Berry-Esséen Inequality for Kaplan-Meier L-Estimator

Qi-Hua Wang[1] Li-Xing Zhu[2]

[1]Chinese Academy of Sciences

[2]Chinese Academy of Sciences and Catholic University of Louvain

Abstract

Let $\widehat{F}_n$ be the Kaplan-Meier estimator of distribution function F. Let $J(\cdot)$ be measurable real-valued function. In this paper, U-statistic representation for Kaplan-Meier L-estimator, $T(\widehat{F}_n) = \int xJ(\widehat{F}_n(x))\,d\widehat{F}_n(x)$, is derived. Furthermore the representation is also used to establish a Berry-Esséen inequality for $T(\widehat{F}_n)$.

Short title: Kaplan-Meier L-Estimator

AMS 1991 Subject Classifications. Primary 62G05, Secondary 62E20

Key Words and Phrases. Berry-Esséen Inequality, Kaplan-Meier L-estimator, U-statistic representation.

1 Introduction

Let $X_1, X_2, \cdots, X_n$ be nonnegative independent random variables representing survival times with common distribution function F. In survival analysis or medical follow-up studies, one does not observe the true random survival times $X_1, X_2, \cdots, X_n$ of the n patients who entered the study, but rather observes only the right censored version of these variables due to dropping out of the patients from the study or living with drawl of patients from the study. That is, associated with X_i there are an independent nonnegative censoring variable Y_i which are iid r.v. with d.f. G, one observes only $(Z_1, \delta_1), \cdots, (Z_n, \delta_n)$, where

$$Z_i = \min(X_i, Y_i), \quad \delta_i = I[X_i \le Y_i]$$

where $i = 1, 2, \cdots, n$, $I[\cdot]$ is the indicator function for some event.

It is noted that some important statistic characters can be expressed as $T(F) = \int xJ(F(x))\,dF(x)$. This also happen in survival analysis identically (see, for example, Reid(1981)). In the uncensored case, a L-estimator can be defined replacing F in $T(F)$ by its empirical estimator F_n. But, in the case of censored data, it fails to define the empirical estimator of F. Reid (1981) defined a Kaplan-Meier L-estimator $T(\widehat{F}_n)$ replacing F in $T(F)$ by the Kaplan-Meier estimator $\widehat{F}_n$ of F, where $\widehat{F}_n$ is defined as

[1]Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing, China. The research was supported by the Postdoctoral Programme Science Foundation.

[2]Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing, China and Institute of Statistics, Catholic University of Louvain, Louvain-la-Neuve, Belgium . The research was supported by the National Natural Science Foundation of China and the contract " Projet d'Actions de Recherche Concertées " (PARC No. 93/98-164) of the Belgian Government.

follows

$$1-\widehat{F}_n(x)=\begin{cases}\prod_{j=1}^{n}\left(\frac{N^+(Z_j)}{1+N^+(Z_j)}\right)^{I[Z_j\le t,\delta_j=1]}, & t\le Z_{(n)}\\ 0, & \text{otherwise.}\end{cases} \tag{1.1}$$
$$N^+(Z_j)=\sum_{j=1}^{n} I[Z_i>Z_j]$$

As a special case in her paper, Reid(1981) calculated the influence curve of the Kaplan-Meier L-estimator and gave some examples.

In this paper, we represent $T(\widehat{F}_n)$ as U-statistic plus a remainder, and employ the representation to establish a Berry-Esséen inequality for it. The main results will be formulated in Section 2 and proved in Section 3.

For the sake of simplicity, denote with c a constant and c_0 a positive absolute constant which may take different values for each appearance.

2 The Main Results

Due to the complicated structure of representation of $T(\widehat{F}_n)$ described below, we first introduce some notations which will be used frequently in the proof of the theorems for conveniently reading.

2.1 Some Notations

Let $H(x)=P(Z_1\le x), H_1=P(Z_1\le x,\delta_1=1), \widetilde{H}_1(x)=P(Z_1>x,\delta_1=1), \bar{F}(x)=1-F(x)$, $\bar{G}(x)=1-G(x), \bar{H}(x)=1-H(x)=\bar{F}\bar{G}$ and

$$\begin{aligned}
&\widetilde{g}(Z_i,\delta_i;x)=-\bar{H}^{-1}(Z_i)I[0\le Z_i\le x]-\int_0^{x\wedge Z_i}\bar{H}^{-2}\,d\bar{H}_1,\\
&\xi(Z_i,\delta_i;x)=\bar{F}(x)\widetilde{g}(Z_i,\delta_i;x),\\
&\widetilde{\psi}(Z_i,\delta_i,Z_j,\delta_j;x)=\bar{H}^{-2}(Z_i)I[0\le Z_i\le x,Z_i<Z_j,\delta_i=1]+\int_0^{x\wedge Z_i\wedge Z_j}\bar{H}^{-3}\,d\bar{H}_1\\
&\qquad -E\Big[\bar{H}^{-2}(Z_i)I[0\le Z_i\le x,Z_i<Z_j,\delta_i=1]+\int_0^{x\wedge Z_i\wedge Z_j}\bar{H}^{-3}\,d\bar{H}_1\Big|Z_i,\delta_i\Big],\\
&\widetilde{h}(Z_i,\delta_i,Z_j,\delta_j;x)=\widetilde{g}(Z_i,\delta_i;x)+\widetilde{g}(Z_j,\delta_j;x)+\widetilde{\psi}(Z_i,\delta_i,Z_j,\delta_j;x)+\widetilde{\psi}(Z_j,\delta_j,Z_i,\delta_i;x),\\
&h(Z_i,\delta_i;Z_j,\delta_j)=\int_0^{\infty}J(F(x))\bar{F}(x)[\widetilde{h}(Z_i,\delta_i;Z_j,\delta_j;x)+\widetilde{g}(Z_i,\delta_i;x)\widetilde{g}(Z_j,\delta_j;x)]\,dx\\
&\qquad +\int_0^{\infty}J'(F(x))\xi(Z_i,\delta_i;x)\xi(Z_j,\delta_j;x)dx
\end{aligned}$$

and a U-statistic

$$U_n=\frac{1}{n^2}\sum_{i<j}h(Z_i,\delta_i;Z_j,\delta_j).$$

Furthermore let $\tau_F=\inf\{t:F(t)=1\}$. τ_G and τ_H are denoted similarly. In this paper, we assume $\tau_F\le\tau_G$ throughout.

