T_1 \leq t \wedge Y_1),$$

and the empirical d.f. of F_* is defined by

$$F_{*n}(t) := \frac{1}{n} \sum_{k=1}^{n} I(Z_k \leq t, \delta_k = 1) = \frac{N_n(t)}{n}.$$

We have then

$$F_*(t) = \int_0^{\infty} F(t \wedge z)\,\mathrm{d}G(z) = \int_0^{t} \bar{G}(z)\,\mathrm{d}F(z),$$

and

$$\mathrm{d}F_*(t) = \bar{G}(t)\,\mathrm{d}F(t).$$

Another question of interest in survival analysis is the estimation of the hazard function h defined as follows when it is further assumed that F has a density f:

$$h(x) := \frac{\mathrm{d}}{\mathrm{d}x}\left(-\log \bar{F}(x)\right) = f(x)/\bar{F}(x) \quad \text{for } F(x) < 1,$$

with $\bar{F} = 1 - F$. The quantity

$$H(x) = -\log \bar{F}(x) = \int_0^x \frac{dF(s)}{\bar{F}(s)} = \int_0^x \frac{dF_*(s)}{\bar{L}(s)} \tag{1.1}$$

is called the cumulative hazard function. The empirical cumulative hazard function $\hat{H}_n(x)$ is given by

$$\hat{H}_n(x) := \int_0^x \frac{dN_n(s)}{Y_n(s)} = \int_0^x \frac{dF_{*n}(s)}{\bar{L}_n(s)}, \tag{1.2}$$

where $\bar{L}_n(s) = 1 - L_n(s)$.

Since $N_n(t)$ is a step function, and $dN_n(Z_{(k)}) = \delta_{(k)}$, $k = 1, 2, \ldots, n$, it can be easily seen that

$$\hat{H}_n(x) = \sum_{k=1}^{n} \frac{I(Z_{(k)} \le x, \delta_{(k)} = 1)}{n - k + 1}, \tag{1.3}$$

and

$$\hat{F}_n(x) = 1 - \prod_{k=1}^{n} \left(1 - \frac{\delta_{(k)}}{n - k + 1}\right)^{I(Z_{(k)} \le x)} = 1 - \prod_{k=1}^{n} \left(\frac{n - k}{n - k + 1}\right)^{I(\delta_{(k)} = 1, Z_{(k)} \le x)}, \tag{1.4}$$

where $Z_{(1)} \le Z_{(2)} \le \cdots \le Z_{(n)}$ denote the order statistics of $Z_1, Z_2, \ldots, Z_n$, and $\delta_{(i)}$ is the concomitant of $Z_{(i)}$.

There is extensive literature on the Kaplan-Meier and the hazard estimator $\hat{F}_n(x)$ and $\hat{H}_n(x)$ for censored independent observations. We refer to papers by Breslow and Crowley [2], Foldes and Rejto [3] and Gu and Lai [4]. Martingale methods for analyzing properties of $\hat{F}_n(x)$ are described in the monograph by Gill [5]. However, the censored dependent data appear in a number of applications. For example, repeated measurements in survival analysis follow this pattern, see Kang and Koehler [6] or Wei *et al.* [7]. In the context of censored time series analysis, Shumway *et al.* [8] considered (hourly or daily) measurements of the concentration of a given substance subject to some detection limits, thus being potentially censored from the right. Ying and Wei [9], Lecoutre and Ould-Saïd [10], Cai [11] and Liang and Uña-Álvarez [12] studied the convergence of $\hat{F}_n(x)$ for the stationary α-mixing data.

The main purpose of this paper is to study the strong convergence rates and strong representation of the Kaplan-Meier estimator and the hazard estimator based on censored data when the survival and the censoring times form the NA (see the following definition) sequences. Under certain regularity conditions, we find strong convergence rates of the Kaplan-Meier and hazard estimator, and the expression of the Kaplan-Meier estimator and the hazard estimator as the mean of random variables, with the remainder of order $n^{-1/2} \ln^{1/2} n$ a.s.

