Electronic Journal of Statistics
Vol. 6 (2012) 843–860
ISSN: 1935-7524
DOI: 10.1214/12-EJS697

Smooth confidence intervals for the survival function under random right censoring

Dimitrios Bagkavos

Accenture, 17 Rostoviou Str, 11526, Athens, Greece
e-mail: dimitrios.bagkavos@accenture.com

and

Dimitrios Ioannides

University of Macedonia, 156 Egnatia Str, 54006, Thessaloniki, Greece

Abstract: The present article presents a methodological advance which contributes to the area of nonparametric survival analysis under random right censoring. The central idea is to develop pointwise confidence intervals for the survival function by means of a central limit theorem for an, already existing in the literature, kernel smooth survival estimate. Numerical simulations reveal the progress in coverage accuracy offered by the suggested confidence intervals over the proposals already existing in the literature.

AMS 2000 subject classifications: Primary 62G05; secondary 62N02.
Keywords and phrases: Survival function, confidence interval, censoring, kernel.

Received April 2011.

1. Introduction

This paper considers the development of smooth, pointwise confidence intervals for the survival function defined by $S(x) = P(T > x)$ where T is a random variable representing the observed lifetime of an individual. Practically $S(x)$ expresses the probability that a continuous duration will exceed a specified time point x. Naturally then, it has many applications in fields as diverse as reliability, medicine demography and insurance to name but a few.

Development of the confidence intervals is based on the kernel survival function estimate of [4], discussed in detail in Section 2. In the same article, Gulati and Padgett improved the coverage accuracy of the confidence bands of [8], by utilizing the smaller Mean Square Error (MSE) of kernel smooth estimates compared to that of their empirical counterparts. Both approaches though are still dependent on the empirical estimate of the Brownian motion probabilities (as exhibited in remark 2 below) which determines the width of the intervals and the confidence level. As a result, the bias of the estimated probabilities undermines the coverage accuracy of the obtained intervals. The purpose of the present research is to remedy this by quantifying the difference between the true and the estimated survival function as a normally distributed random variable. As a result, the estimated Brownian motion probabilities of the two aforementioned approaches are replaced by the usual normal distribution quantiles. This leads in significant improvement in coverage accuracy as it is evident from the replication of the simulation study of [4] in Section 4 below.

The methodological contributions of the present research include a central limit theorem for the smooth estimate of [4] as well as establishment of the rate with which uniform consistency of the estimate is achieved.

The rest of the paper is organized as follows. The framework of study together with definition of the smooth estimate is given in Section 2. Section 3 is devoted to the central limit theorem for the estimate, its consistency and the suggested confidence interval. The simulation study on the coverage accuracy of the suggested confidence interval is given in Section 4. Proofs of all theorems are deferred for Section 5 while Section 6 contains the proofs of auxiliary lemmas.

2. Notation and preliminaries

Let $T_1, T_2, \ldots, T_n$ be a sample of n i.i.d. survival times censored at the right by n i.i.d. random variables $U_1, U_2, \ldots, U_n$, independent of the T_i's. Let f and F be the density and distribution function of the T_i's and H the distribution function of the U_i's. The observed data are then the pairs (X_i, Δ_i), $i = 1, 2, \ldots, n$ with $X_i = \min\{T_i,\ U_i\}$ and $\Delta_i = 1_{\{T_i \le U_i\}}$ where $1_{\{\cdot\}}$ is the indicator random variable of the event $\{\cdot\}$. The observed data form an i.i.d. sample with probability density g and distribution function G which satisfies $1-G = (1-F)(1-H)$. An estimate of the unknown survival function can be defined by $\hat{S}(x) = 1 - \hat{F}(x)$ where

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i}{1-H(X_i)} W\left(\frac{x-X_i}{h}\right),$$

$$W(x) = \int_{-\infty}^{x} K(u)\, du.$$

The real-valued function K is called kernel and integrates to 1, while h is called bandwidth and controls the amount of smoothing applied to the estimate. Estimator $\hat{S}(x)$ cannot be used directly in practice as it involves the unknown censoring distribution $H(x)$. One solution is to reverse the intuitive role played by T_i and U_i and estimate $1 - H(x)$ by the (sightly modified) [10] estimator,

$$1 - \hat{H}(x) = \begin{cases} 1, & 0 \le x \le Z_1 \\ \prod_{i=1}^{k-1} \left(\frac{n-i+1}{n-i+2}\right)^{1-\Lambda_i}, & Z_{k-1} < x \le Z_k, k = 2, \ldots, n \\ \prod_{i=1}^{n} \left(\frac{n-i+1}{n-i+2}\right)^{1-\Lambda_i}, & Z_n < x, \end{cases}$$

where (Z_i, Λ_i) are the ordered X_i's, along with their censoring indicators Δ_i, $i = 1, \ldots, n$. This gives rise to the practically useful estimator

$$\hat{S}_n(x) = 1 - \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i}{1-\hat{H}(X_i)} W\left(\frac{x-X_i}{h}\right).$$

Estimator $\hat{S}_n(x)$ was employed in the context of survival function estimation in [4] where its asymptotic mean square error, optimal (with respect to MSE) bandwidth, strong uniform consistency and convergence to a mean zero Gaussian process were established. In addition, [13] and [15] respectively provided the conditions under which the estimate has superior Mean Square Error (MSE) and Mean Integrated Absolute Error compared to the Kaplan-Meier Estimate. Applications of $\hat{S}_n(x)$ include the works of [7, 12] on kernel hazard rate estimation. Moreover, the distribution estimate $1 - \hat{S}_n(x)$ has been applied extensively on quantile estimation under random right censorship, see [16] and the references therein.

3. Assumptions and main results

The main results of this research are presented in this section in the form of theorems 2–3 below. These concern, quantification of the asymptotic distribution of $\hat{S}(x)$ and the rate of its uniform consistency with respect to the true curve. Furthermore, theorem 2 is applied in construction of smooth confidence intervals for the true survival curve. Prior to that, the necessary notation, assumptions and the already established in the literature asymptotic properties of $\hat{S}(x)$ are given.

First, denote with $\mu_i(K)$ the ith moment, $i = 0, 1, 2$ of the function K and with $R(K)$ the integral of the real function K^2 over its domain. The following conditions are assumed throughout

1. $S(x)$ is twice differentiable and $S''(x)$ is bounded and uniformly continuous.
2. For $l = 0, 1, 2$, the lth derivative of K, $K^{(l)}$, is bounded and absolutely integrable with finite second moments.
3. $R(K) < +\infty$ and $\mu_0(K) = 1, \mu_1(K) = 0, \mu_2(K) < +\infty$, i.e. the kernel K is of order 2.
4. There exists small enough h such that $W((y - x)h^{-1})/(1 - G(y))$ is uniformly bounded for $|y - x| > M$, for any $M > 0$.