2.2 Theorems and Remarks

Theorem 1. Assume that F and G are continuous and that $J(\cdot)$ satisfies

(i) $J(\cdot)$ is real-value measurable, and vanishes in $[1-\beta, 1]$ for some $0<\beta<1$.

(ii) $J'(\cdot)$ exists in $(0, 1-\beta)$, and is a bounded function satisfying Lipschitz condition, i.e., there exists c_0 such that

$$|J'(x_1) - J'(x_2)| \le c_0 |x_1 - x_2|, \forall x_1, x_2 \in (0, 1-\beta),$$

then

$$T(\widehat{F}_n) - T(F) = U_n + \triangle_n + R_n \tag{2.1}$$

where $\triangle_n = -n^{-1}\int_0^\infty \left[J(F(x)\bar{F}(x)\int_0^x \bar{H}^{-2}\, d\bar{H}_1\right] dx$ and R_n satisfy

$$P(\sqrt{n}|R_n| > cn^{-\frac{1}{2}}) \le cn^{-1}. \tag{2.2}$$

Remark 1. Note that uniform convergence of $\widehat{F}_n$ to F is only valid in $[0, T]$ for any $T < \tau_H$ (see, e.g., (2.28) and (2.39) below), hence we restrict $J(\cdot)$ to vanish in $[1-\beta, 1]$ for some $0<\beta<1$(the same condition is also employed in Reid(1981)). Indeed, this is also why Sander(1975) investigated Kaplan-Meier mean value on $[0, T]$, $\int_0^T x\, d\widehat{F}_n(x)$, instead of $\int_0^\infty x\, d\widehat{F}_n(x)$, and Gijbels and Veraverbeke (1991) considered a class of functionals of Kaplan-Meier estimator, $\int_0^T \cdots \int_0^T h(x_1, \cdot, x_m)\prod_{i-1}^m d\widehat{F}_n(x_i)$, instead of $\int_0^\infty \cdots \int_0^\infty h(x_1, \cdot, x_m)\prod_{i-1}^m d\widehat{F}_n(x_i)$.

Remark 2. Some examples of $J(\cdot)$ satisfying the conditions of Theorem 1 are given in Serfling(1980) and Reid (1981).

Denote $\tau_0 = \inf\{x : F(x) = 1-\beta\}$. We now established a Berry-Esséen inequality for $T(\widehat{F}_n)$.

Theorem 2. Under the conditions of Theorem 1, we have

$$\sup_x |P(\sqrt{n}\sigma^{-1}(T(\widehat{F}_n) - T(F) \le x) - \Phi(x)| \le c_0 B n^{-\frac{1}{2}},$$

where

$$\sigma^2 = \int_0^{\tau_0}\int_0^{\tau_0} J(F(x)J(F(y))\bar{F}(x)\bar{F}(y)\Big[\int_0^{x\wedge y} \bar{H}^{-2}(s)\, dH_1(s)\Big]\, dx\, dy, \tag{2.3}$$

$$\begin{aligned} B = \quad & \left[\frac{\int_0^{\tau_0} J(F(x))\bar{F}(x)\bar{H}^{-2}(x)\, dx}{\sigma}\right]^3 \\ & + \frac{\left(\int_0^{\tau_0} J(F(x))\bar{F}(x)\bar{H}^{-2}(x)\, dx\right)^2 + \left(\int_0^{\tau_0} J'(F(x))\bar{F}^2(x)\bar{H}^{-2}(x)\, dx\right)^2}{\sigma^2} \\ & + \frac{\int_0^{\tau_0} J(F(x))\bar{F}(x)\bar{H}^{-2}(x)\, dx}{\sigma}. \end{aligned} \tag{2.4}$$

We leave all proofs of the theorems in the next section.

3 The proof of Theorems

Let $\hat{\bar{F}}_n(x) = 1 - \hat{F}_n(x)$, $\tilde{r}_n(x) = \log \hat{\bar{F}}_n(x) - \log \bar{F}(x) - U_n - \tilde{\triangle}_n$ where $\tilde{\triangle}_n = n^{-1}\int_0^x \bar{H}^{-2}\,d\bar{H}_1$ and U_n is as defined in subsection 2.1.

Lemma 1. Assume that F and G are continuous, we then have

$$E\tilde{r}_n^2(x) \le c_0 \bar{H}^{-8}(x) n^{-3} \tag{3.1}$$

for any $0 \le x < \tau_H$.

Proof. Let

$$\begin{gathered}\bar{H}_{1n}(x) = \tfrac{1}{n}\textstyle\sum_{i=1}^n I[Z_i > x, \delta_i = 1], \quad \bar{H}_n(x) = \tfrac{1}{n}\textstyle\sum_{i=1}^n I[Z_i > x],\\ \theta_n(x) = \log \hat{\bar{F}}_n(x) - \textstyle\int_0^x \frac{d\bar{H}_{1n}(s)}{\bar{H}_n(s)}.\end{gathered} \tag{3.2}$$

By Chang and Rao(1989), $\tilde{r}_n(x)$ admits the following decomposition:

$$\tilde{r}_n(x) = h_1^{(n)}(x) + h_2^{(n)}(x) + h_3^{(n)}(x) + \theta_n(x), \tag{3.3}$$

with

$$E(h_1^{(n)}(x))^2 \le c_0 \bar{H}^{-3}(x) n^{-3}, \tag{3.4}$$

$$E(h_3^{(n)}(x))^2 \le c_0 \bar{H}^{-6}(x) n^{-3}, \tag{3.5}$$

$$P(|h_2^{(n)}(x)| > s) \le c_0\Big[\exp\Big\{-\tfrac{n^{\frac{2}{3}}\bar{H}^2(u)s^{\frac{1}{3}}}{2}\Big\} + \exp\Big\{-\tfrac{n\bar{H}^2(x)}{2}\Big\}\Big]. \tag{3.6}$$

where $h_1^{(n)}$ $h_2^{(n)}$ $h_3^{(n)}$ are as defined in Chang and Rao(1989). We omit their forms here since it is not necessary to specify them at all for our aims. By Lemma 1 in Breslow and Crowley(1974), it follows that

$$P(|\theta_n(x)| > s) \le c_0 \exp\big\{-\frac{n}{2}\bar{H}^2(x)\big\}, \qquad s \ge \frac{4}{n^2\bar{H}^2(x)}. \tag{3.7}$$