Definition Random variables $X_1, X_2, \ldots, X_n$, $n \ge 2$ are said to be negatively associated (NA) if for every pair of disjoint subsets A_1 and A_2 of $\{1, 2, \ldots, n\}$,

$$\operatorname{cov}\big(f_1(X_i; i \in A_1), f_2(X_j; j \in A_2)\big) \le 0,$$

where f_1 and f_2 are increasing for every variable (or decreasing for every variable) so that this covariance exists. A sequence of random variables $\{X_i; i \geq 1\}$ is said to be NA if every finite subfamily is NA.

Obviously, if $\{X_i; i \geq 1\}$ is a sequence of NA random variables, and $\{f_i; i \geq 1\}$ is a sequence of nondecreasing (or non-increasing) functions, then $\{f_i(X_i); i \geq 1\}$ is also a sequence of NA random variables.

This definition was introduced by Joag-Dev and Proschan [13]. A statistical test depends greatly on sampling. The random sampling without replacement from a finite population is NA, but is not independent. NA sampling has wide applications such as in multivariate statistical analysis and reliability theory. Because of the wide applications of NA sampling, the limit behaviors of NA random variables have received more and more attention recently. One can refer to Joag-Dev and Proschan [13] for fundamental properties, Matula [14] for the three series theorem, and Wu and Jiang [15, 16] for the strong convergence.

We give two lemmas, which are helpful in proving our theorems.

Lemma 1.1 (Yang [17], Lemma 1) *Let $\{X_i; i \geq 1\}$ be a sequence of negatively associated random variables with zero means and $|X_i| \leq b_i$, a.s. $(i = 1, 2, \ldots)$. Let $t > 0$ be such that $t \max_{1 \leq i \leq n} b_i \leq 1$. Then, for all $\varepsilon > 0$,*

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

Lemma 1.2 *Let $\{X_i; i \geq 1\}$ be a sequence of NA r.v.s with continuous d.f. F, and let $F_n(x) := \frac{1}{n}\sum_{i=1}^{n} I(X_i < x)$ be the empirical d.f. based on the segments $X_1, \ldots, X_n$. Then*

$$\sup_{x \in \mathbb{R}} \left|F_n(x) - F(x)\right| = O\left(n^{-1/2} \ln^{1/2} n\right) \quad a.s.$$

Proof Similar to the proof of Lemma 4 in Yang [17], we can prove Lemma 1.2. □

Theorem 1.3 *Let $\{T_n; n \geq 1\}$ and $\{Y_n; n \geq 1\}$ be two sequences of NA random variables. Suppose that the sequences $\{T_n; n \geq 1\}$ and $\{Y_n; n \geq 1\}$ are independent. Then, for any $0 < \tau < \tau_L$,*

$$\sup_{0 \leq t \leq \tau} \left|\hat{H}_n(t) - H(t)\right| = O(a_n) \quad a.s. \tag{1.5}$$

and

$$\sup_{0 \leq t \leq \tau} \left|\hat{F}_n(t) - F(t)\right| = O(a_n) \quad a.s., \tag{1.6}$$

here and in the sequel, $a_n = n^{-1/2}(\ln n)^{1/2}$.

For positive reals z and t, and δ taking value 0 or 1, let

$$\xi(z, \delta, t) = g(z \wedge t) - I(z \leq t, \delta = 1)/\bar{L}(z), \tag{1.7}$$

where $g(x) = \int_0^x \frac{\mathrm{d}F_*(s)}{\bar{L}^2(s)}$.

Theorem 1.4 *Assume that the conditions of Theorem* 1.3 *hold. Then*

$$\hat{H}_n(t) - H(t) = -\frac{1}{n}\sum_{i=1}^{n}\xi(Z_i, \delta_i, t) + r_{1n}(t) \tag{1.8}$$

and

$$\hat{F}_n(t) - F(t) = -\frac{\bar{F}(t)}{n}\sum_{i=1}^{n}\xi(Z_i, \delta_i, t) + r_{2n}(t), \tag{1.9}$$

where $\sup_{0\le t\le\tau} |r_{in}(t)| = O(a_n)$ *a.s.* $i = 1, 2$, $0 < \tau < \tau_L$.