A consequence of condition 2 is that $W(x)$ is bounded, while condition 3 and particularly $\mu_0(K) = 1$ implies

$$\lim_{x \to -\infty} W(x) = 0, \text{ and } \lim_{x \to +\infty} W(x) = 1.$$

If, in contrast to condition 3, a kernel with order higher than 2 is used, then, while this will be asymptotically beneficiary in terms of bias, the resulting estimate may be negative which does not make any sense from a practical perspective. Conditions 1–3 are satisfied by virtually all kernels in use in practice, see for example [12]. Condition 4 essentially means that there should be enough censored data at the right end of the estimation region for the asymptotics to apply. It has to be noted that it is automatically satisfied when the kernel has bounded support. The following theorem summarizes the asymptotic properties of $\hat{S}(x)$ and it is a direct consequence of lemma 1 in [16], see also [4].

Theorem 1. *Let $n \to +\infty$, $h \to 0$. Then, under conditions 1–4,*

$$\mathbb{E}\left\{\hat{S}(x)\right\} - S(x) = S''(x)\frac{h^2}{2}\mu_2(K) + o(h^2)$$
$$\mathbb{V}\mathrm{ar}\left\{\hat{S}(x)\right\} = \frac{1}{n}\int_0^x \frac{f(y)}{1-H(y)}\,dy - \frac{S(x)^2}{n} - \frac{2h}{n}\frac{f(x)}{1-H(x)}\int yK(y)W(y)\,dy + O(h^2n^{-1}).$$

The next theorem establishes the pointwise asymptotic normality of $\hat{S}(x)$.

Theorem 2. *Under conditions 1–4, provided that $nh \to +\infty$ as $h \to 0$ as $n \to +\infty$,*

$$\sqrt{\mathbb{V}\mathrm{ar}\left\{\hat{S}(x)\right\}}^{-1}\left(\hat{S}(x) - \mathbb{E}\left\{\hat{S}(x)\right\}\right) \xrightarrow{d} N(0,1). \tag{1}$$

The next theorem provides the rate with which uniform consistency of $\hat{S}(x)$ is achieved.

Theorem 3. *Define $T = \sup\{x : F(x) < 1-\varepsilon\}$ for a small $\varepsilon > 0$. Under conditions 1–4 we have*

$$\sup_{0<x<T}|\hat{S}(x) - S(x)| = O_p(h^2 + n^{-\frac{1}{2}}).$$

Using the strong convergence of the Kaplan - Meier estimator to the true survival function it is easy to show that

$$\hat{S}_n(x) = \hat{S}(x) + o_p\left(n^{-1/2}\right) \tag{2}$$

which means that $\hat{S}(x)$ and $\hat{S}_n(x)$ are asymptotically equivalent.

Corollary 1. *Using (2) and Slutsky's lemma it is readily verified that theorems 1, 2 and 3 hold also for estimator $\hat{S}_n(x)$.*

Now, by theorem 1 we have that

$$B_n = \frac{\mathbb{E}\left\{\hat{S}(x)\right\} - S(x)}{\sqrt{\mathbb{V}\mathrm{ar}\left\{\hat{S}(x)\right\}}} \simeq \frac{S''(x)\frac{h^2}{2}\mu_2(K)}{\sqrt{\mathbb{V}\mathrm{ar}\left\{\hat{S}(x)\right\}}}. \tag{3}$$

Then, combining corollary 1, (1) of theorem (2) and (3), it easily follows that a $100(1-\alpha)\%$ two sided confidence interval for $S(x)$ is

$$\hat{S}_n(x) - \left(z_{a/2} + \hat{S}_n''(x)\frac{h^2}{2}\mu_2(K)\right)\sqrt{\mathbb{V}\mathrm{ar}\left\{\hat{S}_n(x)\right\}} \le S(x) \le \hat{S}_n(x) + \left(z_{a/2} - \hat{S}_n''(x)\frac{h^2}{2}\mu_2(K)\right)\sqrt{\mathbb{V}\mathrm{ar}\left\{\hat{S}_n(x)\right\}}. \tag{4}$$

Remark 1. By theorem 2 a confidence interval for $\mathbb{E}\hat{S}_n(x)$ is

$$\left[\hat{S}_n(x) - z_{a/2}\sqrt{\mathbb{V}\text{ar}\left\{\hat{S}(x)\right\}}, \hat{S}_n(x) + z_{a/2}\sqrt{\mathbb{V}\text{ar}\left\{\hat{S}(x)\right\}}\right]. \tag{5}$$

Note though that in the case where $nh^4 \to 0$, the deterministic term $B_n \to 0$ and therefore the confidence interval (5) is applicable to $S(x)$ as well. In the general case where $nh^4 \to c \neq 0$, by theorem 3 and corollary 1, $\mathbb{E}\hat{S}_n(x)$ can be regarded as the estimable part of $S(x)$ so still it is sensible to regard (5) as an asymptotically valid confidence interval for $S(x)$. The obvious advantage of (5) over (4) is that no derivative estimation is required. But the nominal coverage of (5), as a confidence interval of $S(x)$, is $\Phi(z_{a/2} - B_n) + \Phi(z_{a/2} + B_n) - 1$, which is less than $1 - a$. This means that for medium and below sized samples, (5) will have inferior performance compared to (4).

Remark 2. The confidence bands of [4] for the survival function $S(x)$ are

$$\left[\hat{S}_n(x) - \hat{S}_n(x)\frac{\hat{\lambda}_n}{(nc)^{1/2}(1 - c\hat{d}_n^*(t)[1 + cd_n^*(t)]^{-1})},\right.$$
$$\left.\hat{S}_n(x) + \hat{S}_n(x)\frac{\hat{\lambda}_n}{(nc)^{1/2}(1 - c\hat{d}_n^*(t)[1 + cd_n^*(t)]^{-1})}\right] \tag{6}$$

where $\hat{\lambda}_n$ is such that $Q_{c\hat{d}_n^*(T)}(\hat{\lambda}_n) = 1 - \alpha$, $Q_T(\lambda)$ denotes the probability that the standard Brownian motion process, say $B(x)$, lies between $\pm\lambda(1 + x)$ for all $x \in (0, T)$ and

$$\hat{d}_n^*(t) = -\int_0^t (\hat{S}_n(u)\hat{H}(u))^{-1}\, d\hat{S}_n(u)$$

is the limit variance of the process $\sqrt{n}(\hat{S}_n(x) - S(x))S^{-1}(x)$. From theorem 3.1 of [8], $\hat{\lambda}_n$ determines the confidence level and since it is estimated empirically it inherits loss in precision. This constitutes a substantial difference with (5) (and (4)) where due to theorem 2 the confidence level is determined by $z_{a/2}$.

Remark 3. The bootstrap (see [14] and the references therein for an overview and applications) version of the proposed confidence intervals can be obtained in a straightforward manner by combining the suggestions of [2] on resampling from censored data and those of [6] on constructing bootstrap-t confidence intervals in the density setting.

First, the bootstrap samples can be obtained by resampling with replacement the pairs (X_i, Δ_i), $i = 1, 2, \ldots, n$, with probability n^{-1} for each doublet to be drawn. This leads to the sample (X_i^*, Δ_i^*), $i = 1, 2, \ldots, n$ in each iteration where inevitably some pairs are contained more than once while others are omitted. According to [2], this is equivalent to constructing the Kaplan-Meier estimates for both the survival and the censoring populations and treating them as the true distributions for the purpose of drawing second stage samples.