Let $R_n^* = \hat{\bar{F}}_n(x) - \bar{F}(x) - \bar{F}(x)\tilde{U}_n - \bar{F}(x)\tilde{\triangle}_n$, where $\tilde{U}_n = \frac{1}{n^2}\sum_{i<j}\tilde{h}(Z_i,\delta_i;Z_j,\delta_j)$. By Taylor's expansion for $exp(\log \hat{\bar{F}}_n(x) - \log \bar{F}(x))$, we have

$$\hat{\bar{F}}_n(x) - \bar{F}(x) = \bar{F}(x)\tilde{U}_n + \bar{F}(x)\tilde{\triangle}_n + \bar{F}(x)\tilde{r}_n(x) + \frac{1}{2}\hat{\bar{F}}_n^{\theta_1}(x)\bar{F}^{1-\theta_1}(x)(\log \hat{\bar{F}}_n(x) - \log \bar{F}(x)), \tag{3.8}$$

where $0 < \theta_1 < 1$. By (3.8) R_n^* then becomes that

$$R_n^* = \bar{F}(x)\tilde{r}_n(x) + \frac{1}{2}\hat{\bar{F}}_n^{\theta_1}(x)\bar{F}^{1-\theta_1}(x)(\log \hat{\bar{F}}_n(x) - \log \bar{F}(x))^2 \tag{3.9}$$

which implies that

$$\{|\bar{F}(x)\tilde{r}_n(x)| > s\} \subseteq \{|R_n^*(x)| > s\}. \tag{3.10}$$

Note that $|\hat{\bar{F}}_n(x) - \bar{F}(x)| < 1$, hence it is easily seen by (3.8) and (3.9) that for $s > 2$

$$\{|R_n^*(x)| > s\} \subseteq \{|\bar{F}(x)\tilde{U}_n + \bar{F}(x)\tilde{\triangle}_n| > \frac{s}{2}\} \subseteq \{|\tilde{U}_n| > \frac{s}{4}\}. \tag{3.11}$$

Recall that $n^2\widetilde{U}_n/\begin{pmatrix} n \\ 2 \end{pmatrix}$ is an U-statistic and note that $|\widetilde{h}(Z_1,\delta_1;Z_2,\delta_2;x)| \le 6\bar{H}^{-2}(x)$ and $E\widetilde{h}(Z_1,\delta_1;Z_2,\delta_2) = 0$, thus by Berry-Esséen inequality for U-statistic (see,e.g., Serfling 1980, p.201) and (3.10) and (3.11) it follows that

$$P(|\bar{F}(x)\widetilde{r}_n(x) > s) \le P(|\widetilde{U}_n| \ge \frac{s}{4}) \le 2\exp\big\{-\frac{n\bar{H}^4(x)}{576}s^2\big\}, \quad s > 2, \tag{3.12}$$

Note that

$$\begin{aligned} \bar{F}^2(x)E|\widetilde{r}_n(x)|^2 \quad &= 2\Big[\int_0^2 sP(\bar{F}(x)|\widetilde{r}_n(x)| > s)\,ds + \int_2^\infty sP(\bar{F}(x)|\widetilde{r}_n(x)| > s)\,ds\Big] \\ &\triangleq 2[E_{n1} + E_{n2}] \end{aligned} \tag{3.13}$$

It can be derived that

$$\begin{aligned} E_{n1} \quad &\le 16\int_0^\infty \frac{s}{4}P(|h_1^{(n)}(x)| > \frac{s}{4})\,d\frac{s}{4} + 8\int_0^2 P(|h_2^{(n)}(x)| > \frac{s}{4})\,d\frac{s}{4} \\ &+ 16\int_0^\infty \frac{s}{4}P(|h_3^{(n)}(x)| > \frac{s}{4})\,d\frac{s}{4} + 16\int_{\frac{4}{n^2\bar{H}^2(x)}}^2 \frac{s}{4}P(|\theta_n(x)| > \frac{s}{4})\,d\frac{s}{4} \\ &+ \int_0^{\frac{4}{n^2\bar{H}^2(x)}} sP(|\theta_n(x)| > \frac{s}{4})\,ds. \end{aligned} \tag{3.14}$$

Hence, (3.3)-(3.7) and (3.14) together yield the following inequality

$$E_{n1} \le c_0\bar{H}^{-6}(x)n^{-3}. \tag{3.15}$$

By (3.12), it is easy to obtain that

$$E_{n2} \le 288\bar{H}^{-4}(x)n^{-1}\exp\big\{-\frac{n\bar{H}^4(x)}{144}\big\}. \tag{3.16}$$

Combining (3.13)-(3.16), Lemma 1 is proved.