2 Proofs

Proof of Theorem 1.3 It is easy to see from Property P_7 of Joag-Dev and Proschan [13] that $\{Z_n; n \ge 1\}$ and $\{(Z_n, \delta_n); n \ge 1\}$ are also two sequences of NA r.v.s. Therefore

$$\sup_{t\ge 0}\bigl|L_n(t) - L(t)\bigr| = O(a_n) \quad \text{a.s.} \tag{2.1}$$

and

$$\sup_{t\ge 0}\bigl|F_{*n}(t) - F_*(t)\bigr| = O(a_n) \quad \text{a.s.} \tag{2.2}$$

follow from Lemma 1.2 and the fact that both L_n and F_{*n} are empirical distribution functions of L and F_*.

Now, by (1.1) and (1.2), let us write

$$\begin{aligned}
\hat{H}_n(t) - H(t) &= \int_0^t \frac{\mathrm{d}F_{*n}(s)}{\bar{L}_n(s)} - \int_0^t \frac{\mathrm{d}F_*(s)}{\bar{L}(s)} \\
&= \int_0^t \left(\frac{1}{\bar{L}_n(s)} - \frac{1}{\bar{L}(s)}\right)\mathrm{d}F_*(s) + \int_0^t \frac{\mathrm{d}(F_{*n}(s) - F_*(s))}{\bar{L}_n(s)} \\
&= \int_0^t \frac{\bar{L}(s) - \bar{L}_n(s)}{\bar{L}_n(s)\bar{L}(s)}\,\mathrm{d}F_*(s) + \frac{F_{*n}(t) - F_*(t)}{\bar{L}_n(t)} \\
&\quad - \int_0^t \bigl(F_{*n}(s) - F_*(s)\bigr)\,\mathrm{d}\bigl(\bar{L}_n(s)\bigr)^{-1}.
\end{aligned} \tag{2.3}$$

Therefore, by the combination of equations (2.1) and (2.2), and $\bar{L}_n(\tau) \to \bar{L}(\tau) > 0$, for $0 < \tau < \tau_L$, we obtain

$$\begin{aligned}
\sup_{0\le t\le\tau}\bigl|\hat{H}_n(t) - H(t)\bigr| \le{}& \frac{\sup_{t\ge 0}|\bar{L}_n(t) - \bar{L}(t)|}{\bar{L}_n(\tau)\bar{L}(\tau)}\bigl(F_*(\tau) - F_*(0)\bigr) \\
&+ \frac{\sup_{t\ge 0}|F_{*n}(t) - F_*(t)|}{\bar{L}_n(\tau)} \\
&+ \sup_{t\ge 0}\bigl|F_{*n}(t) - F_*(t)\bigr|\left(\frac{1}{\bar{L}_n(\tau)} - \frac{1}{\bar{L}_n(0)}\right)
\end{aligned}$$

$$\begin{aligned}&\le \frac{\sup_{t\ge 0}|\bar{L}_n(t)-\bar{L}(t)|}{\bar{L}_n(\tau)\bar{L}(\tau)}+\frac{2\sup_{t\ge 0}|F_{*n}(t)-F_*(t)|}{\bar{L}_n(\tau)}\\ &= O(a_n).\end{aligned}$$

Thus, (1.5) holds.

Now we prove (1.6). By (1.3) and (1.4),

$$\begin{aligned}-\hat{H}_n(t)-\ln\bigl(1-\hat{F}_n(t)\bigr)&=-\sum_{i=1}^n\frac{I(\delta_{(i)}=1,Z_{(i)}\le t)}{n-i+1}-\sum_{i=1}^n I(\delta_{(i)}=1,Z_{(i)}\le t)\ln\frac{n-i}{n-i+1}\\ &=\sum_{i=1}^n I(\delta_{(i)}=1,Z_{(i)}\le t)\biggl(\ln\frac{n-i+1}{n-i}-\frac{1}{n-i+1}\biggr).\end{aligned}$$