Construction then of a confidence interval for the survival function is based on approximating the sampling distribution of the adjusted by the standard deviation difference between the estimated and the true curve. For this, let $L^*(x) = (\hat{S}_n^*(x) - \hat{S}_n(x))(\sigma^*(x))^{-1}$ where $\hat{S}_n^*(x)$ denotes $\hat{S}_n(x)$ using the resample $(X_i^*, \Delta_i^*), i = 1, 2, \ldots, n$ and $\sigma^*(x)$ is its standard deviation. Define the bootstrap estimate of the ath quantile, $\hat{u}_a$, of the distribution of $L^*(x)$, by

$$P(L^*(x) \leq \hat{u}_a | (X_i^*, \Delta_i^*), i = 1, 2, \ldots, n) = a.$$

Then, by arguments entirely similar to [6], bootstrap-t confidence intervals for $S(x)$ with explicit bias correction are given by

$$\Big[\hat{S}_n(x) - \hat{u}_a \sqrt{\mathbb{V}\mathrm{ar}\left\{\hat{S}_n(x)\right\}} - \hat{S}_n''(x)\frac{h^2}{2}\mu_2(K),$$
$$\hat{S}_n(x) - \hat{u}_{(1-a)/2} \sqrt{\mathbb{V}\mathrm{ar}\left\{\hat{S}_n(x)\right\}} - \hat{S}_n''(x)\frac{h^2}{2}\mu_2(K)\Big]. \quad (7)$$

Another alternative is to deliberately undersmooth $\hat{S}_n(x)$ in constructing the above confidence interval so that its bias be identically 0. In this case the corresponding confidence interval for $S(x)$ would be

$$\left[\hat{S}_n(x) - \hat{u}_a \sqrt{\mathbb{V}\mathrm{ar}\left\{\hat{S}_n(x)\right\}}, \hat{S}_n(x) - \hat{u}_{(1-a)/2} \sqrt{\mathbb{V}\mathrm{ar}\left\{\hat{S}_n(x)\right\}}\right] \quad (8)$$

where now $\hat{S}_n(x)$ is using bandwidth ch where $0 < c < 1$.

Practical implementation of the suggested confidence intervals as well as their performance is discussed in the next section.

4. Simulations

In this section the performance of the suggested confidence intervals is investigated through numerical examples. For this purpose the simulation examples of [4] have been replicated so as to compare the coverage probabilities of all approaches discussed here, see [4], Section 4 for full implementation details.

In addition, estimation of $1 - H(x)$ is done by $1 - \hat{H}(x)$, defined in Section 2. The density $f(x)$ is estimated by estimator $f_n(x)$ of [17] (equation (2.1) there). The bandwidth of $f_n(x)$ is determined by the ISE optimal rule of [17], i.e. by minimization of the cross validation criterion, page 1526, [17], over the region $(0, X_{(n)}/2)$ where $X_{(n)}$ is the largest sample observation. The routine `nlminb` of S-plus is used to minimize the cross validation function. Moreover, in each example $f_n(x)$ uses the density equivalent kernel of $\hat{S}_n(x)$. In all expressions that involve $\hat{S}_n''(x)$, its value is obtained by

$$S_n''(x) = -\frac{1}{na^2} \sum_{i=1}^{n} \frac{\Delta_i}{1 - \hat{H}(X_i)} K'\left(\frac{x - X_i}{a}\right)$$

where a is the bandwidth resulting from minimization of the same cross validation criterion as in the case of density estimation but with $f_n(x)$ replaced by $S''_n(x)$, $f_{n,i}(x)$ by $S''_{n,i}(x)$ (the leave the ith observation out versions of $f_n(x)$ and $S''_n(x)$ respectively) and K' is the first derivative of the kernel K. The confidence intervals (7), (8) are implemented as follows. For each of the 2000 (or 1000) iterations used in [4], 800 bootstrap iterations are used for every implementation of (7) and (8). For each bootstrap sample, $L^*(x)$ is being calculated and then definition 7 of [9] is applied on the 800 $L^*(x)$ values to determine $\hat{u}_a$ for the desired level a. Particularly for (8), $\hat{S}_n(x)$ is undersmoothed in each instance of calculation by a grid search; that is $\hat{S}_n(x)$ is calculated 100 times, one for each ch, $c = 0.01, 0.02, \ldots, 1$ and h determined as indicated by [4] and the version of which corresponds to minimum bias at x, compared to the true survival function is used.

Particular focus is given here on the right tail of the area of estimation as this is the area with most room for improvement. The reason is that, as also noted by [4], sparseness of the data on the right tail together with censoring make it more difficult to estimate accurately the required percentiles on the right tail. The present simulation confirmed the good performance of the Gulati and Padgett, [4], suggestion on the left tail and showed that both approaches are rather equivalent there, with the confidence intervals obtained by (4) being slightly better in terms of coverage probabilities. Consequently, only coverage probabilities that concern the right tail are presented here. Moreover, it should be clarified that the confidence bands (6) are used as confidence intervals here just as a benchmark to examine the performance of (4). Use of (4) is more appropriate for point estimations such as distribution parameters or in general inference proportions or impact numbers (measures) used for inference. On the contrary, (6) would be more appropriate when a continuous estimate is required such as when estimating the survival function as a whole.

The results are presented in table 1. Column n is the sample size, the second column (D) is the distribution used to generate the data, the third column (PC) is the percentage of censoring, the fourth column (CD) is the censoring distribution, column CL is the confidence level, column NR is the number of replications, column K denotes the kernel used in the implementation (T is for Triangular and N for Normal), column t is the point of estimation. The columns titled CP1 and CP2 are the achieved coverage probabilities that correspond to the approach of [4] and (4) respectively. As, in contrast with the approach of [4], the coverage probabilities obtained by (4) are independent of the c parameter, only the best coverage probabilities in terms of the c, for each estimation point t and distribution are compared the with present suggestion. In the reported coverage probabilities in table 1 all values in the CP1 column correspond to $c = 1.5$ in the tables of [4]. In addition, the proposed confidence intervals are compared with those obtained by using the Kaplan Meier estimate with the variance correction proposed by [19]. This approach is readily implemented in the survival package of R by the `survfit` routine and the results are contained in column CP3 of table 1. The reasoning behind this comparison is that it gives a feel on how the proposed confidence intervals perform when compared with