The proof of Theorem 1. Let $K(u) = \int_0^u J(t)\,dt$. Then, we have

$$\begin{aligned} &T(\widehat{F}_n) - T(F) \\ &= \int_0^\infty x\,d(K(\widehat{F}_n) - K(F)) \\ &= \int_0^\infty x\,d[J(F(x))(\widehat{F}_n(x) - F(x)) + \frac{1}{2}J'(F(x))(\widehat{F}_n(x) - F(x))^2] \\ &+ \int_0^\infty x\,d(K(\widehat{F}_n(x)) - K(F(x))(\widehat{F}_n(x) - F(x)) - \frac{1}{2}J'(F(x))(\widehat{F}_n(x) - F(x))^2] \\ &\triangleq m_n + R_T^{(n)}. \end{aligned} \tag{3.17}$$

Integrating by part, it follows that

$$m_n = -\int_0^\infty J(F(x))(\widehat{F}_n(x) - F(x))\,dx - \int_0^\infty J'(F(x))(\widehat{F}_n(x) - F(x))^2\,dx \triangleq I + II. \tag{3.18}$$

By three term Taylor's expansion for $exp(\log \widehat{\bar{F}}_n(x) - \log \bar{F}(x))$, it follows that

$$\begin{aligned}
I &= \int_0^\infty J(F(x))\bar{F}(x)[\log \widehat{\bar{F}}_n(x) - \log \bar{F}(x) + \frac{1}{2}(\log \widehat{\bar{F}}_n(x) - \log \bar{F}(x))^2]\, dx \\
&+ \int_0^\infty J(F(x))\frac{\widehat{\bar{F}}_n^{\theta_2} \bar{F}^{1-\theta_2}(x)}{3!}(\log \widehat{\bar{F}}_n(x) - \log \bar{F}(x))^3\, dx \\
&\triangleq I_1 + R_I^{(n)}, \qquad (3.19)
\end{aligned}$$

where $0 < \theta_2 < 1$. Denote $r_n(x) = \log \widehat{\bar{F}}_n(x) - \log \bar{F}(x) - \frac{1}{n}\sum_{i=1}^n \widetilde{g}(Z_i, \delta_i; x)$, then

$$\begin{aligned}
I_1 &= n^{-2}\sum_{i<j}\int_0^\infty J(F(x))\bar{F}(x)[\widetilde{h}(Z_i, \delta_i; Z_j, \delta_j; x) + \widetilde{g}(Z_i, \delta_i; x)\widetilde{g}(Z_j, \delta_j; x)]\, dx \\
&+ \triangle_n + n^{-2}\sum_{i=1}^n \int_0^\infty J(F(x))\bar{F}(x)\widetilde{g}^2(Z_i, \delta_i; x)\, dx \\
&+ 2n^{-1}\sum_{i=1}^n \int_0^\infty J(F(x))\bar{F}(x) r_n(x)\widetilde{g}(Z_i, \delta_i; x)\, dx \\
&+ \int_0^\infty J(F(x))\bar{F}(x) r_n^2(x)\, dx + \int_0^\infty J(F(x))\bar{F}(x)\widetilde{r}_n(x)\, dx \\
&\triangleq I_{11} + \triangle_n + \sum_{i=1}^4 R_{iI_1}^{(n)}, \qquad (3.20)
\end{aligned}$$

where $\widetilde{r}_n$ is defined in the beginning of this section and the definitions of $\widetilde{h}$ and $\widetilde{g}$ are in the subsection 2.1. By two term Taylor's expansion for $exp(\log \widehat{\bar{F}}_n) - exp(\log \bar{F})$, we have

$$\begin{aligned}
II = &-\frac{1}{2}\int_0^\infty J'(F(x))\bar{F}^2(x)(\log \widehat{\bar{F}}_n(x) - \log \bar{F}(x))^2\, dx \\
&-\frac{1}{2}\int_0^\infty J'(F(x))\widehat{\bar{F}}_n^{\theta_3}(x)\bar{F}^{2-\theta_3}(x)(\log \widehat{\bar{F}}_n(x) - \log \bar{F}(x))^3\, dx \\
&-\frac{1}{8}\int_0^\infty J'(F(x))\widehat{\bar{F}}_n^{\theta_3}(x)\bar{F}^{2(1-\theta_3)}(x)(\log \widehat{\bar{F}}_n(x) - \log \bar{F}(x))^4\, dx \\
= &\; II_1 + R_{1II}^{(n)} + R_{2II}^{(n)}, \qquad (3.21)
\end{aligned}$$

where $0 < \theta_3 < 1$. Recalling the definition of $r_n(x)$, it follows that

$$\begin{aligned}
II_1 = &-n^2\sum_{i<j}\int_0^\infty J'(F(x))\bar{F}^2(x)\widetilde{g}(Z_i, \delta_i; x)\widetilde{g}(Z_j, \delta_j; x)\, dx \\
&- 2n^{-2}\sum_{i=1}^n \int_0^\infty J'(F(x))\bar{F}^2(x)\widetilde{g}^2(Z_i, \delta_i; x)\, dx \\
&- n^{-1}\sum_{i=1}^n \int_0^\infty J'(F(x))\bar{F}^2(x)\widetilde{g}(Z_i, \delta_i; x) r_n(x)\, dx \\
&- 2\int_0^\infty J'(F(x))\bar{F}^2(x) r_n^2(x)\, dx \\
\triangleq &\; II_{11} + \sum_{i=1}^3 R_{iII_1}^{(n)} \qquad (3.22)
\end{aligned}$$

Recalling the definition of U_n, we have

$$I_{11} + II_{11} = U_n. \qquad (3.23)$$

Hence, Combining (3.17) to (3.23), it follows that

$$T(\widehat{F}_n) - T(F) = U_n + \triangle_n + R_n,$$

where

$$R_n = R_T^{(n)} + R_I^{(n)} + \sum_{i=1}^{4} R_{iI_1}^{(n)} + \sum_{i=1}^{3} R_{iII_1}^{(n)} + R_{1II}^{(n)} + R_{2II}^{(n)}.$$

For getting the conclusion in Theorem 1, all we need to do is to bound R_n with the following inequality

$$P(\sqrt{n}|R_n(x)| > cn^{-\frac{1}{2}}) \leq cn^{-1}.$$

We now bound every term in R_n.

1) To bound $R_T^{(n)}$.