Therefore, by combining the inequality $0<\ln\frac{x+1}{x}-\frac{1}{x+1}<\frac{1}{x(x+1)}$, $x>0$, and (2.1), for $0<\tau<\tau_L$, $0\le t\le\tau$, we get that

$$\begin{aligned}0 &< -\hat{H}_n(t)-\ln\bigl(1-\hat{F}_n(t)\bigr)\le\sum_{i=1}^n I(\delta_{(i)}=1,Z_{(i)}\le t)\frac{1}{(n-i)(n-i+1)}\\ &\le\sum_{i;Z_{(i)}\le t}\frac{1}{(n-i)(n-i+1)}=\sum_{i=1}^{n-Y_n(t)}\biggl(\frac{1}{n-i}-\frac{1}{n-i+1}\biggr)\\ &=\frac{1}{Y_n(t)}-\frac{1}{n}\le\frac{1}{n}\frac{1}{\frac{Y_n(t)}{n}}=\frac{1}{n}\frac{1}{\bar{L}_n(t)}\\ &=O\biggl(\frac{1}{n}\biggr).\end{aligned}\tag{2.4}$$

By (1.1),(1.6) and (2.4), using the Taylor expansion, $e^x=1+x+o(x)$, we obtain

$$\begin{aligned}\hat{F}_n(t)-F(t)&=1-F(t)-\bigl(1-\hat{F}_n(t)\bigr)=\bigl(e^{-H(t)}-e^{-\hat{H}_n(t)}\bigr)+\bigl(e^{-\hat{H}_n(t)}-e^{\ln(1-\hat{F}_n(t))}\bigr)\\ &=e^{-H(t)}\bigl(1-e^{-\hat{H}_n(t)+H(t)}\bigr)+e^{\ln(1-\hat{F}_n(t))}\bigl(e^{-\hat{H}_n(t)-\ln(1-\hat{F}_n(t))}-1\bigr)\\ &=e^{-H(t)}\bigl(\hat{H}_n(t)-H(t)+o\bigl(\hat{H}_n(t)-H(t)\bigr)\bigr)\\ &\quad+\bigl(1-\hat{F}_n(t)\bigr)\bigl(-\hat{H}_n(t)-\ln\bigl(1-\hat{F}_n(t)\bigr)+o\bigl(-\hat{H}_n(t)-\ln\bigl(1-\hat{F}_n(t)\bigr)\bigr)\bigr)\\ &=e^{-H(t)}\bigl(\hat{H}_n(t)-H(t)\bigr)+o(a_n)+O\biggl(\frac{1}{n}\biggr)\\ &=\bar{F}(t)\bigl(\hat{H}_n(t)-H(t)\bigr)+o(a_n).\end{aligned}\tag{2.5}$$

Thence, the combination (1.5), (1.6) holds. This completes the proof of Theorem 1.3. □

Proof of Theorem 1.4 By (2.1),

$$\begin{aligned}\frac{1}{\bar{L}_n(s)}-\frac{1}{\bar{L}(s)}&=\frac{\bar{L}(s)-\bar{L}_n(s)}{\bar{L}^2(s)}-\frac{2}{\bar{L}(s)}+\frac{\bar{L}_n(s)}{\bar{L}^2(s)}+\frac{1}{\bar{L}_n(s)}\\ &=\frac{\bar{L}(s)-\bar{L}_n(s)}{\bar{L}^2(s)}+\frac{(\bar{L}_n(s)-\bar{L}(s))^2}{\bar{L}^2(s)\bar{L}_n(s)}=\frac{1}{\bar{L}(s)}-\frac{\bar{L}_n(s)}{\bar{L}^2(s)}+O\bigl(a_n^2\bigr).\end{aligned}$$

Thus, by the combination of (2.3),

$$\begin{aligned}
\hat{H}_n(t) - H(t) &= \int_0^t \left(\frac{1}{\bar{L}_n(s)} - \frac{1}{\bar{L}(s)}\right) \mathrm{d}F_*(s) + \int_0^t \frac{\mathrm{d}(F_{*n}(s) - F_*(s))}{\bar{L}(s)} \\
&\quad + \int_0^t \left(\frac{1}{\bar{L}_n(s)} - \frac{1}{\bar{L}(s)}\right) \mathrm{d}\big(F_{*n}(s) - F_*(s)\big) \\
&= \left(\int_0^t \frac{\mathrm{d}F_{*n}(s)}{\bar{L}(s)} - \int_0^t \frac{\bar{L}_n(s)}{\bar{L}^2(s)} \mathrm{d}F_*(s)\right) \\
&\quad + \int_0^t \left(\frac{1}{\bar{L}_n(s)} - \frac{1}{\bar{L}(s)}\right) \mathrm{d}\big(F_{*n}(s) - F_*(s)\big) + O\big(a_n^2\big) \\
&:= I_1(t) + I_2(t) + O\big(a_n^2\big).
\end{aligned} \tag{2.6}$$