TABLE 1

Coverage probabilities for the confidence intervals obtained by (6) (column CP1, the Gulati and Padgett approach in [4]), (4) (column CP2, the suggested approach resulting from theorem 2), the `survfit` in `R` (column CP3) and (5) (column CP4, the simplified – compared to (4) – approach stemming from using $\mathbb{E}S_n(x)$ instead of $S(x)$) and the bootstrap confidence intervals of (7) and (8) (columns (CP5, CP6)

n	D	PC	CD	CL	NR	K	t	CP1	CP2	CP3
25	exp(1)	25	exp(3)	0.9	2000	T	2.1	0.927	0.965	0.966
25	exp(1)	25	exp(3)	0.9	2000	T	2.1	0.969	0.989	0.925
50	exp(1)	50	exp(1)	0.9	2000	T	1.7	0.869	0.956	0.913
50	exp(1)	50	exp(1)	0.9	2000	T	2.1	0.758	0.944	0.832
50	exp(1)	50	exp(1)	0.95	2000	T	1.5	0.872	0.968	0.938
50	exp(1)	50	exp(1)	0.95	2000	T	1.8	0.787	0.954	0.84
50	exp(1)	50	exp(1)	0.95	2000	N	1.5	0.868	0.981	0.892
50	exp(1)	50	exp(1)	0.95	2000	N	1.8	0.785	0.964	0.862
30	W(0.5,1)	30	exp(0.375)	0.95	1000	T	3.1	0.861	0.99	0.887
30	W(0.5,1)	30	exp(0.375)	0.95	1000	T	3.6	0.805	0.959	0.892
50	W(2,1)	50	exp(0.865)	0.95	1000	T	1.65	0.944	0.983	0.991
n	D	PC	CD	CL	NR	K	t	CP4	CP5	CP6
25	exp(1)	25	exp(3)	0.9	2000	T	2.1	0.984	0.968	0.966
25	exp(1)	25	exp(3)	0.9	2000	T	2.1	0.995	0.978	0.984
50	exp(1)	50	exp(1)	0.9	2000	T	1.7	0.978	0.962	0.958
50	exp(1)	50	exp(1)	0.9	2000	T	2.1	0.954	0.941	0.939
50	exp(1)	50	exp(1)	0.95	2000	T	1.5	0.989	0.97	0.966
50	exp(1)	50	exp(1)	0.95	2000	T	1.8	0.978	0.958	0.955
50	exp(1)	50	exp(1)	0.95	2000	N	1.5	0.997	0.989	0.982
50	exp(1)	50	exp(1)	0.95	2000	N	1.8	0.976	0.959	0.955
30	W(0.5,1)	30	exp(0.375)	0.95	1000	T	3.1	0.98	0.978	0.977
30	W(0.5,1)	30	exp(0.375)	0.95	1000	T	3.6	0.972	0.964	0.961
50	W(2,1)	50	exp(0.865)	0.95	1000	T	1.65	0.986	0.979	0.979

standard methods of survival analysis. Furthermore the proposed confidence intervals are compared to the confidence intervals obtained by (5) (column CP4 of table 1) so as to exhibit the difference in precision of the two methods in practice. Finally, columns CP5 and CP6 correspond to the bootstrap confidence intervals obtained by (7) and (8) respectively.

It is evident from table 1 that on the examples considered, utilizing the approach suggested by (4) leads to an improvement of approximately 11% compared to (6), on average across all examples. The improvement is driven by the independence of (4) of percentile estimation. A side indication of this comparison though is that the proposed methodology produces consistently larger coverage probabilities compared to the [4] approach. Also, the confidence intervals produced by (4) appear to have closest to the nominal coverage probabilities, in general, compared to those produced by the R/`survfit` approach, (5), (7) and (8). It has to be noted that especially (7) and (8) produce actually results slightly inferior compared to (4), but one has to take into account that they are obtained by a quite computer intensive manner. Finally, in table 2, on exactly the same settings as in table 1 the average widths of the confidence intervals are displayed.

Table 2

Coverage probabilities for the confidence intervals obtained by (6) (column CP1, the Gulati and Padgett approach in [4]), (4) (column CP2, the suggested approach resulting from theorem 2), the `survfit` in `R` (column CP3) and (5) (column CP4, the simplified – compared to (4) – approach stemming from using $\mathbb{E}S_n(x)$ instead of $S(x)$) and the bootstrap confidence intervals of (7) and (8) (columns (CP5, CP6)

n	D	PC	CD	CL	NR	K	t	CP1	CP2	CP3
25	exp(1)	25	exp(3)	0.9	2000	T	2.1	0.1969	0.2009	0.1992
40	exp(1)	25	exp(3)	0.9	2000	T	2.1	0.1914	0.1933	0.1934
40	exp(1)	50	exp(1)	0.9	2000	T	1.7	0.1915	0.1958	0.1846
40	exp(1)	50	exp(1)	0.9	2000	T	2.1	0.2005	0.1985	0.1995
25	exp(1)	50	exp(1)	0.95	2000	T	1.5	0.2087	0.2134	0.1866
25	exp(1)	50	exp(1)	0.95	2000	T	1.8	0.2035	0.2096	0.2197
25	exp(1)	50	exp(1)	0.95	2000	N	1.5	0.2075	0.2164	0.208
25	exp(1)	50	exp(1)	0.95	2000	N	1.8	0.214	0.2064	0.2182
25	W(0.5,1)	30	exp(0.375)	0.95	1000	T	3.1	0.2162	0.2081	0.213
25	W(0.5,1)	30	exp(0.375)	0.95	1000	T	3.6	0.2183	0.2259	0.2124
25	W(2,1)	50	exp(0.865)	0.95	1000	T	1.65	0.2104	0.2235	0.22
n	D	PC	CD	CL	NR	K	t	CP4	CP5	CP6
25	exp(1)	25	exp(3)	0.9	2000	T	2.1	0.201	0.1976	0.1972
40	exp(1)	25	exp(3)	0.9	2000	T	2.1	0.2005	0.1996	0.2009
40	exp(1)	50	exp(1)	0.9	2000	T	1.7	0.2228	0.1964	0.1956
40	exp(1)	50	exp(1)	0.9	2000	T	2.1	0.1883	0.1921	0.1917
25	exp(1)	50	exp(1)	0.95	2000	T	1.5	0.2132	0.198	0.1972
25	exp(1)	50	exp(1)	0.95	2000	T	1.8	0.2204	0.1956	0.1949
25	exp(1)	50	exp(1)	0.95	2000	N	1.5	0.1945	0.2019	0.2005
25	exp(1)	50	exp(1)	0.95	2000	N	1.8	0.2126	0.1958	0.1949
25	W(0.5,1)	30	exp(0.375)	0.95	1000	T	3.1	0.1935	0.1996	0.1994
25	W(0.5,1)	30	exp(0.375)	0.95	1000	T	3.6	0.2014	0.1968	0.1962
25	W(2,1)	50	exp(0.865)	0.95	1000	T	1.65	0.22	0.1998	0.1998

5. Proofs

5.1. Proof of theorem 2

Estimator $\hat{F}(x)$ can be written as follows

$$S = \hat{F}(x) = \sum_{i=1}^{n} S_i, \quad S_i = \frac{1}{n}\frac{\Delta_i}{1 - H(X_i)} W\left(\frac{x - X_i}{h}\right).$$

The proof is based on lemma 4.1, [5] according to which the standardized versions of S and its projection on the subspace of all such independent terms, say $\hat{S}$, have the same asymptotic distribution. Then, it suffices to establish the asymptotic normality of the standardized $\hat{S}$ via the Lyapunov theorem. The approximation of S is given by

$$\hat{S} = \sum_{i=1}^{n} \mathbb{E}(S|Y_i) - (n-1)\mathbb{E}S.$$

The following conditions, which are easy to verify must hold

$$\mathbb{E}\hat{S} = \mathbb{E}S \text{ and } \mathbb{E}(S - \hat{S})^2 = \mathrm{Var}(S) - \mathrm{Var}(\hat{S}). \tag{9}$$