Let $\tau_n = \inf_x\{x : \widehat{F}_n(x) = 1-\beta\}$. τ_0 is as defined right before Theorem 2. Integrating by parts on $R_T^{(n)}$, and then employing two term Taylor's expansion for $K(\widehat{F}_n(x))$ as well as the Lipschitz condition for $J'(\cdot)$, it can be verified that

$$|R_T^{(n)}| \leq c_0 \int_0^{(2\tau_n)\vee\tau_0} |\widehat{F}_n(x) - F(x)|^3\, dx. \tag{3.24}$$

Hence

$$\begin{aligned}
&P(\sqrt{n}|R_T^{(n)}(x)| > cn^{-\frac{1}{2}}) \\
&\leq P(\sqrt{n}|R_T^{(n)}(x)| > cn^{-\frac{1}{2}},) \\
&+P(\sqrt{n}|R_T^{(n)}(x)| > cn^{-\frac{1}{2}}) \\
&\leq 2P(\int_0^{4\tau_0} |\widehat{F}_n(x) - F(x)|^3\, dx > cn^{-1}) + P(\tau_n > 2\tau_0) \\
&\triangleq r_{n1} + r_{n2}.
\end{aligned} \tag{3.25}$$

Similar to Wang(1996, formulae (2.50)), we have for $0 \leq x < \tau_H, p \geq 2$

$$E|\log \widehat{\bar{F}}_n(x) - \log \bar{F}(x)|^p \leq c_0 n^{-\frac{p}{2}} \bar{H}^{-2p}(x) \tag{3.26}$$

Hence, by the Tchebychev inequality, the inequality $|x-y| \leq |\log x - \log y|$, for $0 < x, y < 1$ and (3.26), we have

$$r_{n1} \leq cn^2 \int_0^{2\tau_0} E|\widehat{\bar{F}}_n(x) - \bar{F}(x)|^6\, dx \leq cn^{-1}. \tag{3.27}$$

Note that

$$\{\tau_n > 2\tau_0\}$$

$$\subseteq \{\text{There exists} \quad \tau_0 \leq s_0 < \tau_H, \text{and} 0 < a < 1 \quad \text{such that} \sup_{0\leq x\leq s_0} |\widehat{F}_n(x) - F(x)| > a\}.$$

Invoking Major and Rejtö (1988, formulae (2.18)), we have for sufficiently large n

$$r_{n2} \le P(\sup_{0 \le x \le s_0} |\log \widehat{\bar{F}}_n(x) - \log \bar{F}(x)| > a) \le c_0 \exp\{-\lambda \bar{H}(s_0) a^2 n\}, \tag{3.28}$$

where both λ and c_0 are some absolute constants. Clearly, (3.25), (3.27) and (3.28) together yield that

$$P(\sqrt{n}|R_T^{(n)}(x)| > cn^{-\frac{1}{2}}) \le cn^{-1}. \tag{3.29}$$

2). To bound $R_I^{(n)}$.

By the Tchebyschev inequality, Cauchy-Schwartz inequality and (3.26), we have

$$\begin{aligned} & P(\sqrt{n}|R_I^{(n)}(x)| > cn^{\frac{1}{2}}) \\ & \le cn^2 \int_0^{\tau_0} \int_0^{\tau_0} J(F(x))J(F(y)) \\ & \quad \times E^{\frac{1}{2}}(\log \widehat{\bar{F}}_n(x) - \log \bar{F}(x))^6 E^{\frac{1}{2}}(\log \widehat{\bar{F}}_n(y) - \log \bar{F}(y))^6 \, dx \, dy \le cn^{-1}. \end{aligned} \tag{3.30}$$

3). To bound $R_{iII}^{(n)}$.

Similar to $R_I^{(n)}$, we have

$$P(\sqrt{n}|R_{1II}^{(n)}(x)| > cn^{-\frac{1}{2}}) \le cn^{-1}. \tag{3.31}$$

The Markov inequality then immediately yields that

$$P(\sqrt{n}|R_{2II}^{(n)}(x)| > cn^{-\frac{1}{2}}) \le cn^{-1}. \tag{3.32}$$

4). To bound $R_{iI_1}^{(n)}$ and $R_{iI_2}^{(n)}$.

Note that, by Chang and Rao(1989),

$$E\widetilde{g}^2(Z_1, \delta_1; x) = -\int_0^x \bar{H}^{-2}(s) \, d\bar{H}_1(s). \tag{3.33}$$

Again using the Markov inequality, we then have

$$\begin{aligned} & P(\sqrt{n}|R_{1I_1}^{(n)}(x)| > cn^{-\frac{1}{2}}) \\ & \le n^{-1} \int_0^{\tau_0} |J(F(x))| \bar{F}(x) E\widetilde{g}^2(Z_1, \delta_1; x) \, dx \\ & \le \big(c\bar{H}^{-2}(\tau_0) \int_0^{\tau_0} |J(F(x))| \, dx\big) n^{-1} \le cn^{-1}. \end{aligned} \tag{3.34}$$

Similarly, we have

$$P(\sqrt{n}|R_{1II_1}^{(n)}(x)| > cn^{-\frac{1}{2}}) \le cn^{-1}. \tag{3.35}$$

For $R_{2I_1}^{(n)}$, applying the Cauchy-Schwartz inequality and the Dharmadhikar-Jogdeo (D-J) inequality (see, e.g., Rao 1987), it follows that

$$\begin{aligned} ER_{2I_1}^{(n)^2} \quad & \leq c_0 n^{-1} \int_0^{\tau_0} \int_0^{\tau_0} J(F(x)J(F(y))\bar{F}(x)\bar{F}(y) \\ & \times E^{\frac{1}{4}} r_n^4(x) E^{\frac{1}{4}} r_n^4(y) E^{\frac{1}{4}} \tilde{g}^4(Z_1,\delta_1;x) E^{\frac{1}{4}} \tilde{g}^4(Z_1,\delta_1;y)\, dx\, dy \\ & \leq c_0 n^{-1} \bar{H}^{-2}(\tau_0) \Big(\int_0^{\tau_0} J(F(x))\bar{F}(x) E^{\frac{1}{4}} r_n^4(x)\, dx \Big)^2. \end{aligned} \tag{3.36}$$

where $E^b f =: (Ef)^b$ for positive constant b. Let $r_n^*(x) = \widehat{F}_n(x) - \bar{F}(x) - \frac{1}{n}\sum_{i=1}^n \xi(Z_i,\delta_i;x)$ where $\xi(z_i,\delta_i;x)$ is as defined in subsection 1. By Taylor's expansion for $exp(\log \widehat{\bar{F}}_n(x) - \log \bar{F}(x))$ and the same argument as that for (3.9), we obtain