Noting that $F_{*n}(s) = \frac{N_n(s)}{n}$ and $N_n(s)$ is a step function, we get

$$\begin{aligned}
I_1(t) &= \frac{1}{n} \sum_{i;Z_{(i)} \le t} \frac{N_n(Z_i) - N_n(Z_i^-)}{\bar{L}(Z_{(i)})} - \frac{1}{n} \int_0^t \frac{\sum_{i=1}^n I(Z_i \ge s)}{\bar{L}^2(s)} \mathrm{d}F_*(s) \\
&= \frac{1}{n} \sum_{i;Z_{(i)} \le t} \frac{\delta_{(i)}}{\bar{L}(Z_{(i)})} - \frac{1}{n} \sum_{i=1}^n \int_0^{t \wedge Z_i} \frac{\mathrm{d}F_*(s)}{\bar{L}^2(s)} \\
&= \frac{1}{n} \sum_{i=1}^n \frac{I(Z_{(i)} \le t, \delta_{(i)} = 1)}{\bar{L}(Z_{(i)})} - \frac{1}{n} \sum_{i=1}^n g(t \wedge Z_i) \\
&= \frac{1}{n} \sum_{i=1}^n \frac{I(Z_i \le t, \delta_i = 1)}{\bar{L}(Z_i)} - \frac{1}{n} \sum_{i=1}^n g(t \wedge Z_i) \\
&= -\frac{1}{n} \sum_{i=1}^n \xi(Z_i, \delta_i, t).
\end{aligned} \tag{2.7}$$

Therefore, to prove (1.8), it suffices to prove that $\sup_{0 \le t \le \tau} |I_2(t)| = O(a_n)$ for $\tau < \tau_H$. Let us divide the interval $[0, \tau]$ into subintervals $[x_i, x_{i+1}]$, $i = 1, \ldots, k_n$, where $k_n = O(a_n^{-1})$, and $0 = x_1 < x_2 < \cdots < x_{k_n+1} = \tau$ are such that $H(x_{i+1}) - H(x_i) = O(a_n)$. For $0 \le t \le \tau$, it is easy to check that

$$\begin{aligned}
I_2(t) &= \int_0^t \left(\frac{1}{\bar{L}_n(s)} - \frac{1}{\bar{L}(s)}\right) \mathrm{d}\big(F_{*n}(s) - F_*(s)\big) \\
&\le 2 \max_{1 \le i \le k_n} \sup_{y \in [x_i, x_{i+1}]} \left|\bar{L}_n^{-1}(y) - \bar{L}_n^{-1}(x_i) - \bar{L}^{-1}(y) + \bar{L}^{-1}(x_i)\right| \\
&\quad + \sup_{0 \le x \le \tau} \left|\bar{L}_n^{-1}(x) - \bar{L}^{-1}(x)\right| \max_{1 \le i \le k_n} \left|F_{*n}^{-1}(x_{i+1}) - F_{*n}^{-1}(x_i) - F_*(x_{i+1}) + F_*(x_i)\right| \\
&\le c \max_{1 \le i \le k_n} \sup_{y \in [x_i, x_{i+1}]} \left|\bar{L}_n(y) - \bar{L}_n(x_i) - \bar{L}(y) + \bar{L}(x_i)\right| \\
&\quad + c \max_{1 \le i \le k_n} \left|F_{*n}(x_{i+1}) - F_{*n}(x_i) - F_*(x_{i+1}) + F_*(x_i)\right| + O\big(a_n^2\big) \\
&:= I_{21} + I_{22} + O\big(a_n^2\big).
\end{aligned} \tag{2.8}$$