We should also show that

$$\mathrm{Var}(\hat{S})/\mathrm{Var}(S) \to 1 \text{ as } n \to +\infty. \tag{10}$$

Now,

$$\mathbb{E}(S_i|Y_i) = n^{-1}V(X_i), \tag{11}$$

where

$$V(X_i) = \frac{\Delta_i}{1-H(X_i)} W\left(\frac{x-X_i}{h}\right).$$

Using the fact that for $i \neq j$ $\mathbb{E}(S_j|Y_i) = \mathbb{E}S_j$, lemma 1 of [16], (9), (11) and lemma 1, it is easily seen that for a constant $C > 0$

$$\begin{aligned}\mathrm{Var}(\hat{S}) &= \mathrm{Var}\left(\sum_{i=1}^{n}\left\{n^{-1}V(X_i) + n^{-1}\mathbb{E}S + (n-1)n^{-1}C\right\}\right)\\ &= \mathrm{Var}(S) + n\mathrm{Var}(\mathbb{E}S) = \mathrm{Var}(S).\end{aligned}$$

Hence (10) is proved. By (9) and (10)

$$\mathbb{E}\left(\frac{\hat{S}-\mathbb{E}\hat{S}}{\sqrt{\mathrm{Var}(\hat{S})}} - \frac{S-\mathbb{E}S}{\sqrt{\mathrm{Var}(\hat{S})}}\right)^2 = \frac{\mathbb{E}\left(\hat{S}-S\right)^2}{\mathrm{Var}(\hat{S})} = \frac{\mathrm{Var}(\hat{S})-\mathrm{Var}(S)}{\mathrm{Var}(\hat{S})} \to 0$$

which proves that S and $\hat{S}$ have the same asymptotic distribution. Finally, in order to show that the standardized version of $\hat{S}$ converges in distribution to the standard normal distribution, note that the Lyapunov condition is readily established by combining lemma 3 and the fact that $h \to 0$ as $n \to +\infty$. From theorem 1, easily one derives that, $(\mathbb{E}\hat{S}(x) - S(x))/\mathbb{V}\mathrm{ar}(\hat{S}(x))) \to 0$ as $n \to +\infty$ which completes the proof.

5.2. Proof of (2)

By [11], we have that

$$\sup_{0<x<T} |\sqrt{n}(1-\hat{H}(x) - S(x))| = O_p(1). \tag{12}$$

Using (12) in the last step below yields

$$\begin{aligned}\hat{S}_n(x) &= \frac{1}{n}\sum_{i=1}^{n}\frac{\Delta_i}{1-H(X_i)}W\left(\frac{x-X_i}{h}\right)\\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\Delta_i W\left(\frac{x-X_i}{h}\right)\left\{\frac{1}{1-\hat{H}(X_i)} - \frac{1}{1-H(X_i)}\right\}\\ &= \hat{S}(x) + o_p\left((\sqrt{n})^{-1}\right).\end{aligned}$$

which together with theorem 1, theorem 2 and the Slutsky lemma gives the result.

5.3. Proof of theorem 3

First, denote with H the Heaviside function ($H(x) = 1(x > 0)$). Following [1], write

$$W(x) = \int_{-\infty}^{x} K(t)\,dt = \int_{-\infty}^{+\infty} K(t)H(z-t)\,dt = K * H(z)$$

where $*$ denotes convolution. Then the Fourier transform of W is the product of the Fourier transforms of K and H. Denote with $\mathcal{F}(F)$ the Fourier transform of the function F. We have

$$\mathcal{F}(W(x)) = \mathcal{F}(K(x))\mathcal{F}(H(x)) = \kappa(t)\pi\delta(t) + \kappa(t)(it)^{-1}$$

with $\delta(t)$ being the dirac delta function and $\kappa(t)$ the characteristic function of K, i.e.

$$\kappa(t) = \int e^{itx}K(x)\,dx.$$

From the inversion theorem of Fourier transforms,

$$W(x) = \frac{1}{2\pi}\int e^{-itx}\left(\kappa(t)\pi\delta(t) + \kappa(t)(it)^{-1}\right)\,dt.$$

Then, estimator $\hat{F}(x)$ can be written as

$$\begin{aligned}
\hat{F}(x) &= \frac{1}{n}\sum_{i=1}^{n}\frac{\Delta_i}{1-H(X_i)}W\left(\frac{x-X_i}{h}\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{\Delta_i}{1-H(X_i)}\frac{1}{2\pi}\int e^{-it(\frac{x-X_i}{h})}\left(\kappa(t)\pi\delta(t) + \kappa(t)(it)^{-1}\right)\,dt\\
&= \frac{1}{2\pi}\int\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\Delta_i}{1-H(X_i)}e^{\frac{itX_i}{h}}\right)e^{\frac{-itx}{h}}\kappa(t)\left(\pi\delta(t) + (it)^{-1}\right)\,dt\right\}\\
&= \frac{h}{2\pi}\int\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\Delta_i}{1-H(X_i)}e^{isX_i}\right)e^{-isx}\kappa(sh)\left(\pi\delta(sh) + (ish)^{-1}\right)\,ds
\end{aligned}$$

after setting $s = t/h$ in the last step above. Then,

$$\hat{F}(x) = \frac{h}{2\pi}\int e^{-isx}\hat{\phi}(s)\kappa(sh)\left(\pi\delta(sh) + (ish)^{-1}\right)\,ds,$$

where

$$\hat{\phi}(s) = \frac{1}{n}\sum_{i=1}^{n}\frac{\Delta_i}{1-H(X_i)}e^{isX_i}$$

is an empirical estimate of the characteristic function of $\phi(x)$ of $f(x)$,

$$\phi(x) = \int e^{ixy}f(y)\,dy$$

Therefore

$$\hat{F}(x) - \mathbb{E}\hat{F}(x) = \frac{h}{2\pi}\int e^{-itx}\left(\hat{\phi}(s) - \mathbb{E}(\hat{\phi}(s))\right)\kappa(sh)\left(\pi\delta(sh) + (ish)^{-1}\right)\, ds.$$

Now,

$$\begin{aligned}
\mathbb{E}\left(\frac{\Delta_i}{1-H(X_i)}e^{isX_i}\right) &= \mathbb{E}\left(\mathbb{E}\left(\frac{\Delta_i}{1-H(X_i)}e^{isX_i}\Big| X_i = X_{(r)}\right)\right)\\
&= \mathbb{E}\left(\mathbb{E}\left(\frac{\Delta_{(r)}}{1-H(X_{(r)})}e^{isX_{(r)}}\right)\right)\\
&= n\mathbb{E}\int \frac{f(y)(1-H(y))}{(1-H(y))g(y)}e^{isy}\binom{n}{r}G^r(y)(1-G(y))^{n-r}g(y)\, dy\\
&= n\mathbb{E}\int e^{isy}\binom{n}{r}G^r(y)(1-G(y))^{n-r}f(y)\, dy.
\end{aligned}$$