$$r_n(x) = \frac{r_n^*(x)}{\bar{F}(x)} - \frac{\widehat{\bar{F}}_n^{\theta_4}(x)}{\bar{F}^{\theta_4}(x)} (\log \widehat{\bar{F}}_n(x) - \log \bar{F}(x))^2, 0 < \theta_4 < 1, \tag{3.37}$$

where $r_n(x)$ is as defined right after (3.19). Hence, (3.36) and (3.37) together give

$$\begin{aligned} ER_{2I_1}^{(n)^2} \quad & \leq c_0 n^{-1} \Big[E^{\frac{1}{2}} \Big[\sup_{0\leq x\leq \tau_0} r_n^*(x) \Big]^4 \Big(\int_0^{\tau_0} |J(F(x))|\bar{F}(x)\, dx \Big)^2 \\ & + \Big(\int_0^{\tau_0} |J(F(x))| \widehat{\bar{F}}_n^{\theta_4}(x) \bar{F}^{1-\theta_4}(x) E^{\frac{1}{4}} (\log \widehat{\bar{F}}_n(x) - \log \bar{F}(x))^8\, dx \Big)^2 \Big]. \end{aligned} \tag{3.38}$$

By Theorem 1(c) in Gijbels and Wang(1993), we have

$$E[\sup_{0\leq x\leq \tau_0} |r_n^*(x)|^p] \leq c_0 n^{-p} \tag{3.39}$$

for any $p > 0$. Hence, combining (3.38), (3.39) with (3.26) it follows that

$$ER_{2I_1}^{(n)^2} \leq cn^{-3}. \tag{3.40}$$

Clearly, the Tchebyschev inequality and (3.40) together yield that

$$P(\sqrt{n}|R_{2I_1}^{(n)}(x)| > cn^{-\frac{1}{2}}) \leq cn^{-1}. \tag{3.41}$$

Similarly for $R_{2II_1}^{(n)}$. That is

$$P(\sqrt{n}|R_{2II_1}^{(n)}(x)| > cn^{-\frac{1}{2}}) \leq cn^{-2}. \tag{3.42}$$

Employing the Markov inequality and combining (3.37), (3.39) with (3.26)

$$\begin{aligned} & P(\sqrt{n}|R_{3I_1}^{(n)}(x)| > cn^{-\frac{1}{2}}) \\ & \leq cn\Big\{ E[\sup_{0\leq x\leq \tau_0} r_n^{*2}(x)] \int_0^{\tau_0} |J(F(x))|(F(x))^{-1}\, dx \\ & + \int_0^{\tau_0} |J(F(x))| E(\log \widehat{\bar{F}}_n(x) - \log \bar{F}(x))^4 (F(x))^{-1}\, dx \Big\} \leq cn^{-1} \end{aligned} \tag{3.43}$$

where r_n^* is as defined right after (3.36). Similarly, it can proved that

$$P(\sqrt{n}|R_{3II_1}^{(n)}(x)| > cn^{-\frac{1}{2}}) \leq cn^{-1}. \tag{3.44}$$

In virtue of the Markov inequality, the Cauchy-Schwartz inequality and Lemma 1, it follows that

$$P(\sqrt{n}|R_{4I_1}^{(n)}(x)| > cn^{-\frac{1}{2}}) \le cn^{-1}\Big(\int_0^{\tau_0} J(F(x))\bar{F}(x)\bar{H}^{-8}(x)\,dx\Big)^2. \tag{3.45}$$

So far we have proved that

$$P(\sqrt{n}|R_n| > cn^{-\frac{1}{2}}) \le cn^{-1}. \tag{3.46}$$

Completing the proof of Theorem 1.

To prove Theorem 2, we first cite a lemma which was derived by Chang and Rao (1989).

Lemma 2. For any random variable ζ and η and any real value nonnegative constant α, we have

$$\sup_x |P(\zeta+\eta \le x) - \Phi(x)| \le \sup_x |P(\zeta \le x) - \Phi(x)| + \frac{\alpha}{\sqrt{2}\pi} + P(|\eta| > \alpha). \tag{3.47}$$

The proof of Theorem 2. Note that

$$E\widetilde{g}(Z_i, \delta_i; x) = 0 \tag{3.48}$$

where $\widetilde{g}(\cdot,\cdot,\cdot)$ is as defined in subsection 2.1. Then

$$g(Z_1,\delta_1) \triangleq E[h(Z_1,\delta_1;Z_2,\delta_2)|Z_1,\delta_1] = \int_0^\infty J(F(x))\bar{F}(x)\widetilde{g}(Z_1,\delta_1;x)\,dx, \tag{3.49}$$

and

$$\begin{aligned}\psi(Z_1,\delta_1;Z_2,\delta_2) \quad &\triangleq h(Z_1,\delta_1;Z_2,\delta_2) - g(Z_1,\delta_1) - g(Z_2,\delta_2)\\ &= \int_0^\infty J(F(x))\bar{F}(x)[\widetilde{\psi}(Z_1,\delta_1;Z_2,\delta_2;x) + \widetilde{g}(Z_1,\delta_1;x) + \widetilde{g}(Z_2,\delta_2;x)]\,dx\\ &+ \int_0^\infty J'(F(x))\bar{F}^2(x)\widetilde{g}(Z_1,\delta_1;x)\widetilde{g}(Z_2,\delta_2;x)\,dF(x).\end{aligned}$$

By (3.49) we can get that, via some elementary calculation,

$$\begin{aligned}&\sigma^2 \triangleq Eg^2(Z_1,\delta_1)\\ &= E\Big\{\int_0^\infty\int_0^\infty J(F(x))J(F(y))\bar{F}(x)\bar{F}(y)\Big[\bar{H}^{-1}(Z_1)I[0\le Z_1\le x,\delta_1=1]\\ &+ \int_0^{x\wedge Z_1}\bar{H}^{-2}(s)\,d\bar{H}_1(s)\Big]\Big[\bar{H}^{-1}(Z_1)I[0\le Z_1\le y,\delta_1=1]\\ &+ \int_0^{y\wedge Z_1}\bar{H}^{-2}(s)\,d\bar{H}_1(s)\Big]\,dx\,dy\Big\}\\ &= \int_0^\infty\int_0^\infty J(F(x))J(F(y))\bar{F}(x)\bar{F}(y)\Big[\int_0^{x\wedge y}\bar{H}^{-2}(s)\,dH_1(s)\Big]\,dx\,dy.\end{aligned} \tag{3.50}$$