To estimate I_{21}, we further subdivide each $[x_i, x_{i+1}]$ into subintervals $[x_{ij}, x_{i(j+1)}]$, $j = 1, \ldots, b_n$, where $b_n = O(k_n^{1/2}) = O(a_n^{-1/2})$ such that $|\bar{L}(x_{i(j+1)}) - \bar{L}(x_{ij})| = O(a_n^{3/2})$ uniformly in i, j. Now, by (2.1) and $|\bar{L}_n(y) - \bar{L}_n(x_{ij})| \le 1/n \le O(a_n^{3/2})$, for $y \in [x_{ij}, x_{i(j+1)}]$, it follows that

$$\begin{aligned}
I_{21} &= \max_{1\le i\le k_n} \sup_{y\in[x_i,x_{i+1}]} \left|\bar{L}_n(y) - \bar{L}_n(x_i) - \bar{L}(y) + \bar{L}(x_i)\right| \\
&\le \max_{1\le i\le k_n} \max_{1\le j\le b_n} \sup_{y\in[x_{ij},x_{i(j+1)}]} \left|\bar{L}_n(x_{ij}) - \bar{L}_n(x_i) - \bar{L}(x_{ij}) + \bar{L}(x_i)\right| \\
&\quad + \max_{1\le i\le k_n} \max_{1\le j\le b_n} \sup_{y\in[x_{ij},x_{i(j+1)}]} \left(\left|\bar{L}_n(y) - \bar{L}_n(x_{ij})\right| + \left|-\bar{L}(y) + \bar{L}(x_{ij})\right|\right) \\
&\le \max_{1\le i\le k_n} \max_{1\le j\le b_n} \left|\bar{L}_n(x_{ij}) - \bar{L}_n(x_i) - \bar{L}(x_{ij}) + \bar{L}(x_i)\right| + O\left(a_n^{3/2}\right). \qquad (2.9)
\end{aligned}$$

For $1 \le i \le k_n$, $1 \le j \le b_n$, $1 \le k \le n$, let $\eta_k = I(Z_k \ge x_i) - EI(Z_k \ge x_i)$, $\zeta_k = I(Z_k \ge x_{ij}) - EI(Z_k \ge x_{ij})$. Then $\bar{L}_n(x_{ij}) - \bar{L}_n(x_i) - \bar{L}(x_{ij}) + \bar{L}(x_i) = \frac{1}{n}\sum_{k=1}^n(\eta_k + \zeta_k)$, $\{\eta_k\}$ and $\{\zeta_k\}$ are NA sequences with $|\eta_k| \le 1$, $|\zeta_k| \le 1$, $E\eta_k = E\zeta_k = 0$, $E\eta_k^2 \le 1$, $E\zeta_k^2 \le 1$.

Taking $t = a_n$ in Lemma 1.1, yields the following probability bound:

$$\begin{aligned}
&P\left(\max_{1\le i\le k_n} \max_{1\le j\le b_n} \left|\bar{L}_n(x_{ij}) - \bar{L}_n(x_i) - \bar{L}(x_{ij}) + \bar{L}(x_i)\right| \ge 8a_n\right) \\
&\quad\le \sum_{i=1}^{k_n}\sum_{j=1}^{b_n} P\left(\left|\sum_{k=1}^n(\eta_k + \zeta_k)\right| \ge 8na_n\right) \\
&\quad\le \sum_{i=1}^{k_n}\sum_{j=1}^{b_n} P\left(\left|\sum_{k=1}^n\eta_k\right| \ge 4na_n\right) + \sum_{i=1}^{k_n}\sum_{j=1}^{b_n} P\left(\left|\sum_{k=1}^n\zeta_k\right| \ge 4na_n\right) \\
&\quad\le \sum_{i=1}^{k_n}\sum_{j=1}^{b_n} 4\exp\left(-4na_n^2 + na_n^2\right) \\
&\quad= 4k_nb_n\mathrm{e}^{-3\ln n} \le \frac{1}{n^2}.
\end{aligned}$$

Using the bound and the Borel-Cantelli lemma, we deduce that $I_{21} = O(a_n)$ a.s. The estimation of $I_{22} = O(a_n)$ is similar noting that $|F_*(x) - F_*(y)| \le |\bar{L}(x) - \bar{L}(y)|$ for all x and y. Therefore, by (2.6)-(2.9), (1.8) holds. (1.9) follows from (2.5) and (1.8). □

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
QW conceived of the study and drafted, complete the manuscript. PC participated in the discussion of the manuscript. QW and PC read and approved the final manuscript.