Since $r = i$ with probability $1/n$,

$$\begin{aligned}
\mathbb{E}\left(\frac{\Delta_i}{1-H(X_i)}e^{isX_i}\right) &= \frac{1}{n}\sum_{r=1}^{n} n\int e^{isy}\binom{n}{r}G^r(y)(1-G(y))^{n-r}f(y)\, dy\\
&= \int e^{isy}f(y)\, dy.
\end{aligned}$$

Hence

$$\mathbb{E}\hat{\phi}(s) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\frac{\Delta_i}{1-H(X_i)}e^{isX_i}\right) = \int e^{isy}f(y)\, dy = \phi(s)$$

and therefore $\hat{\phi}(s)$ is an unbiased estimator of $\phi(s)$. Then,

$$\hat{F}(x) - \mathbb{E}\hat{F}(x) = \frac{h}{2\pi}\int e^{-itx}\left(\hat{\phi}(s) - \phi(s)\right)\kappa(sh)\left(\pi\delta(sh) + (ish)^{-1}\right)\, ds.$$

Thus

$$|\hat{F}(x) - \mathbb{E}\hat{F}(x)| = \left|\frac{h}{2\pi}\int e^{-itx}\left(\hat{\phi}(s) - \phi(s)\right)\kappa(sh)\left(\pi\delta(sh) + (ish)^{-1}\right)\, ds\right|.$$

Now,

$$\left|e^{-itx}\right| = |\cos tx - i\sin tx| = (\cos^2 tx + \sin^2 tx)^{\frac{1}{2}} = 1$$

and then

$$\sup_x |\hat{F}(x) - \mathbb{E}\hat{F}(x)| \le \frac{h}{2\pi}\int \left|\hat{\phi}(s) - \phi(s)\right| |\kappa(sh)| \left|\pi\delta(sh) + (ish)^{-1}\right|\, ds.$$

Note that the RHS in the equation above does not depend on x and thus no supremum is needed. Now,

$$\mathbb{E}\sup_x |\hat{F}(x) - \mathbb{E}\hat{F}(x)| \le \frac{h}{2\pi}\int \mathbb{E}\left|\hat{\phi}(s) - \phi(s)\right| |\kappa(sh)| \left|\pi\delta(sh) + (ish)^{-1}\right|\, ds.$$

Set $g = \mathbb{E}|\hat{\phi}(s) - \phi(s)|$. Then,

$$\begin{aligned}\mathbb{E}\sup_x |\hat{F}(x) - \mathbb{E}\hat{F}(x)| &\leq \frac{h}{2\pi} g \int |\kappa(sh)| \left|\pi\delta(sh) + (ish)^{-1}\right| \, ds \\ &= \frac{1}{2\pi} g \int |\kappa(t)| \left|\pi\delta(t) + (it)^{-1}\right| \, dt \;\text{ after using } sh = t\end{aligned}$$

From the definition of the delta dirac function and its property that it integrates to 1 over $\mathbb{R}$, the fact that the kernel is a non-negative function, the property $\kappa(0) = 1$ we get

$$\int |\kappa(t)| \, |\pi\delta(t)| \, dt = \int \kappa(t)\pi\delta(t) \, dt = \int \kappa(0)\pi\delta(t) \, dt = \pi \int \delta(t) \, dt = \pi.$$

Also, noting that $\kappa(t)$ is an even function as the fourier transform of an even function,

$$\int \kappa(t)(it)^{-1} \, dt = i^{-1} \int \kappa(t) t^{-1} \, dt = 0$$

as the integral of an odd function over a compact set (since the kernel vanishes outside its compact support and since the product of an odd with an even function is an odd function). Thus we conclude,

$$\mathbb{E}\sup_x |\hat{F}(x) - \mathbb{E}\hat{F}(x)| \leq \frac{1}{2} \mathbb{E}\left|\hat{\phi}(s) - \phi(s)\right|$$

Now,

$$\begin{aligned}\mathbb{E}\left|\hat{\phi}(t) - \phi(t)\right| &= \mathbb{E}\left|\frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i}{1 - H(X_i)} \left(e^{-itX_i} - \mathbb{E}e^{-itX_i}\right)\right| \\ &= \mathbb{E}\Big|\frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i}{1 - H(X_i)} \left(\cos tX_i - \mathbb{E}\cos tX_i\right) \\ &\quad + i\frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i}{1 - H(X_i)} \left(\sin tX_i - \mathbb{E}\sin tX_i\right)\Big|.\end{aligned}$$

Set

$$\begin{aligned}V_1 &= \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i}{1 - H(X_i)} \left(\cos tX_i - \mathbb{E}\cos tX_i\right) \\ V_2 &= \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i}{1 - H(X_i)} \left(\sin tX_i - \mathbb{E}\sin tX_i\right).\end{aligned}$$

Note that $\mathbb{E}V_1 = 0 = \mathbb{E}V_2$ and so $\mathbb{V}\mathrm{ar}(V_1) = \mathbb{E}V_1^2$ and $\mathbb{V}\mathrm{ar}(V_2) = \mathbb{E}V_2^2$. Also it can be easily seen that $\mathbb{E}(V_1^2 + V_2^2)^{\frac{1}{2}} \leq (\mathbb{E}V_1^2 + \mathbb{E}V_2^2)^{\frac{1}{2}}$ by setting $Z^2 = V_1^2 + V_2^2$.

With these in mind we get

$$\begin{aligned}\mathbb{E}\left|\hat{\phi}(t)-\phi(t)\right| &= \mathbb{E}|V_1+iV_2| = \mathbb{E}((V_1+iV_2)(V_1-iV_2))^{\frac{1}{2}} \\ &= \mathbb{E}(V_1^2+V_2^2)^{\frac{1}{2}} \le (\mathbb{E}V_1^2+\mathbb{E}V_2^2)^{\frac{1}{2}} \\ &= (\mathbb{V}\mathrm{ar}\{V_1\}+\mathbb{V}\mathrm{ar}\{V_2\})^{\frac{1}{2}}. \end{aligned} \tag{13}$$

Now,

$$\begin{aligned}\mathbb{V}\mathrm{ar}\{V_1\} &= \mathbb{V}\mathrm{ar}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\Delta_i}{1-H(X_i)}(\cos tX_i-\mathbb{E}\cos tX_i)\right\} \\ &\le \mathbb{V}\mathrm{ar}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\Delta_i}{1-H(X_i)}\cos tX_i\right\} \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}\mathrm{ar}\left\{\frac{\Delta_i}{1-H(X_i)}\cos tX_i\right\} \\ &\quad -\frac{2}{n^2}\sum_{i=1}^{n-1}\mathbb{C}\mathrm{ov}\left\{\frac{\Delta_i}{1-H(X_i)}\cos tX_i,\frac{\Delta_{i+1}}{1-H(X_{i+1})}\cos tX_{i+1}\right\} \\ &\le \frac{\nu}{n^2p}\int_0^x\frac{f(u)}{1-H(u)}\cos^2(tu)\,du+o(n^{-1}).\end{aligned}$$

where $\nu=\{\#\Delta_i=1\}$. Similarly,

$$\mathbb{V}\mathrm{ar}\{V_2\}\le\frac{\nu}{n^2p}\int_0^x\frac{f(u)}{1-H(u)}\sin^2(tu)\,du+o(n^{-1}).$$