Also

$$Eh^2(Z_1,\delta_1;Z_2,\delta_2)$$

$$
\begin{aligned}
&\leq 3\Big\{ \int_0^\infty \int_0^\infty J(F(x))J(F(y))\bar F(x)\bar F(y)E[\widetilde h(Z_1,\delta_1;Z_2,\delta_2;x)\widetilde h(Z_1,\delta_1;Z_2,\delta_2;y)]\,dx\,dy\\
&+\int_0^\infty \int_0^\infty J(F(x))J(F(y))\bar F(x)\bar F(y)\\
&\times E[\widetilde g(Z_1,\delta_1;x)\widetilde g(Z_1,\delta_1;x)\widetilde g(Z_2,\delta_2;x)\widetilde g(Z_2,\delta_2;y)]\,dx\,dy\\
&+\int_0^\infty \int_0^\infty J'(F(x))J'(F(y))E[\xi(Z_1,\delta_1;x)\xi(Z_1,\delta_1;y)\xi(Z_2,\delta_2;x)\xi(Z_2,\delta_2;y)]\,dx\,dy\\
&\triangleq 3(E_{n1}+E_{n2}+E_{n3}).
\end{aligned} \tag{3.51}
$$

Note again that

$$
\begin{aligned}
&E\widetilde g(Z_1,\delta_1;\cdot)=0, \qquad E\widetilde g(Z_1,\delta_1;x)\widetilde g(Z_1,\delta_1;y)=\int_0^{x\wedge y}\bar H^{-2}\,dH_1,\\
&E[\widetilde\psi(Z_1,\delta_1;Z_2,\delta_2;\cdot)|Z_1,\delta_1]=E[\widetilde\psi(Z_2,\delta_2;Z_1,\delta_1;\cdot)|Z_2,\delta_2]\\
&\qquad\qquad\qquad\qquad = E\widetilde\psi(Z_1,\delta_1;Z_2,\delta_2;\cdot)]=0.
\end{aligned}
$$

Hence, combining the independence of $(Z_1,\delta_1),\cdots,(Z_n,\delta_n)$, we have

$$
E_{n2}=\int_0^\infty \int_0^\infty J(F(x))J(F(y))\bar F(x)\bar F(y)\big(\int_0^{x\wedge y}\bar H^{-2}\,dH_1\big)^2\,dx\,dy, \tag{3.52}
$$

$$
E_{n3}=\int_0^\infty \int_0^\infty J'(F(x))J'(F(y))\bar F^2(x)\bar F^2(y)\big(\int_0^{x\wedge y}\bar H^{-2}\,dH_1\big)^2\,dx\,dy, \tag{3.53}
$$

and

$$
\begin{aligned}
&E\widetilde h(Z_1,\delta_1;Z_2,\delta_2;x)\widetilde h(Z_1,\delta_1;Z_2,\delta_2;y)\\
&=2E\widetilde g(Z_1,\delta_1;x)\widetilde g(Z_1,\delta_1;y)+2E\widetilde\psi(Z_1,\delta_1;Z_2,\delta_2;x)\widetilde\psi(Z_1,\delta_1;Z_2,\delta_2;y)\\
&+2E\widetilde\psi(Z_1,\delta_1;Z_2,\delta_2;x)\widetilde\psi(Z_2,\delta_2;Z_1,\delta_1;y)\\
&\leq 2\int_0^{x\wedge y}\bar H^{-2}\,dH_1+4\int_0^{x\wedge y}\bar H^{-3}\,dH_1.
\end{aligned} \tag{3.54}
$$

Consequently, E_{n1} can be bounded by

$$
6\int_0^\infty \int_0^\infty J(F(x))J(F(y))\bar F(x)\bar F(y)\int_0^{x\wedge y}\bar H^{-3}\,dH_1\,dx\,dy. \tag{3.55}
$$

Under the conditions of Theorem 2

$$
\begin{aligned}
Eh^2(Z_1,\delta_1;Z_2,\delta_2) \quad &\leq 21\big[\big(\textstyle\int_0^\infty J(F(x))\bar F(x)\bar H^{-2}(x)\,dx\big)^2\\
&\quad+\big(\textstyle\int_0^\infty J'(F(x))\bar F(x)\bar H^{-2}(x)\,dx\big)^2\big]
\end{aligned} \tag{3.56}
$$

On the other hand

$$
\begin{aligned}
E|g(Z_1,\delta_1)|^3 \qquad &\leq \int_0^\infty \int_0^\infty \int_0^\infty J(F(x))J(F(y))J(F(z))\bar F(x)\bar F(y)\bar F(z)\\
&\quad\times E|\widetilde g(Z_1,\delta_1;x)\widetilde g(Z_1,\delta_1;y)\widetilde g(Z_1,\delta_1;z)|\,dxdydz
\end{aligned} \tag{3.57}
$$

and

$$
E|\widetilde g(Z_1,\delta_1;x)\widetilde g(Z_1,\delta_1;y)\widetilde g(Z_1,\delta_1;z)|
$$

$$
\begin{aligned}
\le\ & -\int_0^{x\wedge y\wedge z} \bar H^{-3}\,d\bar H_1 + \int_0^{x\wedge y} \big(\bar H^{-2}\int_0^{y\wedge s} \bar H^{-2}\,d\bar H_1\big)\,d\bar H_1(s) \\
& + \int_0^{x\wedge z} \big(\bar H^{-2}(s)\int_0^{y\wedge s} \bar H^{-2}\,d\bar H_1\big)\,d\bar H_1(s) \\
& + \int_0^{y\wedge z} \big(\bar H^{-2}(s)\int_0^{x\wedge s} \bar H^{-2}\,d\bar H_1\big)\,d\bar H_1(s) \\
& - \int_0^{x} \bar H^{-1}(s)\big(\int_0^{y\wedge s} \bar H^{-2}\,d\bar H_1\big)\big(\int_0^{z\wedge s} \bar H^{-2}\,d\bar H_1\big)\,d\bar H_1(s) \\
& - \int_0^{y} \bar H^{-1}(s)\big(\int_0^{x\wedge s} \bar H^{-2}\,d\bar H_1\big)\big(\int_0^{z\wedge s} \bar H^{-2}\,d\bar H_1\big)\,d\bar H_1(s) \\
& - \int_0^{z} \bar H^{-1}(s)\big(\int_0^{x\wedge s} \bar H^{-2}\,d\bar H_1\big)\big(\int_0^{y\wedge s} \bar H^{-2}\,d\bar H_1\big)\,d\bar H_1(s) \qquad (3.58)
\end{aligned}
$$