Author details
[1]College of Science, Guilin University of Technology, Guilin, 541004, P.R. China. [2]Department of Mathematics, Ji'nan University, Guangzhou, 510630, P.R. China.

Authors' information
Qunying Wu, Professor, Doctor, working in the field of probability and statistics.

Acknowledgements
Supported by the National Natural Science Foundation of China (11061012), project supported by Program to Sponsor Teams for Innovation in the Construction of Talent Highlands in Guangxi Institutions of Higher Learning ([2011] 47), and the Support Program of the Guangxi China Science Foundation (2012GXNSFAA053010, 2013GXNSFDA019001).

Received: 4 February 2013 Accepted: 10 July 2013 Published: 25 July 2013

References

1. Kaplan, EM, Meier, P: Nonparametric estimation from incomplete observations. J. Am. Stat. Assoc. **53**, 457-481 (1958)
2. Breslow, N, Crowley, J: A large sample study of the life table and product limit estimates under random censorship. Ann. Stat. **2**, 437-453 (1974)
3. Földes, A, Rejtö, L: A LIL type result for the product limit estimator. Z. Wahrscheinlichkeitstheor. Verw. Geb. **56**, 75-84 (1981)
4. Gu, MG, Lai, TL: Functional laws of the iterated logarithm for the product-limit estimator of a distribution function under random censorship or truncation. Ann. Probab. **18**, 160-189 (1990)
5. Gill, R: Censoring and Stochastic Integrals. Mathematical Centre Tracts, vol. 124. Math. Centrum, Amsterdam (1980)
6. Kang, SS, Koehler, KJ: Modification of the Greenwood formula for correlated failure times. Biometrics **53**, 885-899 (1997)
7. Wei, LJ, Lin, DY, Weissfeld, L: Regression analysis of multivariate incomplete failure times data by modelling marginal distributions. J. Am. Stat. Assoc. **84**, 1064-1073 (1989)
8. Shumway, RH, Azari, AS, Johnson, P: Estimating mean concentrations under transformation for environmental data with detection limits. Technometrics **31**, 347-356 (1988)
9. Ying, Z, Wei, LJ: The Kaplan-Meier estimate for dependent failure time observations. J. Multivar. Anal. **50**, 17-29 (1994)
10. Lecoutre, JP, Ould-Sad, E: Convergence of the conditional Kaplan-Meier estimate under strong mixing. J. Stat. Plan. Inference **44**, 359-369 (1995)
11. Cai, ZW: Estimating a distribution function for censored time series data. J. Multivar. Anal. **78**, 299-318 (2001)
12. Liang, HY, Uña-Álvarez, J: A Berry-Esseen type bound in kernel density estimation for strong mixing censored samples. J. Multivar. Anal. **100**, 1219-1231 (2009)
13. Joag-Dev, K, Proschan, F: Negative association of random variables with applications. Ann. Stat. **11**(1), 286-295 (1983)
14. Matula, PA: A note on the almost sure convergence of sums of negatively dependent random variables. Stat. Probab. Lett. **15**, 209-213 (1992)
15. Wu, QY, Jiang, YY: A law of the iterated logarithm of partial sums for NA random variables. J. Korean Stat. Soc. **39**(2), 199-206 (2010)
16. Wu, QY, Jiang, YY: Chover's law of the iterated logarithm for NA sequences. J. Syst. Sci. Complex. **23**(2), 293-302 (2010)
17. Yang, SC: Consistency of nearest neighbor estimator of density function for negative associated samples. Acta Math. Appl. Sin. **26**(3), 385-394 (2003)

doi:10.1186/1029-242X-2013-340
Cite this article as: Wu and Chen: **Strong representation results of the Kaplan-Meier estimator for censored negatively associated data.** *Journal of Inequalities and Applications* 2013 **2013**:340.