Note that $\nu\le n$. Then for a positive constant C,

$$\mathbb{V}\mathrm{ar}\{V_1\}+\mathbb{V}\mathrm{ar}\{V_2\}\le\frac{n}{n^2p}\int_0^x\frac{f(u)}{1-H(u)}\,du+o(n^{-1})\le Cn^{-1}$$

and hence (13) becomes,

$$\mathbb{E}\left|\hat{\phi}(t)-\phi(t)\right|\le Cn^{-\frac{1}{2}}.$$

Hence

$$\mathbb{E}\sup_x|\hat{F}(x)-\mathbb{E}\hat{F}(x)|\le C\sqrt{n^{-1}}=O_p\left(n^{-\frac{1}{2}}\right).$$

Using Markov's inequality

$$P\left\{\sup|\hat{F}(x)-\mathbb{E}\hat{F}(x)|>\varepsilon\right\}=O_p\left(n^{-\frac{1}{2}}\right)$$

which implies that

$$\sup_x|\hat{F}(x)-\mathbb{E}\hat{F}(x)|=O_p\left(n^{-\frac{1}{2}}\right) \tag{14}$$

Finally, using (14)

$$\sup_x |\hat{F}(x) - F(x)| \leq \sup_x |\hat{F}(x) - \mathbb{E}\hat{F}(x)| + \sup_x |\mathbb{E}\hat{F}(x) - F(x)|$$
$$= O_p\left(n^{-\frac{1}{2}}\right) + O(h^2) = O_p\left(n^{-\frac{1}{2}} + h^2\right).$$

This means that there is C such that $\sup_{0<x<T} |\hat{F}(x) - F(x)| \leq C(h^2 + n^{-\frac{1}{2}})$. Also, $U_n = O_p(W_n)$ if for each $\varepsilon > 0$ there exist M and N depending on ε such that $P(|U_n| \leq M|W_n|) > 1 - \varepsilon$ for all $n > N$. Hence

$$\sup_{0<x<T} |\hat{F}(x) - F(x)| = O_p(h^2 + 1/\sqrt{n})$$

from which theorem 3 follows immediately.

6. Auxiliary lemmas

First, define $K_h(\cdot) = h^{-1}K(\cdot/h)$.

Lemma 1. *For $i \neq j$,*

$$\mathbb{E}\left(S_j|Y_i\right) = n^{-1}\int K_h\left(x-y\right)F(y)\,dy + O(n^{-1}).$$

Proof. Let $Y_{(r)}, Y_{(k)}$ denote the rth and kth respectively ordered statistics, as these correspond to $X_{(r)}$ and $X_{(k)}$, the rth and kth in order of magnitude observations of the sample, along with their censoring indicators.

$$\begin{aligned}\mathbb{E}(S_j|Y_i) &= \mathbb{E}\left(\mathbb{E}(S_j|Y_i)\Big|Y_j = Y_{(r)}, Y_i = Y_{(k)}\right)\\ &= \frac{1}{n}\mathbb{E}\left(\mathbb{E}\left(\frac{\Delta_{(r)}}{1-H(X_{(r)})}W\left(\frac{x-X_{(r)}}{h}\right)\Big|Y_{(k)}\right)\Big|r<k\right)\\ &\quad+\frac{1}{n}\mathbb{E}\left(\mathbb{E}\left(\frac{\Delta_{(r)}}{1-H(X_{(r)})}W\left(\frac{x-X_{(r)}}{h}\right)\Big|Y_{(k)}\right)\Big|r>k\right).\end{aligned} \tag{15}$$

The density of $X_{(r)}$ is

$$X_{(r)} \sim \begin{cases}\binom{n-2}{r-1}G(X_{(r)})^{r-1}(1-G(X_{(r)}))^{n-(r+1)}, & r<k\\ \binom{n-2}{r-2}G(X_{(r)})^{r-2}(1-G(X_{(r)}))^{n-r}, & r>k.\end{cases}$$

Also, from [18] we have that $\mathbb{E}(\Delta_{(r)}|X_{(r)} = y) = f(y)(1-H(y))/g(y)$. In the case where $r<k$ we have that

$$\begin{aligned}&\frac{1}{n}\mathbb{E}\left(\mathbb{E}\left(\frac{\Delta_{(r)}}{1-H(X_{(r)})}W\left(\frac{x-X_{(r)}}{h}\right)\Big|Y_{(k)}\right)\Big|r<k\right)\\ &= \frac{1}{n}\sum_{r=1}^{n}\int\frac{f(y)(1-H(y))}{(1-H(y))g(y)}W\left(\frac{x-y}{h}\right)\binom{n-2}{r-1}\\ &\qquad\times G^{r-1}(y)(1-G(y))^{n-(r+1)}g(y)\,dy\\ &= \frac{1}{n}\int W\left(\frac{x-y}{h}\right)f(y)\,dy\end{aligned} \tag{16}$$

where we used the fact that

$$\begin{aligned}
&\sum_{r=1}^{n}\binom{n-2}{r-1}G^{r-1}(y)(1-G(y))^{n-(r+1)}\\
&\qquad=\sum_{i=0}^{n-1}\binom{n-2}{i}G^{i}(y)(1-G(y))^{n-i-2},\text{for } i=r-1\\
&\qquad=\sum_{i=0}^{m+1}\binom{m}{i}G^{i}(y)(1-G(y))^{m-i},\text{for } n-2=m\\
&=\sum_{i=0}^{m}\binom{m}{i}G^{i}(y)(1-G(y))^{m-i}+\binom{m}{m+1}G^{i}(y)(1-G(y))^{m-i}=1
\end{aligned}$$

after applying the binomial theorem on the first term in the last step above and using, on the second term, the fact that $\binom{n}{r}=0$ when $n>0$ and $n<r$. In the case that $r>k$ we have

$$\begin{aligned}
&\frac{1}{n}\mathbb{E}\left(\mathbb{E}\left(\frac{\Delta_{(r)}}{1-H(X_{(r)})}W\left(\frac{x-X_{(r)}}{h}\right)\Big|Y_{(k)}\right)\Big|r>k\right)\\
&=\frac{1}{n}\sum_{r=2}^{n}\int\frac{f(y)(1-H(y))}{(1-H(y))g(y)}W\left(\frac{x-y}{h}\right)\binom{n-2}{r-2}G^{r-2}(y)(1-G(y))^{n-r}g(y)\,dy\\
&=\frac{1}{n}\sum_{i=0}^{n}\int\left[\binom{n-2}{i}G^{i}(y)(1-G(y))^{n-i-2}\right]W\left(\frac{x-y}{h}\right)f(y)\,dy\\
&=\frac{1}{n}\int\frac{1}{(1-G(y))^2}W\left(\frac{x-y}{h}\right)f(y)\,dy.
\end{aligned}\tag{17}$$