Substituting (3.58) into (3.57), it is easy to see that

$$
E|g(Z_1,\delta_1)|^3 \le \big(\int_0^{\tau_0} |J(F(x))|\bar F(x)\bar H^{-2}(x)\,dx\big)^3. \qquad (3.59)
$$

By (3.59) and (3.50), we have that

$$
\frac{E|g(Z_1,\delta_1)|^3}{(Eg^2(Z_1,\delta_1))^{\frac{3}{2}}} \le \left[\frac{\int_0^{\tau_0} |J(F(x))\bar F(x)\bar H^{-2}(x)\,dx}{\sigma}\right]^3 \triangleq B_1.
$$

By (3.56) and (3.50), we then get that

$$
\begin{aligned}
& \frac{Eh^2(Z_1,\delta_1;Z_2,\delta_2)}{Eg^2(Z_1,\delta_1)} \\
\le\ & \frac{21\big[\big(\int_0^{\tau_0} J(F(x))\bar F(x)\bar H^{-2}(x)\,dx\big)^2 + \big(\int_0^{\tau_0} J'(F(x))\bar F^2(x)\bar H^{-2}(x)\,dx\big)^2\big]}{\sigma^2} \triangleq B_2
\end{aligned}
$$

Let $\widetilde B = B_1 + B_2/21$. By Helmers and Van Zwet(1982), there exists an absolute constant c_0 such that

$$
\sup_x |P(\frac{n}{n-1}\sqrt{n}\sigma^{-1}U_n \le x) - \Phi(x)| \le c_0\widetilde B n^{-\frac{1}{2}}. \qquad (3.60)
$$

Since

$$
VarU_n = \frac{(n-1)(n-2)}{n^3}\sigma^2 + \frac{n-1}{2n^3}Eh^2(Z_1,\delta_1;Z_2,\delta_2),
$$

hence, we then have for large n

$$
P(|\frac{1}{n-1}\sqrt{n}\sigma^{-1}U_n| > \widetilde B n^{-\frac{1}{2}}) \le cn^{-1} \le \widetilde B n^{-\frac{1}{2}}. \qquad (3.61)
$$

By (3.60), (3.61) and Lemma 2, we come to the conclusion that there exists an absolute constant c_0 such that

$$
\sup_x |P(\sqrt{n}\sigma^{-1}U_n \le x) - \Phi(x)| \le c_0\widetilde B n^{-\frac{1}{2}}. \qquad (3.62)
$$

Recalling the definition of $\triangle_n$ in Theorem 1, by (3.62) it follows that

$$
\begin{aligned}
& \sup_x |P(\sqrt{n}\sigma^{-1}(U_n + \triangle_n) \le x) - \Phi(x)| \\
\le\ & \sup_x |P(\sqrt{n}\sigma^{-1}U_n \le x) - \Phi(x)| + \sup_x |\Phi(x - \sqrt{n}\sigma^{-1}\triangle_n) - \Phi(x)| \\
\le\ & c_0\widetilde B n^{-\frac{1}{2}} + \frac{1}{\sqrt{2\pi}}\sigma^{-1}\big|\int_0^{\tau_0} J(F(x)\bar F(x)\bar H^{-2}(x)\,dx\big|n^{-\frac{1}{2}} \le c_0 B n^{-\frac{1}{2}}, \qquad (3.63)
\end{aligned}
$$

where B is as defined in Theorem 2. Combining (3.63), Theorem 1 and Lemma 2, Theorem 2 is proved.

REFERENCES

1. Breslow, N. and Crowley, J.(1974), A larfe sample study of the life and product-limit estimator under random censorship, *Ann. Statist.*, **2**, 437-453.

2. Chang, M. N. and Rao, P. V.(1989), Berry-Esséen Bound for the Kaplan-Meier estimator, *Commun.Statist. - Theory Meth.*, **18**, 4647-4664.

3. Gijbels, I. and Wang, J.(1993). Strong representations of the survival function estimator for truncated and censored data with applications, *J. Multi. Analy.*, **47**, 210-229.

4. Gijbels, I. and Veraverbeke, N. (1991), Almost sure asymptotic representation for a class of functionals of the Kaplan-Meier estimator, *Ann. Statist.*, **19**, 1457-1470.

5. Helmers, R. and Van Zwet, W. R.(1982), The Berry-Esséen Bound for U-Statistic, III(S. S. Gupta and J. O. Berger, Eds), New York, Academic, I, 497-512.

6. Kaplan, E. L. and Meier, P.(1958), Nonparametric estimation from incomplete observation, *JASA*, **53**, 457-481.

7. Major, P. and Rejtö, L.(1988), Strong embedding of the estimator of the distribution function under random censorship, *Ann. Statist.*, **16**, 1113-1132.

8. Rao, B. L. S. P. (1987), *Asymptotic Theory of Statistical Influence*, John Wiley & Sons, 127-128.

9. Reid, N.(1981), Influence functions for censored data, *Ann. Statist.*, **9**, 78-92.

10. Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New York.

11. Sander, J. M.(1975), Asymptotic normality of linear combinations of functions of order statistics with censored data, Technical report 8, Div. Biostatistics, Standford Univ.

12. Wang, Q. H.(1996), Consistent estimator in random censorship semiparametric models, *Science In China Ser A*, **39**, 163-176.