In view of (3), [12], (16) and (17), (15) becomes

$$\begin{aligned}
\mathbb{E}(S_j|Y_i)&=n^{-1}\int K_h(x-y)F(y)\,dy\\
&\quad+n^{-1}\int\frac{1}{(1-G(y))^2}W\left(\frac{x-y}{h}\right)f(y)\,dy
\end{aligned}$$

from which, under condition 4, the lemma follows immediately. □

Lemma 2. *Under conditions 1-4, for* $m,l>0$ *and for* i *fixed,*

$$\begin{aligned}
&\mathbb{E}\left\{\left(\frac{\Delta_i}{1-H(X_i)}W\left(\frac{x-X_i}{h}\right)\right)^m\left(\frac{\Delta_i}{1-H(X_i)}W\left(\frac{y-X_i}{h}\right)\right)^l\right\}\\
&\quad=h\int W^m(t)W^l\left(\frac{y-x+ht}{h}\right)\frac{f(x-ht)}{(1-H(x-ht))^{m+l-1}}\,dt.
\end{aligned}$$

Proof. The proof is based on conditioning on the number of the uncensored observations of the observed sample $Y_i=(X_i,\Delta_i), i=1,\dots,n$. If N denotes the number of the uncensored observations then $N \sim$Binomial(n, p) where $p = \int f(x)(1 - H(x))\,dx$. For given $N = \nu$, $(X_i : \Delta_i = 1)$ is a set of i.i.d random variables with density $f(x)(1 - H(x))/p$ for $\nu = 1, 2, \ldots, n$. Then,

$$\begin{aligned}
&\mathbb{E}\left\{\left(\frac{\Delta_i}{1 - H(X_i)} W\left(\frac{x - X_i}{h}\right)\right)^m \left(\frac{\Delta_i}{1 - H(X_i)} W\left(\frac{y - X_i}{h}\right)\right)^l\right\} \\
&= \mathbb{E}\left(\mathbb{E}\left\{\left(\frac{\Delta_i}{1 - H(X_i)} W\left(\frac{x - X_i}{h}\right)\right)^m \left(\frac{\Delta_i}{1 - H(X_i)} W\left(\frac{y - X_i}{h}\right)\right)^l \Big| \Delta_i = 1\right\}\right) \\
&= \mathbb{E}\left\{\frac{1}{p}\int \frac{f(z)}{(1 - H(z))^{m+l-1}} W^m\left(\frac{x - z}{h}\right) W^l\left(\frac{y - z}{h}\right) dz\right\} \\
&= h\int W^m(t) W^l\left(\frac{y - x + ht}{h}\right) \frac{f(x - ht)}{(1 - H(x - ht))^{m+l-1}}\,dt.
\end{aligned}$$

Where, in the last step above we used the fact that since i is fixed, $(X_i : \Delta_i = 1)$ is a Bernouli random variable with mean p as well as the change of variable $x - z = ht$. □

Lemma 3. *Under conditions 1–4,*

$$\mathbb{E}V^r = \frac{hf(x)}{(1 - H(x))^{r-1}} a_{r,h} + o(a_{r,h}), \quad r = 1, 2, \ldots.$$

Proof. Apply lemma 2 with $m = r, l = 0$ and expand in Taylor series around x the term $f(x - ht)(1 - H(x - ht))^{-r}$ to get

$$\begin{aligned}
\mathbb{E}V^r &= \frac{hf(x)}{(1 - H(x))^{r-1}} a_{r,h} \\
&\quad + \int_{|x-y| \geq M} \frac{1}{(1 - G(y))^r} W^r\left(\frac{x - y}{h}\right) f(y)(1 - F(y))^r\,dy + o(a_{r,h})
\end{aligned}$$

from which, using condition 4 and choosing M so that $F(x + M) < 1$ the result follows immediately. □

References

[1] Berg, A. and Politis, D. (2009). CDF and survival function estimation with infinite-order kernels. *Electronic Journal of Statistics* **3** 1436–1454. MR2578832

[2] Efron, B. (1981). Censored data and the bootstrap. *J. Amer. Statist. Assoc.*, **76**, 312–319. MR0624333

[3] Fleming, T. H. and Harrington, D. P. (1984). Nonparametric estimation of the survival distribution in censored data. *Comm. in Statistics* **13**, 2469–2486. MR0764844

[4] Gulati, S. and Padgett, W. J. (1996). Families of smooth confidence bands for the survival function under the general random censorship model. *Lifetime Data Anal.* **2** 349–362.

[5] Hajek, J. (1968). Asymptotic normality of simple linear statistics under alternatives. *Ann. Mathemat. Statist.* **39** 325–346. MR0222988
[6] Hall, P. (1992). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Ann. Statist.*, **20**, 675–694. MR1165587
[7] Hall, P. and Claeskens, G. (2002). Data sharpening for hazard rate estimation. *Aust. N.Z. J. Stat*, **44**, 277–283. MR1919194
[8] Hollander, M. and Peña, E. (1989). Families of confidence bands for the survival function under the general random censorship model and the Koziol-Green model. *Can. J. Stat.* **17** 59–74. MR1014091
[9] Hyndman, R. J. and Fan, Y. (1996) Sample Quantiles in Statistical Packages. *American Statist.* **50**, 361–365.
[10] Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481. MR0093867
[11] Karunamuni, R. and Yang, S. (1991) Weak and strong uniform consistency rates of kernel density estimates for randomly censored data. *Can. J. Statist.* **19(4)**, 349–359. MR1166842
[12] Kim, C., Bae, W., Choi, H. and Park, B. U. (2005). Non-parametric hazard function estimation using the Kaplan-Meier estimator. *J. Nonparam. Statist.* **17**, 937–948. MR2192167
[13] Kulasekera, K.B., Williams, C. L., Coffin, M. and Manatuga, A. (2001). Smooth estimation of the reliability function. *Lifetime Data Anal.* **7** 415–433. MR1872568
[14] Leger, C., Politis, D. and Romano, J. (1992). Bootstrap technology and applications. *Technometrics* **34** 378–398.
[15] Lemdani, M. and Ould-Saïd, E. (2003). $\mathcal{L}_1$-deficiency of the Kaplan–Meier estimator. *Statist. Probab. Let.* **63** 145–155. MR1986684
[16] Lio, Y. L. and Padgett, W. J. (1992). Asymptotically optimal bandwidth for a smooth nonparametric quantile estimator under censoring. *Nonparam. Statist.* **1** 219–229. MR1241524
[17] Marron, J.S. and Padgett, W. J. (1987). Asymptotically optimal bandwidth selection for kernel density estimators from randomly right censored samples. *Ann. Statist.* **15**, 1520–1535. MR0913571
[18] Tanner, M. and Wong, W.H. (1983). The estimation of the hazard function from randomly censored data by the kernel method. *Ann. Statist.*, **11**, 989–993. MR0707949
[19] Tsiatis, A. (1981) A large sample study of the estimate for the integrated hazard function in Cox regression model for survival data. *Ann. Statist.* **9**, 93–108. MR0600535